

Adjusting win statistics for dependent censoring

Gaohong Dong¹, Bo Huang², Duolao Wang³, Johan Verbeek⁴,
Jiuzhou Wang⁵ and David C. Hoaglin⁶

Abstract: For composite outcomes whose components can be prioritized on clinical importance, the win ratio, the net benefit and the win odds apply that order in comparing patients pairwise to produce wins and subsequently win proportions. Because these three statistics are derived using the same win proportions and they test the same hypothesis of equal win probabilities in the two treatment groups, we refer to them as *win statistics*. These methods, particularly the win ratio and the net benefit, have received increasing attention in methodological research and in design and analysis of clinical trials. For time-to-event outcomes, however, censoring may introduce bias. Previous work has shown that inverse-probability-of-censoring weighting (IPCW) can correct the win ratio for bias from independent censoring. The present article uses the IPCW approach to adjust win statistics for dependent censoring that can be predicted by baseline covariates and/or time-dependent covariates (producing the CovIPCW-adjusted win statistics). Theoretically and via examples and simulations, we show that the CovIPCW-adjusted win statistics are unbiased estimators of treatment effect in the presence of dependent censoring.

Keywords: Inverse-probability-of-censoring weighting, IPCW, CovIPCW, time-dependent covariates, win ratio, net benefit, win odds

¹ iStats Inc., Long Island City, NY, USA

² Pfizer Inc., Groton, CT, USA

³ Liverpool School of Tropical Medicine, Liverpool, UK

⁴ DSI, I-Biostat, University Hasselt, Hasselt, Belgium

⁵ ImmunoGen Inc., Waltham, MA, USA

⁶ Department of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

1 Introduction

Prioritizing component outcomes in composite endpoints according to their clinical importance has received much attention, especially when the composite endpoint consists of only time-to-event outcomes. Comparisons between treated patients and control patients use the importance order to produce wins for each group (and also ties)¹, following the framework of generalized pairwise comparisons (GPC)². The wins can be conveniently summarized in the win proportions. Because the win ratio (ratio of win proportions)¹, the net benefit (difference in win proportions)² and the win odds (odds of win proportions)³ are derived using the same win proportions and they test the same hypothesis of equal win probabilities in the two treatment groups, we refer to them as *win statistics*. The pioneering work of Finkelstein and Schoenfeld⁴ is equivalent to the test of the difference in the number of wins between the two groups. Verbeek et al.^{5,6} comprehensively compare the win ratio, the net benefit, and other related methods.

The win ratio and the net benefit have received attention in both methodological research and applications. They have been used as the pre-specified primary analysis method in high-profile Phase III clinical studies such as the EMPULSE trial (registration number NCT04157751 in ClinicalTrials.gov) and the ATTR-ACT trial^{7,8}. In EMPULSE (a multicenter, randomized, double-blind study in patients hospitalized for acute heart failure) the primary endpoint is a hierarchical composite of time to death, number of heart-failure events (HFEs), time to the first HFE, and change in KCCQ-CSS (the clinical summary score of the Kansas City cardiomyopathy questionnaire) after 90 days of treatment. ATTR-ACT was a multicenter, international, double-blind, placebo-controlled trial in patients with transthyretin amyloid cardiomyopathy. In the primary analysis, the trialists hierarchically assessed all-cause mortality, followed by frequency of cardiovascular-related hospitalizations^{7,8}. The win ratio, as the primary analysis of the composite

primary endpoint, supported the approval by the US Food and Drug Administration (FDA) of VYNDAQEL® (tafamidis meglumine) and VYNDAMAX™ (tafamidis) for treatment of cardiomyopathy to reduce cardiovascular mortality and cardiovascular-related hospitalization.

One shortcoming of the win ratio is that its calculation ignores ties. In time-to-event data, the ties are mostly due to censored time. To address this problem, Dong et al. formally introduced the win odds³, which divides a tie into two half wins and assigns a half win to each treatment group. This statistic has been discussed by Brunner⁹ and Gasparyan et al.¹⁰, as well as by Peng¹¹ in the setting of non-inferiority clinical trials.

Over the past decade, win statistics have been the subject of rich methodological research¹²⁻²⁵. For time-to-event outcomes, censoring may introduce bias. Methods have been developed to correct for bias in the presence of independent censoring. Péron et al.²³ suggested an extension of Efron's correction²⁶, which reduces the bias for the net benefit. Dong et al.¹⁶ applied inverse-probability-of-censoring weighting (IPCW); they show that the IPCW-adjusted win ratio is an unbiased estimator of treatment effect.

Most work on win statistics focuses on the unmatched approach¹, which compares each patient in the Treatment group with every patient in the Control group, in the setting of independent censoring. Some patient pairs, however, may be influenced by covariates that cause informative censoring. In general, informative censoring may occur when time to event and time to censoring are dependent, either directly or through covariates. When covariates are associated with event time and can predict censoring, this type of censoring is often referred to as *dependent censoring*.

In this article, we adjust win statistics for bias in the presence of dependent censoring, using an IPCW approach based on the Cox model with baseline covariates and/or time-dependent covariates (i.e., the CovIPCW-adjusted win statistics). IPCW compensates for censored patients

by giving more weight to patients with similar characteristics who are not censored^{27,28}. We demonstrate the CovIPCW-adjusted win statistics with examples and simulations.

