Research Article

Statistics in Medicine

Received 10 August 2012, Accepted 20 May 2013 Published online 12 June 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5877

Informative censoring in relative survival

Anamarija Rebolj Kodre and Maja Pohar Perme*†

With changing the age distribution at the time of cancer diagnosis, the administrative censoring due to study end may be informative. This problem has been mentioned frequently in the relative survival field, and an estimator aiming to correct this problem has been developed. In this paper, we review the existing methods for estimation in relative survival, demonstrate their deficiencies, and propose weighting to correct both the recently introduced net survival estimator and the Ederer I estimator. Using simulations and real cancer registry data, we evaluate the magnitude of the informative censoring problem. We clarify the assumptions behind the reviewed methods and provide guidance to their usage in practice. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: relative survival ratio; net survival; informative censoring; inverse probability weighting

1. Introduction

Survival analysis of data with long-term follow-up is often faced with the issue of competing risks: Many deaths occur because of causes other than the disease of interest, so that the overall survival is not particularly informative. Although crucial information would be provided by the cause of death, it is often unreliable or even unavailable. Data from cancer registries give a prominent example. The methodology developed for such data, named *relative survival*, is based on the assumption that the hazard of causes other than the disease under study can be obtained from the general population mortality tables.

One of the main goals of the analysis of cancer registry data is to compare cancer burden in different calendar periods or different subpopulations without being affected by the differences in the general population mortality trends. Therefore, the focus of the analysis is usually not on competing risks probabilities but rather on the estimation of net survival, defined as the survival probability in a hypothetical world where patients could only die of cancer. Estimation of this probability requires strong nontestable assumptions, and the interpretation of the measure as 'what would happen if patients could only die of cancer' is not really of interest. Nevertheless, unlike in the usual competing risks setting where such functionals are avoided (see e.g., [1]), net survival is a key measure in the relative survival field because it is the only quantity independent of the population mortality and thus directly intercomparable. In addition, the net survival curves graphically accompany the excess hazards models [2–5].

Another quantity of interest is the relative survival ratio defined as the ratio of the actual observed survival of the patients and the survival of their population counterparts. In fact, the net survival and the relative survival ratio were until recently thought to be the same quantity, with the ratio of the observed survival and the population survival serving as the estimator of the net survival [6]. As it turns out, the two quantities can in fact differ considerably in practice. Although the ratio has the advantage of a clear interpretation in the real world, it is usually less desirable than net survival because of its strong dependence on the population mortality trends. Despite the common use of the net survival concept, the fact that it is not a real world measure must be kept in mind at all times, and we try to stress this also by clearly separating the two topics in the manuscript.

A further complication of the cancer registry data, mentioned already by Hakulinen [7], is the fact that if the calendar time frame in which the patients are recruited is wide, their covariate distribution

Department of Biostatistics and Medical Informatics, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia
**Correspondence to: Maja Pohar Perme, Department of Biostatistics and Medical Informatics, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia.*
†E-mail: maja.pohar@mf.uni-lj.si

at diagnosis date may vary considerably. As a consequence, administrative censoring at the end of the study is informative, with some subgroups of patients having a higher probability of getting censored than others [8]. An example is the age of the patients. With the ageing general population, the mean age of diagnosed patients is increasing, which in turn implies that with the study closing at the same time for all the patients, the patients who have a shorter maximum potential follow-up are more likely to be older. Valid inference of all basic survival estimators (e.g., Kaplan–Meier) is provided under the assumption that censoring times must be independent of survival times in the sense that the censoring mechanism keeps the risk set representative for what it would have been without censoring [9]. In our case, this assumption is violated because the censored individuals have a higher hazard of dying as those who remain under follow-up; we refer to this issue as the problem of informative censoring.

The goal of this paper is to find a solution for the problem of informative censoring and evaluate it in practice. In Section 2, we review the existing estimators of relative survival ratio and net survival. Section 3 presents a study of the current solution to the problem and a proposal of a new approach. We explore the performance of the estimators and their corrections with simulations in Section 4. In Section 5, we discuss the assumptions posed by different approaches to estimating survival and evaluate their performance in practice in Section 6. We conclude the paper in Section 7.

2. Estimation of relative survival ratio and net survival

The idea of the relative survival field is to use two data sets – the observed data on patients and the general population mortality data of the relevant country or region. The observed data for ith patient comprise the time of follow-up T_i ($i = 1, \ldots, n$, where n denotes the total number of patients) and the censoring indicator δ_i ($\delta_i = 1$ if the follow-up for ith patient ended in death and 0 otherwise). Time of cancer diagnosis is taken as the time origin (time 0). Using this information, we define the at risk indicator $Y_i(t)$ for each individual ($Y_i(t) = 1$ if $T_i \geqslant t$ and 0 otherwise) and the number of events at each time point $\mathrm{d}N_i(t)$ ($\mathrm{d}N_i(t) = 1$ if $T_{i=t}$, $\delta_{i=1}$ and 0 otherwise). The hazard function that generates these data is referred to as the observed hazard

$$\lambda_O(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leqslant t + \Delta t | T > t)}{\Delta t}.$$

Additionally, a vector Z_i of several covariates might be available. The subgroup D_i of these covariates that are used in the population mortality tables and referred to as demographic variables typically includes age, sex, and calendar year and sometimes also race or deprivation factor. The population mortality tables provide the conditional probability of dying within next year (conditional on still being alive at the beginning of the year) for every combination of the demographic variables. We use the population mortality tables to calculate the population hazard λ_{Pi} for an individual with demographic covariates D_i. We denote the hazard of dying due to cancer as λ_{Ei} (excess hazard) and assume that λ_{Pi} describes well the hazard of dying due to all other causes. The observed hazard λ_{Oi} is then

$$\lambda_{Oi}(t) = \lambda_{Pi}(t) + \lambda_{Ei}(t).$$

We use i in the previous formula to stress that the hazards may differ between individuals and are subject to their covariate values, that is, $\lambda_{Pi}(t) = \lambda_P(t|D_i)$, $\lambda_{Ei}(t) = \lambda_E(t|Z_i)$.

