Received: 16 January 2022 | Revised: 20 April 2022 | Accepted: 31 May 2022

DOI: 10.1002/pst.2251

WILEY

MAIN PAPER

Win statistics (win ratio, win odds, and net benefit) can complement one another to show the strength of the treatment effect on time-to-event outcomes

Gaohong Dong[1] | Bo Huang[2] | Johan Verbeeck[3] | Ying Cui[4] | James Song[1] | Margaret Gamalo-Siebers[5] | Duolao Wang[6] | David C. Hoaglin[7] | Yodit Seifu[8] | Tobias Mütze[9] | John Kolassa[10]

[1]BeiGene, Ridgefield Park, New Jersey, USA

[2]Pfizer Inc., Groton, Connecticut, USA

[3]DSI, I-Biostat, University Hasselt, Hasselt, Belgium

[4]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA

[5]Pfizer Inc., Collegeville, Pennsylvania, USA

[6]Liverpool School of Tropical Medicine, Liverpool, UK

[7]Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, Massachusetts, USA

[8]Bristol Myers Squibb, Berkeley Heights, New Jersey, USA

[9]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

[10]Department of Statistics, Rutgers University, Piscataway, New Jersey, USA

Correspondence
Gaohong Dong, BeiGene, 55 Challenger Road, Ridgefield Park, NJ, USA.
Email: gaohong.dong@beigene.com

Abstract

Conventional analyses of a composite of multiple time-to-event outcomes use the time to the first event. However, the first event may not be the most important outcome. To address this limitation, generalized pairwise comparisons and win statistics (win ratio, win odds, and net benefit) have become popular and have been applied to clinical trial practice. However, win ratio, win odds, and net benefit have typically been used separately. In this article, we examine the use of these three win statistics jointly for time-to-event outcomes. First, we explain the relation of point estimates and variances among the three win statistics, and the relation between the net benefit and the Mann–Whitney U statistic. Then we explain that the three win statistics are based on the same win proportions, and they test the same null hypothesis of equal win probabilities in two groups. We show theoretically that the Z-values of the corresponding statistical tests are approximately equal; therefore, the three win statistics provide very similar *p*-values and statistical powers. Finally, using simulation studies and data from a clinical trial, we demonstrate that, when there is no (or little) censoring, the three win statistics can complement one another to show the strength of the treatment effect. However, when the amount of censoring is not small, and without adjustment for censoring, the win odds and the net benefit may have an advantage for interpreting the treatment effect; with adjustment (e.g., IPCW adjustment) for censoring, the three win statistics can complement one another to show the strength of the treatment effect. For calculations we use the R package WINS, available on the CRAN (Comprehensive R Archive Network).

KEYWORDS

IPCW, IPCW-adjusted win statistics, inverse-probability-of-censoring weighting, generalized pairwise comparisons, Mann–Whitney U statistic

1 | INTRODUCTION

For the analysis of prioritized multiple time-to-event outcomes in clinical trials, the common time-to-first-event analysis does not consider the outcomes' priorities. The first event may be due to an outcome of lower clinical importance

(e.g., progressive disease vs. death in oncology studies, or heart failure hospitalization vs. cardiovascular death in chronic heart failure studies). To overcome this limitation, methods that incorporate the order of clinical importance among the outcomes, such as generalized pairwise comparisons (GPC)[1] and the win statistics (win ratio,[2] win odds,[3] and net benefit[1]), have been studied.[1–28] All these methods compare each subject in the Treatment group with every subject in the Control group, and the result of each pairwise comparison is either a win or a tie. Each comparison starts with the highest priority outcome and only takes the next most important outcome into account when the comparison based on the higher priority outcome results in a tie. If the comparison based on each outcome is a tie, the overall result for that pair is a tie. Thus, lower-priority outcomes do not "mask" more important outcomes just because they occur earlier.

The generalized pairwise comparisons and the win statistics have been applied in design and analysis of Phase III clinical trials (e.g., NCT04001504, NCT04847557, and NCT04510493 as registered in ClinicalTrials.gov) and in supporting drug approval by health authorities (e.g., tafamidis for treatment of cardiomyopathy per the ATTR-ACT trial). The stratified win ratio[9] has also been applied to Phase III and Phase IV clinical trials such as the EMPULSE study of the SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure[29] and the ACTION study of therapeutic versus prophylactic anticoagulation for patients admitted to hospital with COVID-19 and elevated D-dimer concentration.[30]

Win ratio, win odds, and net benefit have been extensively studied. When there are no ties, the win odds reduces to the win ratio. There is a rich literature on their variance estimators and on weighted and stratified analyses.[5–12,19–22] Regression analysis,[22,23] sample size and power calculations,[21,24] matched win statistics,[27] recurrent-event analysis,[28] and methods that adjust the win statistics for censoring[13,17,18] have become available.

One shortcoming of the win ratio is that its calculation ignores ties. To address this problem, Dong et al.[3] formally introduced the win odds, which divides a tie into two half wins and assigns a half win to each treatment group. Subsequently, Peng[19] discussed the win odds in the design of non-inferiority trials. Gasparyan et al.[11,21] showed the win odds as an interpretable treatment effect measure, and provided a sample size and power calculation for a univariate outcome. Dong et al.[17,18] showed that the inverse-probability-of-censoring weighting (IPCW) adjusted win odds (or IPCW-adjusted win statistics in general) is an unbiased estimator of treatment effect in the presence of censoring. Brunner et al.[20] further explained the win odds as an adaptation of the win ratio to include ties, and argued that, for count, ordinal and continuous outcomes, and with some discussions on time-to-event outcomes, the win odds is preferable to the win ratio for quantifying the treatment effect in the presence of ties, because the proportion of ties indicates the similarity of the two groups.