2 Notation and regularity conditions

2.1 Notation

We consider a randomized clinical trial with N_t patients in the Treatment group and N_c patients in the Control group. Let T denote event time, C denote censoring time, $Y = \min(T, C)$ be the observed time, and $\delta = I(T < C)$ be the event indicator, where $I(\cdot)$ is the indicator function. The p -dimensional covariate vector $Z(x)$ at time x can include baseline covariates at $x = 0$ and time-dependent covariates at $x > 0$. We define $\bar{Z}(x) = \{Z(t), 0 \leq t \leq x\}$ as the covariate history, associated with the events and predicting censoring. We use $i = 1, 2, \dots, N_t$ for patients in the Treatment group and $j = 1, 2, \dots, N_c$ for patients in the Control group. Therefore, the data at times Y_i and Y_j are $(Y_i, \delta_i, \bar{Z}(Y_i))$ and $(Y_j, \delta_j, \bar{Z}(Y_j))$, respectively. We use $F(\cdot)$ and $G(\cdot)$ to denote the survival functions for the event time T and the censoring time C , respectively.

For a composite endpoint of Q time-to-event outcomes with priority order from most important to least important, indexed by $q = 1, 2, \dots, Q$, the $Y_i, \delta_i, Y_j, \delta_j$ are Q -dimensional, and the data are $(Y_{qi}, \delta_{qi}, \bar{Z}(Y_{qi}))$ and $(Y_{qj}, \delta_{qj}, \bar{Z}(Y_{qj}))$, where $Y_{qi} = \min(T_{qi}, C_{qi})$ is the observed time, and $\delta_{qi} = I(T_{qi} < C_{qi})$.

2.2 Regularity conditions

To develop the CovIPCW-adjusted win statistics, we first consider seven regularity conditions, typically described for time-to-event analyses and the IPCW procedure^{27,28,29,30,31}.

- (R1) The values of the covariates $Z(x)$ are known at any time point x in the patient's observation time, and the covariate history $\bar{Z}(x) = \{Z(t), 0 \leq t \leq x\}$ is associated with

the events and predicts censoring. This condition means that event time T and censoring time C are dependent through the covariate history $\bar{Z}(x)$.

(R2) All covariates that might be associated with events and predict censoring are measured. That is, there are ‘no unmeasured confounders for censoring’.

(R3) Conditionally on the covariates $\bar{Z}(x)$, the censoring at a given time x only depends on the data observed up to x and not on future data to be observed as an event or unobserved as censored.

(R4) The survival probability of the censoring time is uniformly bounded away from 0, namely, $G(x|\bar{Z}(x)) > 0$ uniformly in the covariates $\bar{Z}(x)$.

(R5) Any patients alive at the end of the study are considered censored, i.e., $\text{Prob}(C \geq \tau) = \text{Prob}(C = \tau) > 0$, where τ is the analysis time.

(R6) The probability that a patient survives after τ is positive, i.e., $\text{Prob}(T > \tau) > 0$.

(R7) There is a positive probability of observing an event, i.e., $\text{Prob}(T \leq C | \bar{Z}(x)) > 0$.

3 Win statistics

Let π_t be the probability that a patient in the Treatment group wins over a patient in the Control group, π_c be the probability that a patient in the Control group wins over a patient in the Treatment group, and $\pi_{tie} = 1 - \pi_t - \pi_c$ be the probability that the two patients are tied. The win ratio (WR), the win odds (WO), and the net benefit (NB) can be expressed as follows:

$$WR = \frac{\pi_t}{\pi_c}, \tag{1a}$$

$$WO = \frac{\pi_t + 0.5\pi_{tie}}{\pi_c + 0.5\pi_{tie}} = \frac{\pi_t + 0.5(1 - \pi_t - \pi_c)}{\pi_c + 0.5(1 - \pi_t - \pi_c)}. \tag{1b}$$

$$NB = \pi_t - \pi_c, \tag{1c}$$

These three win statistics, the win ratio, the win odds, and the net benefit, are derived using the same win proportions, and they test the same hypothesis that the win probabilities in the treatment groups are equal, namely $H_0: \pi_t = \pi_c$.

4 CovIPCW-adjusted win statistics

For a composite of Q prioritized outcomes, the counting approach is straightforward and intuitive. The principal idea is to compare each patient in a treatment group with every patient in the other group. Within each pair, the comparison starts with the most important outcome, and uses lower-priority outcomes only if higher-priority outcomes have not occurred or result in a tie. Conditional on the covariates $\bar{Z}(x)$, we define the kernel functions K and L as follows: $K_{ij} = 1$ if patient i wins over patient j , otherwise $K_{ij} = 0$; and $L_{ij} = 1$ if patient j wins over patient i , otherwise $L_{ij} = 0$. We use " $Y_{qi} \approx Y_{qj}$ " to denote that patient i and patient j are tied on outcomes 1 through q .

