An inverse probability weighting method for estimating the net benefit in survival analyses in observational studies

Yasutaka Chiba[1]

(1) Clinical Research Center, Kinki University Hospital, Osaka, Japan

CORRESPONDING AUTHOR: *Yasutaka Chiba, PhD, Clinical Research Center, Kinki University Hospital, 377-2, Ohno-higashi, Osaka-sayama, Osaka, 589-8511, Japan, Tel: +81-72-366-0221 - Fax: +81-72-368-1193, E-mail: chibay@med.kindai.ac.jp*

DOI: ***10.2427/12993***
Accepted on December 5, 2018

ABSTRACT

Background: Recently, in the context of randomized trials, a measure on the difference scale called the net benefit was developed for survival analysis. As this measure does not require the assumption of proportional hazards, it is an attractive new measure of the treatment effect to apply instead of the hazard ratio calculated under this assumption. However, no method for estimating it has been presented in observational studies. Therefore, we describe a simple method to estimate the net benefit adjusted for confounding.
Methods: We reviewed a method for estimating the net benefit in a randomized trial and extended it to a method that adjusts for confounding using inverse probability of treatment weights.
Results: We performed Monte Carlo simulations to test the performance of our method. The results show that our method estimated adjusted net benefits in an unbiased manner regardless of whether the assumption of proportional hazards held. In addition, we illustrated our method using data from an observational study evaluating disease-free survival of Ewing's sarcoma patients. Our method yielded an adjusted net benefit of –0.032, whereas an existing method, used to analyze data from randomized trials, yielded an unadjusted net benefit of 0.284.
Conclusions: In observational studies with a time-to-event outcome, the net benefit adjusted for confounding can readily be estimated using inverse probability of treatment weights.

Key words: confounding, net chance of a longer survival, time-to-event outcome

INTRODUCTION

In studies involving a time-to-event outcome, the treatment effect is usually reported as a hazard ratio (HR), which is calculated using the proportional hazard model under the assumption of proportional hazards, under which HR is constant over time. However, if the assumption is violated, the estimated HR is no longer a reliable measure

of the treatment effect because the actual HR changes over time. Therefore, several alternatives to the HR have been discussed [1]. Recently, a measure called the net benefit was developed [2]. As this measure does not require the proportional hazards assumption, it is an attractive new measure instead of the HR. The net benefit is also sometimes termed "the net chance of a longer survival" and is defined as the probability that a random patient in the treatment group survives longer than a random control patient minus the probability of the opposite situation [3]. This directly addresses the net chance of surviving longer with the treatment than without it.

In randomized trials, the net benefit can readily be estimated without making any assumptions, including the assumption of proportional hazards, and it has been discussed from both theoretical and practical perspectives [1-6]. However, no method for estimating it has been presented in observational studies. Therefore, we describe a simple method for estimating the net benefit adjusted for confounding by applying the inverse probability weight (IPW). We perform Monte Carlo simulations to test the performance of our method, and illustrate it using data from an observational study evaluating disease-free survival of Ewing's sarcoma patients.

METHODS

First, we review a method for estimating the net benefit in randomized trials. Next, we extend this to a method for estimating the net benefit adjusted for confounding in observational studies.

Randomized trials

We assume that the numbers of patients in the treatment and control groups are n_1 and n_0, respectively. We denote $x_i=min(x_i^0,u_i)$ as the observable value of the time-to-event outcome for individual i (i=1,...,n_1) in the treatment group, where x_i^0 denotes the survival time and u_i the censoring time. Here, we discuss only the case of right censoring. Similarly, we denote $y_j= min(y_j^0,v_j)$ as the observable value of the time-to-event outcome for individual j (j=1,...,n_0) in the control group, where y_j^0 denotes the survival time and v_j the censoring time.

