Biometrics 000, 1–21 DOI: 000
000 0000

Doubly robust estimator for net survival rate in analyses of cancer registry data

Sho Komukai

Graduate School of Medicine, Kurume University, Asahi-Machi 67, Kurume City, Fukuoka 830-0011, Japan

email: a214gm008k@std.kurume-u.ac.jp

and

Satoshi Hattori

Biostatistics Center, Kurume University, Asahi-Machi 67, Kurume City, Fukuoka 830-0011, Japan

email: hattori_satoshi@med.kurume-u.ac.jp

Summary: Cancer population studies based on cancer registry databases are widely conducted to address various research questions. In general, cancer registry databases do not collect information on cause of death. The net survival rate is defined as the survival rate if a subject would not die for any causes other than cancer. This counterfactual concept is widely used for the analyses of cancer registry data. Perme, Stare, and Estève (2012) proposed a non-parametric estimator of the net survival rate under the assumption that the censoring time is independent of the survival time and covariates. Kodre and Perme (2013) proposed an inverse weighting estimator for the net survival rate under the covariate-dependent censoring. An alternative approach to estimating the net survival rate under covariate-dependent censoring is to apply a regression model for the conditional net survival rate given covariates. In this paper, we propose a new estimator for the net survival rate. The proposed estimator is shown to be doubly robust in the sense that it is consistent at least one of the regression models for survival time and for censoring time. We examine the theoretical and empirical properties of our proposed estimator by asymptotic theory and simulation studies. We also apply the proposed method to cancer registry data for gastric cancer patients in Osaka, Japan.

Key words: Cancer registry; Covariate-dependent censoring; Doubly robust estimator; Inverse weighting estimator; Net survival rate

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

To address various research questions, many cancer population studies utilize cancer registry databases. The CONCORD study addressed differences in cancer survival rates among nations in various cancer types, including breast, colon, gastric, prostate, and so on (Coleman et al., 2008). Rachet et al. (2009) examined changes of the survival rates for patients diagnosed each year over time in England and Wales with various common cancers. These studies used data from cancer registries. In general, cancer registries do not collect data on the causes of death. It may lead to difficulty in addressing research questions by using the standard survival analysis techniques since survival times recorded in cancer registry data may not be due to cancer death. For example, Coleman et al. (2008) presented relative survival rates in order to make comparisons among nations by adjusting for differences in survival rates of general populations. Relative survival rates were also employed in Rachet et al. (2009) to adjust for differences in baseline mortality rates over time in order to make more relevant comparisons for cancer survival rates over time.

As done by Coleman et al. (2008) and Rachet et al. (2009), relative survival rates are widely used for analyses of cancer registry data. The relative survival rate is defined as the ratio of the survival rate due to any cause to the survival rate for the general population. Non-parametric estimators by Ederer, Axitell, and Cutler (1961) and Hakulinen (1982) are widely used in practice, and in this paper we refer to them as $ED1$ and HK estimators, respectively. An alternative to the relative survival rate is the net survival rate, which is defined as the survival rate if each subject would not die due to causes other than cancer. In general, the relative survival rate cannot be interpreted as a survival rate and does not agree with the net survival rate. Recently, Perme, Stare, and Estève (2012) proposed a simple estimator, which we here call the Pohar Perme (PP) estimator, for the net survival rate based on cancer registry data. Perme et al. (2012) utilized two survival times for their arguments: time-to-

death due to cancer and that due to other causes. The PP estimator consistently estimates the net survival rates under the assumption that the two survival times are conditionally independent given covariates. In general, neither $ED1$ nor HK is consistent for the net survival rate under the conditional independence assumption: (unconditional) independence between the two survival times is required for consistency. Due to its simplicity and fewer required assumptions, the PP estimator has been of practical use in cancer registry data analyses (Monnereau et al., 2013).

Although the PP estimator relies on weaker assumptions than the other methods on independence between two survival times, it still relies on the assumption that the two survival times are independent of the censoring time. However, as pointed out by Hakulinen (1982), this assumption may not hold in practice. In analyses of cancer registry data, the end of follow-up is often administrative (i.e., independent of covariates), but the entry of a subject may be dependent on covariates. If this is the case, the dependence of entry time on covariates induces covariate-dependent censoring, which is often called informative censoring (Danieli et al., 2012). One approach to estimating the net survival rate in the presence of informative censoring is to apply a regression model for the net survival rate. Various regression methods have been developed, including the Cox-type regression (Hakulinen and Tenkanen, 1987; Estève et al., 1990; Sasieni, 1996; Perme, Henderson and Stare, 2009), the additive hazards model (Lambert et al., 2005; Cortese and Scheike, 2008), and spline-based non-proportional hazards models (Bolard et al., 2002; Giorgi et al., 2003). If a regression model for the net survival rate is correctly specified, by averaging the conditional net survival rate out over an entire population, one can estimate the (marginal) net survival rate. We call these regression-based methods the outcome regression (OR) estimator. In contrast to the above regression-based approaches, an alternative method proposed recently by Kodre and Perme (2013) is based on inverse weighting, and here we call it the weighted Pohar

Perme (wPP) estimator in the present paper. The weight is defined by the conditional survival function for the censoring time given covariates, and the wPP estimator relies on a regression model for the censoring time. If the regression model for the censoring time holds, the wPP estimator consistently estimates the net survival rate.

The OR and wPP estimators provide two ways to estimate net survival rates, and they are complementary to each other. The former is based on modeling time-to-death due to cancer (net survival) and the other is based on modeling censoring time. In causal inference literatures, doubly robust estimators have been of great interest and intensively examined (Lunceford and Davidian, 2004). Motivated by the development of doubly robust estimators in causal inference literatures, we propose an estimator by hybridizing the two estimators of OR and wPP that possesses advantages of the two estimators simultaneously. Our estimator has double robustness in the sense that if at least one of two models for net survival rate and censoring time is correctly specified, the proposed estimator consistently estimates the marginal net survival rate and then has an advantage over OR and wPP for practical use. We denote the proposed doubly robust estimator as the DR estimator.

The rest of the paper is organized as follows. In Section 2, we introduce cancer registry data and some notations. In Section 3, we introduce data subject to covariate-dependent (informative) censoring and describe the DR estimator, and then summarize the theoretical properties of the DR, including consistency and asymptotic normality. In Section 4 we report the results of a simulation study, and in Section 5 we apply the proposed method to a real data from a cancer registry in Osaka prefecture, Japan. Some discussions are made in Section 6. We construct the asymptotic variance estimator in Appendix, and all details on the theoretical developments and complicated formula are placed in Web Appendices.

2. Preliminary

2.1 *Cancer registry data*

Cancer registry data often records gender, age, and year when a subject is diagnosed as cancer for each subject. Let Z be a vector of such baseline covariates. In cancer registry, cause of death is not usually recorded. Let T_O be time-to-death due to any cause. We suppose that T_O may be right-censored by the potential censoring time C. Then we observe only $T = \min(T_O, C)$ and $\Delta = I(T_O \leqslant C)$. We suppose n i.i.d. copies of the triple (T, Δ, Z) are available, which are denoted by (T_i, Δ_i, Z_i) with the subscript i for the ith subject for $i = 1, 2, \cdots, n$.

2.2 *Analysis of the cancer registry data under independent censoring*

To describe statistical methods for estimating net survival rates, we introduce a pair of latent time-to-death. Let T_E be the time-to-death due to a cancer from the date when the cancer was diagnosed, and T_P be the time-to-death due to causes other than the cancer. We suppose $T_O = \min(T_P, T_E)$. The survival function for T_E conditional on Z is denoted by $S_E(t|Z) = P(T_E > t|Z)$. Corresponding hazard and cumulative hazard functions are denoted by $\lambda_E(t|Z)$ and $\Lambda_E(t|Z)$, respectively. Those for T_P and T_O are defined in a similar way with the subscript "P" and "O", respectively.