2.1. Relative survival ratio

For a given population of patients, the relative survival ratio S_R is defined as the ratio of the observed survival of these patients S_O and the population survival S_P of their counterparts from the general population with the same values of the demographic covariates. The ratio can be estimated as [10]

$$\hat{S}_R(t) = \frac{\hat{S}_O(t)}{\hat{S}_P(t)}, \tag{1}$$

where $\hat{S}_O(t)$ is the Kaplan–Meier estimate using the observed data on patients and $\hat{S}_P(t)$ is calculated as the average of individual survival curves $\hat{S}_P(t) = \frac{1}{n}\sum_{i=1}^{n} S_{Pi}(t)$ calculated from the population mortality tables. The latter implies estimating the cumulative hazard function Λ as (see [6] for details)

$$\mathrm{d}\hat{\Lambda}_P(t) = \frac{\sum_{i=1}^{n} S_{Pi}(t)\mathrm{d}\Lambda_{Pi}(t)}{\sum_{i=1}^{n} S_{Pi}(t)}. \tag{2}$$

Note that the population survival estimate depends on the sample only as far as the individuals are determined by their demographic covariates and is independent of the actual survival of the patients. The estimator (1) is referred to as the Ederer I estimator.

2.2. Net survival

The net survival is defined as the survival due to the excess hazard λ_E alone; that is, it is a measure defined in the hypothetical world where the population hazard λ_P does not act. For a cohort of individuals with possibly different excess hazards λ_{Ei}, net survival equals

$$S_E(t) = \frac{1}{n}\sum_{i=1}^{n} S_{Ei}(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\left\{-\int_0^t \lambda_{Ei}(t)\mathrm{d}t\right\}.$$

Pohar Perme *et al.* [6] introduced a new estimator of net survival estimator (subsequently denoted as *PP estimator*) that consistently estimates the desired population value. The method estimates the cumulative hazard function as

$$\hat{\Lambda}_E(t) = \int_0^t \frac{\sum_{i=1}^n \mathrm{d}N_i^*(u)}{\sum_{i=1}^n Y_i^*(u)} - \int_0^t \frac{\sum_{i=1}^n Y_i^*(u)\mathrm{d}\Lambda_{Pi}(u)}{\sum_{i=1}^n Y_i^*(u)}, \tag{3}$$

where $\mathrm{d}N_i^*(t) = \mathrm{d}N_i(t)/S_{Pi}(t_i)$ and $Y_i^*(t) = Y_i(t)/S_{Pi}(t)$.
The other approach that can ensure consistent estimation of net survival is to fit a multivariate model so that individual net survival $\hat{S}_{Ei}(t)$ can be predicted for each patient at each follow-up time, and the average of these quantities is then used as the estimate of the overall net survival.

3. Estimators in the presence of informative censoring

Let T_i^* denote the actual time to death for each patient and C_i the time to censoring, so that the observed time is $T_i = \min(T_i^*, C_i)$. In this paper, we distinguish between two sources of censoring. We assume that the initially planned follow-up time, referred to as the potential follow-up time and denoted by G_i, is given for all patients. In the usual case, the potential follow-up time is the difference between the closing date of the study and the diagnosis date of each individual. The actual censoring time is given by $C_i = \min(\widetilde{C}_i, G_i)$, where $\widetilde{C}_i$ is the time to censoring due to any reason but the end of the study. Throughout this work, we shall assume $\widetilde{C}_i$ to be independent of Z_i and T_i, whereas G_i and T_i may depend on Z_i and thus introduce informative censoring. A practical reason for this could be the dependence of the distribution of covariates Z_i on diagnosis date, which in turn determines G_i if the closing date is common for all individuals. Both Ederer I and PP estimators assume noninformative censoring. The effect and importance of informative censoring when estimating net survival have recently been studied in [11], where the screening and inverse screening procedures are presented as a part of informative censoring mechanism. This section explores the possible approaches if the assumption of noninformative censoring does not hold.

3.1. Simulated example

We first illustrate the effect of informative censoring by a simplified example. The cohort under study consists of two equally sized age groups (35 and 55 years, 2500 patients each, all men). All the patients from the older group are diagnosed in 1970, whereas in the younger group, half of the individuals are diagnosed in 1970, and the other half in 1980. If the follow-up period for all patients ends in 1990, this scenario generates informative censoring. The younger patients have a probability 0.5 to be censored after 10 years, whereas all the older patients, who have a considerably higher hazard of dying both due to cancer and due to other causes, stay in the sample for 20 years. The purpose of this example is to show how this censoring mechanism affects the estimators; that is, how the estimated curves on the censored data compare with the ideal case where all individuals were followed up for the entire period of 20 years.

When estimating either relative survival ratio or net survival within the two age groups (dotted lines in Figure 1), any estimator results in the same value because the two groups are homogeneous, and the introduced censoring has no effect beside some random variation (the curves in censored and uncensored case overlap in our case, so only the uncensored cases is plotted). On the contrary, differences arise when