Let us define the censoring indicators as τ_i (τ_i =1 if $x_i=x_i^0$, and τ_i=0 if $x_i=u_i$) in the treatment group and η_j (η_j=1 if $y_j=y_j^0$, and η_j =0 if $y_j=v_j$) in the control group. Let us further define a pairwise scoring indicator for the pair formed by individual i in the treatment group and individual j in the control group as s_{ij}. The value of s_{ij} is defined as in Table 1 [2, 6]. In Table 1, s_{ij} =1 implies that the survival time of individual i is longer than that of individual j. Conversely, s_{ij} = −1 implies that the survival time of individual i is shorter than that of individual j. (τ_i,η_j)=(1,1) implies that both survival times of individuals j and i were observed. Therefore, s_{ij}=0 in the cell with (τ_j,η_j) =(1,1) and $x_i=y_j$ implies that the survival times of individuals j and i are equal. However, s_{ij}=0 in the other cells implies that it is not clear whether the survival time of individual i is longer than that of individual j, because the follow-up of both or either individual i and/or j was censored. This situation is sometimes referred to as uninformative or neutral [6].

Using indicator I_{ij} (I_{ij} =1 if s_{ij}=1, and I_{ij} =0 otherwise), the number of pairs in which a random patient in the treatment group survives longer than a random control patient can be expressed as $\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} I_{ij}$, whereas the total number of pairs can be expressed as $n_1 x n_0$. Therefore, the probability that a random patient in the treatment group survives longer than a random control patient, p_1, is estimated as

$$\widehat{p_1} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} I_{ij}}{n_1 \times n_0}$$

Similarly, using indicator J_{ij} (J_{ij}=1 if s_{ij}= −1, and J_{ij}=0 otherwise), the probability that a random patient in the control group survives longer than a random treatment patient (i.e., the opposite situation), p_0, is estimated as

$$\widehat{p_0} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} J_{ij}}{n_1 \times n_0}$$

Consequently, the net benefit, δ, is estimated as

$$\hat{\delta} = \widehat{p_1} - \widehat{p_0} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} s_{ij}}{n_1 \times n_0}$$

which is interpreted as a 100xδ (%) greater net chance of

TABLE 1. Value of the pairwise scoring indicator s_{ij} in a randomized trial

(τ_i, η_j)	$x_i>y_j$	$x_i<y_j$	$x_i=y_j$
(1, 1)	1	−1	0
(1, 0)	0	−1	0
(0, 1)	1	0	0
(0, 0)	0	0	0

the patient surviving longer with treatment compared with the control. The confidence interval (CI) and *p*-value can be calculated using the bootstrap method.

Observational studies

Here, we assume the positivity [7, 8]; this requires the existence of both exposed and unexposed participants for every combination of the values of observed confounder(s) in the population under study. Formally, if f(C)≠0 then $\Pr(E=1|C=c_k)>0$ for e=0,1, where E indicates exposure status (E=1 if exposed and E=0 if unexposed), C indicates the set of measured confounders, and f(.) is the probability density function. We also assume the conditional exchangeability [7, 8] (i.e., no unmeasured confounder exists after conditioning on C).

To estimate the net benefit adjusted for confounding in observational studies, we apply the IPW using the propensity score $p_k=\Pr(E=1|C=c_k)$ [9], where $k=i$ for patients in the treatment group and k=j for patients in the control group. While various forms of IPW have been proposed [10,11], we recommend using the normalized-stabilized weight [11] that can be expressed as

$$nsw_i \equiv \frac{n_1}{\sum_{i=1}^{n_1}\frac{1}{\Pr(E=1|\mathbf{C}=\mathbf{c}_i)}} \times \frac{1}{\Pr(E=1|\mathbf{C}=\mathbf{c}_i)}$$

for exposed individuals and

$$nsw_j \equiv \frac{n_0}{\sum_{j=1}^{n_0}\frac{1}{\Pr(E=0|\mathbf{C}=\mathbf{c}_j)}} \times \frac{1}{\Pr(E=0|\mathbf{C}=\mathbf{c}_j)}$$

for unexposed individuals. The use of the normalized-stabilized weight guarantees that $\sum_{i=1}^{n_1} nsw_i = n_1$ and $\sum_{j=1}^{n_0} nsw_j = n_0$. When these two equalities do not hold, the variance of the estimated net benefit may be excessively large or small.

