Published in final edited form as:

J Biopharm Stat. 2020 September 02; 30(5): 882–899. doi:10.1080/10543406.2020.1757692.

The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of right censoring

Gaohong Dong^{a,*}, Lu Mao^b, Bo Huang^c, Margaret Gamalo-Siebers^d, Jiuzhou Wang^e, GuangLei Yu^d, David C. Hoaglin^f

^aiStats Inc., Long Island City, NY, USA

^bDepartment of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA

^cPfizer Inc., Groton, CT, USA

^dEli Lilly & Company, Indianapolis, IN, USA

^eImmunoGen Inc., Waltham, MA, USA

^fDepartment of Population and Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

Abstract

The win ratio method has received much attention in methodological research, ad hoc analyses, and designs of prospective studies. As the primary analysis it supported the approval of tafamidis for treatment of cardiomyopathy to reduce cardiovascular mortality and cardiovascular-related hospitalization. However, its dependence on censoring is a potential shortcoming. In this article, we use inverse-probability-of-censoring weighting (IPCW) in adjusting the win ratio to overcome censoring issues. We consider independent censoring, common censoring across endpoints, and right censoring. We develop an asymptotic variance estimator for the logarithm of the IPCW-adjusted win ratio statistic and evaluate it via simulation. Our simulation studies show that, as the amount of censoring increases, the unadjusted win proportions may decrease greatly. Consequently, the bias of the unadjusted win ratio estimate may increase greatly, producing either an overestimate or an underestimate. We demonstrate theoretically and through simulation that the IPCW-adjusted win ratio statistic gives an unbiased estimate of treatment effect.

Keywords

censoring; hazard ratio; IPCW; inverse-probability-of-censoring weighting; win probability; win proportion; win ratio

*Correspondence to: Gaohong Dong, iStats Inc., Long Island City, NY, USA, Gaohong_Dong@iStats.org.

1 Introduction

Since its introduction by Pocock et al. in 2012, the win ratio method has received much attention in methodological research (Dong et al., 2016, 2018, 2019 and 2020; Luo et al., 2015 and 2017; Bebu and Lachin, 2016; Wang and Pocock, 2016; Oakes, 2016; Finkelstein and Schoenfeld, 2019; Mao, 2019), ad hoc analyses, and designs of prospective studies. In the tafamidis Phase III transthyretin amyloid cardiomyopathy study (ATTR-ACT) (Maurer et al., 2018; Pocock et al., 2019), 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. The win ratio and the net benefit (Buyse, 2010; Péron et al., 2016 and 2018) build on generalized pairwise comparisons (GPC), introduced by Buyse (2010). Verbeek et al. (2019 and 2020) comprehensively compare the win ratio, the net benefit and related methods.

Conventional time-to-first-event analyses typically use the hazard ratio (HR) as the summary measure of treatment effect. If, however, the hazards are not proportional, the HR is problematic because it varies over time. In such situations the overall HR from a Cox proportional-hazards model is difficult to interpret, but the win ratio can be a good alternative because it requires no assumptions besides independent censoring.

Dong et al. (2019) explain that the win ratio can be calculated by two approaches: counting and integral. Censoring has an impact on both approaches. For a single time-to-event outcome, the integral approach plugs in the Kaplan-Meier estimates of the survival functions of the event times to estimate the win ratio. When the study endpoints are prioritized multiple outcomes, the marginal survival distribution of the first-priority outcome and the conditional survival distribution of the second-priority outcome given the first-priority outcome (mostly parametric models) can be plugged in. The integral approach is more robust because censoring is handled via risk sets in Kaplan-Meier estimates or parametric models. However, one challenge is to derive the marginal survival distribution of the first-priority outcome and the conditional survival distribution of the second-priority outcome given the first-priority outcome. The counting approach has the advantages that the calculation is simple, and the approach can incorporate project-specific rules defining wins (or losses) and ties. However, similar to the hazard ratio, the estimated win ratio can be greatly biased if censoring is substantial.

In this article, we propose inverse-probability-of-censoring weighting (IPCW) in an adjusted estimator of the win ratio to overcome censoring issues. The IPCW approach was originally proposed to correct for censoring, particularly dependent censoring (Robins, 1993; Robins and Finkelstein, 2000). It compensates for censored subjects by giving more weight to subjects with similar characteristics who are not censored. IPCW has been applied in such analyses as the concordance index (C-statistic) (Uno et al., 2011; Cheung et al., 2019), restricted mean survival time (Tian, Zhao and Wei, 2014), a cumulative incidence function for competing-risk data (e.g., Fine and Gray, 1999; Lok et al., 2018), and Cox regression with right-truncated data (Vakulenko-Lagun, Mandel and Betensky, 2019).

This is the first article to introduce the IPCW-adjusted win ratio statistic. We use win ratio and win ratio statistic interchangeably. We consider independent censoring, common censoring across endpoints, and right censoring, and we show that the IPCW-adjusted win ratio is unbiased. We illustrate this estimator with simulation studies.

2 IPCW-adjusted win ratio of a single time-to-event outcome

2.1 Unadjusted win ratio

Consider a randomized clinical trial with N_t patients in the Treatment group and N_c patients in the Control group, indicated by the subscripts t and c or the superscripts (t) and (c) , respectively. 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. 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. In the counting approach for the win ratio, if any difference in time is considered meaningful, a win for the Treatment group occurs when an observed time $Y_i^{(t)} = \min(T_i^{(t)}, C_i^{(t)})$ in this group is longer than an event time $T_j^{(c)}$ in the Control group; that is, $Y_i^{(t)} > Y_j^{(c)}$ and $\delta_j^{(c)} = 1$ based on the observed data $(Y_i^{(t)}, \delta_i^{(t)})$ and $(Y_j^{(c)}, \delta_j^{(c)})$. This condition can be equivalently expressed as $\min(T_i^{(t)}, C_i^{(t)}, C_j^{(c)}) > T_j^{(c)}$. Then a kernel function K as $K_{ij} = K[(Y_i^{(t)}, \delta_i^{(t)}), (Y_j^{(c)}, \delta_j^{(c)})] = K[(T_i^{(t)}, C_i^{(t)}), (T_j^{(c)}, C_j^{(c)})]$ can be defined by

$$K_{ij} = 1(\text{win for Treatment patient } i), \text{ if } \min(T_i^{(t)}, C_i^{(t)}, C_j^{(c)}) > T_j^{(c)} \\ = 0, \text{ otherwise.} \quad (1a)$$

Similarly, a kernel function L as $L_{ij} = L[(Y_i^{(t)}, \delta_i^{(t)}), (Y_j^{(c)}, \delta_j^{(c)})] = L[(T_i^{(t)}, C_i^{(t)}), (T_j^{(c)}, C_j^{(c)})]$ can be defined by

$$L_{ij} = 1(\text{win for Control patient } j), \text{ if } \min(T_j^{(c)}, C_j^{(c)}, C_i^{(t)}) > T_i^{(t)} \\ = 0, \text{ otherwise.} \quad (1b)$$