For time-to-event outcomes, censoring-induced ties do not necessarily mean that the two patients in a pair have the same value of the outcomes (see details in Section 2). Moreover, at present, win ratio, win odds, and net benefit have typically been used separately; and they have not been compared systematically. Therefore, in this article, we examine the use of these three win statistics jointly for time-to-event outcomes. We explain that the three win statistics test the same null hypothesis of equal win probabilities in the Treatment and Control groups. We show theoretically that the Z-values of the corresponding statistical tests are approximately equal; therefore, the three win statistics provide very similar *p*-values and statistical powers. Through simulation studies and data from a large randomized cardiovascular outcome trial (the CHARM study[31]), we show that the three win statistics can complement one another to show the strength of the treatment effect. Therefore, presenting win proportions, win ratio, win odds, and net benefit together can give a more detailed picture of an analysis. Because the three win statistics are complementary, sometimes it may be enough to use and present only one of them for clinical trial design and analysis. For calculations, we use the R package WINS by Cui and Huang,[32] which is available on the CRAN (Comprehensive R Archive Network).

2 | WIN STATISTICS

2.1 | Test statistics

In generalized pairwise comparisons, each comparison of a pair of two patients (one from the Treatment group and one from the Control group) has three possible results: the patient in the Treatment group wins, the Control patient wins, or the two patients are tied. For example, the Treatment patient would "win" if the Control patient died earlier from a cardiovascular cause. Let π_t, π_c, and π_{tie} be the probabilities corresponding to these three results ($\pi_t + \pi_c + \pi_{tie} = 1.0$). The subscripts t and c denote the Treatment and Control groups, respectively. Win ratio (WR), win odds (WO) and net benefit (NB) are defined as follows.

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

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

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

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. We use $i = 1, 2, ..., N_t$ for patients in the Treatment group and $j = 1, 2, ..., N_c$ for patients in the Control group. We define the kernel function K by $K_{ij} = 1$ if a win for the Treatment group occurs when an observed time $Y_i = \min(T_i, C_i)$ in this group is longer than an event time T_j in the Control group:

$$\begin{aligned} K_{ij} &= 1 \ \ (\text{win for patient } i \text{ in the Treatment group}), \text{if } Y_i > Y_j \text{ and } \delta_j = 1, \\ &= 0, \ \ \text{otherwise}. \end{aligned} \tag{2a}$$

Similarly, the kernel function L takes the value $L_{ij} = 1$ if patient j in the Control group wins over the patient i in the Treatment group:

$$\begin{aligned} L_{ij} &= 1 \ \ (\text{win for patient } j \text{ in the Control group}), \text{if } Y_j > Y_i \text{ and } \delta_i = 1, \\ &= 0, \ \ \text{otherwise}. \end{aligned} \tag{2b}$$

In the N_tN_c pairwise comparisons, the numbers of wins for the Treatment group and the Control group are $n_t = \sum_{i=1}^{N_t}\sum_{j=1}^{N_c} K_{ij}$ and $n_c = \sum_{i=1}^{N_t}\sum_{j=1}^{N_c} L_{ij}$. The corresponding estimates of the win probabilities π_t and π_c are the win proportions: $\widehat{\pi}_t = P_t = n_t/N_tN_c$ and $\widehat{\pi}_c = P_c = n_c/N_tN_c$, respectively; and $P_{tie} = 1 - P_t - P_c$. Therefore, win ratio, win odds and net benefit can be estimated by

$$\widehat{WR} = \frac{P_t}{P_c}, \tag{3a}$$

$$\widehat{WO} = \frac{P_t + 0.5P_{tie}}{P_c + 0.5P_{tie}} = \frac{P_t + 0.5(1-P_t-P_c)}{P_c + 0.5(1-P_t-P_c)} = \frac{P_t + 0.5(1-P_t-P_c)}{1-[P_t + 0.5(1-P_t-P_c)]}, \tag{3b}$$

$$\widehat{NB} = P_t - P_c. \tag{3c}$$

Hence, the win ratio is a ratio of win proportions, the win odds is an odds of win proportions, and the net benefit is a difference in win proportions. Because win ratio, win odds and net benefit are derived from the same win proportions, their corresponding null hypotheses of H_0: $WR = 1.0$, H_0: $WO = 1.0$, and H_0: $NB = 0$ are equivalent to testing the null hypothesis of equal win probabilities in the two treatment groups, H_0: $\pi_t = \pi_c$.

The statistics P_t and P_c are U-statistics and are asymptotically normally (AN) distributed. Therefore, n_t and n_c are also asymptotically normal,

$$\begin{pmatrix} n_t \\ n_c \end{pmatrix} \sim AN\left(\begin{bmatrix} \theta_t \\ \theta_c \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & \sigma_{tc} \\ \sigma_{tc} & \sigma_c^2 \end{bmatrix} \right). \tag{4}$$

By the delta method, $log(WR)$, $log(WO)$, and NB are asymptotically normally distributed with the following variances:

$$\sigma^2_{log(WR)} = \frac{\sigma_t^2}{(\theta_t)^2} - \frac{2\sigma_{tc}}{\theta_t\theta_c} + \frac{\sigma_c^2}{(\theta_c)^2}, \tag{5a}$$

$$\sigma^2_{log(WO)} = \left(\sigma_t^2 - 2\sigma_{tc} + \sigma_c^2\right)\left(\frac{1}{\gamma} + \frac{1}{N_tN_c - \gamma}\right)^2/4, \tag{5b}$$

$$\sigma^2_{NB} = \left(\sigma_t^2 - 2\sigma_{tc} + \sigma_c^2\right)/(N_tN_c)^2, \tag{5c}$$

where $\gamma = \theta_t + 0.5(N_tN_c - \theta_t - \theta_c)$.