In a randomized trial, one individual i in the treatment group and one individual j in the control group form 1x1=1 pair. In an observational study, when we consider the pseudo-population [7] created using the IPW, in which it can be assumed that no confounder exists, we can consider nsw_i individuals in the treatment group and nsw_j individuals in the control group to form nsw_i x nsw_j pairs. Therefore, the net benefit adjusted for confounding can be estimated by applying $s_{ij}^{IPW} = \widehat{nsw_i}\times\widehat{nsw_j}$ rather than s_{ij}=1 and $s_{ij}^{IPW} = -\widehat{nsw_i}\times\widehat{nsw_j}$ instead of s_{ij}= −1 in Table 1; consequently, the adjusted net benefit, δ^{IPW}, is estimated as

$$\widehat{\delta^{IPW}} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} s_{ij}^{IPW}}{n_1\times n_0}$$

The CI and *p*-value can be calculated using the bootstrap method. The adjusted probability that a random patient in the treatment group survives longer than a random control patient, p_1^{IPW}, is estimated as

$$\widehat{p_1^{IPW}} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} I_{ij}^{IPW}}{n_1\times n_0}$$

where $I_{ij}^{IPW}=s_{ij}^{IPW}$ if $s_{ij}^{IPW}>0$, and $I_{ij}^{IPW}=0$ otherwise. Similarly, the adjusted probability of the opposite situation, p_0^{IPW}, is estimated as

$$\widehat{p_0^{IPW}} = \frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_0} J_{ij}^{IPW}}{n_1\times n_0}$$

where $I_{ij}^{IPW}=-s_{ij}^{IPW}$ if $s_{ij}^{IPW}<0$, $I_{ij}^{IPW}=0$ and otherwise.

RESULTS

Monte Carlo simulation

We performed Monte Carlo simulations to test the performance of our method. In the simulations, we wished to determine the true values of the net benefits. However, this was a difficult task, because the adjusted net benefit discussed in this article is a marginal effect. Therefore, we derived a true value distribution using a simulated large dataset consisting of 5,000 patients who have two outcomes of an actual and the counterfactual outcomes. The procedure was as follows.

Step 1 (confounders): We assumed that four confounders existed; these were generated as variables that independently followed normal distributions with mean of zero and unit variance.

Step 2 (treatment group): The true model of the propensity score, $p_k=\Pr(E=1|C=c_k)$, was set as a logistic model

$$\log\frac{p_k}{1-p_k} = -0.2C_{1k} + 0.4C_{2k} + 0.6C_{3k} - 0.8C_{4k}$$

The group for each patient (E_k=1 or 0) was simulated using the Bernoulli distribution with the parameter p_k.

Step 3 (time-to-event outcome): The true hazard, $h_k(t)$, for each patient was set as

$$h_k(t) = \exp\left(\log(\lambda)\times E_k + \log\frac{3}{5}\times C_{1k} + \log\frac{5}{4}\times C_{2k} + \log\frac{5}{3}\right.$$

$$\left.\times C_{3k} + \log\frac{4}{5}\times C_{4k}\right)$$

where λ was the true conditional hazard ratio. We

examined two scenarios each featuring proportional and non-proportional hazards. For the proportional hazards scenario, λ was fixed at $\lambda=2/3$. For the non-proportional hazards scenario, λ was set as $\lambda=2/3$ for half of the patients and as $\lambda=6/5$ for the other half of the patients. The time-to-event outcome of each patent, t_k, was simulated using the exponential distribution with the parameter $h_k(t)$.

Step 4 (counterfactual outcome): We could not directly use the value of λ to determine the true value of the net benefit, because we discuss the net benefit as a marginal effect, estimated by applying the IPW, rather than as a conditional effect. Therefore, we simulated the counterfactual outcome for each patient. For a patient with $E_k=e_k$ ($e_k=0,1$) in Step 2, his or her counterfactual group was $E_k=1-e_k$, and the counterfactual time-to-event outcome $\lambda^{1-2e_k}\times t_k$.