We use the term “kernel” for the functions K and L as they are used below to define win proportions, which are U-statistics with K and L as kernels. One may refer to K and L as winning score functions. Under the assumption of independent censoring (i.e., T and C are independent—regularity condition R1 in Appendix A2), the number of wins (n_t) for the Treatment group can be counted as

$$n_t = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij} = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} I(\min(T_i^{(t)}, C_i^{(t)}, C_j^{(c)}) > T_j^{(c)}) \\ = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} I(T_i^{(t)} > T_j^{(c)}) I(C_i^{(t)} > T_j^{(c)}) I(C_i^{(c)} > T_j^{(c)}), \quad (2)$$

where $I(\cdot)$ is the indicator function. The win proportion (P_t) for the Treatment group can be calculated as

$$P_t = \frac{n_t}{N_t N_c} \quad (3)$$

Similarly, the number of wins (n_c) and win proportion (P_c) for the Control group are

$$n_c = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} L_{ij} = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} I(T_j^{(c)} > T_i^{(t)}) I(C_j^{(c)} > T_i^{(t)}) I(C_i^{(t)} > T_i^{(t)}) \quad (4)$$

$$P_c = \frac{n_c}{N_t N_c} \quad (5)$$

Therefore, the win ratio (WR) can be estimated as

$$\widehat{WR} = \frac{P_t}{P_c}. \quad (6)$$

Appendix A.1 briefly describes the asymptotic normality of n_t and n_c and of P_t and P_c , as well as $\log(\widehat{WR})$.

2.2 IPCW-adjusted win ratio

In this section, we show that P_t and P_c are biased estimators of win probabilities (they are unbiased estimators of $E(K_{ij})$ and $E(L_{ij})$, respectively), whereas the IPCW-adjusted win proportions are unbiased estimators of win probabilities. Consequently, the IPCW-adjusted win ratio is unbiased. Following Dong et al. (2019), the win probability for the Treatment group by time τ is

$$\tilde{\pi}_t(\tau) = \text{Prob}(\min(T_i^{(t)}, C_i^{(t)}, C_j^{(c)}, \tau) > T_j^{(c)}), \quad (7a)$$

and the win proportion, P_t , expressed in (3) is an estimate of this probability.

However, our primary interest is in the event time, T (not the censoring time C), and the probability that a patient in the Treatment group “wins” over a patient in the Control group by time τ is

$$\pi_t(\tau) = \text{Prob}(\min(T_i^{(t)}, \tau) > T_j^{(c)}), \quad (7b)$$

as expressed in Oakes (2016). The “global” win probability $\text{Prob}(T_i^{(t)} > T_j^{(c)})$ is not estimable, as the distributions of the event time beyond τ are not identifiable (regularity condition R4 in Appendix A.2). The time τ can be the maximum length of follow-up, in order to maximize the information gathered from the trial (regularity condition R2, described in Appendix A.2). In the Discussion, we provide some suggestions for choosing τ in practice. Since τ is a constant value (required for the regularity conditions) and we use the observed data $(Y_i^{(t)}, \delta_i^{(t)})$

and $(Y_j^{(c)}, \delta_j^{(c)})$ by time τ to estimate the win ratio and for inference on it, we simplify notation by using $\tilde{\pi}_t$ for $\tilde{\pi}_t(\tau)$ and π_t for $\pi_t(\tau)$, and suppress τ in other similar notations.

Therefore, the win proportion, P_w is an unbiased estimator for $\tilde{\pi}_t$, namely $E(K_{ij})$ by the theory of U-statistics, but it is biased for π_t because the probability in (7a) involves the censoring times, C , for both groups. In the following we show that the IPCW-adjusted win proportion is an unbiased estimator for the win probability in (7b).

Let $F(\cdot)$ and $G(\cdot)$ denote the survival functions for the event time T and the censoring time C , respectively. We assume that C is supported on $[0, \tau]$ with a positive probability mass on τ (regularity condition R2). We use $G^{(t)}(x)$ and $G^{(c)}(x)$ to denote the survival functions for C at time x in the Treatment and Control groups, respectively. $G^{(t)}(x)$ and $G^{(c)}(x)$ are positive almost surely, so they can be used as inverse weights. Under the assumption of independent censoring (i.e., T and C are independent), the expectation of K_{ij} is

$$\begin{aligned}\tilde{\pi}_t &= E(K_{ij}) = E\left\{I\left(\min(T_i^{(t)}, C_i^{(t)}, C_j^{(c)}) > T_j^{(c)}\right)\right\} \\ &= E\left\{I\left(T_i^{(t)} > T_j^{(c)}\right)I\left(C_i^{(t)} > T_j^{(c)}\right)I\left(C_j^{(c)} > T_j^{(c)}\right)\right\} \\ &= \text{Prob}\left(T_i^{(t)} > T_j^{(c)}\right)G^{(t)}\left(T_j^{(c)}\right)G^{(c)}\left(T_j^{(c)}\right).\end{aligned}$$

Therefore, $\frac{K_{ij}}{G^{(t)}(T_j^{(c)})G^{(c)}(T_j^{(c)})}$ is an unbiased estimator for the win probability π_t ; that is,

$$E\left(\frac{K_{ij}}{G^{(t)}(T_j^{(c)})G^{(c)}(T_j^{(c)})}\right) = \pi_t \quad (8a)$$

The Kaplan-Meier estimator is asymptotically unbiased (e.g., its bias converges to zero at an exponential rate as the sample size $n \rightarrow \infty$ (Khan and Shaw, 2016)). In the sequel, we use “unbiased” as shorthand for “asymptotically unbiased”. By replacing the survival probabilities of the censoring times in (8a) with the corresponding Kaplan-Meier estimators $\hat{G}^{(t)}$ and $\hat{G}^{(c)}$, we obtain an unbiased estimator for π_t as

$$E\left(\frac{K_{ij}}{\hat{G}^{(t)}(T_j^{(c)})\hat{G}^{(c)}(T_j^{(c)})}\right) = \pi_t. \quad (8b)$$

$\frac{1}{\hat{G}^{(t)}(T_j^{(c)})}$ and $\frac{1}{\hat{G}^{(c)}(T_j^{(c)})}$ are inverse-probability-of-censoring weights.

Now we use superscript A to denote adjustment. We define the IPCW-adjusted kernel functions as

$$K_{ij}^A = \frac{K_{ij}}{\hat{G}^{(t)}(T_j^{(c)})\hat{G}^{(c)}(T_j^{(c)})} \quad (9a)$$

$$L_{ij}^A = \frac{L_{ij}}{\hat{G}^{(t)}(T_i^{(t)})\hat{G}^{(c)}(T_i^{(t)})} \quad (9b)$$

Then the IPCW-adjusted numbers of wins and win proportions are

$$n_t^A = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij}^A \quad (10a)$$

$$n_c^A = \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} L_{ij}^A \quad (10b)$$

$$P_t^A = \frac{n_t^A}{N_t N_c} \quad (10c)$$

$$P_c^A = \frac{n_c^A}{N_t N_c} \quad (10d)$$

Finally, the IPCW-adjusted win ratio estimator is

$$\widehat{WR}^A = \frac{P_t^A}{P_c^A}, \quad (11a)$$

and so

$$E(\widehat{WR}^A) = \frac{E(P_t^A)}{E(P_c^A)} = \frac{\pi_t}{\pi_c} = WR. \quad (11b)$$

Therefore, the IPCW-adjusted estimator, $\widehat{WR}^A$, is an unbiased estimator for the win ratio, and Kaplan-Meier estimators of survival probabilities of censoring can be plugged into (11a). Under the null hypothesis of equal treatment effect between the two groups, $E(\widehat{WR}^A) = 1$.