Under the null hypothesis H_0: $\pi_t = \pi_c$, θ_t and θ_c can be estimated by

$$\widehat{\theta}_t = \widehat{\theta}_c = (n_t + n_c)/2. \tag{6}$$

Then the variances of $log(WR)$, $log(WO)$, and NB under the null hypothesis can be estimated by

$$\widehat{\sigma}^2_{log(WR)} = \frac{\widehat{\sigma}_t^2 - 2\widehat{\sigma}_{tc} + \widehat{\sigma}_c^2}{[(n_t + n_c)/2]^2}, \tag{7a}$$

$$\widehat{\sigma}^2_{log(WO)} = \frac{\widehat{\sigma}_t^2 - 2\widehat{\sigma}_{tc} + \widehat{\sigma}_c^2}{(N_tN_c/2)^2}, \tag{7b}$$

$$\widehat{\sigma}^2_{NB} = \frac{\widehat{\sigma}_t^2 - 2\widehat{\sigma}_{tc} + \widehat{\sigma}_c^2}{(N_tN_c)^2}. \tag{7c}$$

The calculations for $\widehat{\sigma}_t^2$, $\widehat{\sigma}_c^2$, and $\widehat{\sigma}_{tc}$ are given by Dong et al.[8,9] and Bebu and Lachin.[7]

Definitions (2a) and (2b) are for a single time-to-event outcome. The setting for prioritized multiple outcomes and for inverse-probability-of-censoring weighting (IPCW) adjustment for censoring can be formulated similarly (see details in Dong et al.[17,18]).

2.2 | Point estimates

Since the win odds considers a tie as a half win for the Treatment group and a half win for the Control group, the win odds, as defined in (1b), is always closer to the null value of 1.0 than the win ratio.[3,20] When there are no ties, the win odds reduces to the win ratio.

The win ratio, win odds and net benefit have the following relations.

$$NB = \frac{WR - 1}{WR + 1}(1 - \pi_{tie}), \tag{8a}$$

$$NB = \frac{WO - 1}{WO + 1}, \tag{8b}$$

$$WO = \frac{1 + NB}{1 - NB}, \tag{8c}$$

$$WO = \frac{WR - 0.5\pi_{tie}(WR - 1)}{1 + 0.5\pi_{tie}(WR - 1)}. \tag{8d}$$

Correspondingly, $\widehat{NB}=\frac{\widehat{WR}-1}{\widehat{WR}+1}(1-P_{tie})=\frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{n_t+n_c}{N_tN_c}$, $\widehat{NB}=\frac{\widehat{WO}-1}{\widehat{WO}+1}$, $\widehat{WO}=\frac{1+\widehat{NB}}{1-\widehat{NB}}$, and $\widehat{WO}=\frac{\widehat{WR}-0.5P_{tie}\left(\widehat{WR}-1\right)}{1+0.5P_{tie}\left(\widehat{WR}-1\right)}$. These relations indicate that the win odds increases (or decreases) as the net benefit increases (or decreases) regardless of ties. However, the relations between the win odds and the net benefit vs. the win ratio depend on ties. We further discuss these relations in Section 4.1.

2.3 | Variances

The estimated variance of the win odds ($\widehat{\sigma}^2_{log(WO)}$ per (7b)) is always smaller than or equal to that of the win ratio ($\widehat{\sigma}^2_{log(WR)}$ per (7a)) because the total number of wins cannot exceed the total number of comparisons: $n_t+n_c\le N_tN_c$, with equality only if there are no ties. Consequently, the confidence interval for the win odds is always narrower than that for the win ratio, and their confidence intervals coincide when there are no ties. For a large clinical trial with low event rates (i.e., a high proportion of ties), the point estimate of the win odds can be much closer to 1.0 and its confidence interval can be very narrow compared with the win ratio, as illustrated in CHARM application in Section 5.

2.4 | Net benefit as a direct transformation of the Mann–Whitney U statistic

Without ties (i.e., $n_t+n_c=N_tN_c$), the numbers of wins, n_t and n_c, are Mann–Whitney U statistics.[33] With ties, $U=n_t+0.5(N_tN_c-n_t-n_c)$ is also a Mann–Whitney U statistic, and its variance is[34]

$$\sigma_U^2=\frac{N_tN_c(N+1)}{12}-\frac{N_tN_c\sum_{k=1}^{K}\left(d_k^3-d_k\right)}{12N(N-1)}, \tag{9}$$

where $N=N_t+N_c$, K is the number of distinct observations and d_k is the number of times the kth tied observation is repeated ($k=1, 2, ..., K$). Expression (9) shows that the variance of U increases as the number of ties decreases.

Verbeeck et al.[35] explained that the net benefit is a direct transformation of U:

$$NB=\frac{2U}{N_tN_c}-1, \tag{10}$$

where the quantity $\frac{U}{N_tN_c}$ is an estimator of the probabilistic index. Therefore, the variance of the net benefit also increases as the number of ties decreases.

2.5 | Approximate equivalence of statistical tests for the three win statistics

Because the null hypothesis of equal win probabilities in the Treatment and Control groups underlies the null hypotheses of H_0: $WR=1.0$, H_0: $WO=1.0$, and H_0: $NB=0$ for the three win statistics, the Z-values of the three corresponding statistical tests are approximately equal as we show below. Therefore, the three win statistics provide very similar p-values and statistical powers.