Step 5 (true value): For 1,000 datasets, the net benefit, δ, was estimated using 5,000 simulated patients and their 5,000 counterfactuals, corresponding to 10,000 randomized patients; we thus examined 5,000 x 5,000=25 million pairs to derive the net benefit. The range of 1,000 $\hat{\delta}$ would be narrow, and thus the distribution would be close to the true value of the net benefit.

To test our method, in addition to considering both proportional and non-proportional hazards, we considered two scenarios about the regression model for estimating the propensity score p_k; full model to derive unbiased propensity scores, which was the same as the logistic model in the above Step 2, and reduced model to derive biased propensity scores, in which the fourth confounder, C_4, was removed from the logistic model. We also considered two scenarios about censoring of the time-to-event outcome t_k; no censoring, under which no patient suffered from censoring, and censoring, under which patients suffered from censoring following the Bernoulli distribution with the parameter 0.3. For the total 2x2x2=8 scenarios, 1,000 datasets were simulated; each contained 200 patients. The procedure was the same as the above procedure up to and including Step 3. Rather than proceeding to Steps 4 and 5, we estimated the adjusted net benefit, $\widehat{\delta^{IPW}}$, using our method.

The results of Monte Carlo simulations are displayed as box-whisker plots in Figure 1. Figure 1(a) shows the results of the proportional hazards scenario, and Figure 1(b) the results of the non-proportional hazards scenario. On both horizontal axes, "True" indicates $\hat{\delta}$ and S1-S4 indicates $\widehat{\delta^{IPW}}$ under the following scenarios:

- S1: full model of the propensity score p_k and no censoring of the time-to-event outcome t_k,
- S2: full model of p_k and censoring of t_k,
- S3: reduced model of p_k and no censoring of t_k, and
- S4: reduced model of p_k and censoring of t_k.

The box-whisker plots of Figures 1(a) and 1(b) are similar for the same scenarios. Furthermore, in both Figures 1(a) and 1(b), the net benefit under scenario "S1" is close to that under "True". The results reveal that our method does not require the proportional hazards assumption.

In scenario "S2", the net benefit was underestimated. Even in randomized trials, estimation of the net benefit varies by the censoring pattern imposed on the observations [4]. We note that the simulation of Péron et al. [6] (in the context of randomized trials) showed that the higher the proportion of censored data, the greater the underestimation of net benefit.

The results under scenario "S3" reveal that estimation of the net benefit is biased when the propensity score is biased. Moreover, the results under scenario "S4" reveal

FIGURE 1. Box-whisker plots of δ and δ^{IPW} for 1,000 simulated datasets when (a) proportional hazards hold and (b) proportional hazards do not hold, where "True" indicates δ and S1-S4 indicate (δ^{IPW}) (S1: full model of p_k and no censoring of t_k, S2: full model of p_k and censoring of t_k, S3: reduced model of p_k and no censoring of t_k, S4: reduced model of p_k and censoring of t_k)