The net survival rate at time t is defined as $S_E(t) = P(T_E > t)$, which is the survival rate at t if a subject diagnosed as a cancer would not die due to causes other than the cancer. This is an alternative quantity to the relative survival rate $S_O(t)/S_P(t)$, where $S_O(t) = P(T_O > t)$ and $S_P(t) = P(T_P > t)$. In this paper, we focus on the estimation of the net survival rate. Define the counting process for the observed failure time T_O and the at-risk process by $N(t) = I(t \geqslant T, \Delta = 1)$ and $Y(t) = I(t \leqslant T)$, respectively. Note that we use the subscript i for the observation of the ith subject for any quantities. For example, we denote $N(t)$ and $Y(t)$ for an ith individual by $N_i(t)$ and $Y_i(t)$, respectively. The PP estimator (Perme et al.,

2012) is defined as

$$\hat{\Lambda}_E^{PP}(t) = \sum_{i=1}^{n} \int_0^t \frac{\frac{1}{S_P(u|Z_i)}}{\sum_{j=1}^{n} \frac{Y_j(u)}{S_P(u|Z_j)}} \{dN_i(u) - Y_i(u) d\Lambda_P(u|Z_i)\}. \quad (1)$$

In the PP estimator and other estimators for the net survival function, which we will discuss later, the conditional population survival function $S_P(t|Z)$ is calculated by an external database for population mortality and is regarded as known. The PP estimator is a consistent estimator for the cumulative hazard function $\Lambda_E(t)$ under the following conditions $(C1)$ and $(C2)$:

$$(C1) \quad T_E \perp T_P | Z$$

$$(C2) \quad C \perp \{T_E, T_P, Z\},$$

where for any events A, B, and C, $A \perp B|C$ implies the conditional independence of A and B given C. In this paper, we refer to condition $(C2)$ as the independent censoring.

3. Inference under covariate-dependent censoring

3.1 *Covariate-dependent censoring in cancer registry data*

In analyses of cancer registry data, covariate-dependent censoring (often called informative censoring in literatures) may arise in practice (Hakulinen, 1982; Kodre and Perme, 2013). We follow the situation discussed by Kodre and Perme (2013). Consider two potential censoring times G and $\tilde{C}$ and suppose $C = \min(G, \tilde{C})$. We denote G and $\tilde{C}$ for the ith individual by using the subscript i, such as in G_i and $\tilde{C}_i$. We suppose that G_i is observed for all subjects. Consider the following conditions:

$$(C3\text{-}a) \; G \perp \{T_E, T_P\} | Z$$

$$(C3\text{-}b) \; \tilde{C} \perp \{T_E, T_P, Z\}.$$

As argued in Kodre and Perme (2013), this situation often arises in practice. In their arguments, G is a potential follow-up time, which is the time from the entry to the end of follow-up, and $\tilde{C}$ is a time to censoring due to any reasons other than the end of follow-

up. Since the entry time of a subject depends on the covariates Z but the end of follow-up is often administrative, it may be natural that G depends on the covariates Z. The conditions $(C3\text{-}a)$ and $(C3\text{-}b)$ do not imply the condition $(C2)$. Then the PP estimator loses its validity under $(C3\text{-}a)$ and $(C3\text{-}b)$.

3.2 *Existing inference procedures under covariate-dependent censoring*

Kodre and Perme (2013) proposed the following estimator,

$$\hat{\Lambda}_E^{wPP}(t) = \sum_{i=1}^{n} \int_0^t \frac{\frac{1}{\hat{S}_G(u|Z_i)S_P(u|Z_i)}}{\sum_{j=1}^{n} \frac{Y_j(u)}{\hat{S}_G(u|Z_j)S_P(u|Z_j)}} \{dN_i(u) - Y_i(u)d\Lambda_P(u|Z_i)\}, \tag{2}$$

where $\hat{S}_G(t|Z)$ is an estimator of $S_G(t|Z) = P(G > t|Z)$. The estimator (2) is defined by replacing the counting process and the at-risk process in the PP estimator with their inversely weighted versions, respectively. We call the estimator (2) the weighted Pohar Perme (wPP) estimator. To realize the wPP estimator, we need to model $S_G(t|Z)$ with a regression model. In this paper, we employ the Cox regression for the censoring time G,

$$\lambda_G(t|Z) = \underline{\lambda}_G(t) \exp(\beta_G^T Z), \tag{3}$$

where $\underline{\lambda}_G(t)$ is a baseline hazard function and β_G is a vector of regression coefficients. One can estimate β_G and $\underline{\Lambda}_G(t) = \int_0^t \underline{\lambda}_G(u)du$ by the standard maximum partial likelihood method and the Breslow estimator, which are denoted by $\hat{\beta}_G$ and $\hat{\underline{\Lambda}}_G(t)$, respectively. Then, $S_G(t|Z)$ is estimated by $\hat{S}_G(t|Z) = \exp\{-\hat{\underline{\Lambda}}_G(t) \exp(\hat{\beta}_G^T Z)\}$. As suggested in Kodre and Perme (2013), the wPP estimator consistently estimates the net cumulative hazard function $\Lambda_E(t)$ if the conditions $(C1)$, $(C3\text{-}a)$, and $(C3\text{-}b)$ hold and the model for $S_G(t|Z)$ is correctly specified.

Alternatively to the wPP estimator, using a regression model for T_E, the net cumulative hazard function $\Lambda_E(t)$ is estimated by

$$\hat{\Lambda}_E^{OR}(t) = \int_0^t \frac{\sum_{i=1}^{n} \hat{S}_E(u|Z_i)d\hat{\Lambda}_E(u|Z_i)}{\sum_{j=1}^{n} \hat{S}_E(u|Z_j)} \tag{4}$$

or,

$$\hat{\Lambda}_E^{OR}(t) = -\log \frac{1}{n} \sum_{i=1}^{n} \hat{S}_E(t|Z_i), \tag{5}$$

where $\hat{S}_E(t|Z)$ and $\hat{\Lambda}_E(t|Z)$ are estimators of $S_E(t|Z)$ and $\Lambda_E(t|Z)$, respectively, based on a regression model for T_E. We call $\hat{\Lambda}_E^{OR}(t)$ the Outcome Regression (OR) estimator. If the conditions ($C1$), ($C3$-a) and ($C3$-b) hold and the regression model for the estimation of $S_E(t|Z)$ is correctly specified, the OR estimator is a consistent estimator of $\Lambda_E(t)$.