From (7a) and (7c), we obtain $\widehat{\sigma}^2_{NB}=\frac{(n_t+n_c)^2}{4(N_tN_c)^2}\widehat{\sigma}^2_{log(WR)}$. Applying $\widehat{NB}=\frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{n_t+n_c}{N_tN_c}$ yields the Z-value for the net benefit:

$$Z_{NB}=\frac{\widehat{NB}}{\widehat{\sigma}_{NB}}=\frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{n_t+n_c}{N_tN_c}\frac{1}{\widehat{\sigma}_{\log(WR)}}\frac{2N_tN_c}{n_t+n_c}=\frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{2}{\widehat{\sigma}_{\log(WR)}}.$$

Following the series expansion $\log(x)=2\left[\frac{x-1}{x+1}+\frac{1}{3}\left(\frac{x-1}{x+1}\right)^3+\frac{1}{5}\left(\frac{x-1}{x+1}\right)^5+\cdots\right]$ for $x>0$, $\log(x)\approx\frac{2(x-1)}{x+1}$. Therefore, $\log\left(\widehat{WR}\right)\approx\frac{2\left(\widehat{WR}-1\right)}{\widehat{WR}+1}$. Hence, the Z-values of the statistical tests for *NB* and *log(WR)* are approximately equal:

$$Z_{NB}\approx\frac{\log\left(\widehat{WR}\right)}{\widehat{\sigma}_{\log(WR)}}=Z_{\log(WR)}. \tag{11a}$$

Similarly, since $\widehat{NB}=\frac{\widehat{WO}-1}{\widehat{WO}+1}$ and $\widehat{\sigma}^2_{NB}=\frac{1}{4}\widehat{\sigma}^2_{log(WO)}$ per (7b) and (7c), the Z-values of the statistical tests for the net benefit and *log(WO)* are also approximately equal.

$$Z_{NB}\approx\frac{\log\left(\widehat{WO}\right)}{\widehat{\sigma}_{\log(WO)}}=Z_{\log(WO)}. \tag{11b}$$

3 | CENSORING FOR TIME-TO-EVENT OUTCOMES

When, ideally, no event times are censored, the win probabilities at time x can be calculated by $\pi_t(x)=-\int_0^x S_t dS_c$ and $\pi_c=-\int_0^x S_c dS_t$, where S_t and S_c are the survival functions of time to event in the Treatment and Control groups, respectively; and the true values of the three win statistics can be calculated from $\pi_t(x)$ and $\pi_c(x)$. In practice, S_t and S_c are unknown, and Kaplan–Meier estimators of S_t and S_c, can be used to estimate the win statistics,[4] or equivalently the U-statistics approach described in Section 2.1 can be used. The latter approach can be applied regardless of whether there is censoring.

As seen from the kernel functions defined in (2a) and (2b), censoring of time-to-event outcomes could impact the win statistics. Censoring can occur when (a) some patients dropped out without experiencing an event of interest (i.e., early dropout) and (b) at the time of the data cutoff for the analysis, some patients have not experienced an event (i.e., administrative censoring or end-of-study censoring). To illustrate censoring bias, assume that patients A and B both had a death event (Figure 1A). Patient B is the winner because patient A died earlier. However, if patient A was censored (Figure 1B), or patient B was censored (Figure 1C) or both patients were censored before the death of patient A (Figure 1D), a "win" cannot be determined for this pair of patients. In calculating win statistics, this situation is typically considered a "tie", and it introduces a bias. In fact, for time-to-event outcomes, the censoring-induced ties do not necessarily mean that the two patients in a pair have the same value of the outcomes.

Therefore, censoring has an impact on the win statistics. As demonstrated by Dong et al.,[16] the win probabilities can be calculated by $\pi_t(x)=-\int_0^x S_t G_t G_c dS_c$ and $\pi_c(x)=-\int_0^x S_c G_c G_t dS_t$, where G_t and G_c are the survival functions of time to censoring in the Treatment and Control groups, respectively. Statistical methods that adjust the win statistics for censoring are available. For example, Péron et al.[13] suggested an adaptation of Efron's scoring to adjust the win statistics, and Dong et al.[17,18] applied the IPCW approach to adjust the win statistics for independent censoring (i.e., IPCW-adjusted win statistics) and dependent censoring such that covariates (baseline covariates and/or time-dependent covariates) are associated with event time and can predict censoring (i.e., CovIPCW-adjusted win statistics)