2.3 Asymptotic variance estimator for the IPCW-adjusted win ratio

An asymptotic variance estimator for the IPCW-adjusted win ratio can easily be derived from the variance-covariance estimator for the logarithm of the win ratio (Dong et al., 2016 and 2018; Bebu and Lachin, 2016) by replacing the kernel functions K_{ij} and L_{ij} with the IPCW-adjusted kernel functions K_{ij}^A and L_{ij}^A . We provide the formulas in Appendix A.3.

2.4 Illustrative simulation examples

We use three scenarios to illustrate differences between the unadjusted win ratio and the IPCW-adjusted win ratio. Figure 1 shows the three scenarios without censoring.

- Scenario a: The event time $T^{(t)}$ follows an exponential distribution with parameter $\lambda = 0.0693$ in the Treatment group, i.e., $T^{(t)} \sim \text{Exp}(0.0693)$; and $T^{(c)} \sim \text{Exp}(0.1155)$ in the Control group. Therefore, the hazards are proportional with hazard ratio (HR) = 0.6 (win ratio = 1.67).
- Scenario b: Before Month 4, $T \sim \text{Exp}(0.1155)$ in both groups, resulting in HR = 1.0. After Month 4, $T^{(t)} \sim \text{Exp}(0.0693)$, $T^{(c)} \sim \text{Exp}(0.1155)$, and HR = 0.6. This pattern mimics the well-known phenomenon of delayed treatment effect in immune-oncology trials.
- Scenario c: Before Month 2, $T^{(t)} \sim \text{Exp}(0.0693)$, $T^{(c)} \sim \text{Exp}(0.1155)$ and HR = 0.6; between Month 2 and Month 6, $T \sim \text{Exp}(0.0693)$ and HR = 1.0; then after Month 6, $T^{(t)} \sim \text{Exp}(0.0693)$, $T^{(c)} \sim \text{Exp}(0.03465)$, and HR = 2.0. This pattern mimics a situation in which the two survival curves cross.

For each scenario, we consider a hypothetical trial with a sample size of 200 patients per group. For censoring, we use the same distribution in the two groups: less censoring, $\text{Exp}(0.05)$, and more censoring, $\text{Exp}(0.09)$. Following Huang and Kuan (2018), we use piecewise exponential functions to generate 1000 simulated datasets. Table 1 shows the unadjusted and IPCW-adjusted win proportions and win ratios. For all scenarios, as the amount of censoring increases, the win proportions decrease substantially. For example, in scenario a, the win proportion for the Treatment group decreases from 55.5% when there is no censoring to 39.0% when censoring is 29% and to 31.0% when censoring is 44%. This is not surprising since heavy censoring produces more inconclusive comparisons (i.e., ties) and fewer pairs with determinate wins, and consequently the win proportions decrease.

However, since win proportions in both the numerator and the denominator are simultaneously reduced, censoring can cause the unadjusted win ratio to be either underestimated (Scenario b) or overestimated (Scenario c). These scenarios illustrate differences between the unadjusted win ratio and the IPCW-adjusted win ratio, but they do not comprehensively describe situations in which the unadjusted win ratio is overestimated or underestimated. The IPCW-adjusted win ratio can correct for bias from censoring and produce a point estimate almost equal to the win ratio in the absence of censoring (i.e., the first row of Table 1 for each scenario). However, the 95% confidence interval for the IPCW-adjusted win ratio is slightly wider than the corresponding interval when there is no censoring.

- Scenario a: Since the event hazards remain constant for both groups over time (i.e., the proportional-hazards assumption holds), censoring has no impact on the win-ratio estimate (see details in Dong et al., 2019). However, the 95% confidence interval of the unadjusted win ratio becomes wider as the amount of censoring increases. Interestingly, the 95% confidence intervals for the IPCW-adjusted win ratio are narrower than the corresponding unadjusted intervals, but slightly wider than the interval without censoring.

- Scenario b: Since the event hazard for the Treatment group becomes smaller after Month 4, the censoring in the unadjusted analysis has relatively more impact (i.e., reduced win proportions) on the win ratio results after Month 4 than before Month 4. Therefore, because of the delayed effect, the unadjusted win ratio is biased and becomes smaller with more censoring (i.e., the treatment effect is underestimated). In contrast, the IPCW-adjusted win ratio appropriately overcomes the censoring bias, and the point estimate is the same as the win ratio when censoring does not occur. The 95% confidence interval of the IPCW-adjusted win ratio is slightly wider.
- Scenario c: Since the event hazard for the Control group becomes smaller over time, similar to Scenario b, censoring has relatively more impact on the win ratio at later times. Therefore, because of the crossing hazards, the unadjusted win ratio is biased and becomes larger with more censoring (i.e., the treatment effect is overestimated). In contrast, the IPCW-adjusted win ratio corrects for the censoring bias, and the point estimate is almost the same as the win ratio when censoring does not occur. The 95% confidence interval for the IPCW-adjusted win ratio is slightly wider.

3 IPCW-adjusted win ratio of prioritized multiple time-to-event outcomes

3.1 IPCW-adjusted win ratio

Consider a clinical trial in which a composite endpoint has Q outcomes with priority order from most important to least important, $q = 1, 2, 3, \dots, Q$, with event times $T_{qi}^{(t)}$ and $T_{qj}^{(c)}$ and censoring times $C_{qi}^{(t)}$ and $C_{qj}^{(c)}$. As discussed in the previous section, for illustration and simplicity, we consider any difference in time to be meaningful. Each pairwise 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. The kernel functions defined in (1a) and (1b) are extended to the setting of prioritized multiple time-to-event outcomes in (12a) and (12b) below. To ease notation, we omit the transition condition of ties from the higher-priority outcomes when we determine K_{ij} and L_{ij} for the q th outcome.

$$K_{ij} = \begin{cases} 1 & \begin{cases} \text{if } \min(T_{1i}^{(t)}, C_{1i}^{(t)}, C_{1j}^{(c)}) > T_{1j}^{(c)} \\ \text{else if } \min(T_{2i}^{(t)}, C_{2i}^{(t)}, C_{2j}^{(c)}) > T_{2j}^{(c)} \\ \vdots \\ \text{else if } \min(T_{Qi}^{(t)}, C_{Qi}^{(t)}, C_{Qj}^{(c)}) > T_{Qj}^{(c)} \end{cases} \\ 0, & \text{otherwise.} \end{cases} \quad (12a)$$

$$L_{ij} = \begin{cases} 1 & \text{if } \min(T_{1j}^{(c)}, C_{1j}^{(c)}, C_{1i}^{(t)}) > T_{1i}^{(t)} \\ & \text{else if } \min(T_{2j}^{(c)}, C_{2j}^{(c)}, C_{2i}^{(t)}) > T_{2i}^{(t)} \\ & \vdots \\ & \text{else if } \min(T_{Qj}^{(c)}, C_{Qj}^{(c)}, C_{Qi}^{(t)}) > T_{Qi}^{(t)} \\ 0 & \text{otherwise.} \end{cases} \quad (12b)$$