$$K_{ij}(Y_i, \delta_i, Y_j, \delta_j) = I(Y_{1i} > Y_{1j})\delta_{1j} + I(Y_{1i} \approx Y_{1j})I(Y_{2i} > Y_{2j})\delta_{2j} \\ + I(Y_{2i} \approx Y_{2j})I(Y_{3i} > Y_{3j})\delta_{3j} + \cdots + I(Y_{Q-1,i} \approx Y_{Q-1,j})I(Y_{Qi} > Y_{Qj})\delta_{Qj} \quad (2a)$$

$$L_{ij}(Y_i, \delta_i, Y_j, \delta_j) = I(Y_{1j} > Y_{1i})\delta_{1i} + I(Y_{1j} \approx Y_{1i})I(Y_{2j} > Y_{2i})\delta_{2i} \\ + I(Y_{2j} \approx Y_{2i})I(Y_{3j} > Y_{3i})\delta_{3i} + \cdots + I(Y_{Q-1,j} \approx Y_{Q-1,i})I(Y_{Qj} > Y_{Qi})\delta_{Qi} \quad (2b)$$

Dong et al.¹⁶ introduced the IPCW-adjusted win ratio, for which the IPCW-adjusted kernel functions K^A and L^A are

$$K_{ij}^A(Y_i, \delta_i, Y_j, \delta_j) = \frac{I(Y_{1i} > Y_{1j})\delta_{1j}}{G^{(t)}(Y_{1j})G^{(c)}(Y_{1j})} + I(Y_{1i} \approx Y_{1j}) \frac{I(Y_{2i} > Y_{2j})\delta_{2j}}{G^{(t)}(Y_{2j})G^{(c)}(Y_{2j})} \\ + I(Y_{2i} \approx Y_{2j}) \frac{I(Y_{3i} > Y_{3j})\delta_{3j}}{G^{(t)}(Y_{3j})G^{(c)}(Y_{3j})} + \cdots + I(Y_{Q-1,i} \approx Y_{Q-1,j}) \frac{I(Y_{Qi} > Y_{Qj})\delta_{Qj}}{G^{(t)}(Y_{Qj})G^{(c)}(Y_{Qj})}, \quad (3a)$$

$$L_{ij}^A(Y_i, \delta_i, Y_j, \delta_j) = \frac{I(Y_{1j} > Y_{1i})\delta_{1i}}{G^{(t)}(Y_{1i})G^{(c)}(Y_{1i})} + I(Y_{1j} \approx Y_{1i}) \frac{I(Y_{2j} > Y_{2i})\delta_{2i}}{G^{(t)}(Y_{2i})G^{(c)}(Y_{2i})} \\ + I(Y_{2j} \approx Y_{2i}) \frac{I(Y_{3j} > Y_{3i})\delta_{3i}}{G^{(t)}(Y_{3i})G^{(c)}(Y_{3i})} + \cdots + I(Y_{Q-1,j} \approx Y_{Q-1,i}) \frac{I(Y_{Qj} > Y_{Qi})\delta_{Qi}}{G^{(t)}(Y_{Qi})G^{(c)}(Y_{Qi})}, \quad (3b)$$

where $\frac{1}{\hat{G}^{(t)}(\cdot)\hat{G}^{(c)}(\cdot)}$ is an IPCW adjustment. It can be estimated by plugging in the Kaplan-Meier survival estimates $\hat{G}^{(t)}(\cdot)$ and $\hat{G}^{(c)}(\cdot)$ of censoring. Dong et al.¹⁶ showed that the IPCW-adjusted win proportions $P_t^A = \frac{n_t^A}{N_t N_c} = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij}^A$ and $P_c^A = \frac{n_c^A}{N_t N_c} = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} L_{ij}^A$ are unbiased estimators of the win probabilities π_t and π_c for the Treatment and Control groups, in the presence of independent censoring.

In this article, we incorporate baseline covariates and time-dependent covariates in the IPCW adjustment via the time-dependent Cox model. We refer to this adjustment as CovIPCW. Then we apply the CovIPCW adjustment to derive the three win statistics. Conditional on the covariates $\bar{Z}(x)$, the CovIPCW-adjusted kernel functions K^Z and L^Z can be expressed as

$$\begin{aligned} K_{ij}^Z[(Y_i, \delta_i, Y_j, \delta_j) | \bar{Z}(Y_i), \bar{Z}(Y_j)] &= \frac{I(Y_{1i} > Y_{1j})\delta_{1j}}{G^{(t)}(Y_{1j} | \bar{Z}(Y_{1j}))G^{(c)}(Y_{1j} | \bar{Z}(Y_{1j}))} \\ &+ I(Y_{1i} \approx Y_{1j}) \frac{I(Y_{2i} > Y_{2j})\delta_{2j}}{G^{(t)}(Y_{2j} | \bar{Z}(Y_{2j}))G^{(c)}(Y_{2j} | \bar{Z}(Y_{2j}))} \\ &+ \dots + I(Y_{Q-1,i} \approx Y_{Q-1,j}) \frac{I(Y_{Qi} > Y_{Qj})\delta_{Qj}}{G^{(t)}(Y_{Qj} | \bar{Z}(Y_{Qj}))G^{(c)}(Y_{Qj} | \bar{Z}(Y_{Qj}))}, \end{aligned} \quad (4a)$$

$$\begin{aligned} L_{ij}^Z[(Y_i, \delta_i, Y_j, \delta_j) | \bar{Z}(Y_i), \bar{Z}(Y_j)] &= \frac{I(Y_{1j} > Y_{1i})\delta_{1i}}{G^{(t)}(Y_{1i} | \bar{Z}(Y_{1i}))G^{(c)}(Y_{1i} | \bar{Z}(Y_{1i}))} \\ &+ I(Y_{1j} \approx Y_{1i}) \frac{I(Y_{2j} > Y_{2i})\delta_{2i}}{G^{(t)}(Y_{2i} | \bar{Z}(Y_{2i}))G^{(c)}(Y_{2i} | \bar{Z}(Y_{2i}))} \\ &+ \dots + I(Y_{Q-1,j} \approx Y_{Q-1,i}) \frac{I(Y_{Qj} > Y_{Qi})\delta_{Qi}}{G^{(t)}(Y_{Qi} | \bar{Z}(Y_{Qi}))G^{(c)}(Y_{Qi} | \bar{Z}(Y_{Qi}))}. \end{aligned} \quad (4b)$$