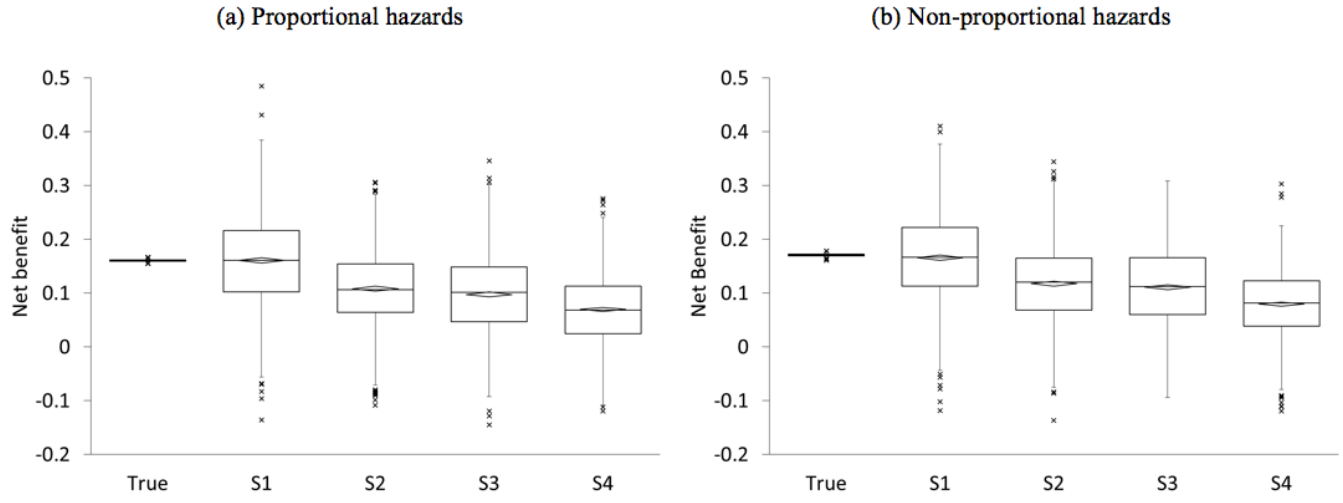


FIGURE 2. (a) Unadjusted disease-free survival curves and (b) inverse probability weighted (adjusted) disease-free survival curves of the treatment and control groups

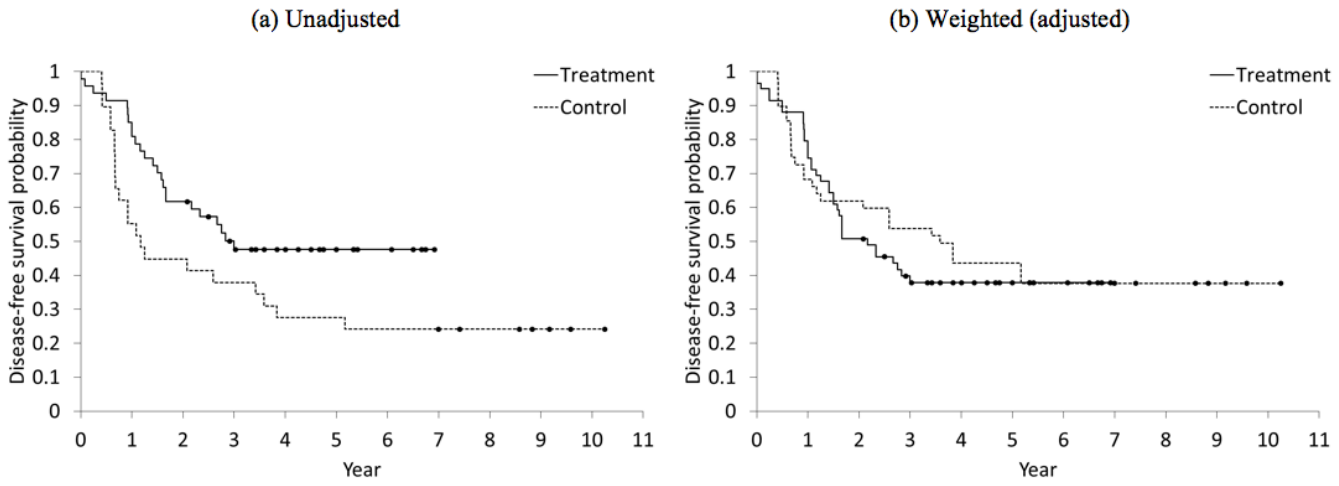
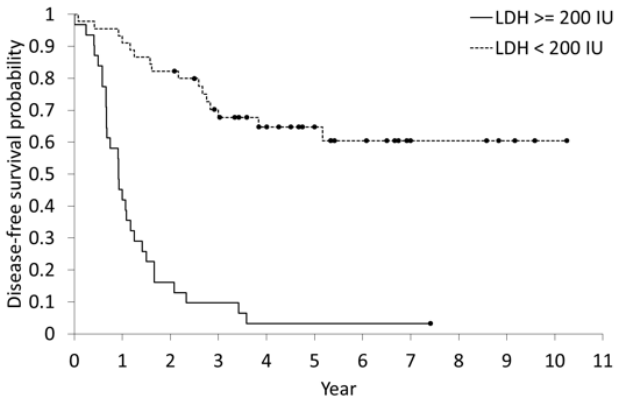


FIGURE 3. Unadjusted disease-free survival curves for patients with LDH levels ≥ 200 IU and < 200 IU



that both censoring of the time-to-event outcome and biased estimation of the propensity score may seriously bias the estimation of net benefit.