The IPCW-adjusted kernel functions defined in (9a) and (9b) can be extended to

$$K_{ij}^A = \begin{cases} \frac{1}{\hat{G}_1^{(t)}(T_{1j}^{(c)})\hat{G}_1^{(c)}(T_{1j}^{(c)})}, & \text{if } \min(T_{1i}^{(t)}, C_{1i}^{(t)}, C_{1j}^{(c)}) > T_{1j}^{(c)} \\ \frac{1}{\hat{G}_2^{(t)}(T_{2j}^{(c)})\hat{G}_2^{(c)}(T_{2j}^{(c)})}, & \text{else if } \min(T_{2i}^{(t)}, C_{2i}^{(t)}, C_{2j}^{(c)}) > T_{2j}^{(c)} \\ & \vdots \\ \frac{1}{\hat{G}_Q^{(t)}(T_{Qj}^{(c)})\hat{G}_Q^{(c)}(T_{Qj}^{(c)})}, & \text{else if } \min(T_{Qi}^{(t)}, C_{Qi}^{(t)}, C_{Qj}^{(c)}) > T_{Qj}^{(c)} \\ 0 & \text{otherwise;} \end{cases} \quad (13a)$$

$$L_{ij}^A = \begin{cases} \frac{1}{\hat{G}_1^{(t)}(T_{1i}^{(t)})\hat{G}_1^{(c)}(T_{1i}^{(t)})}, & \text{if } \min(T_{1j}^{(c)}, C_{1j}^{(c)}, C_{1i}^{(t)}) > T_{1i}^{(t)} \\ \frac{1}{\hat{G}_2^{(t)}(T_{2i}^{(t)})\hat{G}_2^{(c)}(T_{2i}^{(t)})}, & \text{else if } \min(T_{2j}^{(c)}, C_{2j}^{(c)}, C_{2i}^{(t)}) > T_{2i}^{(t)} \\ & \vdots \\ \frac{1}{\hat{G}_Q^{(t)}(T_{Qi}^{(t)})\hat{G}_Q^{(c)}(T_{Qi}^{(t)})}, & \text{else if } \min(T_{Qj}^{(c)}, C_{Qj}^{(c)}, C_{Qi}^{(t)}) > T_{Qi}^{(t)} \\ 0 & \text{otherwise,} \end{cases} \quad (13b)$$

where $\hat{G}_q(q = 1, 2, \dots, Q)$ is the marginal estimate of the survival function for the censoring time of the q th outcome.

For a single time-to-event outcome, the Kaplan-Meier estimator can be used to estimate the survival function of the censoring time. However, for multivariate time-to-event outcomes, there is no simple analog of the Kaplan-Meier estimator for marginal survival functions. Therefore, we follow Fine and Gray (1999) and assume that (a) the censoring time and event time of all outcomes are independent, and (b) the censoring of all outcomes is the same. Then we can apply the common estimates $\hat{G}^{(t)}$ and $\hat{G}^{(c)}$ to (13a) and (13b) as follows:

$$K_{ij}^A = \begin{cases} \frac{1}{\hat{G}^{(t)}(T_{1j}^{(c)})\hat{G}^{(c)}(T_{1j}^{(c)})}, & \text{if } \min(T_{1i}^{(t)}, C_{1i}^{(t)}, C_{1j}^{(c)}) > T_{1j}^{(c)} \\ \frac{1}{\hat{G}^{(t)}(T_{2j}^{(c)})\hat{G}^{(c)}(T_{2j}^{(c)})}, & \text{else if } \min(T_{2i}^{(t)}, C_{2i}^{(t)}, C_{2j}^{(c)}) > T_{2j}^{(c)} \\ \vdots \\ \frac{1}{\hat{G}^{(t)}(T_{Qj}^{(c)})\hat{G}^{(c)}(T_{Qj}^{(c)})}, & \text{else if } \min(T_{Qi}^{(t)}, C_{Qi}^{(t)}, C_{Qj}^{(c)}) > T_{Qj}^{(c)} \end{cases} \quad (14a)$$

$$= 0, \text{ otherwise;}$$

$$L_{ij}^A = \begin{cases} \frac{1}{\hat{G}^{(t)}(T_{1i}^{(t)})\hat{G}^{(c)}(T_{1i}^{(t)})}, & \text{if } \min(T_{1j}^{(c)}, C_{1j}^{(c)}, C_{1i}^{(t)}) > T_{1i}^{(t)} \\ \frac{1}{\hat{G}^{(t)}(T_{2i}^{(t)})\hat{G}^{(c)}(T_{2i}^{(t)})}, & \text{else if } \min(T_{2j}^{(c)}, C_{2j}^{(c)}, C_{2i}^{(t)}) > T_{2i}^{(t)} \\ \vdots \\ \frac{1}{\hat{G}^{(t)}(T_{Qi}^{(t)})\hat{G}^{(c)}(T_{Qi}^{(t)})}, & \text{else if } \min(T_{Qj}^{(c)}, C_{Qj}^{(c)}, C_{Qi}^{(t)}) > T_{Qi}^{(t)} \end{cases} \quad (14b)$$

$$= 0, \text{ otherwise.}$$

The formulas defined in (10a) through (11) can be used to calculate the IPCW-adjusted numbers of wins, the win proportions, and the win ratio, under the setting of prioritized multiple time-to-event outcomes using the IPCW-adjusted kernel functions K_{ij}^A and L_{ij}^A defined in (14a) and (14b). Alternatively, following the decomposition of win probabilities described by Mao (2019), the IPCW-adjusted win proportion can be expressed as the sum of the adjusted win proportions ($P_{t,q}^A$) for the prioritized multiples outcomes ($q = 1, 2, \dots, Q$):

$$P_t^A = P_{t,1}^A + P_{t,2}^A + \dots + P_{t,Q}^A \quad (15)$$

where $P_{t,1}^A = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \frac{I(\min(T_{1i}^{(t)}, C_{1i}^{(t)}, C_{1j}^{(c)}) > T_{1j}^{(c)})}{\hat{G}^{(t)}(T_{1j}^{(c)})\hat{G}^{(c)}(T_{1j}^{(c)})}$ and

$$P_{t,q}^A = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \frac{I(\min(T_{qi}^{(t)}, C_{qi}^{(t)}, C_{qj}^{(c)}) > T_{qj}^{(c)})}{\hat{G}^{(t)}(T_{qj}^{(c)})\hat{G}^{(c)}(T_{qj}^{(c)})}$$

provided that the comparison between patients i and j is inconclusive (i.e., is a tie) for the higher-priority outcomes. Then

$$E(\widehat{WR}^A) = \frac{E(P_t^A)}{E(P_c^A)} = \frac{E(\sum_{q=1}^Q P_{t,q}^A)}{E(\sum_{q=1}^Q P_{c,q}^A)} = \frac{\pi_t}{\pi_c} = WR \quad (16)$$

under the setting of prioritized multiple time-to-event outcomes. An asymptotic variance estimator for the IPCW-adjusted win ratio can be calculated as described in Section 2.3. We provide the formulas in Appendix A.3.