Many studies have proposed inference procedures of regression models for T_E. For example, inference of the Cox-type model $\lambda_E(t|Z) = \underline{\lambda}_E(t) \exp(\beta^T Z)$ was discussed by Sasieni (1996) and Perme et al. (2009), where $\underline{\lambda}_E(t)$ is an unspecified baseline hazard function and β is a regression coefficient vector. Alternatively, one can employ some non-proportional hazards models such as the additive hazards model (Cortese and Scheike, 2008). On the other hand, there are disadvantages to the use of the regression models for T_E. One disadvantage is that such models require special softwares. Further disadvantage is that tools helpful for model identification, such as goodness-of-fit tests, are less developed. We introduce an alternative method to overcome these disadvantages. Under the condition ($C1$), it holds that

$$S_E(t|Z) = \frac{S_O(t|Z)}{S_P(t|Z)}. \tag{6}$$

We note that the standard relative survival rate is defined as $S_E(t) = S_O(t)/S_P(t)$, which is based on marginal probabilities. On the other hand, the equation (6) is based on conditional probabilities given covariates, and implies that the conditional version of the relative survival rate can be regarded as the conditional net survival rate under the condition ($C1$). This conditional relative survival rate (6) is useful to estimate $\Lambda_E(t)$. One can estimate $\Lambda_E(t)$ with (4) or (5) by estimating $S_E(t|Z)$ with $\hat{S}_O(t|Z)/S_P(t|Z)$ or $\Lambda_E(t|Z)$ with $\hat{\Lambda}_O(t|Z) - \Lambda_P(t|Z)$, where $S_P(t|Z)$ is extracted from an external database for a general population; we are available $S_P(t|Z)$ for $t = 1, 2, \cdots$ which are recorded in the external databases, such as a life-table, and get it for other t by linear extrapolation, and $\hat{S}_O(t|Z)$ is based on a regression

model for T_O. To estimate $S_O(t|Z)$, noting that the data (T, Δ, Z) are the standard right-censored data for T_O, one can employ any regression models for the standard right-censored data, which have been extensively developed. Here, we employ the Cox regression model for T_O,

$$\lambda_O(t|Z) = \underline{\lambda}_O(t) \exp(\beta_O^T Z), \tag{7}$$

where $\underline{\lambda}_O(t)$ is a baseline hazard function and β_O is a vector of regression coefficients. One can estimate β_O and $\underline{\Lambda}_O(t) = \int_0^t \underline{\lambda}_O(u)du$ by the standard maximum partial likelihood method and the Breslow estimator, which are denoted by $\hat{\beta}_O$ and $\hat{\underline{\Lambda}}_O(t)$, respectively, and $S_O(t|Z)$ is estimated by $\hat{S}_O(t|Z) = \exp\{-\hat{\underline{\Lambda}}_O(t) \exp(\hat{\beta}_O^T Z)\}$. Then, $\Lambda_E(t|Z)$ is estimated by $\hat{\Lambda}_E(t|Z) = \hat{\underline{\Lambda}}_O(t) \exp(\hat{\beta}_O^T Z) - \Lambda_P(t|Z)$. If the regression model for $S_O(t|Z)$ is correctly specified, the OR estimator is consistent. For the model (7), various techniques for model-identification have been developed (Lin, 1991; Lin and Wei, 1991; Lin, Wei and Ying, 1993 among others). In particular, the model-checking procedure based on the cumulative martingale residuals by Lin et al. (1993) can be easily implemented with PHREG procedure in SAS (SAS institute) and *timereg* package in R (Martinussen and Scheike, 2006). Then we can identify a model more accurately than relying on a model for T_E.

3.3 *Doubly robust estimator*

Finally, we propose a new estimator of double robustness by combining ideas of the wPP and OR estimators. The DR estimator is defined as

$$\hat{\Lambda}_E^{DR}(t) = \hat{\Lambda}_E^{wPP}(t) - \hat{\Gamma}(t) + \hat{\Lambda}_E^{OR}(t), \tag{8}$$

where

$$\hat{\Gamma}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{I(G_i \geqslant u)}{\hat{S}_G(u|Z_i)} \hat{S}_E(u|Z_i) d\hat{\Lambda}_E(u|Z_i)}{\sum_{j=1}^n \frac{I(G_j \geqslant u)}{\hat{S}_G(u|Z_j)} \hat{S}_E(u|Z_j)}. \tag{9}$$

The first and third terms of the right-hand side of (8) are the wPP estimator (2) and the OR estimator (4), respectively. Note that the second term (9) is regarded as the inverse-weighted

version of the OR estimator (4). Then, if the model for $S_G(t|Z)$ is correctly specified, the second and third terms of the right-hand-side of (8) converge to the same limit and are canceled each other. Therefore, the DR estimator is asymptotically equivalent to the wPP estimator, which is consistent. On the other hand, when the model $S_O(t|Z)$ correctly specified $\Pr(t < T_O|Z)$, the second term $\hat{\Gamma}(t)$ is asymptotically equivalent to

$$\sum_{i=1}^{n} \int_0^t \frac{\frac{1}{\hat{S}_G(u|Z_i)S_P(u|Z_i)}}{\sum_{j=1}^{n} \frac{Y_j(u)}{\hat{S}_G(u|Z_j)S_P(u|Z_j)}} Y_i(u) d\Lambda_E(u|Z_i).$$

Then the first and second terms of (8) are asymptotically equivalent to

$$\sum_{i=1}^{n} \int_0^t \frac{\frac{1}{\hat{S}_G(u|Z_i)S_P(u|Z_i)}}{\sum_{j=1}^{n} \frac{Y_j(u)}{\hat{S}_G(u|Z_j)S_P(u|Z_j)}} dM_i(u),$$

where $M_i(t) = N_i(t) - \int_0^t Y_i(u)\{d\Lambda_E(u|Z_i) + d\Lambda_P(u|Z_i)\}$ is the counting process martingale for $N_i(t)$. Then, from the standard counting process martingale arguments (Fleming and Harrington, 1991), it converges in probability to 0 and then (8) has the same limit as the OR estimator, which is consistent in this case. Thus, the DR estimator (8) consistently estimates the net cumulative hazard function $\Lambda_E(t)$ if at least one of $S_G(t|Z)$ and $S_O(t|Z)$ is correctly specified. That is, the DR estimator has a double robustness. A formal proof of the double robustness is presented in Web Appendix A. In general, $\hat{\Lambda}_E^{DR}(t)$ converges in probability to some limit, which is denoted by $\Lambda_E^*(t)$. In Web Appendix B, we show that $n^{1/2}\left\{\hat{\Lambda}_E^{DR}(t) - \Lambda_E^*(t)\right\}$ has a normal distribution asymptotically, and its asymptotic variance can be consistently estimated by $n^{-1}\sum_{i=1}^{n} \hat{k}_i^{DR}(t;\hat{\theta})^2$, where the definition of $\hat{k}_i^{DR}(t;\hat{\theta})$ is given in Appendix. Due to the double robustness, $\Lambda_E^*(t)$ agrees with $\Lambda_E(t)$ if at least one of $S_G(t|Z)$ and $S_O(t|Z)$ is correctly specified. Then, one can construct a pointwise confidence interval of $\Lambda_E(t)$ for a given t according to the asymptotic normality.

4. Simulation study

We conducted a simulation study to examine the behavior of the proposed estimator. We considered three covariates, *age*, *gender*, and *year*, which were the age at diagonosis, the

gender, and the year of diagnosis. *Age* and *gender* were generated from the normal distribution $N(60, 10^2)$ and the Bernoulli distribution $B(1/2)$, respectively. We generated the potential follow-up time G from the exponential distribution with hazard rate $\lambda_G(t|Z) = 0.12\exp\{\log 0.7 \times st(age) + \log 1.7 \times gender + \log 0.7 \times st(age)^2\}$, and then *year* was calculated by $e_f - G$, where $st(age) = (age - 60)/10$ and e_f was the date of the end of the follow-up period. We considered three settings in generating T_E. The failure time T_E was generated from the exponential distribution with the hazard rate $\lambda_E(t|Z) = 0.1\exp(\beta^T Z)$, where $\beta = (\log 2, \log 0.5, \log 2)^T$, and Z was as follows;

$$
\begin{aligned}
&\textit{Dataset } 1: \quad Z = \{st(age), gender, st(age)^2\}^T \\
&\textit{Dataset } 2: \quad Z = \{st(age), gender, st(age) \times gender\}^T \\
&\textit{Dataset } 3: \quad Z = \{st(age), gender, st(age)^2 \times gender\}^T.
\end{aligned}
$$

To calculate $S_P(t|Z)$, we employed the life-table based on the National Cancer Center in Japan. It records $S_P(t|Z)$ for $t = 1, 2, \cdots$, and for other t, we extrapolated linearly. We ganarated T_P from this model, asuuming that T_E and T_P were conditionally independent given the covariates Z. The potential censoring time $\tilde{C}$ was generated from the uniform distribution from 0 to 50. We set the number of subjects $n = 1,000$, and 1,000 datasets were simulated. The true net survival function $S_E(t) = E_Z[\exp\{-t\lambda_E(t|Z)\}]$ was calculated by $20,000^{-1}\sum_{m=1}^{20,000}\exp\{-t\lambda_E(t|Z_m)\}$, where Z_m was the mth sets of covariates.