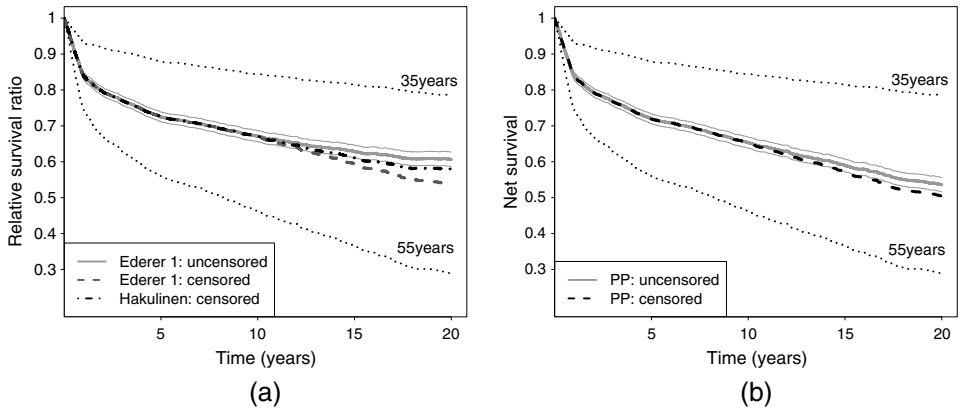

Figure 1. Performance of existing estimators in presence of informative censoring (simulated data) : (a) Ederer I and Hakulinen estimators of relative survival ratio and (b) PP estimator of net survival. The thin grey curves denote the 95% confidence intervals for the uncensored case; the dotted curves represent the estimates within the age subgroups.

considering the overall results. Both the Ederer I (1) and the PP estimators (3) in the censored case underestimate their respective values in the uncensored case – the curves start diverging after 10 years, when the censoring mechanism starts acting and causes the older group to be over represented in the sample.

3.2. The Hakulinen estimator

If censoring is informative, the Kaplan–Meier estimate of the observed survival $\hat{S}_O(t)$ is biased, and hence, the relative survival ratio estimated by the Ederer I method (1) is biased as well. Realizing this fact, Hakulinen [7] proposed a correction of this estimator that aims to introduce a similar bias in the denominator of (1), by allowing each individual to contribute to the overall curve only up to his potential follow-up time. The population cumulative hazard function of the Hakulinen estimator is given by

$$\mathrm{d}\hat{\Lambda}_P^H(t) = \frac{\sum_{i=1}^n Y_i^H(t) S_{Pi}(t) \mathrm{d}\Lambda_{Pi}(t)}{\sum_{i=1}^n Y_i^H(t) S_{Pi}(t)}, \tag{4}$$

where

$$Y_i^H(t) = \begin{cases} I(t \leqslant C_i) \ \ if & \delta_i = 0 \\ I(t \leqslant G_i) \ \ if & \delta_i = 1. \end{cases}$$

The Hakulinen estimator was designed for the situation where no patients are censored before the end of the study, that is, $C_i = G_i$. Such a situation may be realistic in many cancer registries where follow-up is practically complete. In such situation, $Y_i^H(t) = I(t \leqslant G_i)$, and thus, all individuals are included in calculation (4) until their potential follow-up time.

If G does not depend on covariates, the indicator Y_H has the same distribution for all individuals i, and its expected value cancels out in equation (4), so that the Hakulinen and Ederer I estimate the same quantity. Because the Kaplan–Meier estimator of the observed survival part in the numerator of the relative survival ratio (1) is consistent under noninformative censoring, both estimators are consistent.

If G depends on covariates, the Kaplan–Meier estimator of S_O in the numerator of equation (1) is biased; that is, the expected value of the estimates obtained on the censored and uncensored data is no longer equal. The Hakulinen estimator of S_P aims to diminish the subsequent bias of the ratio by introducing a similar bias into the denominator. As simple algebra shows, to entirely remove the bias, we should have

$$\text{bias in the numerator} = S_R(t) \times \text{bias in the denominator}. \tag{5}$$

There is no reason why equation (5) would be true and no general scenarios can be found to ensure the Hakulinen estimator to work perfectly. In particular, the bias in the numerator depends on the effect of any covariate on excess hazard, whereas the bias introduced in the denominator depends only on the

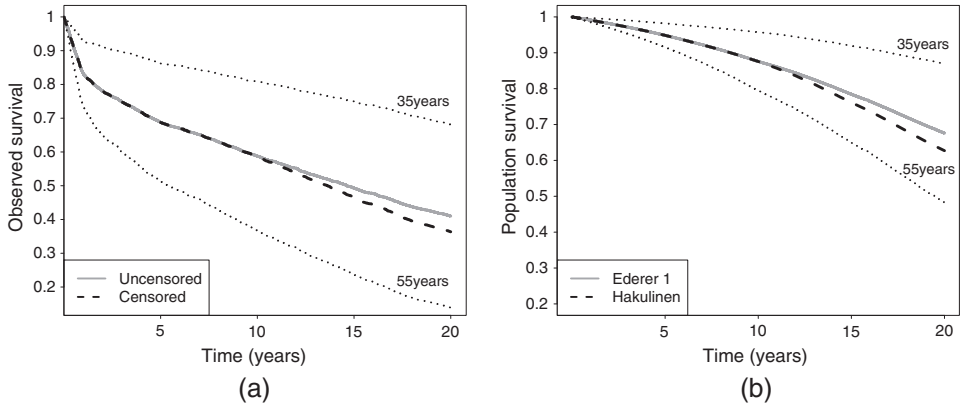

Figure 2. Differences between estimators of the numerator and denominator in the relative survival ratio (simplified example, simulated data) : (a) observed survival, numerator and (b) population survival, denominator. The dotted curves represent the estimates within the age subgroups.

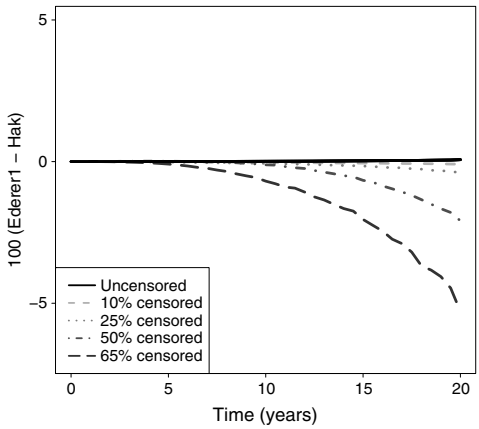

Figure 3. Effect of noninformative censoring on differences between the Ederer I and the Hakulinen estimators (data from Cancer registry of Slovenia, thyroid cancer).

age, sex, and diagnosis year of the patients. The two biases may even be of opposite signs, which would mean that the Hakulinen estimator is even more biased than Ederer I.