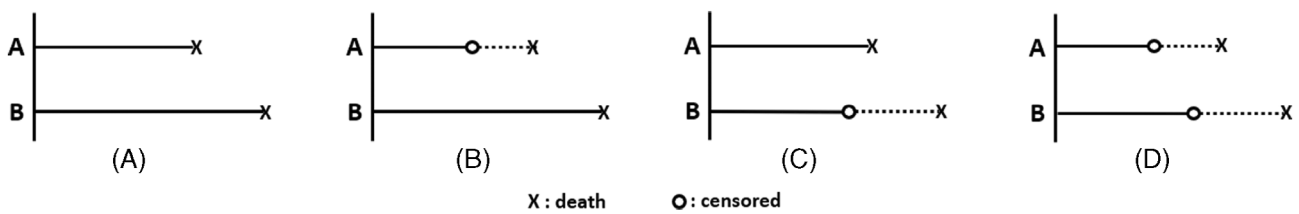

FIGURE 1 Illustration of censoring bias

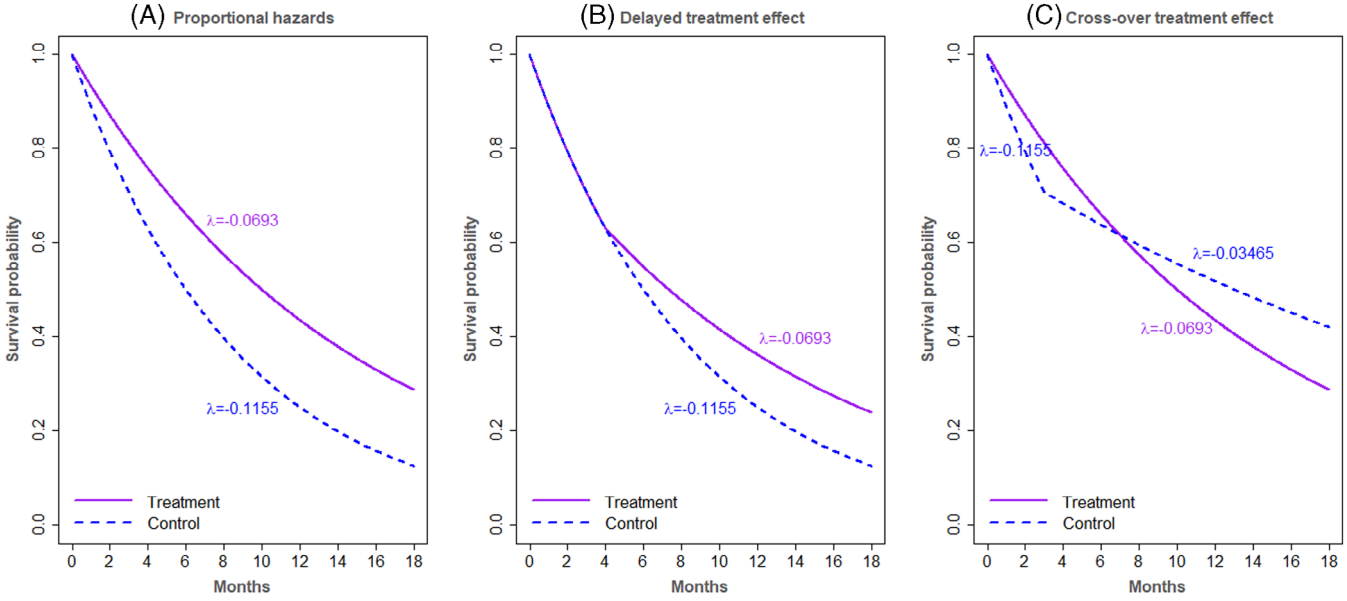

FIGURE 2 Assumed hazard rates (λ) and survival curves for simulations

4 | SIMULATION STUDIES

We extend the simulation studies presented in Dong et al.[17] to investigate win ratio, win odds and net benefit in the setting of time-to-event outcomes.

4.1 | Simulation study 1

In this simulation study, we analyze a single time-to-event outcome without censoring. We use three scenarios that arise in practice: (a) proportional hazards in the two groups, (b) delayed treatment effect in the Treatment group, and (c) cross-over treatment effect between the two groups. For all three scenarios, following Huang and Kuan,[36] we use exponential or piecewise exponential functions to generate 1000 simulated datasets with 200 patients per group. Figure 2 shows the hazard rates and survival curves for each scenario.

Table 1 presents win proportions and win statistics for the three scenarios. For each scenario, the *p*-values for the three win statistics are very similar at all timepoints, as the Z-values of the corresponding statistical tests are approximately equal (as theoretically shown in Section 2.5).

In Scenario (a), proportional hazards in the two groups (Table 1), the win ratio is the reciprocal of the hazard ratio (HR).[4,15,16] Therefore, the estimated win ratio (the median over the 1000 simulated datasets) is close to the true value of 1.67 at all timepoints. The width of the estimated 95% confidence interval (i.e., the 95% percentile interval [2.5th percentile, 97.5th percentile] from 1000 simulations) is 2.68 at month 1, which is relatively wide because few events are observed and the evidence of the treatment effect is not yet strong (*p*-value = 0.13155). The estimated 95% confidence interval narrows to width 0.81 at month 18, as most events have been observed and the evidence of the treatment effect has become very strong (*p*-value = 0.00005). In contrast, the point estimates of the win odds and the net benefit increase over time as the evidence of the treatment effect becomes stronger. These results make sense. Since $WR = 1/HR$ is constant for this scenario and the proportion of ties, P_{tie}, decreases as more events are observed, as expressed in Section 2.2, $\widehat{NB} = \frac{\widehat{WR}-1}{\widehat{WR}+1}(1-P_{tie})$ increases with decreasing ties, and $\widehat{WO} = \frac{1+\widehat{NB}}{1-\widehat{NB}}$ also increases over time.

The widths of their 95% confidence intervals for the win odds and the net benefit also increase over time. For example, the estimated win odds is 1.08 at month 1, reflecting that a considerable proportion of pairwise comparisons result in ties and so provide little evidence for an effect of the treatment. Similarly, at month 1, the 95% confidence interval is narrowest. As more events are observed and the evidence of the treatment effect becomes stronger, the point estimate of the win odds and the lower limit of the confidence interval move away from the null value of 1.0. Although the 95% confidence interval becomes wider over time, the *p*-value of the win odds decreases from 0.13940 at month 1 to 0.00005 at month 18.

TABLE 1 Win statistics and 95% confidence intervals (CIs) for the three Scenarios