Illustration

We applied our method to data from an observational study evaluating the disease-free survival in 76 Ewing's sarcoma patients; the dataset is from Cole and Hernán [12]. Of the 76 patients, 47 received a novel treatment (treatment group), while 29 received one of three standard treatments (control group). Figure 2(a) shows the unadjusted disease-free survival curves for the treatment and control groups. The unadjusted HR was estimated as 0.534 (95% CI: 0.299, 0.955; p = 0.035) using the proportional hazard model. The unadjusted net benefit was estimated as (95% bootstrap CI: 0.027, 0.534; p = 0.031), where $\widehat{p_1}$=0.569 and $\widehat{p_0}$=0.285.

The dataset also includes the serum lactic acid dehydrogenase (LDH) level as a binary confounder (LDH ≥ 200 IU [abnormally high] or LDH < 200 IU). The pre-treatment LDH level was strongly prognostic of recurrence, as seen in Figure 3 (HR = 7.613; 95% CI: 3.995, 14.508). Furthermore, high LDH levels indicated a lower likelihood of assignment to the novel treatment instead of a standard treatment, where the proportions of patients with LDH ≥ 200 IU were 25.5% (12/47) in the treatment group and 65.5% (19/29) in the control group. Therefore, we need to adjust for LDH to estimate the treatment effect in an unbiased manner. As only one binary confounder is in play, the normalized-stabilized weight can be expressed as

$$\frac{n_e}{\sum_{c=0}^{1}\frac{n_e \Pr(C=c|E=e)}{\Pr(E=e|C=c)}} \times \frac{1}{\Pr(E=e|C=c)} = \frac{\Pr(E=e)}{\Pr(E=e|C=c)}$$

where e=1 for patients in the treatment group and e=0 for those in the control group, and where c=1 if the LDH level is ≥ 200 IU and c=0 if the LDH level is < 200 IU. Here, we assume that no other confounder or residual confounding was in play (i.e., the conditional exchangeability assumption held), because we sought to illustrate our method. We also note that the above two percentages (25.5% and 65.5%) imply that the positivity assumption holds.

Figure 2(b) shows the disease-free survival curves adjusted for the LDH levels using normalized-stabilized weights, estimated by applying the above formula. The proportional hazard model with the normalized-stabilized weight yielded an HR of 1.094 (95% bootstrap CI: 0.671, 1.779; p = 0.691), where the bootstrap method was used to calculate the CI and p-value following the

consideration of Austin [13]. However, we consider that the estimated HR is not a reliable measure of the treatment effect, because the two Kaplan-Meier curves of Figure 2(b) cross. Crossing of Kaplan-Meier curves indicates a clear departure from the proportional hazards assumption [14].

Our method yielded a net benefit of $\widehat{\delta^{IPW}}$=–0.032 (95% bootstrap CI: –0.230, 0.201; p = 0.723), where $\widehat{p_1^{IPW}}$=0.395 and $\widehat{p_0^{IPW}}$=0.428. The estimated net benefit is a more reliable measure of the treatment effect in comparison to the HR, because the assumption of proportional hazards is not required to estimate the net benefit, as evident in the above Monte Carlo simulations.

DISCUSSION

To date, the net benefit has been discussed only in the context of randomized trials. However, in observational studies, methods developed to assess randomized trials generally cannot estimate the net benefit in an unbiased manner because of confounding. Here, we have presented a method to estimate the net benefit adjusted for confounding using the IPW. The method is a simple extension of an existing method developed in randomized trials. To estimate the adjusted net benefit using our method, only the effort to derive the IPW is imposed over the method in randomized trials. It will be feasible for researchers to analyze their data using our method.