3.2 Examples

Example 1: Monoclonal gammopathy of undetermined significance—A monoclonal gammopathy of undetermined significance (MGUS) occurs in up to 2 percent of persons 50 years of age or older (Kyle et al., 2002). Among residents of southeastern Minnesota, MGUS was diagnosed in 1384 patients at the Mayo Clinic from 1960 through 1994 and followed through 1999. The dataset is available in the R package *Survival* (<https://cran.r-project.org/web/packages/survival>). We analyzed the composite of progression to a plasma cell malignancy (PCM) or death, with death as the more important outcome. We applied the IPCW-adjusted win ratio to the MGUS data and compared males vs. females. For illustration, we first excluded 409 patients who were censored, leaving 975 uncensored patients for analysis, so that we could estimate the win ratio without bias from censoring. Then, to demonstrate the performance of the IPCW-adjusted win ratio, we artificially applied independent exponentially distributed censoring, $Exp(0.0015)$, $Exp(0.0060)$, and $Exp(0.0135)$, corresponding to 10%, 30% and 50% censoring, respectively, and generated 1000 datasets for each censoring scheme. Because we first removed all censored patients and then imposed artificial censoring, one should not interpret this example and the next as actual clinical trial results; they only illustrate our proposed method.

Figure 2 shows the Kaplan-Meier estimates of the survival functions for the 975 patients with PCM or death outcomes without censoring applied, and Figure 3 shows the Kaplan-Meier estimates of the survival functions for the censoring ($C \sim Exp(0.0135)$) in one simulated dataset. Table 2 summarizes the results, comparing the unadjusted vs. the IPCW-adjusted win ratio without censoring and with censoring at different degrees. Similar to the simulated examples in Section 2.4, as the amount of censoring increases, the unadjusted win proportion decreases, and the confidence intervals for the unadjusted win ratio become wider. In contrast, the IPCW-adjusted win proportion and the IPCW-adjusted win ratio remain stable even under heavy censoring (e.g., 50% censoring). This example further supports the robustness of the IPCW-adjustment approach.

For the IPCW-adjusted win ratio, the 95% confidence interval is narrower than the corresponding unadjusted intervals, but only slightly wider than the interval without censoring (Table 2). This indicates that the IPCW adjustment can greatly reduce the impact of censoring and correct for bias from censoring. In our simulations, the empirical variance is very similar to the median asymptotic variance of the logarithm of the win ratio, and the coverage of 95% confidence intervals of the IPCW-adjusted win ratio is very close to the nominal 95% (Table 3). Therefore, the asymptotic variance estimator for the IPCW-adjusted win ratio described in Appendix A.3 is appropriate. In this example, the biases of both the unadjusted and the IPCW-adjusted win ratios are small.

Example 2: Cardiovascular trial data—For a second example, we extracted data from clinical trials in cardiovascular (CV) disease with the composite of death and hospitalization as the primary endpoint. We selected the first 800 patients, used the data up to 3 years, and excluded patients who dropped out prior to Year 3, so that we could estimate the win ratio in the absence of censoring. For the conventional analysis, which uses the first event in the composite, 67 (60.0%) and 63 (52.5%) of 111 and 120 deaths in the Treatment

group and Control group, respectively, were counted in the composite endpoint (Table 4). In contrast, the win ratio method uses all death information. In Figure 4, the Kaplan-Meier estimates of the survival curves for the composite endpoint show that the hazards in the two groups are nonproportional, without a particular pattern.

To demonstrate the impact of the IPCW adjustment, we artificially applied independent exponentially distributed censoring, $Exp(0.0004)$ and $Exp(0.001)$, corresponding to 25% and 50% censoring, respectively, and generated 1000 datasets for each censoring scheme. Table 5 presents the unadjusted and IPCW-adjusted win ratios. As in the simulated examples in Section 2.4, the IPCW-adjusted win proportion and the IPCW-adjusted win ratio remain stable even under heavy censoring (e.g., 50% censoring). The 95% confidence intervals for the IPCW-adjusted win ratio are narrower than the corresponding unadjusted intervals, but slightly wider than the interval without censoring. As shown in Table 6, the empirical variance is very similar to the median asymptotic variance of the logarithm of the win ratio, and the coverage of 95% confidence intervals for the IPCW-adjusted win ratio is very close to the nominal 95%. However, the coverage of the 95% confidence interval for the unadjusted win ratio is only 84.0% when the censoring is about 50%, far below the nominal level of 95%. As shown in Table 6, the median bias of the unadjusted win ratio is 0.059 and 0.167 when the amount of censoring is 25% and 50%, respectively. It is not surprising that the bias of the unadjusted win ratio may increase greatly as the amount of censoring increases. However, the bias of IPCW-adjusted win ratio is very small (e.g., bias = 0.008 and 0.038 for the scenarios with censoring of 25% and 50%, respectively).

4 Discussion

While the win ratio has been used mainly for study design and for analysis of prioritized multiple time-to-event outcomes, it can be also a useful analytic method for single time-to-event outcomes and time-to-first-event outcomes. In fact, it is a good alternative method when the hazards are not proportional, specifically because it requires no assumptions apart from independent censoring. The method can be implemented through a simple counting approach, which can incorporate project-specific rules defining wins (or losses) and ties that are based on what is deemed clinically meaningful. For example, a two-week difference in time can be meaningful for an outcome in one disease area, but not meaningful for the same outcome in another disease area. For simplicity, we have considered any difference in time meaningful.

Similar to the hazard ratio, the estimated win ratio can be heavily biased if censoring is substantial. In fact, assuming independent censoring, common censoring across endpoints, and right censoring, we have shown that as the amount of censoring increases, the unadjusted win proportions may decrease greatly. Therefore, the win proportions may underestimate the win probabilities, and the unadjusted win ratio can be either an overestimate or an underestimate. To correct for this bias, we propose the IPCW-adjusted win ratio. Theoretical results demonstrate that the IPCW-adjusted win ratio statistic gives an unbiased estimate of treatment effect. This is also supported by the results of simulations in which various patterns of hazard ratios reflect what arises in clinical practice. The results of the simulations also show that, although the 95% confidence interval of the IPCW-adjusted

win ratio is much narrower than that of the unadjusted win ratio, it can be slightly wider than the corresponding interval when there is no censoring.

In this article introducing the IPCW-adjusted win ratio, we consider independent censoring, common censoring across endpoints, and right censoring. For simplicity and ease of exposition, we have not included adjustments for covariates. However, if informative censoring is expected in a clinical trial, adjustment for baseline covariates and time-dependent covariates is usually necessary to reduce the variation in the estimated weights for the uncensored observations. Therefore, we expect that the 95% confidence interval for the IPCW-adjusted win ratio would become narrower when covariate adjustment is incorporated. We look forward to further research on the performance of the IPCW-adjusted win ratio with covariate adjustment in the presence of informative censoring.