Time	Win proportion (%)		Win ratio			Win odds			Net benefit (%)		
	Treatment	Control	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value
Scenario (a): proportional hazards in the two groups											
Month 1	10.2	6.1	1.67 (0.86, 3.54)	2.68	0.13155	1.08 (0.97, 1.22)	0.25	0.13940	4.1 (−1.3, 9.9)	11.2	0.13967
Month 3	25.9	15.7	1.68 (1.11, 2.50)	1.39	0.01927	1.22 (1.04, 1.45)	0.41	0.02019	10.0 (2.2, 18.3)	16.1	0.02066
Month 6	41.3	24.9	1.66 (1.21, 2.32)	1.11	0.00152	1.40 (1.13, 1.71)	0.58	0.00172	16.5 (6.2, 26.2)	20.0	0.00190
Month 9	50.2	30.3	1.66 (1.27, 2.19)	0.92	0.00037	1.50 (1.21, 1.85)	0.64	0.00041	20.0 (9.6, 29.7)	20.1	0.00049
Month 12	55.3	33.3	1.67 (1.28, 2.16)	0.88	0.00013	1.58 (1.25, 1.98)	0.73	0.00014	22.2 (11.0, 32.9)	21.9	0.00018
Month 18	60.0	36.1	1.66 (1.31, 2.12)	0.81	0.00005	1.63 (1.30, 2.06)	0.76	0.00005	23.9 (12.9, 34.7)	21.9	0.00007
Scenario (b): delayed treatment effect in the Treatment group											
Time	Win proportion (%)		Win ratio			Win odds			Net benefit (%)		
	Treatment	Control	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value
Month 1	10.0	10.0	1.00 (0.63, 1.55)	0.91	0.49486	1.00 (0.92, 1.09)	0.17	0.49565	0.0 (−4.4, 4.3)	8.7	0.49569
Month 3	24.6	24.7	1.00 (0.78, 1.29)	0.52	0.51161	1.00 (0.88, 1.14)	0.25	0.51182	0.0 (−6.2, 6.4)	12.6	0.51188
Month 6	37.4	34.5	1.08 (0.88, 1.33)	0.45	0.38331	1.06 (0.91, 1.23)	0.32	0.38348	2.9 (−4.5, 10.4)	14.9	0.38365
Month 9	44.8	39.0	1.14 (0.96, 1.38)	0.42	0.14723	1.12 (0.97, 1.31)	0.34	0.14742	5.7 (−1.6, 13.5)	15.1	0.14786
Month 12	49.2	41.6	1.18 (1.00, 1.40)	0.41	0.06308	1.16 (1.00, 1.36)	0.36	0.06320	7.5 (0.0, 15.4)	15.4	0.06370
Month 18	52.8	43.9	1.21 (1.02, 1.42)	0.40	0.02776	1.20 (1.02, 1.40)	0.38	0.02779	9.0 (1.0, 16.8)	15.8	0.02822
Scenario (c): cross-over treatment effect between the two groups											
Time	Win proportion (%)		Win ratio			Win odds			Net benefit (%)		
	Treatment	Control	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value
Month 1	10.2	6.1	1.69 (0.86, 3.62)	2.76	0.13530	1.08 (0.97, 1.21)	0.23	0.13370	4.1 (−1.2, 9.5)	10.7	0.13393
Month 3	26.1	15.6	1.67 (1.12, 2.53)	1.41	0.01516	1.23 (1.05, 1.46)	0.40	0.01718	10.5 (2.6, 18.6)	15.9	0.01757
Month 6	31.4	26.0	1.21 (0.88, 1.69)	0.81	0.25176	1.11 (0.93, 1.35)	0.42	0.25283	5.5 (−3.7, 14.8)	18.5	0.25332
Month 9	35.3	33.6	1.04 (0.79, 1.45)	0.66	0.49996	1.03 (0.85, 1.29)	0.44	0.50014	1.5 (−8.2, 12.5)	20.7	0.50032
Month 12	38.0	39.2	0.96 (0.74, 1.30)	0.56	0.49736	0.97 (0.80, 1.22)	0.43	0.49751	−1.5 (−11.4, 10.2)	21.5	0.49770
Month 18	41.5	45.8	0.91 (0.70, 1.16)	0.46	0.37702	0.92 (0.73, 1.14)	0.41	0.37715	−4.1 (−15.8, 6.5)	22.3	0.37755

In Scenario (b), delayed treatment effect in the treatment group (Table 1), because the survival curves start to separate at month 4, with hazard ratio = 0.60, all three win statistics increase over time after month 4 as the evidence of the treatment effect increases. The 95% confidence interval for the win ratio becomes narrower over time, and those for the win odds and the net benefit become wider.

In Scenario (c), cross-over treatment effect between the two groups (Table 1), before month 3, the Treatment group performs better, with hazard ratio = 0.60; after month 3, the Control group performs better, with hazard ratio = 2.0. Therefore, the point estimates of the win ratio, the win odds and the net benefit first increase and then decrease from month 3. The 95% confidence intervals for the win ratio become narrower over time, whereas those for the net benefit become wider. Interestingly, the width of the 95% confidence intervals for the win odds increases first, then decreases from month 9. However, the *p*-values for the three win statistics are similar at all timepoints.

In all three scenarios, at month 18, the win ratio and the win odds get closer to each other since the majority of events have been observed and there are few ties; consequently, their confidence intervals also get closer to each other.

In summary, as follow-up time increases, more events are observed, and evidence of the treatment effect becomes stronger. For all three scenarios the win proportions in both groups increase over time, and the 95% confidence interval for the win ratio becomes narrower. The 95% confidence interval for the net benefit becomes wider over time. This is not surprising since increase in variance when the number of ties decreases is a well-known property of the Mann–Whitney statistic,[34,35] and the net benefit is a direct transformation of the Mann–Whitney U statistic (as explained in Section 2.4). In Scenario (c), however, the width of the 95% confidence intervals for the win odds may first increase and then decrease, corresponding to the cross-over pattern of the treatment effect.

Nevertheless, this simulation without censoring shows that the three win statistics can complement one another to show the strength of the treatment effect. Therefore, it may be helpful to present the three win statistics together when there is no (or little) censoring. Because the three win statistics are complementary, sometimes it may be enough to use and present only one of them for clinical trial design and analysis.