We have discussed the net benefit defined in the Introduction, which addresses the net chance of surviving longer with treatment than without it. This definition can be generalized to the net benefit by at least m months, which addresses the net chance of surviving for at least m months longer with the treatment than without it [3]. Methods in this article can deal with the general definition simply by replacing $x_i>y_j$, $x_i<y_j$, and $x_i=y_j$ in Table 1 with $x_i-y_j\geq m$, $x_i-y_j\leq -m$, and $|x_i-y_j|<m$, respectively.

Our Monte Carlo simulation showed that the net benefit was underestimated when some patients suffer from censoring of the time-to-event outcome. Similar results were obtained for simulations in the context of randomized trials [6]. In many real-world studies, the net benefit may be underestimated when some patients suffer from the censoring. In both simulations, the time-to-event outcome for each patient was simulated following an exponential distribution, and censored patients were randomly selected. Therefore, more censoring occurred earlier than later. However, in many real-world prospective studies, less censoring occurs earlier rather than later. Thus, compared to our simulation, as real-world studies have fewer uninformative (neutral) situations in calculating the net benefit, the extent of underestimation will be less. As is often remarked, early censoring reflects a problem with study quality, rather than underestimation of a net benefit.

Saad et al. [1] summarized the advantages and disadvantages of several measures that can be used in survival analysis. For example, the restricted mean survival time (RMST) [15], which is the area under the survival curve, is similar to the net benefit in that it does not depend on the assumption of proportional hazards and can be readily interpreted. However, if survival curve lies remote from 0 at the last follow-up time, the RMST may not be readily interpreted, because the entire survival curve to a chosen time is considered. On the other hand, the net benefit lacks this disadvantage. Saad et al. [1] noted only one disadvantage of the net benefit; it was: "Recently proposed, hence little experience."

Traditionally, survival curves and hazard ratios have been reported by studies featuring time-to-event outcomes. In observational studies, it would be appropriate to present weighted (adjusted) survival curves, such as that of Figure 2(b), and the adjusted net benefits that can be estimated using our method.

Funding

This work was supported partially by Grant-in-Aid for Scientific Research (No. 15K00057) from Japan Society for the Promotion of Science.

Acknowledgements

The author thanks Dr. Koji Haratani, who is a medical doctor. His request motivated this work. The author also thanks the reviewers for helpful comments.

References

1. Saad ED, Zalcberg JR, Péron J, et al. Understanding and communicating measures of treatment effect on survival: can we do better?. J Natl Cancer Inst 2018; 110:232-40.
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. Stat Med 2010; 29:3245-57.
3. Péron J, Roy P, Ozenne B, Roche L, Buyse M. The net chance of a longer survival as a patient-oriented measure of treatment benefit in randomized clinical trials. JAMA Oncol 2016; 2:901-5.
4. Latta RB. Generalized Wilcoxon statistics for the two sample problem with censored data. Biometrika 1977; 63:633-5.
5. Péron J, Roy P, Ding K, et al. Assessing the benefit-risk of new treatments using generalized pairwise comparisons: the case of erlotinib in pancreatic cancer. Br J Cancer 2015;112(6):971–976.
6. Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. Stat Methods Med Res 2018; 27:1230-9.
7. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health 2006; 60:578-86.
8. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol 2008; 168:656-64.
9. Rosenbaum PR, Rubin DB. The central role of the propensity score in

observational studies for causal effects. Biometrika 1983; 70:41-55.
10. Hirano K, Imbens GW, Ridder G. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 2003; 71:1161-89.
11. Xiao Y, Abrahamowicz M, Moodie EE. Accuracy of conventional and marginal structural Cox model estimators: a simulation study. Int J Biostat 2010; 6(2):Article 13.
12. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Meth Prog Bio 2004; 75:45-9.
13. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat Med 2016; 35:5642-55.
14. Logan BR, Klein JP, Zhang M. Comparing treatments in the presence of crossing survival curves: an application to bone marrow transplantation. Biometrics 2008; 64:733-40.
15. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol 2014; 32:2380-5.

*