When $\bar{Z}(x)$ is multidimensional, there is no nonparametric approach to estimate survival functions³¹, analogous to the simple and popular Kaplan-Meier estimator. As detailed in Willems et al.²⁹ for the IPCW procedure, we estimate $G^{(t)}(\cdot | \bar{Z}(x))$ and $G^{(c)}(\cdot | \bar{Z}(x))$ using the time-dependent Cox model.

The CovIPCW-adjusted numbers of wins and win proportions can be derived as

$$n_t^Z = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij}^Z [(Y_i, \delta_i, Y_j, \delta_j) | \bar{Z}(Y_i), \bar{Z}(Y_j)], \quad (5a)$$

$$n_c^Z = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} L_{ij}^Z [(Y_i, \delta_i, Y_j, \delta_j) | \bar{Z}(Y_i), \bar{Z}(Y_j)], \quad (5b)$$

$$P_t^Z = \frac{n_t^Z}{N_t N_c}, \quad (6a)$$

$$P_c^Z = \frac{n_c^Z}{N_t N_c}. \quad (6b)$$

Therefore, the CovIPCW-adjusted estimators are

$$\widehat{WR}^Z = \frac{P_t^Z}{P_c^Z}, \quad (7a)$$

$$\widehat{NB}^Z = P_t^Z - P_c^Z, \quad (7b)$$

$$\widehat{WO}^Z = \frac{P_t^Z + 0.5(1 - P_t^Z - P_c^Z)}{P_c^Z + 0.5(1 - P_t^Z - P_c^Z)}, \quad (7c)$$

It is straightforward to show that the CovIPCW-adjusted win statistics are unbiased estimators of the win ratio, the win odds and the net benefit, in the presence of dependent censoring, namely,

$$E(\widehat{WR}^Z) = \frac{E(P_t^Z)}{E(P_c^Z)} = \frac{\pi_t}{\pi_c} = WR, \quad (8a)$$

similarly,

$$E(\widehat{WO}^Z) = WO, \quad (8b)$$

$$E(\widehat{NB}^Z) = NB. \quad (8c)$$

5 Asymptotic variance of CovIPCW-adjusted win statistics

The win proportions P_t^Z and P_c^Z are two-sample U-statistics of degree (1,1) with kernel functions K^Z and L^Z based on the data $(Y_i, \delta_i, \bar{Z}(Y_i))$ and $(Y_j, \delta_j, \bar{Z}(Y_j))$. Consequently, n_t^Z and n_c^Z are asymptotically normal (AN).

$$\begin{pmatrix} n_t^Z \\ n_c^Z \end{pmatrix} \sim AN \left(\begin{bmatrix} \theta_t \\ \theta_c \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & \sigma_{tc} \\ \sigma_{tc} & \sigma_c^2 \end{bmatrix} \right), \quad (9)$$

where $\theta_t = N_t N_c \theta_{KZ}$, $\theta_{KZ} = E\{K_{ij}^Z\}$, $\theta_c = N_t N_c \theta_{LZ}$, and $\theta_{LZ} = E\{L_{ij}^Z\}$. Under the null hypothesis

H_0 , θ_{KZ} , θ_{LZ} , θ_t , and θ_c can be estimated by

$$\hat{\theta}_{KZ} = \hat{\theta}_{LZ} = \frac{n_t^Z + n_c^Z}{2N_t N_c}, \quad (10a)$$

$$\hat{\theta}_t = \hat{\theta}_c = \frac{n_t^Z + n_c^Z}{2}, \quad (10b)$$

and σ_t^2 , σ_c^2 and σ_{tc} can be estimated by

$$\hat{\sigma}_t^2 = \frac{N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} [K_{ij}^Z - \hat{\theta}_{KZ}] [K_{ij'}^Z - \hat{\theta}_{KZ}] + \frac{N_t}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} [K_{ij}^Z - \hat{\theta}_{KZ}] [K_{i'j}^Z - \hat{\theta}_{KZ}], \quad (11a)$$

$$\hat{\sigma}_c^2 = \frac{N_t}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} [L_{ij}^Z - \hat{\theta}_{LZ}] [L_{i'j}^Z - \hat{\theta}_{LZ}] + \frac{N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} [L_{ij}^Z - \hat{\theta}_{LZ}] [L_{ij'}^Z - \hat{\theta}_{LZ}], \quad (11b)$$

$$\hat{\sigma}_{tc} = \frac{N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} [K_{ij}^Z - \hat{\theta}_{KZ}] [L_{ij'}^Z - \hat{\theta}_{LZ}] + \frac{N_t}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} [K_{ij}^Z - \hat{\theta}_{KZ}] [L_{i'j}^Z - \hat{\theta}_{LZ}]. \quad (11c)$$