4.2 | Simulation study 2

As described in Dong et al.,[17] we selected 800 patients from clinical trials in cardiovascular (CV) disease with the composite of death and hospitalization as the primary endpoint. We used the data up to 3 years, and excluded patients who dropped out prior to year 3, so that we could estimate the win statistics without bias from censoring

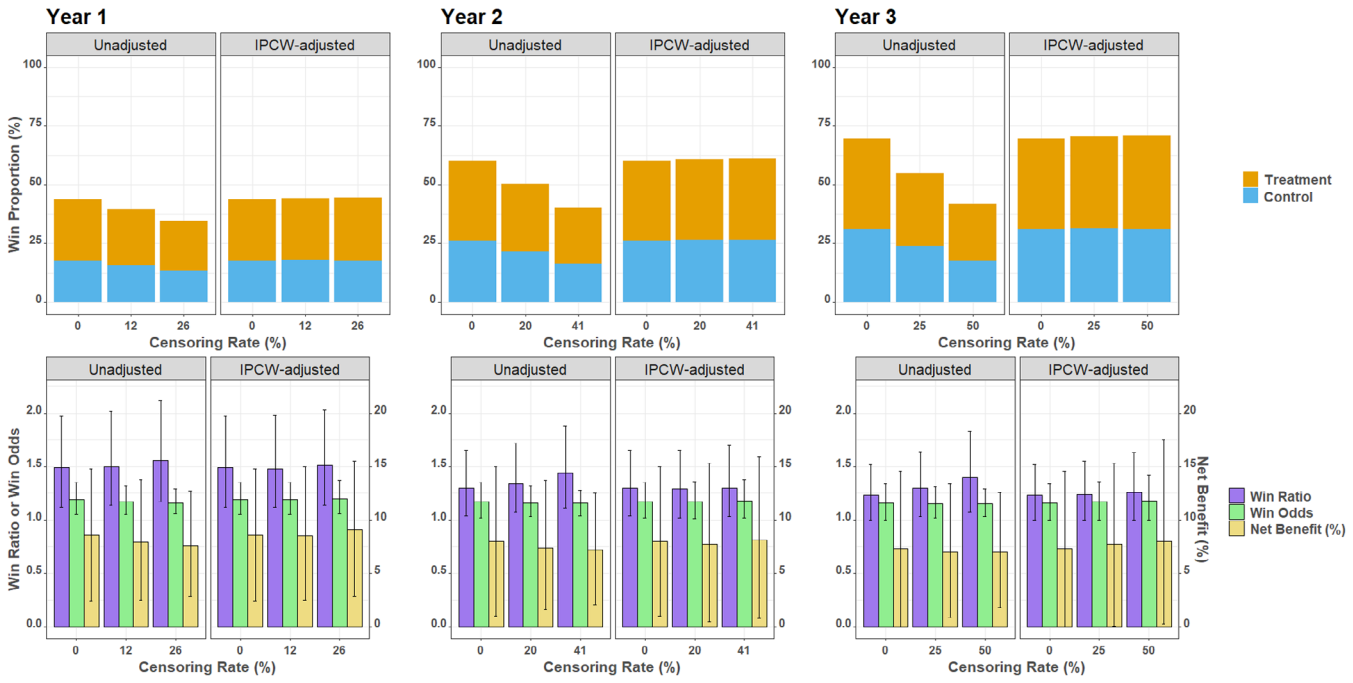

FIGURE 3 Unadjusted versus IPCW-adjusted win proportions and win statistics with 95% confidence intervals (shown with whiskers)

TABLE 2 Win statistics and 95% confidence intervals (CIs) for the CV example

(A) Unadjusted

		Win proportion (%)		Win ratio			Win odds			Net benefit (%)		
Time	Censoring (%)	Treatment	Control	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value
Year 1	0%	26.3	17.7	1.49 (1.12, 1.97)	0.85	0.0056	1.19 (1.05, 1.35)	0.30	0.0062	8.6 (2.4, 14.8)	12.4	0.0063
	12%	23.8	15.9	1.50 (1.14, 2.02)	0.88	0.0059	1.17 (1.05, 1.32)	0.27	0.0065	7.9 (2.5, 13.8)	11.3	0.0067
	26%	21.1	13.4	1.56 (1.17, 2.12)	0.95	0.0040	1.16 (1.06, 1.29)	0.23	0.0045	7.6 (2.9, 12.7)	9.8	0.0046
Year 2	0%	34.1	26.1	1.30 (1.04, 1.65)	0.61	0.0236	1.17 (1.02, 1.35)	0.33	0.0241	8.0 (1.0, 15.0)	14.0	0.0244
	20%	28.8	21.4	1.34 (1.07, 1.72)	0.65	0.0179	1.16 (1.03, 1.32)	0.29	0.0185	7.4 (1.6, 13.7)	12.1	0.0183
	41%	23.8	16.4	1.44 (1.11, 1.88)	0.77	0.0076	1.16 (1.04, 1.28)	0.24	0.0081	7.2 (2.1, 12.5)	10.4	0.0084
Year 3	0%	38.4	31.1	1.23 (1.00, 1.52)	0.52	0.0505	1.16 (1.00, 1.34)	0.34	0.0509	7.3 (0.0, 14.6)	14.6	0.0514
	25%	31.0	23.9	1.30 (1.03, 1.64)	0.61	0.0291	1.15 (1.02, 1.31)	0.29	0.0297	7.0 (0.9, 13.4)	12.5	0.0300
	50%	24.4	17.5	1.40 (1.07, 1.83)	0.76	0.0100	1.15 (1.04, 1.28)	0.24	0.0106	7.1 (2.0, 12.3)	10.2	0.0107