To understand why the Hakulinen estimator nevertheless seems to have at least partially desired effect, consider again our simplified example presented in Figure 1. Because of informative censoring, the observed survival calculated with the Kaplan–Meier method in the numerator is underestimated (Figure 2a) because the older patients are over represented in the sample after 10 years. Because the older patients are also over represented when calculating the general population survival at the same time points, the denominator contains a similar bias (Figure 2b), and the ratio (the Hakulinen estimate) in Figure 1a is closer to the value on the uncensored data set than the Ederer I estimate.

A further problem of the Hakulinen estimator is its performance under noninformative censoring. Although the situation with no censoring before the study end may be common in cancer registry data, we can expect many individuals to be lost to follow up in any other long term study where relative survival could also be of use. Consider the situation where all patients have the same potential follow-up time ($G_i = \tau$) but are subject to noninformative censoring with censoring times $\widetilde{C}_i$ equally distributed for all individuals. The Kaplan–Meier estimator produces an unbiased estimator of the numerator, but the correction in the denominator of the Hakulinen estimator still changes its value. We present the behavior of the Hakulinen estimator under noninformative censoring in Figure 3. We considered a subset of patients diagnosed with thyroid cancer between 1970 and 1990 with 20 years complete follow-up time (actual data from Cancer Registry of Slovenia) and censored them with exponentially distributed

censoring times $\widetilde{C}_i$ ($\widetilde{C}_i \sim \exp(\lambda)$). The differences between the Ederer I and the Hakulinen estimators increase with increasing proportion of randomly censored patients.

To understand how the indicator $Y_i^H(t)$ changes the estimator, we calculate its expected value at a given time t. An individual can either still be at risk ($T_i \geqslant t, C_i \geqslant t$) or already out of the risk set because of death ($T_i \leqslant C_i, T_i < t$) or censoring ($C_i < T_i, C_i < t$). In all cases except the last, the value of the indicator $Y_i^H(t)$ equals 1. The expected value of the indicator is thus

$$E\left(Y_i^H(t)\right) = P(T_i \leqslant C_i, T_i < t, C_i < t) + P(C_i \geqslant t). \tag{6}$$

It is clear that the expected value of this indicator depends both on the censoring and on the survival time distribution of an individual. Therefore, if the survival time distribution differs among individuals (this could be either due to different excess hazards or due to different population hazards), the expected values of Y_i^H differ, and $\mathrm{d}\hat{\Lambda}_P^H(t)$ does not estimate the true $\mathrm{d}\Lambda_P(t)$.

The Hakulinen method thus always introduces a bias in the estimator. If censoring is informative, this might help in reducing the bias of the ratio, but if censoring is noninformative and the numerator is estimated without bias, the ratio is biased again.

3.3. *Weighted estimators*

We propose a general solution to the problem of informative censoring applicable both to the Ederer I estimator for the relative survival ratio and to the PP estimator for the net survival. If censoring is non-informative, the time to death and the time to censoring are independent; the patients lost from the sample because of censoring are a random subsample and thus do not introduce bias to either of the estimators. But if censoring times G_i depend on a covariate, the subgroup of individuals that are still at risk at a certain time point no longer represents a random sample of those who would be at risk if there were no censoring. A possible solution is therefore to appropriately weight the individuals, that is, to use weights proportional to $1/S_{Gi}(t-)$, where $S_{Gi}(t-) = P(G_i \geqslant t)$, and thus ensure that the sample still at risk has the desired distribution. As in [6], the weights follow the idea of inverse probability weighting [12–14] and are used to weight both the at risk indicator Y_i and the number of events on each individual $\mathrm{d}N_i$.

3.3.1. Weighted relative survival ratio estimator. In the case of Ederer I estimator (1), the denominator $\hat{S}_P(t)$ depends only on the sample distribution at the time of diagnosis and is thus not affected by informative censoring. Therefore, only the numerator $\hat{S}_O(t)$ is affected, and the proposed correction is introducing the inverse probability of censoring weights to the Nelson–Aalen estimator of the observed cumulative hazard:

$$\mathrm{d}\hat{\Lambda}_O^w(t) = \frac{\sum_{i=1}^n \frac{\mathrm{d}N_i(t)}{S_{Gi}(t-)}}{\sum_{i=1}^n \frac{Y_i(t)}{S_{Gi}(t-)}}. \tag{7}$$

The corrected Ederer I estimator is then the ratio of the survival function $\hat{S}_O^w(t)$ that corresponds to (7) and the population survival corresponding to the cumulative hazard in (2).

3.3.2. Weighted net survival estimator. In the PP estimator, the weights using the population survival function in the cumulative hazard function estimate is multiplied by the weights that correct for informative censoring:

$$\hat{\Lambda}_E^w(t) = \int_0^t \frac{\sum_{i=1}^n \frac{\mathrm{d}N_i(u)}{S_{Pi}(u)S_{Gi}(u-)}}{\sum_{i=1}^n \frac{Y_i(u)}{S_{Pi}(u)S_{Gi}(u-)}} - \int_0^t \frac{\sum_{i=1}^n \frac{Y_i(u)}{S_{Pi}(u)S_{Gi}(u-)}\mathrm{d}\Lambda_{Pi}(u)}{\sum_{i=1}^n \frac{Y_i(u)}{S_{Pi}(u)S_{Gi}(u-)}}, \tag{8}$$

3.3.3. Weights calculation. Note that S_{Gi} has to be evaluated at $t-$, the time just before t, but this is not important in the case of S_{Pi} because it is continuous. Also, the weights S_{Gi} are not given and must be estimated from the sample. Nevertheless, an important advantage over other applications of inverse probability weighting (see e.g., [12–14]) is that the G_i are determined with the time of diagnosis and are thus not affected by the patient's actual survival time, which simplifies the estimation.