By the delta method, $\log(\widehat{WR}^Z) = \log(P_t^Z / P_c^Z)$, $\log(\widehat{WO}^Z)$ and $\widehat{NB}^Z = P_t^Z - P_c^Z$ are asymptotically normally distributed:

$$\log(\widehat{WR}^Z) \sim AN(\theta_{\log(WR)}, \sigma_{\log(WR)}^2), \quad (12a)$$

$$\log(\widehat{WO}^Z) \sim AN(\theta_{\log(WO)}, \sigma_{\log(WO)}^2), \quad (12b)$$

$$\widehat{NB}^Z \sim AN(\theta_{NB}, \sigma_{NB}^2). \quad (12c)$$

$\sigma_{\log(WR)}^2$, $\sigma_{\log(WO)}^2$ and σ_{NB}^2 can be estimated by

$$\hat{\sigma}_{\log(\widehat{WR}^Z)}^2 = \frac{\hat{\sigma}_t^2}{(\hat{\theta}_t)^2} + \frac{\hat{\sigma}_c^2}{(\hat{\theta}_c)^2} - \frac{2\hat{\sigma}_{tc}}{\hat{\theta}_t \hat{\theta}_c}, \quad (13a)$$

$$\hat{\sigma}_{\log(\widehat{WO}^Z)}^2 = (\hat{\sigma}_t^2 - 2\hat{\sigma}_{tc} + \hat{\sigma}_c^2) \left(\frac{1}{\hat{\gamma}} + \frac{1}{N_t N_c - \hat{\gamma}} \right)^2, \quad (13b)$$

$$\hat{\sigma}_{\widehat{NE}^Z}^2 = \frac{\hat{\sigma}_t^2 + \hat{\sigma}_c^2 - 2\hat{\sigma}_{tc}}{(N_t N_c)^2}, \quad (13c)$$

where $\hat{\gamma} = \hat{\theta}_t + 0.5(N_t N_c - \hat{\theta}_t - \hat{\theta}_c)$.

6 Examples

6.1 Cardiovascular trial of Candesartan in Heart Failure Assessment of Reduction in Mortality and Morbidity (CHARM)

The CHARM trial was a randomized, double-blind, controlled trial comparing candesartan with placebo in patients with chronic heart failure³². The primary outcome was cardiovascular death or hospitalizations due to chronic heart failure. In this outcome, cardiovascular death is clinically more important than hospitalizations due to chronic heart failure, and win statistics take this importance order into account. In contrast, the traditional survival analysis methods such as the Kaplan-Meier method, the log-rank test and the Cox proportional hazards model, use the time to the first occurrence of cardiovascular death or hospitalization due to chronic heart failure, and the more frequent outcome of hospitalization could dominate the result.

In this analysis, we explored all important predictors of mortality and morbidity identified in a previous investigation in the CHARM program³³. These variables were age (per 10 years over age 60 years), left ventricular ejection fraction (per 5 percentage-point decrease below 45%), diabetes mellitus (none [reference], insulin treated, and oral therapy or diet only), male sex, New York Heart Association class (II [reference], III, and IV), current smoking, bundle-branch block, cardiomegaly, previous hospitalization for HF (none [reference], ≤6 months, >6 months), diastolic blood pressure (per 10-mm Hg decrease), duration of HF >2 years, previous myocardial infarction, edema, heart rate (per 10-bpm increase), pulmonary crackles, pulmonary edema, mitral regurgitation, atrial fibrillation, and rest dyspnea.

Table 1 presents the results from the win statistics analyses. Win statistics from the IPCW-adjusted and CovIPCW-adjusted analyses produced very similar point and interval estimates. This is because overall the various baseline covariates did not predict censoring in this large randomized study with a total of 7599 patients. Hence, the CovIPCW estimated from the Cox model with these

covariates does not differ much from the corresponding IPCW based on the Kaplan-Meier method. As expected, the unadjusted win proportions are lower than the IPCW-adjusted and CovIPCW-adjusted win proportions. However, the unadjusted win statistics differ only slightly from the IPCW-adjusted and CovIPCW-adjusted analyses since the hazard of cardiovascular death as well as that of hospitalizations due to chronic heart failure were in general proportional between the two treatment groups. This finding is consistent with Oakes²², Finkelstein and Schoenfeld¹⁷, and Dong et al¹⁵ on win statistics under the proportional hazards model.

6.2 Bone marrow transplant

The bone marrow transplant dataset is a classic example in survival analysis from Klein and Moeschberger³⁴. The study compares disease-free survival after bone marrow transplant in leukemia risk groups. In this example, we compared the high-risk acute myelocytic leukemia (AML) patients (n=45) with the acute lymphoblastic leukemia (ALL) patients (n=38). To mimic the clinical trial setting, we analyzed the data up to 1 year. We excluded 1 ALL patient who was censored early, so that the analysis dataset has no censored observations at 1 year, and we considered the estimates in absence of censoring as the “true” values. The Cox model gave very small survival estimates of censoring times at some covariate levels, because the small sample led to a small risk set. Therefore, we triplicated the dataset, so that we have 111 patients in the ALL group and 135 patients in the high-risk AML group.