In analyses, we fitted the model (3) for G with $G1 : Z = \{st(age), gender, st(age)^2\}$ or $G2 : Z = \{st(age), gender\}^T$ for G. In all the datasets, the model with $G1$ was correctly specified, and the model with $G2$ was misspecified. For each dataset, we fitted the model (7)

for T_O with the following $O1$ or $O2$;

$$\begin{aligned}
Dataset\ 1 \quad O1: &\quad Z = \{st(age), gender, st(age)^2\}^T \\
O2: &\quad Z = \{st(age), gender\}^T \\
Dataset\ 2 \quad O1: &\quad Z = \{st(age), gender, st(age) \times gender\}^T \\
O2: &\quad Z = \{st(age), gender\}^T \\
Dataset\ 3 \quad O1: &\quad Z = \{st(age), gender, st(age)^2 \times gender\}^T \\
O2: &\quad Z = \{st(age), gender\}^T.
\end{aligned}$$

In each dataset, the model $O1$ held and $O2$ was misspecified. We estimated the net survival rate by the DR estimator with combination of $O1$ or $O2$ with $G1$ or $G2$. In Analysis 1, we employed $G1$ and $O1$ and then both models were correctly specified. In Analysis 2, we employed $G2$ and $O1$, with which the model for G was misspecified and that for T_O was correctly specified. In Analysis 3, we employed $G1$ and $O2$, with which the model for G was correctly specified and that for T_O was misspecified. In Analysis 4, we employed $G2$ and $O2$, neither of which was correctly specified. We evaluated empirical biases, mean squared errors (MSE), and coverage probabilities (CP) for 5-year, 7-year, and 10-year net survival rates for the DR, wPP, OR, and PP estimators.

The results for Datesets 1, 2, and 3 are summarized in Tables 1, 2, and 3, respectively. Except for the result for $t = 10$ in Table 1, the proposed DR estimator worked as expected. That is, we observed that

1) In Analysis 1, as expected, the biases were negligible in the DR, wPP, and OR estimators, and the PP had a considerable bias in all Datasets. The OR estimator had the smallest MSE.

2) In Analysis 2, the biases and MSEs for the DR and OR estimators were smaller than those for the wPP and PP estimators.

3) In Analysis 3, the biases and MSEs for the DR and wPP estimators were smaller than

those for the others.

4) The DR estimator had only negligible biases if at least one of models (3) and (7) was correctly specified.

5) Empirical coverage probabilities of the DR estimator were close to the nominal level of 95% when at least one of the models was correctly specified.

6) As seen in Analysis 4, the DR estimator did not necessarily outperform when both models were misspecified.

On the other hand, the result for $t = 10$ in Table 1 indicated that the OR estimator outperformed the DR estimator in all the cases including Analysis 3, in which the OR estimator was misspecified. Whereas the DR and wPP estimators had the biases smaller than the OR estimator as expected, the OR estimator was the best performance with respect to MSEs. Furthermore, we observed that in Analysis 4 (both models were misspecified), the DR estimator had substantially larger biases and MSEs than the OR estimator. To examine why the DR estimator did not work well, we checked the weights in inverse weighting when the model was misspecified. There were extremely large weights $S_G(t|Z)^{-1}$ at $t = 10$; the weights ranged from 1.29 to 58.10 at $t = 10$, whereas they ranged from 1.15 to 8.40 at $t = 5$ and from 1.20 to 19.00 at $t = 7$ in Dataset 1. In Datasets 2 and 3, these ranges were similar to those in Dataset 1. Poor performance of the doubly robust estimators due to extremely large weights were reported in the causal inference (Kang and Schafer, 2007), and then we speculate that the poor performance of the proposed DR estimator at $t = 10$ in Table 1 was due to the presence of extremely large weights. Although, in general, our DR estimator has advantages over the other methods, it is recommended to check the distribution of the weights in practice.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

5. Example

We apply the proposed method to the population-based cancer registry data in Osaka prefecture, Japan. This registry consisted of data on 2,023 male patients diagnosed as gastric cancer with the adjacent organs from 1990 to 2000. We analyzed the follow-up data at 2000, at which 1,739 patients had died and 284 patients had been censored. These data had two covariates: age at diagnosis and the calender year at that time of diagnosis. We applied Cox regression models (3) and (7) with standardized age, which is denoted by $st(age)$, as explanatory variables to estimate the distribution of the potential follow-up time and of the failure time, respectively. To calculate the population mortality, we used a life-table from the National Cancer Center in Japan.

Table 4 shows estimates of 5-year and 7-year net survival rates and their 95% confidence intervals. Figure 1 plots the estimated net survival function by each estimator. For reference, the Kaplan-Meier estimator for $S_O(t)$ is also shown, which is referred to as OS. As anticipated, the OS underestimates the net survival rate. The estimates and the confidence intervals by the DR, wPP, and PP estimators were almost the same as each other. On the other hand, the estimate by the OR estimator was lower than those by the others. We applied the model-checking procedure based on the cumulative Martingale residuals (Lin et al., 1993) for the models (3) and (7). In Figure 2, we showed the observed cumulative Martingale residuals (left panel) and the observed score process (right panel) with their randomly selected 50 simulated null processes and p-values of the supremum-type tests for the model (3) (upper panel) and the model (7) (lower panel). Figure 2 indicates that the Cox model (3) seems to fit well, whereas the Cox model (7) seems to be misspecified. Corresponding to this observation, the OR estimator is far from the others. We observed that the regression coefficient for age in the model (3) was not statistically significant, indicating

that there is not a strong association between censoring time and age. Then the PP estimator also may work well, and thus the PP and wPP estimates were close.

[Table 4 about here.]

[Figure 1 about here.]

[Figure 2 about here.]

6. Discussion

In this paper, we proposed the DR estimator that is applicable under covariate-dependent censoring. Recently, the PP estimator has been gaining popularity in practice. However, as observed in our simulation study, it may have considerable biases in the presence of covariate-dependent censoring. Thus, methods valid under covariate-dependent censoring are recommended to be applied in practice at least for sensitivity analysis. Due to its double robustness, the DR estimator has potential for practical use.

While the net survival rate has a counterfactual nature, it is regarded as a survival rate due to cancer. However, as observed in our example and reported previosly (Perme et al., 2012), estimates for the net survival rate may not be a decreasing function. The relative survival rate is an alternative to the net survival rate. From the definition of the relative survival rate, it allows to be increasing and an useful alternative to the net survival rate. We are developing a doubly robust estimator for the relative survival rate and will report the results in the future.