One can use many approaches to estimate S_{Gi}; in simple cases, Kaplan–Meier estimates within subgroups of patients or the Cox model may suffice, but one may need more flexible models such as the

Aalen model [15–17] in general [14]. Because the censoring indicator of the event times G_i equals 1 for all individuals, one can use other approaches from outside of the survival field as well.

4. The performance of the estimators under censoring

In this section, we turn to studying the possible bias of Ederer I and PP methods under different patterns of censoring and the performance of the methods that try to correct this bias (Hakulinen, weighted Ederer I, and weighted PP). As a basis for our simulations, we use the organization of the scenarios proposed in [7], the fundamental paper on the effect of informative censoring in the relative survival field. As a guidance to choosing realistic values of hazard, we use the Slovenian data for thyroid and melanoma cancer for the period of diagnosis between 1970 and 1990: The excess hazard strongly depends on patient age in thyroid cancer but is only slightly lower for younger patients with melanoma. Both primary sites were presented as examples of informative-censoring affected cancers in [7]. We fitted an additive model [2] with age as a covariate and a piecewise constant baseline function to the real data and then, for the purpose of simulation, generated data from this model.

With both cancer sites, excess hazard is higher for older patients, but the effect is much stronger with thyroid cancer (coefficient equal to 0.08 per year as compared with 0.01 per year for melanoma). In 20 years of follow-up, the baseline hazard function drops from 0.17 to 0.03 for melanoma and from 0.14 to 0.02 for thyroid cancer. We perform all the simulations on samples of size $n = 2000$ and 500 simulation runs. The simulated patients are all male; half of them are 35 years old, and half 55 years old at the time of diagnosis. We take the population mortality hazards from Slovenian population tables for the relevant period.

Following Hakulinen's study, we introduce two patterns for generating the diagnosis date:

W_0: 5% of the patients are admitted each year over a period of 20 years (1970–1989);
W_2: 3.1% of the patients are admitted during the first year; each year, this percentage is increased by 0.2% (3.3%, 3.5%,...,6.9%).

With a common closing date of the study in 1990, the chosen patterns determine potential censoring times. If the pattern is not equal for both age subgroups, such censoring is informative. We use three different scenarios – the diagnosis of both age groups is determined with W_0 (denoted as W_0/W_0), the 55 age group is admitted with pattern W_2 (W_0/W_2), and the 35 age group is admitted with pattern W_2 (W_2/W_0). In [7], the data generation was simplified so that neither excess nor population hazard changed with diagnosis year. To obtain comparable results, using the yearly Slovenian population mortality tables for data (that change with calendar year), we invert the situation and let all patients be diagnosed on the same date (in 1970). The closing date is then not common for all individuals, but we set the potential follow-up times so that the censoring schemes stay the same (e.g., W_2 denotes the scenario in which 6.9% of the patients are followed up for 1 year, 6.7% for two, etc.). In this way, we obtain scenarios 1–3. We describe the actual dependence on age and year of diagnosis that arises in these scenarios in Table I.

In each simulation run, we generate a data set by using the described model, with follow-up equal to 20 years for all patients. The estimates given by the Ederer I and PP methods on such data set are referred to as reference values. Subsequently, we censor the data set by using one of the scenarios and recalculate the values. We calculate the weights for the proposed weighted versions of the Ederer I and the PP methods by using the Kaplan–Meier estimate on potential follow-up times in each of the age groups. We perform the simulations by using *relsurv* R package [18, 19].

Table I. Simulation scenarios.

Scenario	35/55	Age	Diagnosis year
1	W_0/W_0	Independent	Independent
2	W_0/W_2	Older patients censored earlier	Independent
3	W_2/W_0	Younger patients censored earlier	Independent
4	W_0/W_0	Independent	Later diagnosis censored earlier
5	W_0/W_2	Older patients censored earlier	Later diagnosis censored earlier
6	W_2/W_0	Younger patients censored earlier	Later diagnosis censored earlier

We report the results in Tables II and III. We give all values as $100(\widehat{S}_{\text{cens}}(t) - \widehat{S}_{\text{ref}}(t))$, where $t = 5, 10, 15$, or 20 years, $\widehat{S}_{\text{ref}}$ denotes the estimates on the uncensored data set, and $\widehat{S}_{\text{cens}}$ on the censored one. The results of scenarios 2 and 3 are those that are comparable with [7], but differences may arise because our patients are observed 20 years later (complete Slovenian data are not available before 1970), and thus, both cancer survival and population survival follow different trends. For better readability of the provided tables, we denote the differences that are significant at 0.05 level (using the Bonferonni correction within each curve) with a star.

Scenario 1 in Table II is the only scenario with noninformative censoring, and bias is negligible with all the estimators. Scenarios 2 and 3 introduce censoring dependent on age. The effect is of equal size in both scenarios but in opposite directions; hence, the biases are comparable in size and opposite in sign. In scenario 2, we censor the older patients earlier, so that the overall values approach those of the younger group and are thus overestimated. The effect is much more pronounced in the case of thyroid cancer with a stronger age effect. Age affects both the excess hazard and the population hazard, and although relative survival ratio depends on both, only the excess hazard determines the difference between the age groups in case of net survival. Therefore, the bias of the PP estimator is considerably smaller.

In scenarios 4–6 (Table III), censoring also depends on diagnosis year. Because the diagnosis year in our simulation only affects the population hazard component and the excess hazard, the bias of the net survival estimator remains the same as in scenarios 1–3. Contrariwise, relative survival ratio estimation obtains an additional 'source' of informative censoring. The bias due to informative censoring is thus present even in scenario 4: We censor the patients who are diagnosed later and have a lower population hazard earlier, and the Ederer I estimator thus underestimates the true value.