In the multivariate Cox regression model, patient age and time to platelet recovery (time-dependent) are significantly associated (at the .05 level) with the survival outcome (Table 2). Therefore, we artificially generated dependent censoring based on these covariates to evaluate the three win statistics. We applied two scenarios of censoring. For each scenario, we generated 1000 simulation datasets. Because we triplicated the dataset, removed a patient and then imposed

artificial censoring, one should not interpret this example as actual clinical trial results; it serves only as an illustration of our proposed methods.

- Scenario 1: Rather than using patient age ranging from 7 to 52 years, we generated dependent censoring based on the square root of patient age since the square root narrows the range in values. As detailed in Willems et al.²⁹, we artificially generated simulation datasets as follows.
 - a) Determine the hazards for censoring, $h_c(x|Z)$, for each patient in the dataset depending on the square root of patient age (Z), by the Cox model $h_c(x|Z) = h_{0c}(x)\exp(\beta Z)$, where $h_{0c}(x)$ is the baseline hazard and β is the regression coefficient for Z .
 - b) Sample the censoring times, C , from an exponential distribution, with rates $h_c(x|Z)$.
 - c) For each patient, determine the observed time $Y = \min(T, C)$ and the event indicator $\delta = I(T < C)$.

We set $\beta = -1.18$ and -1.42 corresponding to 20% and 40% of the patients being censored in the first year, respectively. This censoring led to younger patients having higher probability of being censored.

- Scenario 2: The second scenario induces time-dependent censoring using the variable, time to platelet recovery. In this scenario, 20% and 40% of the patients who achieved platelet recovery were artificially censored by the first year. We did not censor any patients who did not achieve platelet recovery, as the number of these patients was small. We set the censoring time to be uniformly distributed between the time of platelet recovery and the relapse or death event or the cutoff time at 1 year.

The results are shown in Table 3 and Table 4. In both censoring scenarios, the unadjusted simulation results (win probabilities, win ratio, win odds, and net benefit) differed substantially

from the true values (in the absence of censoring), indicating the impact of censoring on win statistics and the potential bias introduced by dependent censoring. As expected, the impact was exacerbated by higher censoring (40% vs 20%).

The IPCW-adjusted method alleviates the issue caused by heavy censoring. The win probabilities are closer to the truth, leading to win statistics closer to the true values in the absence of censoring. However, the differences between the IPCW-adjusted win statistics and the corresponding true values were still relatively large, presumably because of the bias caused by dependent censoring. In both scenarios, the extent of censoring (20% and 40%) does not seem to have much impact on the win probabilities from the IPCW-adjusted method.

Both the baseline CovIPCW-adjusted (Table 3) and time-dependent CovIPCW-adjusted (Table 4) methods produced results that are much closer to the true values in the absence of censoring. The win probabilities and win statistics are very similar to the corresponding true values for both 20% censoring and 40% censoring. These results suggest that the CovIPCW-adjusted methods not only address the heavy censoring issue, which was criticized in the literature for the class of generalized comparison methods; they also effectively reduce the bias caused by dependent censoring.

Last, since the three win statistics are derived with the same win proportions and test the same null hypothesis of equal win probabilities in the two treatment groups, the three tests should theoretically give the same p-values. As our simulations showed, the p-values from the three tests were identical or very close to one another (Table 3 and Table 4).

7 Discussion

Win statistics analyze composite outcomes with different types (e.g., time-to-event, continuous, categorical measurements), and accounts for the relative priority of components (e.g.,

cardiovascular death, hospitalization, stroke). The win ratio and the net benefit have been recently used as the pre-specified primary analysis method in some registered trials. The EMPULSE trial (registration number NCT04157751 in ClinicalTrials.gov) and the ATTR-ACT trial^{7,8} are high-profile examples of Phase III clinical studies.

One shortcoming is that win statistics do not allow adjustments for covariates, although the stratified win ratio, as suggested by Dong et al.¹⁴, can take strata into account. In other words, when the two groups have different patient characteristics at baseline or on time-dependent covariates, the use of win statistics may be questionable. Earlier we applied the IPCW approach to the win ratio¹⁶. Now we have extended it to the win odds and the net benefit, and we have applied covariate adjustments based on the Cox model (i.e., the CovIPCW-adjusted approach) to the three win statistics in the presence of dependent censoring. The IPCW approach deals with tied observations, caused by censored time, by using the predicted probability from Kaplan-Meier estimates, whereas the CovIPCW-adjusted approach compensates for censored patients using the predicted probability from the time-dependent Cox model, giving more weight to patients with similar characteristics who are not censored^{27,28}.

Our examples and simulations show that the CovIPCW-adjusted win statistics effectively correct for bias from dependent censoring. Even under heavy censoring (e.g., 40%), the CovIPCW-adjusted win probabilities and the CovIPCW-adjusted win statistics are very close to the corresponding true values. However, when the sample size is too small, small risk sets at some covariate levels may rule out the CovIPCW adjustment based on the Cox model. This problem is not unique to the CovIPCW-adjusted win statistics; it is a general issue of modelling with covariates in small samples. As indicated in regularity condition R2, the CovIPCW adjustment assumes that all covariates that might be associated with events and predict censoring are

measured. This may be a strong assumption for some studies, but the assumption itself is untestable.