In this paper, we employed the Cox regression for both the failure time and the censoring time. In practice, the Cox regression may not fit well. In principle, one may utilize any regression models, including semiparametric non-proportional hazards models such as the additive hazards model (Lin and Ying, 1994) and the linear transformation model (Chen, Jin and Ying, 2002). To this end, one would need to modify our variance estimators.

As observed in our simulation studies, our proposed DR estimator does not necessarily perform well when both models are misspecified. This phenomenon is observed in the doubly robust estimator for the average causal effect in causal inference leteratures (Kang and Schafer, 2007). In causal inference, several proposals have been made to overcome this important issue (Cao, Tiatis and Davidian, 2009; Han and Wang, 2013). It is worth while to develop estimators for the net survival rate, which are robust when both models are misspecified.

7. Supplementary Materials

Web Appendices A and B referenced in Section 3, is available with this paper at the Biometrics website on Wiley Online Library. A R source code implementing our proposed and related methods is also available with a set of sample datasets.

Acknowledgements

The authors thank to Dr. Yuri Ito in the Osaka Medical Center for Cancer and Cardiovascular Diseases for the providing the Osaka cancer registry data. The authors are also grateful to Dr. Bernard Rachet in the Non-Communicable Disease Epidemiology Unit, London School of Hygiene and Tropical Medicine, London, UK for giving us the motive comments of this study.

References

Bolard, P., Quantin, C., Abrahamowicz, M., Estève, J., Giorgi, R., Chadha-Boreham, H. et al. (2002). Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology and Prevention* **7,** 113–122.

Cao, W., Tiatis, A. A. and Davidian, M. (2009). Improveing efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* 1–12.

Chen, K., Jin, Z. and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89,** 659–668.

Coleman, M. P., Quaresma, Q., Berrino, F., Lutz, J., Angelis, R. D., Capocaccia, R., et al. (2008). Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncology* **9,** 730–756.

Cortese, G. and Scheike, T. H. (2008). Dynamic regression hazards models for relative survival. *Statistics in Medicine* **27,** 3563–3584.

Danieli, C., Remontet, L., Bossard, N., Roche, L. and Belot, A. (2012). Estimating net survival: the importance of allowing for informative censoring. *Statistics in Medicine* **31,** 775–786.

Ederer, F., Axitell, L. M., and Cutler, S. J. (1961). The relative survival rate: a statistical methodology. *National Cancer Institute Monograph* **6,** 101–121.

Estève, J., Benhamou, E., Croasdale, M., and Raymond, L. (1990). Relative survival and the estimation of net survival: elements for further discussion. *Statistics in Medicine* **9,** 529–538.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Hoboken, NJ:John Wiley and Sons, Inc.

Gorgi, R., Abrahamowicz, M., Quantin, C., Bolard, P., Estève, J., Gouvernet, J. and Faivre, J. (2003). A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* **22,** 2767–2784.

Hakulinen, T. (1982). Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* **38,** 933–942.

Hakulinen, T. and Tenkanen, L. (1987). Regression analysis of relative survival rates. *Journal*

of the Royal Statistical Society, Series C. **36,** 309–317.

Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika* **100,2,** 417–430.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, Second Edition. Hoboken, NJ:John Wiley and Sons, Inc.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22,** 523–539.

Kodre, A. R. and Perme, M. P. (2013). Informative censoring in relative survival. *Statistics in Medicine* **32,** 4791–4802.

Lambert, P. C., Smith, L. K., Jones, D. R. and Botha J. L. (2005). Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine* **24,** 3871–3885.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the cox proportional hazard model. *Journal of the American Statistical Association* **84,** 1074–1078.

Lin, D. Y. and Wei, L. J. (1991). Goodness-of-fit tests for the general cox regression model. *Statistica Sinica* **1,** 1–17.

Lin, D. Y. (1991). Goodness-of-fit analysis for the cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association* **86,** 725–728.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika* **80,** 557–572.

Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81,** 61–71.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society* **62,**

711–730.

Lunceford, J. K. and Davidian, M. (2004). Stratication and weighting via the propensity score in estimation of causal treatment eects: a comparative study. *Statistics in Medicine* **23,** 2937–2960.

Martinussen, T. and Scheike, T. H. (2006). *Dynamic Regression Models for Survival Data*. Springer.

Monnereau, A., Troussard, X., Belot, A., Guizard, A., Woronoff, A., Bara, S., et al. (2013). Unbiased estimates of long-term net survival of hematological malignancy patients detailed by major subtypes in France. *International Journal of Cancer* **132,** 2378–2387.

Perme, M. P., Henderson, R., and Stare J.(2009). An approach to estimation in relative survival regression. *Biostatistics* **10,** 136–146.

Perme, M. P., Stare, J., and Estève, J. (2012). On estimation in relative survival. *Biometrics* **68,** 113–120.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Hayward: Institute of Mathematical Statistics.

Rachet, B., Maringe, C., Nur, U., Quaresma, M., Shah, A., Woods, L. M., et al. (2009). Population-based cancer survival trend in England and Wales up to 2007: an assessment of the NHS cancer plan for England. *Lancet Oncology* **10,** 351–369.

Sasieni, P. D. (1996). Proportional excess hazards. *Biometrika* **83,** 127–141.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73,** 363–369.

Appendix : Definition of the variance estimator

As introduced in Section 3, the variance of $\sqrt{n}\{\hat{\Lambda}_E^{DR}(t) - \Lambda_E^*(t)\}$ is consistently estimated by $n^{-1}\sum_{i=1}^n \hat{k}_i^{DR}(t;\hat{\theta})^2$, where $\Lambda_E^*(t)$ was defined in (A.4) of Web Appendix A and $\hat{k}_i^{DR}(t;\theta)$ is obtained from $k_i^{DR}(t;\theta)$, by replacing all theoretical quantities by their respective empirical

counterparts. Although the definition of $k_i^{DR}(t;\theta)$ is given in Web Appendix B, in this appendix, we summarize the definition of $\hat{k}_i^{DR}(t;\theta)$ since it is very complicated. Define $N_{Gi}(t) = I(G_i \leqslant t)$, $\overline{N}_G(t) = n^{-1}\sum_{i=1}^n N_{Gi}(t)$, and $\overline{N}(t) = n^{-1}\sum_{i=1}^n N_i(t)$. Let $V_O^{(r)}(\beta_O, t) = n^{-1}\sum_{i=1}^n Y_i(t) Z_i^{\otimes r} e^{\beta_O^T Z_i}$ and $V_G^{(r)}(\beta_G, t) = n^{-1}\sum_{i=1}^n I(G_i \geqslant t) Z_i^{\otimes r} e^{\beta_G^T Z_i}$ for $r = 0, 1, 2$, where $Z^{\otimes 2} = ZZ^T$, $Z^{\otimes 1} = Z$ and $Z^{\otimes 0} = 1$. We denote $\hat{\Theta}_{Gi}(t;\beta_G)^T = \{\hat{q}_{Gi}(t;\beta_G), \hat{w}_{Gi}(\beta_G)^T\}$, $\hat{\Theta}_{Oi}(t;\beta_O)^T = \{\hat{q}_{Oi}(t;\beta_O), \hat{w}_{Oi}(\beta_O)^T\}$, and $\hat{\Theta}_i(t;\beta_G,\beta_O)^T = \{\hat{\Theta}_{Gi}(t;\beta_G)^T, \hat{\Theta}_{Oi}(t;\beta_O)^T\}$, where