In scenario 5, we censor the older patients earlier, and the net survival thus is overestimated – it approaches the survival of the young. The difference between the reference values estimated by the PP method and the values obtained on the censored data set is thus positive. The effect is again more pronounced in the case of thyroid cancer, which has a stronger age effect. The relative survival ratio is now affected by the two sources of informative censoring acting in opposite directions – although the net survival is overestimated, the population contribution is underestimated because patients with lower population hazard (diagnosed later) is censored earlier. The two effects practically cancel out in the case of skin melanoma, but the bias persists in thyroid cancer, where net survival differences are larger.

Table II. Differences between reference values (uncensored data with complete 20 years follow-up time, estimated by the Ederer I for relative survival ratio or PP method for net survival) and values obtained by Ederer I, Hakulinen, weighted Ederer I, PP method, and weighted PP method using censored data with heterogeneous follow-up time due to common entry to the study and different study end.

			Skin melanoma				Thyroid cancer			
	Estimator		5 years	10 years	15 years	20 years	5 years	10 years	15 years	20 years
Scenario 1	Ratio	Ed1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	−0.1
		Hak	0.0	0.1	0.1	−0.1	0.0	0.0	0.0	−0.1
		wEd1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	−0.1
	Net	MPP	0.0	0.1*	0.1	0.1	0.0	0.0	0.0	−0.1
		wMPP	0.0	0.1*	0.1	0.1	0.0	0.0	0.0	−0.1
Scenario 2	Ratio	Ed1	0.1*	0.4*	0.9*	1.8*	0.3*	0.9*	1.8*	3.2*
		Hak	0.1*	0.2*	0.2*	0.3	0.2*	0.5*	0.8*	0.8*
		wEd1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0
	Net	MPP	0.1*	0.2*	0.2*	0.4*	0.2*	0.6*	1.1*	1.7*
		wMPP	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Scenario 3	Ratio	Ed1	−0.1*	−0.4*	−0.8*	−1.7*	−0.2*	−0.9*	−1.9*	−3.5*
		Hak	−0.1*	−0.1*	−0.1	−0.2	−0.2*	−0.5*	−0.9*	−1.2*
		wEd1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2
	Net	MPP	−0.1*	−0.1*	−0.1	−0.1	−0.2*	−0.5*	−1.1*	−1.6*
		wMPP	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.2

Ed1, Ederer I; Hak, Hakulinen; wEd1, weighted Ederer I; MPP, PP method; wMPP, weighted PP method.

Statistics in Medicine

Table III. Differences between reference values (uncensored data with complete 20 years follow-up time, estimated by Ederer I for relative survival ratio or PP method for net survival) and values obtained by Ederer I, Hakulinen, weighted Ederer I, PP method, and weighted PP method using censored data with heterogeneous follow-up due time to delayed entry to the study and common study end.

			Skin melanoma				Thyroid cancer			
	Estimator		5 years	10 years	15 years	20 years	5 years	10 years	15 years	20 years
Scenario 4	Ratio	Ed1	0.0	−0.1*	−0.3*	−0.8*	0.0	−0.1*	−0.6*	−1.4*
		Hak	0.0	0.0	0.2*	0.3	0.0	0.0	0.1	0.4*
		wEd1	0.0	−0.1*	−0.3*	−0.8*	0.0	−0.1*	−0.6*	−1.4*
	Net	MPP	0.0	0.0	0.2*	0.3	0.0	0.0	0.0	0.2
		wMPP	0.0	0.0	0.2*	0.3	0.0	0.0	0.0	0.2
Scenario 5	Ratio	Ed1	0.1*	0.2*	0.3*	0.4*	0.2*	0.7*	1.2*	1.5*
		Hak	0.1*	0.1*	0.1	0.2	0.2*	0.5*	0.9*	1.1*
		wEd1	0.0	−0.1*	−0.6*	−1.2*	0.0	−0.2*	−0.6*	−1.7*
	Net	MPP	0.1*	0.1*	0.1	0.3	0.2*	0.6*	1.1*	1.6*
		wMPP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Scenario 6	Ratio	Ed1	−0.1*	−0.5*	−1.3*	−2.9*	−0.3*	−1*	−2.5*	−5*
		HaK	0.0	−0.1*	−0.1	−0.4*	−0.2*	−0.5*	−0.8*	−1*
		wED1	0.0	−0.1*	−0.4*	−1.3*	0.0	−0.1*	−0.5*	−1.5*
	Net	MPP	0.0	−0.1*	−0.1	−0.4*	−0.2*	−0.5*	−1*	−1.6*
		wMPP	0.0	0.0	0.1	−0.2	0.0	0.0	0.1	0.2

Ed1, Ederer I; Hak, Hakulinen; wEd1, weighted Ederer I; MPP, PP method; wMPP, weighted PP method.

In scenario 6, both sources of informative censoring act in the same way on the relative survival ratio, and the bias in Ederer I estimator is the highest. As expected, the magnitude of the bias of the PP estimator is similar to that in scenario 5; only the sign is changed.

We now turn to the results given by estimators correcting for informative censoring. In scenarios 4–6, the Hakulinen estimator seems to offer a reasonable correction and is, even though sometimes still significantly different from the true value, closer to it than the Ederer I estimator.

The weighted Ederer I estimator performs perfectly in scenarios 2–3; all the bias is due to age affecting the censoring pattern, and this is exactly what we accounted for with the weights. The Hakulinen estimator performs worse – it removes some but not all of the bias of the Ederer I estimates. Scenarios 4–6 are more problematic because one cannot correct for the fact that the patients with lower population hazard were censored – no data are available on any representative of the censored groups after they are censored. In these scenarios, the Hakulinen estimator seems to offer a better guess.