From the perspective of hypothesis testing, the three win statistics have similar performance, as the nearly identical p-values in our simulation results demonstrate. The win ratio and the net benefit have drawn attention in both methodological research and applications, possibly because they have meaningful clinical interpretations and simple statistical calculations. These two methods quantify the relative treatment benefit (win ratio) and absolute treatment benefit (net benefit). For the purpose of estimation in the estimands framework, this coherent dual presentation is an advantage of win statistics over the hazard ratio, even under proportional hazards. The win odds was formally introduced recently by Dong et al³. Because it takes ties into account, the win odds may have an advantage when the number of ties is substantial, in particular in the setting of non-inferiority trials¹¹.

Data availability statement

The bone marrow transplant data that support the findings of this study are available in the textbook: Klein J and Moeschberger M. Survival analysis techniques for censored and truncated data. New York: Springer. 2003.

References

1. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal*. 2012; 33(2):176-182.
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*. 2010; 29(30):3245-3257.

3. Dong G, Hoaglin DC, Qiu J, Matsouaka RA, Chang Y, Wang J, Vandemeulebroecke M. The win ratio: On interpretation and handling of ties. *Statistics in Biopharmaceutical Research*. 2020; 12(1): 99-106.
4. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine*. 1999; 18:1341–1354.
5. Verbeeck J, Spitzer E, de Vries T, van Es GA, Anderson WN, Van Mieghem NM, Leon MB, Molenberghs G, Tijssen J. Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints. *Statistics in Medicine*. 2019; 38(30):5641-5656.
6. Verbeeck J, Ozenne B, Anderson WN. Evaluation of inferential methods for the net benefit and win ratio statistics. *Journal of Biopharmaceutical Statistics*. 2020; 30(5):765-782.
7. Maurer MS, Schwartz JH, Gundapaneni B, et al., ATTR-ACT Study Investigators. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *New England Journal of Medicine*. 2018; 379(11):1007-1016.
8. Pocock SJ, Collier TJ. Statistical appraisal of 6 recent clinical trials in cardiology: JACC state-of-the-art review. *Journal of the American College of Cardiology*. 2019; 73(21):2740-2755.
9. Brunner E. The Success-Odds - A Modified Win-Ratio. 2020. <https://arxiv.org/abs/2002.09273>.
10. Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch G. Adjusted win ratio with stratification: Calculation methods and interpretation. *Statistical Methods in Medical Research*. 2020; DOI: 10.1177/0962280220942558.
11. Peng L. The use of the win odds in the design of non-inferiority clinical trials. *Journal of Biopharmaceutical Statistics*. 2020; 30(5):941-946.

12. Bebu I, Lachin JM. Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*. 2016; 17(1):178-187.
13. Dong G, Li D, Ballerstedt S, Vandemeulebroecke M. A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components. *Pharmaceutical Statistics*. 2016; 15(5): 430–437.
14. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *Journal of Biopharmaceutical Statistics*. 2018; 28(4):778-796.
15. Dong G, Huang B, Chang Y, Seifu Y, Song J, Hoaglin DC. The win ratio: Impact of censoring and follow-up time and use with non-proportional hazards. *Pharmaceutical Statistics*. 2020; 19(3):168-177.
16. Dong G, Mao L, Huang H, Gamalo-Siebersd M, Wang J, Yu G, Hoaglin DC. The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of right censoring. *Journal of Biopharmaceutical Statistics*. 2020; 30(5): 882-899.
17. Finkelstein DM, Schoenfeld DA. Graphing the win ratio and its components over time. *Statistics in Medicine*. 2019; 38(1):53-61.
18. Luo X, Tian H, Mohanty S, Tsai WY. An alternative approach to confidence interval estimation for the win ratio statistic. *Biometrics*. 2015; 71(1):139-145.
19. Luo X, Qiu J, Bai S, Tian H. Weighted win loss approach for analyzing prioritized outcomes. *Statistics in Medicine*. 2017; 36(15):2452-2465.
20. Luo X, Quan H. Some meaningful weighted log-rank and weighted win loss statistics. *Statistics in Biosciences*. 2020; 12(2): 216–224.
21. Mao L. On the alternative hypotheses for the win ratio. *Biometrics*. 2019; 75(1):347-351.

22. Oakes D. On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika*. 2016; 103(3):742–745.
23. Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons of prioritized outcomes in the presence of censoring. *Statistical Methods in Medical Research*. 2018; 27(4):1230-1239.
24. 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 Oncology*. 2016; 2(7):901-905.
25. Wang D., Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharmaceutical Statistics* 2016; 15(3):238-245.
26. Efron B. The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967; 4:831–853.
27. Robins JM. Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. In: 1993 Proceedings of the Biopharmaceutical Section, Alexandria, Virginia: American Statistical Association. pp. 24–33.
28. Robins JM and Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000; 56:779–788
29. Willems SJW, Schat A, van Noorden MA, Fiocco M. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Statistical Methods in Medical Research*. 2018; 27(2):323–335.
30. Wu J. *Statistical methods for survival trial design: with applications to cancer clinical trials using R*. New York: Chapman and Hall/CRC 2018

31. Lok JJ, Yang S, Sharkey B, Hughes MD. Estimation of the cumulative incidence function under multiple dependent and independent censoring mechanisms. *Lifetime Data Analysis* 2018; 24(2):201-223.
32. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, Michelson EL, Olofsson B, Ostergren J, Yusuf S for the CHARM Investigators and Committees. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 2003; 362:759–766.
33. Pocock SJ, Wang D, Pfeffer MA, Yusuf S, McMurray JJ, Swedberg KB, Ostergren J, Michelson EL, Pieper KS, Granger CB. Predictors of mortality and morbidity in patients with chronic heart failure. *European Heart Journal*. 2006; 27:65–75.
34. Klein J, Moeschberger M. *Survival analysis techniques for censored and truncated data*. New York: Springer. 2003.