$$
\begin{aligned}
\hat{w}_{Gi}(\beta_G) &= \left[\int_0^\tau \left\{-\frac{V_G^{(2)}(\beta_G,u)}{V_G^{(0)}(\beta_G,u)} + \frac{V_G^{(1)}(\beta_G,u)^{\otimes 2}}{V_G^{(0)}(\beta_G,u)^2}\right\} d\overline{N}_G(u)\right]^{-1} \\
&\quad \times \int_0^\tau \left\{Z_i - \frac{V_G^{(1)}(\beta_G,u)}{V_G^{(0)}(\beta_G,u)}\right\}\left\{dN_{Gi}(u) - \frac{I(G_i \geqslant u) e^{\beta_G^T Z_i}}{V_G^{(0)}(\beta_G,u)} d\overline{N}_G(u)\right\}, \\
\hat{q}_{Gi}(t;\beta_G) &= \left\{\int_0^t -\frac{V_G^{(1)}(\beta_G,u)^T}{V_G^{(0)}(\beta_G,u)^2} d\overline{N}_G(u)\right\} \hat{w}_{Gi}(\beta_G) \\
&\quad + \int_0^t \frac{1}{V_G^{(0)}(\beta_G,u)}\left\{dN_{Gi}(u) - \frac{I(G_i \geqslant u) e^{\beta_G^T Z_i}}{V_G^{(0)}(\beta_G,u)} d\overline{N}_G(u)\right\}, \\
\hat{w}_{Oi}(\beta_O) &= \left[\int_0^\tau \left\{-\frac{V_O^{(2)}(\beta_O,u)}{V_O^{(0)}(\beta_O,u)} + \frac{V_O^{(1)}(\beta_O,u)^{\otimes 2}}{V_O^{(0)}(\beta_O,u)^2}\right\} d\overline{N}(u)\right]^{-1} \\
&\quad \times \int_0^\tau \left\{Z_i - \frac{V_O^{(1)}(\beta_O,u)}{V_O^{(0)}(\beta_O,u)}\right\}\left\{dN_i(u) - \frac{Y_i(u) e^{\beta_O^T Z_i}}{V_O^{(0)}(\beta_O,u)} d\overline{N}(u)\right\}, \\
\hat{q}_{Oi}(t;\beta_O) &= \left\{\int_0^t -\frac{V_O^{(1)}(\beta_O,u)^T}{V_O^{(0)}(\beta_O,u)^2} d\overline{N}(u)\right\} \hat{w}_{Oi}(\beta_O) \\
&\quad + \int_0^t \frac{1}{V_O^{(0)}(\beta_O,u)}\left\{dN_i(u) - \frac{Y_i(u) e^{\beta_O^T Z_i}}{V_O^{(0)}(\beta_O,u)} d\overline{N}(u)\right\}.
\end{aligned}
$$

Furthermore, we denote $\theta_O^T = (\underline{\Lambda}_O(.), \beta_O^T)$, $\theta_O^T(t) = (\underline{\Lambda}_O(t), \beta_O^T)$, $\hat{\theta}_O^T = (\hat{\underline{\Lambda}}_O(.), \hat{\beta}_O^T)$, and $\hat{\theta}_O^T(t) = (\hat{\underline{\Lambda}}_O(t), \hat{\beta}_O^T)$. Similar notations are used for model (3) with the subscript G. In addition, we denote $\theta^T = (\theta_G^T, \theta_O^T)$, $\theta^T(t) = (\theta_G^T(t), \theta_O^T(t))$, $\hat{\theta}^T = (\hat{\theta}_G^T, \hat{\theta}_O^T)$, and $\hat{\theta}^T(t) = (\hat{\theta}_G^T(t), \hat{\theta}_O^T(t))$. Then, $\hat{k}_i^{DR}(t;\hat{\theta})$ is defined as

$$
\hat{k}_i^{DR}(t;\hat{\theta}) = \hat{k}_i^{wPP}(t;\hat{\theta}_G) - \hat{k}_i^{C}(t;\hat{\theta}) + \hat{k}_i^{OR}(t;\hat{\theta}_O), \tag{A.1}
$$

where each term of the right-hand side above equation is given as follows:

The first term of (A.1).

$$\begin{aligned}\hat{k}_i^{wPP}(t;\theta_G) &= \int_0^t \frac{\hat{\Theta}_{Gi}(u;\beta_G)^T}{\overline{Y}^w(u;\theta_G)}\left\{d\overline{H}'(u;\theta_G) - \overline{Y}^{w\prime}(u;\theta_G)d\hat{\Lambda}_E^{wPP}(u;\theta_G)\right\} \\ &+ \int_0^t \frac{1}{\overline{Y}^w(u;\theta_G)}\left\{dH_i(u;\theta_G) - Y_i^w(u;\theta_G)d\hat{\Lambda}_E^{wPP}(u;\theta_G)\right\},\end{aligned}$$

where

$$H_i(t;\theta_G) = \int_0^t \frac{dN_i(u) - Y_i(u)d\Lambda_P(u|Z_i)}{\exp\{-\underline{\Lambda}_G(u)e^{\beta_G^T Z_i}\}S_P(u|Z_i)},\ \overline{H}(t;\theta_G) = n^{-1}\sum_{i=1}^n H_i(t;\theta_G),$$

$$\overline{H}'(t;\theta_G) = n^{-1}\sum_{i=1}^n \int_0^t \frac{-\frac{\partial}{\partial\theta_G(u)}\exp\{-\underline{\Lambda}_G(u)e^{\beta_G^T Z_i}\}}{\exp\{-\underline{\Lambda}_G(u)e^{\beta_G^T Z_i}\}}dH_i(u;\theta_G),$$

$$Y_i^w(t;\theta_G) = \frac{Y_i(t)}{\exp\{-\underline{\Lambda}_G(t)e^{\beta_G^T Z_i}\}S_P(t|Z_i)},\ \overline{Y}^w(t;\theta_G) = n^{-1}\sum_{i=1}^n Y_i^w(t;\theta_G),$$

and

$$\overline{Y}^{w\prime}(t;\theta_G) = \frac{\partial}{\partial\theta_G(t)}\overline{Y}^w(t;\theta_G).$$

The second term of (A.1).

$$\begin{aligned}\hat{k}_i^C(t;\theta) &= \int_0^t \frac{\hat{\Theta}_i(u;\beta_G,\beta_O)^T}{\overline{R}(u;\theta)}\left\{d\overline{B}^{(1)}(u;\theta) - \overline{R}'(u;\theta)d\hat{\Gamma}(u;\theta)\right\} \\ &+ \int_0^t \frac{1}{\overline{R}(u;\theta)}\left[d\{\hat{\Theta}_{Oi}(u;\beta_O)^T\overline{B}^{(2)}(u;\theta)\} - \hat{\Theta}_{Oi}(u;\beta_O)^T d\overline{B}^{(3)}(u;\theta)\right] \\ &+ \int_0^t \frac{1}{\overline{R}(u;\theta)}\left\{dB(u|Z_i;\theta) - R(u|Z_i;\theta)d\hat{\Gamma}(u;\theta)\right\},\end{aligned}$$

where

$$S_E(t|Z;\theta_O) = \frac{\exp\{-\underline{\Lambda}_O(t)e^{\beta_O^T Z}\}}{S_P(t|Z)},\ \Lambda_E(t|Z;\theta_O) = \underline{\Lambda}_O(t)e^{\beta_O^T Z} - \Lambda_P(t|Z),$$

$$\Lambda_E'(t|Z;\theta_O) = \frac{\partial}{\partial\theta_O(t)}\Lambda_E(t|Z;\theta_O),$$