This problem does not arise with the weighted PP estimator – because net survival does not change with diagnosis year, the correction is perfect in all cases. This is of course not true in general, and we used the assumption of no effect of diagnosis year on the excess hazard in our simulations only to simplify the scenarios to be able to track the cause of the bias. In an additional simulation (results not reported), where we introduced also an effect of the diagnosis year on the excess hazard, the weighted estimator was not be able to entirely correct the bias because no information was available beyond the last observation date. This is not a problem of the estimator; any extrapolation beyond the last observation date is just a guess.

In our simulation results, the bias exceeds 2% only after 15 or 20 years and only with thyroid cancer, which has a rather extreme effect of age (and diagnosis year) on the excess hazard. Bias shall be the highest, when all sources of informative censoring work in the same direction. Because relative survival ratio also depends on population hazard differences, the bias of the PP estimator in these most extreme examples can never be as large as with the Ederer I estimator.

5. Assumptions

When a correction for informative censoring is necessary, one should above all be careful about the assumptions implied by the methods. We shall compare three different options of correcting:

- Age-standardized estimator [20, 21] : Calculate either Ederer I or the PP estimator in subgroups of patients (defined with respect to age) and then calculate the weighted average, with weights determined by subgroup sizes at time 0.
- Weighted estimator: Estimate $S_{Gi}(t)$ for each individual and use it as a weight in the Ederer I or the PP estimator.
- Multivariate model: Use a multivariate model to estimate $S_{Ei}(t)$ for each individual and average it to estimate the net survival. Analogously, to obtain an estimate of the relative survival ratio, estimate individual $S_{Oi}(t)$ and average the values to calculate the numerator of the Ederer I estimator.

In the simplified scenarios of the previous section, only two values of the age variable were possible. In this case, the age-standardized estimator and the weighted estimator that uses the Kaplan–Meier estimates of S_{Gi} in each group give very similar results. The results are not identical, but the assumptions are. Both assume that patients that are censored within a group would have the same value (of net survival or relative survival ratio) as those that remained in the risk set.

In a more realistic example, where patients' age can vary more, both methods assume that patients who were censored within a subgroup would not change the average net survival or relative survival ratio if they remained at risk.

Suppose now that an age defined subgroup of patients consist of both men and women and that women have a lower excess hazard and hence a better net survival and relative survival ratio.

Further, suppose that half of these women and no men are censored after 5 years of follow-up. If both methods correct for informative censoring only on the level of the age subgroup, they both make the same mistake – after 5 years, the value of the estimate becomes closer to that of men and thus underestimates the true value. The correction is thus unbiased only if the subgroups are chosen so that either all the patients within each subgroup have the same net survival or relative survival ratio or all patients within a subgroup have the same censoring pattern. The age-standardized option is therefore often too coarse, and a model should be used either for the observed (excess) hazard (multivariate model option) or the censoring experience (the weighted estimator option).

There are some further important differences between the multivariate and the weighting option. We first consider the option with no censoring. In this case, no correction with additional weighting is needed in either the PP or the Ederer I estimator, but if the multivariate model is chosen to estimate net survival, it has to be flexible enough to estimate properly the effects of covariates in all nonhomogeneous subgroups. Similarly, when dealing with simple patterns of informative censoring, it is probably easier to model the censoring experience and use it as weights in the estimator.

The other main difference between the alternatives arises when the entire group of patients is censored before the end of the study. In this case, the assumptions of the two methods are very different. The multivariate model assumes that the difference between the groups would stay the same as it was before the time of censoring – if a whole group is lost due to censoring, the multivariate model predicts that the hazard remains the same for the rest of the follow-up period. In the case of the weighted estimator, we simply exclude the group (no representative of the group is left to be weighted); the weighted estimators thus assume that the subgroup was a random subgroup with respect to the net survival or relative survival ratio.

As an example of a situation where the estimated relative survival ratio using multivariate model and weighted estimators will differ, we consider scenario 4, with patients entering the study throughout the 20-year period but with a common final follow-up date. We thus followed the patients with the lowest population hazard for the shortest period. The problem of this informative censoring is that if subgroups are formed with respect to diagnosis year, all patients within each subgroup have the same maximum potential follow-up. Because no patients from a subgroup are left after a certain time, the weights apply to no one, and the weighing does not affect the estimator at all. In other words, knowing that the population hazard decreases with calendar year, we know that the relative survival ratio will be underestimated by our estimator. On the other hand, if a multivariate model is used to model the observed survival, the estimate of the relative survival ratio depends on the assumption of the model – if a time-fixed effect of diagnosis year has been assumed, we assume that the hazard ratio between groups diagnosed in different calendar years remains the same even after some groups are no longer observed. This may seem like a more reasonable assumption, but one must keep in mind that any assumption after the end of follow-up is unverifiable and should in fact best be avoided.

Table IV. Real data example from Slovenian Cancer registry data: estimated survival values for uncensored data with complete 20 years follow-up period and censored data with heterogeneous follow-up time due to common study end on 1 January 1990.

		Relative survival ratio				Net survival		
		Follow-up time						
		Complete	Censored			Complete	Censored	
		Method						
Primary site	Follow-up (years)	Ed1	Ed1	Hak	wEd1	PP	PP	wPP
Skin melanoma	5	57.3	56.7	56.7	56.7	56.9	56.2	56.2
	10	49.8	48.7	48.8	48.7	49.0	47.9	47.8
	15	48.4	48.7	48.7	48.4	46.9	47.2	47.2
	20	49.1	44.5	44.3	45.2	47.3	42.6	43.9
Thyroid cancer	5	73.0	71.6	71.7	71.7	72.3	70.9	71.0
	10	68.7	66.0	66.4	66.4	66.8	64.5	64.8
	15	65.2	62.4	64.1	63.0	61.1	59.0	59.3
	20	64.2	44.6	49.7	53.5	57.2	34.0	43.0

Ed1, Ederer I; Hak, Hakulinen; wEd1, weighted Ederer I; wPP, weighted PP.