As discussed in Howe et al. (2011), IPCW relies on the exchangeability assumption, which implies that, given the measured common predictors of the outcome of interest and censoring, censored subjects have the same prognosis for the outcome of interest as do subjects who are not censored. In scenarios where heavy censoring occurs, especially in the tail of the Kaplan-Meier curves, the number of subjects is small, and the outcomes observed among the subjects who are not censored are unlikely to be representative of the unobserved outcomes among the censored subjects, even if one appropriately measured all common predictors and accounted for them. Further research is warranted on the performance of the IPCW-adjusted win ratio in the presence of heavy censoring.

When the number of patients at risk in the tail of the survival curves of censoring times is too small, the IPCW could be inflated disproportionately (e.g., some weights of IPCW at the tail of the survival curve can be very large, as Robins, Hernan and Brumback (2000) pointed out), and the method could break down. Therefore, one should choose τ so that the number of patients at risk in the tail of the survival curves is not too small. Furthermore, since dependent censoring occurs mostly in early follow-up, it can be analytically feasible to restrict the IPCW analysis to a shorter interval in lieu of the entire follow-up period while still achieving the goal of minimizing bias. The question of choosing the timepoint for analyzing time-to-event outcomes is not unique to the win ratio method. For example, Tian et al. (2020) comprehensively discuss choices of time window for the restricted mean survival time, and Cheung et al. (2019) suggest requiring at least 10% of patients in the risk set to estimate IPCW for the setting of a concordance index. For the IPCW-adjusted win ratio, one may choose the 90th percentile of observed follow-up times (or 10% of patients in the risk set) as the cut-off point for τ , but in general it should be driven by the mechanism of censoring in the study to balance the information discarded after τ and the robustness of the estimation of IPCW toward the upper end of the support of the censoring distribution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The authors thank Stuart J. Pocock and David Oakes for comments of an early version of this article, Junshan Qiu for discussions, and two reviewers for their insightful and constructive comments. Lu Mao's work is supported by NIH grant R01HL149875.

Appendix

A.1 Asymptotic normality of n_t and n_c , and P_t and P_c , as well as unadjusted $\log(\widehat{WR})$.

From equations (1) through (5), the win proportions P_t and P_c are

$$P_t = \frac{n_t}{N_t N_c} = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij},$$

$$P_c = \frac{n_c}{N_t N_c} = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} L_{ij}.$$

P_t and P_c are two-sample U-statistics of degree (1,1) with kernel functions K and L , respectively. From the theory of U-statistics, $E(P_t) = E(K_{ij})$ and $E(P_c) = E(L_{ij})$, and P_t and P_c are asymptotically normal. Then it is straightforward to obtain the asymptotic normality of the numbers of wins n_t and n_c as well as the asymptotic normality of the logarithm of the unadjusted win ratio WR , via the delta method (Dong et al., 2016 and 2018; Bebu and Lachin, 2016). The asymptotic variance estimators for the unadjusted $\log(\widehat{WR})$ are consistent under the null hypothesis of equal treatment effects between the two groups (i.e., $H_0: WR = 1$), based on the theory of U-statistics. The asymptotic variance-covariance estimator of Dong et al. (2016 and 2018) is based on Wei and Johnson (1985).

In principle, U-statistics are generalizations of sample means. The letter “U” stands for unbiased. U-statistics are an efficient way to obtain unbiased estimators. For the theory of U-statistics, we refer to the seminal paper by Hoeffding (1948), Lee (1991), Chapter 12 of van der Vaart (2000), and Chapter 6 of Lehmann (1999).

A.2 Regularity conditions

In developing the IPCW-adjusted win ratio, we use four regularity conditions. In the absence of covariates, these conditions are relatively simple, compared with those typically described for time-to-event analyses, such as in Wu (2018). We use τ to denote the maximum follow-up time. In the Discussion, we provide some suggestions for choosing τ in practice.

(R1) Event time T and censoring time C are independent.

(R2) Any patients alive at the end of the study are considered censored, i.e., $\text{Prob}(C = \tau) = P(C = \tau) > 0$.

(R3) The probability that a patient survives beyond τ is positive, i.e., $\text{Prob}(T > \tau) > 0$.

(R4) There is a positive probability of an event, i.e., $Prob(T = C) > 0$.

Condition (R1) is critical for the derivation of the IPCW-adjusted win ratio statistic. Condition R2, positive probability of censoring, is important to ensure that the inverse probability of censoring is proper. Conditions R3 and R4 are purely for regularity.

A.3 Asymptotic variance estimator for the logarithm of the IPCW-adjusted win ratio

An asymptotic variance estimator for the IPCW-adjusted win ratio can easily be derived from the variance-covariance estimator for the logarithm of the win ratio (Dong et al., 2016 and 2018; Bebu and Lachin, 2016) by replacing the kernel functions K_{ij} and L_{ij} with the IPCW-adjusted kernel functions K_{ij}^A and L_{ij}^A .

The IPCW-adjusted win proportion for the Treatment group is

$$P_t^A = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} K_{ij}^A = \frac{1}{N_t N_c} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \frac{K_{ij}}{G^{(t)}(T_j^{(c)}) G^{(c)}(T_j^{(t)})}.$$

Therefore, P_t^A is a two-sample U-statistic of degree (1,1) with kernel function

$$K^A[(Y_i^{(t)}, \delta_i^{(t)}), (Y_j^{(c)}, \delta_j^{(c)})] = K^A[(T_i^{(t)}, C_i^{(t)}), (T_j^{(c)}, C_j^{(c)})] = \frac{K_{ij}}{G^{(t)}(T_j^{(c)}) G^{(c)}(T_j^{(t)})}$$

based on the observed data $(Y_i^{(t)}, \delta_i^{(t)})$ and $(Y_j^{(c)}, \delta_j^{(c)})$. Similarly, P_c^A is a two-sample U-statistic with kernel function L_{ij}^A . Thus, P_t^A and P_c^A are asymptotically normal, and n_t^A and n_c^A are also asymptotically normal (AN).

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

Where $\theta_t = N_t N_c \theta_{K^A}$,

$$\theta_{K^A} = E\{K_{ij}^A\} = E\left\{K^A[(Y_i^{(t)}, \delta_i^{(t)}), (Y_j^{(c)}, \delta_j^{(c)})]\right\},$$

$$\sigma_t^2 = \frac{1}{N_t} \sigma_{t1}^2 + \frac{1}{N_c} \sigma_{t2}^2,$$

$$\sigma_{t1}^2 = \frac{N_t N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} [K_{ij}^A - \theta_{K^A}] [K_{ij'}^A - \theta_{K^A}],$$

$$\sigma_{t2}^2 = \frac{N_t N_c}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} \left[K_{ij}^A - \theta_{K^A} \right] \left[K_{i'j}^A - \theta_{K^A} \right].$$