$$R(t|Z;\theta) = \frac{I(G \geqslant t)}{\exp\{-\underline{\Lambda}_G(t)e^{\beta_G^T Z}\}}S_E(t|Z;\theta_O),\ \overline{R}(t;\theta) = n^{-1}\sum_{i=1}^n R(t|Z_i;\theta),$$

$$R'(t|Z;\theta) = \frac{\partial}{\partial\theta(t)}R(t|Z;\theta),\ \overline{R}'(t;\theta) = n^{-1}\sum_{i=1}^n R'(t|Z_i;\theta),$$

$$B(t|Z;\theta) = \int_0^t R(u|Z;\theta) d\Lambda_E(u|Z;\theta_O),\ \overline{B}(t;\theta) = n^{-1}\sum_{i=1}^n B(t|Z_i;\theta),$$

$$\overline{B}^{(1)}(t;\theta) = n^{-1}\sum_{i=1}^n \int_0^t R'(u|Z_i;\theta) d\Lambda_E(u|Z_i;\theta_O),$$

$$\overline{B}^{(2)}(t;\theta) = n^{-1}\sum_{i=1}^n \Lambda'_E(t|Z_i;\theta_O) R(t+|Z_i;\theta),$$

and

$$\overline{B}^{(3)}(t;\theta) \ = \ n^{-1}\sum_{i=1}^n \int_0^t \Lambda'_E(u|Z_i;\theta_O) dR(u+|Z_i;\theta).$$

The third term of (A.1).

$$\begin{aligned}
\hat{k}_i^{OR}(t;\theta_O) &= \int_0^t \frac{\hat{\Theta}_{Oi}(u;\beta_O)^T}{\overline{S}_E(u;\theta_O)} \left\{ d\overline{F}^{(1)}(u;\theta_O) - \overline{S}'_E(u;\theta_O) d\hat{\Lambda}_E^{OR}(u;\theta_O) \right\} \\
&+ \int_0^t \frac{1}{\overline{S}_E(u;\theta_O)} \left[d\{\hat{\Theta}_{Oi}(u;\beta_O)^T \overline{F}^{(2)}(u;\theta_O)\} - \hat{\Theta}_{Oi}(u-;\beta_O)^T d\overline{F}^{(3)}(u;\theta_O) \right] \\
&+ \int_0^t \frac{S_E(u|Z_i;\theta_O)}{\overline{S}_E(u;\theta_O)} \left\{ d\Lambda_E(u|Z_i;\theta_O) - d\hat{\Lambda}_E^{OR}(u;\theta_O) \right\}.
\end{aligned}$$

where

$$S'_E(t|Z;\theta_O) = \frac{\partial}{\partial\theta_O(t)} S_E(t|Z;\theta_O),\ \overline{S}'_E(t;\theta_O) = n^{-1}\sum_{i=1}^n S'_E(t|Z_i;\theta_O),$$

$$\overline{S}_E(t;\theta_O) = n^{-1}\sum_{i=1}^n S_E(t|Z_i;\theta_O),$$

$$\overline{F}^{(1)}(t;\theta_O) = n^{-1}\sum_{i=1}^n \int_0^t S'_E(u|Z_i;\theta_O) d\Lambda_E(u|Z_i;\theta_O),$$

$$\overline{F}^{(2)}(t;\theta_O) = n^{-1}\sum_{i=1}^n \Lambda'_E(t|Z_i;\theta_O) S_E(t|Z_i;\theta_O),$$

and

$$\overline{F}^{(3)}(t;\theta_O) = n^{-1}\sum_{i=1}^n \int_0^t \Lambda'_E(u-|Z_i;\theta_O) dS_E(u|Z_i;\theta_O).$$

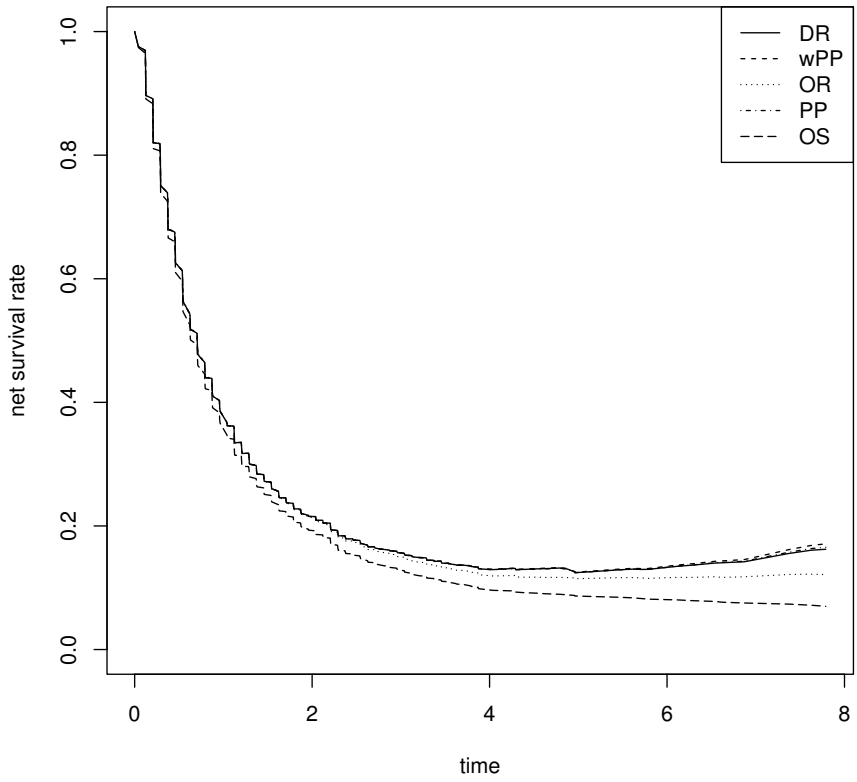

Figure 1. The net survival curves estimated by the DR, wPP, OR and PP estimators and the overall survival (OS) curve.

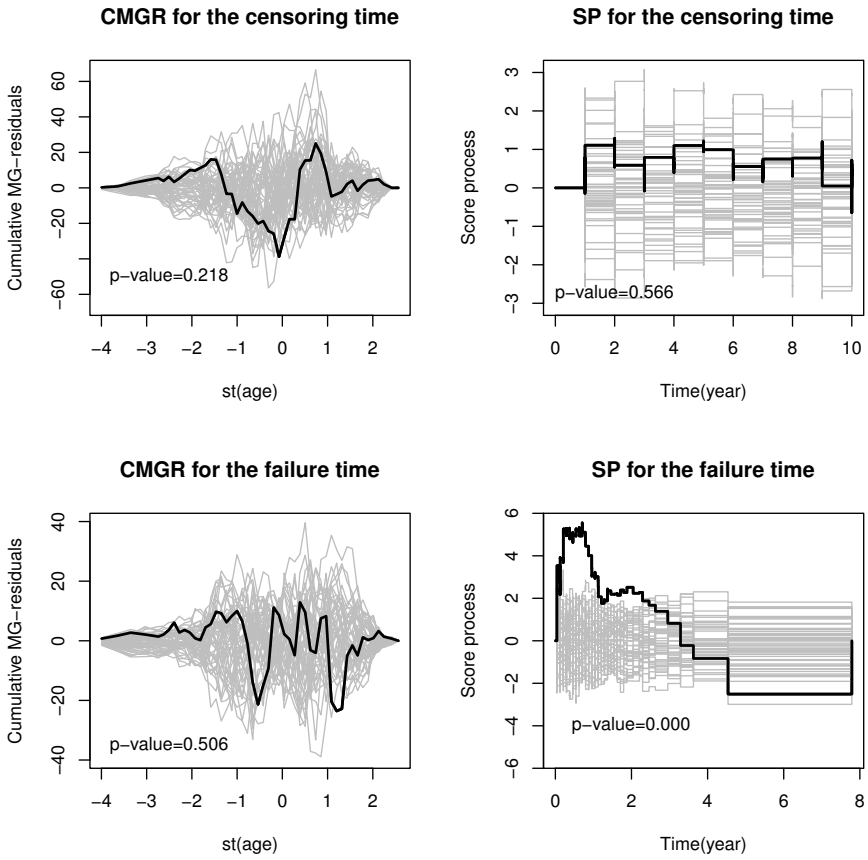

Figure 2. Plot of the cumulative Martingale residuals (CMGR) versus $st(age)$ in the left panels and the score processes (SP) versus time for $st(age)$ in the right panels with their 50 simulated realizations from the null distributions; the upper panels show the censoring time and the lower panels show the failure time; p-values pertain to the supremum-type tests.