Table 1 Win statistics from analyses of the CHARM trial

Method	Win proportion (%)		\overline{WR} (95 % CI)	\overline{WO} (95 % CI)	\overline{NB} (%) (95 % CI)
	Candesartan	Placebo			
Unadjusted	28.1	23.6	1.19 (1.10, 1.29)	1.09 (1.05, 1.14)	4.5 (2.4, 6.6)
IPCW-adjusted	32.6	27.6	1.18 (1.09, 1.28)	1.10 (1.05, 1.16)	5.0 (2.5, 7.5)
Baseline CovIPCW-Adjusted	32.0	27.1	1.18 (1.09, 1.28)	1.10 (1.05, 1.16)	4.9 (2.4, 7.4)

Table 2 Multivariate Cox regression on relapse-free survival

Parameter	Parameter Estimate	Standard Error	Hazard Ratio	P-value
Treatment: ALL vs high risk AML	-0.29	0.20	0.75	0.150
Patient age	0.20	0.07	1.22	0.004
Square root of patient age	-1.98	0.73	0.14	0.007
Time to platelet recovery	-0.66	0.26	0.52	0.013

Table 3 Win statistics from analyses of the bone marrow transplant data with a baseline covariate

Cens-oring (%)	Method	Median win proportion (%)		\overline{WR}		\overline{WO}		\overline{NB} (%)	
		ALL	High-risk AML	Median (95 % CI)	P-value Median	Median (95 % CI)	P-value Median	Median (95 % CI)	P-value Median
0		50.6	28.9	1.75 (1.22, 2.51)	0.002	1.55 (1.17, 2.07)	0.002	21.7 (7.5, 36.0)	0.003
20	Unadjusted	39.6	24.2	1.66 (1.14, 2.43)	0.010	1.37 (1.08, 1.75)	0.011	15.7 (3.9, 27.3)	0.011
	IPCW-adjusted	49.4	30.8	1.61 (1.11, 2.37)	0.012	1.46 (1.09, 1.97)	0.012	18.7 (4.2, 32.6)	0.013
	Baseline CovIPCW-Adjusted	50.7	28.9	1.76 (1.22, 2.59)	0.004	1.56 (1.17, 2.10)	0.004	22.1 (8.0, 35.5)	0.004
40	Unadjusted	29.2	18.2	1.59 (1.06, 2.55)	0.031	1.24 (1.02, 1.54)	0.033	10.8 (1.3, 21.2)	0.033
	IPCW-adjusted	48.8	32.9	1.48 (1.01, 2.32)	0.051	1.38 (1.01, 1.97)	0.051	15.9 (0.3, 32.6)	0.053
	Baseline CovIPCW-Adjusted	50.6	29.0	1.74 (1.15, 2.80)	0.013	1.55 (1.12, 2.24)	0.013	21.5 (5.6, 38.2)	0.014

95% CI (confidence interval) is constructed as the 95% percentile interval (2.5th percentile, 97.5th percentile).

Table 4 Win statistics from analyses of the bone marrow transplant data with a time-dependent covariate

Cens-oring (%)	Method	Median win proportion (%)		\overline{WR}		\overline{WO}		\overline{NB} (%)	
		ALL	High-risk AML	Median (95 % CI)	P-value Median	Median (95 % CI)	P-value Median	Median (95 % CI)	P-value Median
0		50.6	28.9	1.75 (1.22, 2.51)	0.002	1.55 (1.17, 2.07)	0.002	21.7 (7.5, 36.0)	0.003
20	Unadjusted	40.4	22.6	1.78 (1.21, 2.68)	0.004	1.43 (1.13, 1.83)	0.005	17.7 (6.0, 29.3)	0.005
	IPCW-adjusted	48.5	27.8	1.74 (1.19, 2.61)	0.005	1.53 (1.14, 2.05)	0.005	20.7 (6.5, 34.4)	0.005
	Time-dependent CovIPCW-Adjusted	50.1	28.9	1.73 (1.17, 2.60)	0.006	1.54 (1.11, 2.11)	0.006	21.1(5.3, 35.6)	0.006
40	Unadjusted	33.1	18.3	1.82 (1.18, 2.81)	0.006	1.35 (1.08, 1.66)	0.007	15.0 (4.1, 24.9)	0.007
	IPCW-adjusted	46.7	26.8	1.75 (1.15, 2.67)	0.008	1.50 (1.11, 2.02)	0.009	19.9 (5.2, 33.9)	0.009
	Time-dependent CovIPCW-Adjusted	49.9	28.6	1.74 (1.10, 2.72)	0.011	1.51 (1.07, 2.14)	0.012	20.4 (3.7, 36.4)	0.012

95% CI (confidence interval) is constructed as the 95% percentile interval (2.5th percentile, 97.5th percentile).