Similarly, θ_c and σ_c^2 with kernel function $L_{ij}^A = L^A \left[\left(Y_i^{(t)}, \delta_i^{(t)} \right), \left(Y_j^{(c)}, \delta_j^{(c)} \right) \right]$ can be defined as follows:

$$\theta_c = N_t N_c \theta_{L^A},$$

$$\theta_{L^A} = E \left\{ L_{ij}^A \right\} = E \left\{ L^A \left[\left(Y_i^{(t)}, \delta_i^{(t)} \right), \left(Y_j^{(c)}, \delta_j^{(c)} \right) \right] \right\},$$

$$\sigma_c^2 = \frac{1}{N_c} \sigma_{c1}^2 + \frac{1}{N_t} \sigma_{c2}^2,$$

$$\sigma_{c1}^2 = \frac{N_t N_c}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} \left[L_{ij}^A - \theta_{L^A} \right] \left[L_{i'j}^A - \theta_{L^A} \right],$$

$$\sigma_{c2}^2 = \frac{N_t N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} \left[L_{ij}^A - \theta_{L^A} \right] \left[L_{ij'}^A - \theta_{L^A} \right].$$

The covariance of n_t^A and n_c^A is

$$\sigma_{tc}^2 = \frac{1}{N_t} \sigma_{tc1}^2 + \frac{1}{N_c} \sigma_{tc2}^2,$$

$$\sigma_{tc1}^2 = \frac{N_t N_c}{(N_c - 1)} \sum_{i=1}^{N_t} \sum_{j=1}^{N_c} \sum_{\substack{j'=1 \\ j' \neq j}}^{N_c} \left[K_{ij}^A - \theta_{K^A} \right] \left[L_{ij'}^A - \theta_{L^A} \right],$$

$$\sigma_{tc2}^2 = \frac{N_t N_c}{(N_t - 1)} \sum_{j=1}^{N_c} \sum_{i=1}^{N_t} \sum_{\substack{i'=1 \\ i' \neq i}}^{N_t} \left[K_{ij}^A - \theta_{K^A} \right] \left[L_{i'j}^A - \theta_{L^A} \right].$$

By the delta method, the logarithm of $\widehat{WR}^A = P_t^A / P_c^A$ is asymptotically normally distributed:

$$\log(\widehat{WR}^A) \sim AN(\theta_{\log(WR)}, \sigma_{\log(WR)}^2).$$

Where $\sigma_{\log(WR)}^2 = \frac{\sigma_t^2}{\theta_t} + \frac{\sigma_c^2}{\theta_c} - \frac{2\sigma_{tc}}{\theta_t\theta_c}$. Under the null hypothesis H_0 , θ_t and θ_c can be estimated by:

$$\hat{\theta}_t = \hat{\theta}_c = \frac{n_t^A + n_c^A}{2},$$

and θ_{K^A} and θ_{L^A} can be estimated by

$$\hat{\theta}_{K^A} = \hat{\theta}_{L^A} = \frac{n_t^A + n_c^A}{2N_tN_c}.$$

By using $\hat{\theta}_t$, $\hat{\theta}_c$, $\hat{\theta}_{K^A}$ and $\hat{\theta}_{L^A}$ in $\sigma_{\log(WR)}^2$ to replace θ_t , θ_c , θ_{K^A} and θ_{L^A} , respectively, and plugging in Kaplan-Meier survival estimators of censoring to replace $G^{(t)}(x)$ and $G^{(c)}(x)$, we can obtain the asymptotic variance estimate for the logarithm of the IPCW-adjusted win ratio, $\hat{\sigma}_{\log(WR^A)}^2$.

A.4 SAS program for the IPCW-adjusted win ratio

A SAS program for the IPCW-adjusted win ratio is available in the Supplemental Material on the publisher's website.

References

1. 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. [PubMed: 26353896]
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine*. 2010; 29(30):3245–3257. [PubMed: 21170918]
3. Cheung LC, Pan Q, Hyun N, Katki HA. Prioritized concordance index for hierarchical survival outcomes. *Statistics in Medicine*. 2019 7 10;38(15):2868–2882. [PubMed: 30957257]
4. 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. [PubMed: 27485522]
5. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *Journal of Biopharmaceutical Statistics*. 2018; 28(4):778–796. [PubMed: 29172988]
6. 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.
7. 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*. 2019. doi: 10.1002/pst.1977
8. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine*. 1999; 18:1341–1354. [PubMed: 10399200]
9. Finkelstein DM, Schoenfeld DA. Graphing the win ratio and its components over time. *Statistics in Medicine*. 2019; 38(1):53–61. [PubMed: 30206956]
10. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 1999; 94(446):496–509

11. Hoeffding W. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* 1948; 19:293–325.
12. Howe CJ, Cole SR, Chmiel JS, Muñoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *Am J Epidemiol.* 2011;173(5):569–77. [PubMed: 21289029]
13. Huang B, Kuan PF. Comparison of the restricted mean survival time with the hazard ratio in superiority trials with a time-to-event end point. *Pharmaceutical Statistics.* 2018 5;17(3):202–213. [PubMed: 29282880]
14. Khan MHR, Shaw JEH Robust bias estimation for Kaplan–Meier survival estimator with jackknifing. *Journal of Statistical Theory and Practice.* 2016; 10(1):7–19
15. Kyle RA, Therneau TM, Rajkumar SV, Offord JR, Larson DR, Plevak MF, & Joseph Melton L A long-term study of prognosis in monoclonal gammopathy of undetermined significance. *New England Journal of Medicine,* 2002;346(8): 564–569. [PubMed: 11856795]
16. Lee AJ *U-Statistics: Theory and Practice.* Marcel Dekker, New York, NY (1990)
17. Lehmann EL *Elements of Large-Sample Theory.* Springer, New York 1999.
18. 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. [PubMed: 28238045]
19. 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. [PubMed: 25156540]
20. Luo X, Qiu J, Bai S, Tian H. Weighted win loss approach for analyzing prioritized outcomes. *Statistics in Medicine.* 2017; 36(15):2452–2465. [PubMed: 28343373]
21. Mao L. On the alternative hypotheses for the win ratio. *Biometrics.* 2019; 75(1):347–351. [PubMed: 30096729]
22. 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. [PubMed: 30145929]
23. Oakes D On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika.* 2016; 103(3):742–745.
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. [PubMed: 27124210]
25. Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. *Statistical Methods in Medical Research.* 2018; 27(4):1230–1239. [PubMed: 27487842]
26. 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. [PubMed: 21900289]
27. 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. [PubMed: 31060767]
28. Rauch G, Jahn–Eimermacher A, Brannath W, Kieser M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Statistics in Medicine.* 2014; 33:1104–1120. [PubMed: 24122841]
29. Rauch G, Kunzmann K, Kieser M, Wegscheider K, König J, Eulenburg C. A weighted combined effect measure for the analysis of a composite time-to-first-event endpoint with components of different clinical relevance. *Statistics in Medicine.* 2018; 37(5):749–767. [PubMed: 29205425]
30. 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.
31. 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 [PubMed: 10985216]