Table 1

Results of Dataset 1 *in the simulation study: MSE implies mean squard error* ($\times 10^3$) *and CP implies empirical coverage probability(%) for a 95% nominal level.*

				5-year net survival rate True = 0.521			7-year net survival rate True = 0.434			10-year net survival rate True = 0.338		
Analysis	S_G	S_O	Method	Average	MSE	CP	Average	MSE	CP	Average	MSE	CP
1	True	True	DR	0.521	0.330	96.2	0.433	0.446	95.2	0.336	0.617	96.4
			wPP	0.520	0.336	96.5	0.432	0.459	96.1	0.336	0.649	95.1
			OR	0.522	0.316	96.4	0.434	0.413	95.9	0.336	0.525	96.0
			PP	0.498	0.857	83.3	0.403	1.395	74.5	0.299	2.054	72.0
2	False	True	DR	0.521	0.332	96.9	0.433	0.449	96.5	0.336	0.605	97.0
			wPP	0.507	0.548	91.9	0.415	0.828	87.7	0.314	1.188	87.8
			OR	0.522	0.316	96.4	0.434	0.413	95.9	0.336	0.525	96.0
			PP	0.498	0.857	83.3	0.403	1.395	74.5	0.299	2.054	72.0
3	True	False	DR	0.520	0.337	96.4	0.432	0.460	95.6	0.334	0.646	95.5
			wPP	0.520	0.336	96.5	0.432	0.459	96.1	0.336	0.649	94.9
			OR	0.504	0.587	89.1	0.420	0.588	93.1	0.330	0.566	95.6
			PP	0.498	0.857	83.3	0.403	1.395	74.5	0.299	2.054	72.0
4	False	False	DR	0.505	0.605	89.9	0.409	1.124	81.2	0.298	2.223	68.4
			wPP	0.507	0.548	90.8	0.415	0.828	87.6	0.314	1.188	87.9
			OR	0.504	0.587	89.1	0.420	0.588	93.1	0.330	0.566	95.6
			PP	0.498	0.857	83.3	0.403	1.395	74.5	0.299	2.054	72.0

Table 2

Results of Dataset 2 in the simulation study: MSE implies mean squard error ($\times 10^3$) and CP implies empirical coverage probability(%) for a 95% nominal level.

				5-year net survival rate True = 0.632			7-year net survival rate True = 0.551			10-year net survival rate True = 0.459		
Analysis	S_G	S_O	Method	Average	MSE	CP	Average	MSE	CP	Average	MSE	CP
1	True	True	DR	0.632	0.335	97.0	0.549	0.427	96.3	0.456	0.635	95.7
			wPP	0.631	0.345	97.3	0.549	0.439	96.5	0.458	0.667	96.1
			OR	0.632	0.326	97.4	0.549	0.400	95.5	0.456	0.584	95.3
			PP	0.610	0.860	84.3	0.519	1.465	80.1	0.418	2.405	74.1
2	False	True	DR	0.631	0.335	96.7	0.549	0.426	94.9	0.455	0.628	94.5
			wPP	0.629	0.351	97.1	0.549	0.426	96.3	0.466	0.633	96.5
			OR	0.632	0.326	97.4	0.549	0.400	95.5	0.456	0.584	95.3
			PP	0.610	0.860	84.3	0.519	1.465	80.1	0.418	2.405	74.1
3	True	False	DR	0.631	0.335	97.2	0.549	0.430	96.3	0.455	0.649	96.2
			wPP	0.631	0.345	97.5	0.549	0.439	96.3	0.458	0.667	96
			OR	0.628	0.350	96.0	0.543	0.472	94.9	0.445	0.805	92.2
			PP	0.610	0.860	84.3	0.519	1.465	80.1	0.418	2.405	74.1
4	False	False	DR	0.630	0.341	97.0	0.546	0.446	94.9	0.450	0.701	92.8
			wPP	0.629	0.351	97.1	0.549	0.426	96.4	0.466	0.633	96.4
			OR	0.628	0.350	96.0	0.543	0.472	94.9	0.445	0.805	92.2
			PP	0.610	0.860	84.3	0.519	1.465	80.1	0.418	2.405	74.1

Table 3

Results of Dataset 3 in the simulation study: MSE implies mean squard error ($\times 10^3$) and CP implies empirical coverage probability(%) for a 95% nominal level.

				5-year net survival rate True = 0.593			7-year net survival rate True = 0.505			10-year net survival rate True = 0.404		
Analysis	S_G	S_O	Method	Average	MSE	CP	Average	MSE	CP	Average	MSE	CP
1	True	True	DR	0.593	0.347	96.0	0.504	0.468	95.3	0.401	0.671	94.9
			wPP	0.592	0.351	96.0	0.504	0.478	95.6	0.402	0.700	95.0
			OR	0.594	0.339	95.4	0.505	0.432	95.9	0.401	0.591	94.4
			PP	0.574	0.723	88.9	0.479	1.130	84.4	0.372	1.684	84.1
2	False	True	DR	0.592	0.347	96.5	0.503	0.460	96.4	0.401	0.652	95.2
			wPP	0.584	0.443	94.8	0.493	0.595	93.9	0.392	0.795	93.9
			OR	0.594	0.339	95.4	0.505	0.432	95.9	0.401	0.591	94.4
			PP	0.574	0.723	88.9	0.479	1.130	84.4	0.372	1.684	84.1
3	True	False	DR	0.592	0.353	95.8	0.502	0.483	95.2	0.399	0.704	94.4
			wPP	0.592	0.351	96.1	0.504	0.478	95.6	0.402	0.700	95.0
			OR	0.581	0.508	92.9	0.491	0.641	92.2	0.388	0.831	90.8
			PP	0.574	0.723	88.9	0.479	1.130	84.4	0.372	1.684	84.1
4	False	False	DR	0.582	0.485	93.3	0.487	0.808	89.5	0.374	1.608	79.4
			wPP	0.584	0.443	94.7	0.493	0.595	93.8	0.392	0.795	93.8
			OR	0.581	0.508	92.9	0.491	0.641	92.2	0.388	0.831	90.8
			PP	0.574	0.723	88.9	0.479	1.130	84.4	0.372	1.684	84.1

Table 4
Estimates of the 5-year and 7-year net survival rates and their 95% confidence intervals for patients diagnosed with gastric cancer with adjacent organs in cancer registry data in Osaka.

	5-year net survival rate	7-year net survival rate
DR	0.124(0.100,0.148)	0.142(0.090,0.194)
wPP	0.125(0.101,0.150)	0.146(0.091,0.200)
OR	0.115(0.094,0.135)	0.118(0.093,0.143)
PP	0.124(0.102,0.147)	0.143(0.114,0.172)