6. Informative censoring on real data examples

In this section, we study the importance and the impact of informative censoring in real data. We use the Slovene data for thyroid and melanoma cancer. To study data consistent with our simulations, we use a subsample involving 925 and 453 patients aged under 65 years, diagnosed with malignant melanoma of skin and malignant neoplasm of thyroid gland, respectively. The patients are diagnosed between 1 January 1970 and 31 December 1989 and followed up until 1 January 1990. We compare the estimators obtained on these data and the corrections using our proposed method with the reference value calculated with complete follow-up, that is, 20 years for everyone. We calculate the weights by using the Kaplan–Meier estimator in two age groups (below 45 years and 45–65 years). We report survival values obtained by all examined estimators in Table IV.

The resulting values on the censored data set behave similarly for the relative survival ratio as for the net survival. The estimators that do not take into account the informative censoring (Ederer I and PP) underestimate the values obtained when no censoring is present, but it takes quite a long period before these differences may become substantial. The Hakulinen estimator reduces the difference observed when estimating the ratio, but not as much as the weighted Ederer I option. But even the weighted versions as our best attempt to correct for the informative censoring still remain quite far from the results on the uncensored data, implying that the bias due to the improvement of survival in the years beyond 1990, for which we have no information in the censored data set and we thus cannot correct for, is substantial.

7. Discussion

The issue of informative censoring has been present in the relative survival field for many years; in fact, it was the main reason for introducing the Hakulinen estimator. As we have shown in this paper, the corrected relative survival ratio estimator obtained with this method is often reasonable in practice, but only rarely perfect. It diminishes the bias by an arbitrary degree that cannot be relied upon, and there exists no clear set of assumptions that would ensure unbiasedness. Furthermore, the estimator fails in the simplest scenario when censoring exists but is noninformative, and there exists no guidance when the Hakulinen estimator can be trusted.

We propose a new method for correcting both the relative survival ratio and the net survival estimator. The method requires the same information as the Hakulinen's approach, that is, the knowledge of potential follow-up times for all patients, but ensures consistent results.

Excluding the Hakulinen estimator from the set of options, we require modeling to ensure consistency, and one may choose between modeling the survival or the censoring probability. We believe that the second option may usually be simpler in practice. In particular, when using the net survival, the multivariate

model implies modeling in the relative survival setting, whereas the weighting option only requires a classical survival model, so more options to fit a flexible model are readily available.

The other message of this paper, which has been somehow obscured by the simplified examples in the [7], is that in practice, the bias usually consists of two parts – the bias due to informative censoring that can be corrected and the one that arises if the entire group of patients is censored beyond a certain point. In the latter scenario, extrapolation beyond the final observation date is required, so any corrections are subject to pure guessing. If either the population hazard in a certain country or the disease-specific hazard changes considerably with the year of diagnosis, this bias may prevail, so the usefulness of correcting for informative censoring is highly questionable.

To conclude, we believe that analysis of data with large differences in the potential follow-up time requires untestable assumptions, which should be taken into account when interpreting the results.

Acknowledgement

The authors thank to the Cancer Registry of Slovenia for providing the Slovene cancer registry data.

References

1. Andersen PK, Keiding N. Interpretability and importance of functionals in competing risks and multistate models. *Statistics in Medicine* 2012; **31**(11–12):1074–1088. DOI: 10.1002/sim.4385.
2. Estève J, Benhamou E, Croasdale M, Raymond M. Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* 1990; **9**:529–538.
3. Hakulinen T, Tenkanen L. Regression analysis of relative survival rates. *Journal of the Royal Statistical Society — Series C* 1987; **36**:309–317.
4. Andersen PK, Vaeth M. Simple parametric and nonparametric models for excess and relative mortality. *Biometrics* 1989; **45**:523–535.
5. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004; **23**:51–64. DOI: 10.1002/sim.1597.
6. Pohar Perme M, Stare J, Estève J. On estimation in relative survival. *Biometrics* 2012; **68**(1):113–120. DOI: 10.1111/j.1541-0420.2011.01640.x.
7. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; **38**:933–942.
8. Chiang CL. A stochastic study of the life table and its applications. III. The follow-up study with the consideration of competing risks. *Biometrics* 1961; **17**(1):57–78.
9. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. Springer-Verlag: New York, 1993.
10. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: a statistical methodology. *National Cancer Institute Monograph* 1961; **6**:101–121.
11. Danieli C, Remontet L, Bossard N, Roche L, Belot A. Estimating net survival: the importance of allowing for informative censoring. *Statistics in Medicine* 2012; **31**(8):775–786. DOI: 10.1002/sim.4464.
12. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the American Statistical Association - Biopharmaceutical Section*, San Francisco, California, 1993; 24–33.
13. Satten GA, Datta S. The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician* 2001; **55**(3):207–210.
14. Satten GA, Datta S, Robins J. Estimating the marginal survival function in the presence of time dependent covariates. *Statistics and Probability Letters* 2001; **54**(4):397–403. DOI: 10.1016/S0167-7152(01)00113-4.
15. Aalen O. A linear regression model for the analysis of life times. *Statistics in Medicine* 1989; **8**:907–925.
16. Zahl P. A linear non-parametric regression model for the excess intensity. *Scandinavian Journal of Statistics* 1996; **23**(3):353–364.
17. Cortese G, Scheike TH. Dynamic regression hazards models for relative survival. *Statistics in Medicine* 2008; **27**(18):3563–3548. DOI: 10.1002/sim.3242.
18. Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2006; **81**:272–278.
19. Pohar M, Stare J. Making relative survival analysis relatively easy. *Computers in Biology and Medicine* 2007; **37**:1741–1749.
20. Hakulinen T. On long-term relative survival rates. *Journal of Chronological Disease* 1977; **30**:431–443.
21. Hakulinen T, Seppä K, Lambert PC. Choosing the relative survival method for cancer survival estimation. *European Journal of Cancer* 2011; **47**(14):2202–2210. DOI: 10.1016/j.ejca.2011.03.011.