32. Tian L, Zhao L, Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics*. 2014;15(2): 222–233 [PubMed: 24292992]
33. Tian L, Jin H, Uno H, Lu Y, Huang B, Anderson KM, Wei LJ. On the empirical choice of the time window for restricted mean survival time. *Biometrics*. 2020 2 15. doi: 10.1111/biom.13237
34. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*. 2011;30(10):1105–17. [PubMed: 21484848]
35. Vakulenko-Lagun B, Mandel M, Betensky RA. Inverse probability weighting methods for Cox regression with right-truncated data. *Biometrics*. 2019 10 17. doi: 10.1111/biom.13162
36. Van der Vaart AW Asymptotic statistics. Cambridge University Press, Vol. 3 2000.
37. Verbeek J, Spitzer E, de Vries T, van Es GA, Anderson WN, Van Mieghem NM, Leon MB, Molenberghs G, Tijssen J. Stat Med. Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints. *Statistics in Medicine*. 2019; 38(30):5641–5656 [PubMed: 31659790]
38. Verbeek J, Ozenne B, Anderson WN. Evaluation of inferential methods for the net benefit and win ratio statistics. *Journal of Biopharmaceutical Statistics*. 2020 2 25:1–18. doi: 10.1080/10543406.2020.1730873.
39. Wang D, Pocock S A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharmaceutical Statistics* 2016; 15(3):238–245. [PubMed: 26970432]
40. Wei LJ, Johnson WE. Combining dependent tests with incomplete repeated measurements. *Biometrika* 1985; 72(2):359–364.
41. Wu J. Statistical methods for survival trial design: with applications to cancer clinical trials using R. New York: Chapman and Hall/CRC 2018

Figure 1.
Three illustrative scenarios.

Figure 2.
Kaplan-Meier estimates of survival functions for the composite of PCM or death (975 uncensored patients).

Figure 3. Kaplan-Meier estimates of survival functions for the censoring ($C \sim \text{Exp}(0.0135)$) in one simulated dataset.

Figure 4.
Kaplan-Meier estimates of survival functions for the composite of death or hospitalization.

Table 1

Unadjusted vs IPCW-adjusted win ratio in three illustrative scenarios

Scenario	Censoring (%)			Unadjusted			IPCW-adjusted		
	Treatment group	Control group	Overall	P_t (%) Median	P_c (%) Median	$\widehat{WR}$ Median (95 % CI)	P_t^A (%) Median	P_c^A (%) Median	$\widehat{WR}^A$ Median (95 % CI)
a	0	0	0	55.5	33.5	1.67 (1.28, 2.16)			
	32	26	29	39.0	23.6	1.66 (1.24, 2.25)	55.4	33.5	1.66 (1.26, 2.20)
	48	40	44	31.0	18.7	1.65 (1.21, 2.35)	55.4	33.6	1.65 (1.22, 2.29)
b	0	0	0	49.2	41.6	1.18 (1.00, 1.40)			
	28	26	27	35.1	31.6	1.12 (0.92, 1.37)	49.3	41.6	1.18 (0.99, 1.42)
	42	40	41	28.3	26.1	1.08 (0.88, 1.35)	49.2	41.7	1.18 (0.98, 1.45)
c	0	0	0	40.0	38.4	1.04 (0.78, 1.39)			
	32	30	31	30.1	25.8	1.17 (0.84, 1.64)	40.0	38.5	1.05 (0.77, 1.44)
	48	46	47	25.1	20.1	1.25 (0.88, 1.81)	40.0	38.5	1.04 (0.76, 1.45)

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

For each scenario: 1st row: no censoring; 2nd row: censoring time $C \sim \text{Exp}(0.05)$; 3rd row: censoring time $C \sim \text{Exp}(0.09)$.

Table 2

Unadjusted vs IPCW-adjusted win ratio in the MGUS data with simulated censoring

Censoring		Unadjusted			IPCW-adjusted		
Distribution	%	P_t (%) Median	P_c (%) Median	$\widehat{WR}$ Median (95 % CI)	P_t^A (%) Median	P_c^A (%) Median	$\widehat{WR}^A$ Median (95 % CI)
No censoring	0	55.1	44.3	1.24 (1.07, 1.44)			
Exp(0.0015)	10	49.5	39.4	1.26 (1.07, 1.47)	55.3	44.3	1.24 (1.08, 1.44)
Exp(0.0060)	30	37.6	30.0	1.26 (1.05, 1.52)	55.2	44.7	1.24 (1.05, 1.45)
Exp(0.0135)	50	27.4	21.9	1.26 (1.01, 1.57)	55.2	44.7	1.24 (1.05, 1.46)

Table 3

Empirical variance, coverage and bias of unadjusted vs IPCW-adjusted win ratio in the MGUS data with simulated censoring

Censoring		Unadjusted				IPCW-adjusted			
Distribution	%	Empirical variance of $\log(WR)$	Median of $\hat{\sigma}_{\log(WR)}^2$	Coverage of 95% CI of WR	Median bias in $\widehat{WR}$	Empirical variance of $\log(WR)$	Median of $\hat{\sigma}_{\log(WR)}^2$	Coverage of 95% CI of WR	Median bias in $\widehat{WR}$
Exp(0.0015)	10	0.0061	0.0064	95.6%	0.013	0.0057	0.0059	95.9%	0.004
Exp(0.0060)	30	0.0088	0.0085	94.1%	0.017	0.0069	0.0068	94.6%	-0.007
Exp(0.0135)	50	0.0121	0.0118	94.5%	0.012	0.0088	0.0088	94.5%	-0.013

Bias is assessed by assuming that the $\widehat{WR}$ obtained where there is no censoring is the true WR.

Table 4

Summary of deaths and the composite endpoint in the CV data

	Treatment group	Control group
Number of patients	419	381
Number of deaths	111	120
Number (%) of deaths counted as the event for the composite endpoint	67 (60.0%)	63 (52.5%)
Number of patients with an event on the composite endpoint	176	181

Table 5

Unadjusted vs IPCW-adjusted win ratio in the CV data with simulated censoring

Censoring		Unadjusted			IPCW-adjusted		
Distribution	%	P_t (%) Median	P_c (%) Median	$\widehat{WR}$ Median (95 % CI)	P_t^A (%) Median	P_c^A (%) Median	$\widehat{WR}^A$ Median (95 % CI)
No censoring	0	38.4	31.1	1.23 (1.00, 1.52)			
Exp(0.0004)	25	31.0	23.9	1.30 (1.03, 1.64)	39.1	31.5	1.24 (1.00, 1.55)
Exp(0.001)	50	24.4	17.5	1.40 (1.07, 1.83)	39.6	31.3	1.26 (1.00, 1.63)

Table 6

Empirical variance, coverage and bias of unadjusted vs IPCW-adjusted win ratio in the CV data with simulated censoring

Censoring		Unadjusted				IPCW-adjusted			
Distribution	%	Empirical variance of $\log(WR)$	Median of $\hat{\sigma}_{\log(WR)}^2$	Coverage of 95% CI of WR	Median bias in $\widehat{WR}$	Empirical variance of $\log(WR)$	Median of $\hat{\sigma}_{\log(WR)}^2$	Coverage of 95% CI of WR	Median bias in $\widehat{WR}$
Exp(0.0004)	25	0.0129	0.0137	93.9%	0.059	0.0121	0.0131	95.6%	0.008
Exp(0.001)	50	0.0169	0.0175	84.0%	0.167	0.0161	0.0160	93.7%	0.038

Bias is assessed by assuming that the $\widehat{WR}$ obtained where there is no censoring is the true WR.