(B) IPCW-adjusted

		Win proportion (%)		Win ratio			Win odds			Net benefit (%)		
Time	Censoring (%)	Treatment	Control	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value	Median (95% CI)	Width of 95% CI	*p*-value
Year 1	0%	26.3	17.7	1.49 (1.12, 1.97)	0.85	0.0056	1.19 (1.05, 1.35)	0.30	0.0062	8.6 (2.4, 14.8)	12.4	0.0063
	12%	26.3	17.8	1.48 (1.12, 1.98)	0.86	0.0081	1.19 (1.05, 1.35)	0.30	0.0083	8.5 (2.5, 15.0)	12.5	0.0090
	26%	26.7	17.7	1.51 (1.14, 2.03)	0.89	0.0070	1.20 (1.06, 1.37)	0.31	0.0076	9.1 (2.9, 15.5)	12.6	0.0078
Year 2	0%	34.1	26.1	1.30 (1.04, 1.65)	0.61	0.0236	1.17 (1.02, 1.35)	0.33	0.0241	8.0 (1.0, 15.0)	14.0	0.0244
	20%	34.4	26.6	1.29 (1.02, 1.65)	0.64	0.0400	1.17 (1.01, 1.36)	0.35	0.0406	7.7 (0.5, 15.3)	14.8	0.0411
	41%	34.7	26.6	1.30 (1.03, 1.70)	0.65	0.0455	1.18 (1.02, 1.38)	0.36	0.0464	8.1 (0.8, 15.9)	15.1	0.0468
Year 3	0%	38.4	31.1	1.23 (1.00, 1.52)	0.52	0.0505	1.16 (1.00, 1.34)	0.34	0.0509	7.3 (0.0, 14.6)	14.6	0.0514
	25%	39.1	31.5	1.24 (1.00, 1.55)	0.55	0.0566	1.17 (1.00, 1.36)	0.36	0.0571	7.7 (0.1, 15.3)	15.2	0.0576
	50%	39.6	31.3	1.26 (1.00, 1.63)	0.62	0.0548	1.18 (1.00, 1.41)	0.40	0.0554	8.0 (0.5, 17.0)	16.5	0.0561

(i.e., we considered these estimated values as true win statistics) up to year 3. Then we applied independent exponentially distributed censoring, Exp(0.0004) and Exp(0.001), corresponding to 25% and 50% censoring, respectively, at year 3. As discussed in Dong et al.,[17] the Treatment group performs better over time than the Control group, and the hazards in the two groups are nonproportional without a particular pattern. We apply the inverse-probability-of-censoring weighting (IPCW) approach to adjust win statistics for independent censoring. Because we first removed all censored patients and then imposed artificial censoring, this example does not represent actual clinical trial results; it serves only as an illustration.

Figure 3 presents unadjusted versus IPCW-adjusted win proportions and win statistics. As also shown in Table 2, regardless of censoring and adjustment (unadjusted vs. IPCW-adjusted), the *p*-values for the three win statistics are very similar, as the Z-values of the corresponding statistical tests are approximately equal regardless of censoring pattern (as theoretically shown in Section 2.5). The win statistics decrease slightly from year 1 to year 3. This pattern means that a slightly larger treatment effect occurs early in the study.

Only under proportional hazards, the estimate of the win ratio is not impacted by censoring.[4,15,16] Otherwise, in general, censoring may force the observed treatment effect away from the true value in either a positive or a negative direction. In this particular example, as the amount of censoring increases, the *p*-values of the three win statistics become smaller at year 2 and year 3 compared with the corresponding *p*-values in absence of censoring (Table 2A). This means that the unadjusted win statistics are biased to a positive direction. Compared with the unadjusted win ratio, the point estimates and the 95% confidence intervals for the unadjusted win odds and net benefit are relatively stable as the amount of censoring increases. For example, at year 3, the width of the 95% confidence interval for the unadjusted win odds is 0.34, 0.29, and 0.24 for 0%, 25%, and 50% censoring. This indicates that, in presence of censoring, particularly when the proportion of censoring is not small, the win odds and the net benefit may have an advantage over the win ratio for interpreting the treatment effect.

With IPCW-adjustment (Table 2B), both the point estimates and the width of the 95% confidence intervals for the win odds and the net benefit are generally more stable over time than those for the corresponding unadjusted win odds and net benefit, as shown by the ranges of both the point estimates and interval widths. This indicates that the IPCW adjustment may be more effective at correcting bias from censoring. The correction is especially evident for the more variable win ratio (Table 2B): after the IPCW adjustment the width of the 95% confidence interval at year 3 is 0.55 with 25% censoring and 0.62 with 50% censoring, much closer to 0.52 with no censoring than the 0.61, 0.76 without the adjustment. Moreover, with 25% and 50% censoring at year 3, the *p*-values for the three IPCW-adjusted win statistics are close to the corresponding *p*-values in absence of censoring. Therefore, the IPCW adjustment corrects for bias, and it also aligns the confidence interval width to the width under no censoring. This indicates that, with an adjustment (e.g., IPCW adjustment) for censoring, the three win statistics can complement one another to show the strength of the treatment effect.

5 | APPLICATION TO CHARM STUDIES

The CHARM trial[31] was a randomized, double-blind, placebo-controlled study comparing candesartan with placebo in patients with chronic heart failure. The primary endpoint was a composite of cardiovascular death or hospitalizations for chronic heart failure. A total of 7599 patients were randomized to the two groups.

Because only a small number of patients dropped out prior to year 3, the unadjusted and IPCW-adjusted win statistics are very similar. Table 3 and Figure 4 present the IPCW-adjusted win proportions and win statistics. As in Scenarios (a) and

TABLE 3 IPCW-adjusted win statistics and 95% confidence intervals (CIs) for the CHARM program

	Win proportion (%)		Win ratio		Win odds		Net benefit (%)	
Time	Candesartan	Placebo	Win ratio (95% CI)	Width of 95% CI	Win odds (95% CI)	Width of 95% CI	Net Benefit (95% CI)	Width of 95% CI
Month 6	10.4	7.1	1.46 (1.26, 1.70)	0.44	1.07 (1.04, 1.10)	0.06	3.3 (2.0, 4.6)	2.6
Year 1	15.8	11.7	1.35 (1.20, 1.52)	0.32	1.09 (1.05, 1.12)	0.07	4.1 (2.5, 5.7)	3.2
Year 2	23.5	18.9	1.25 (1.14, 1.37)	0.23	1.10 (1.06, 1.14)	0.09	4.7 (2.7, 6.6)	3.9
Year 3	28.9	24.2	1.19 (1.10, 1.29)	0.20	1.10 (1.05, 1.15)	0.10	4.7 (2.5, 6.8)	4.4

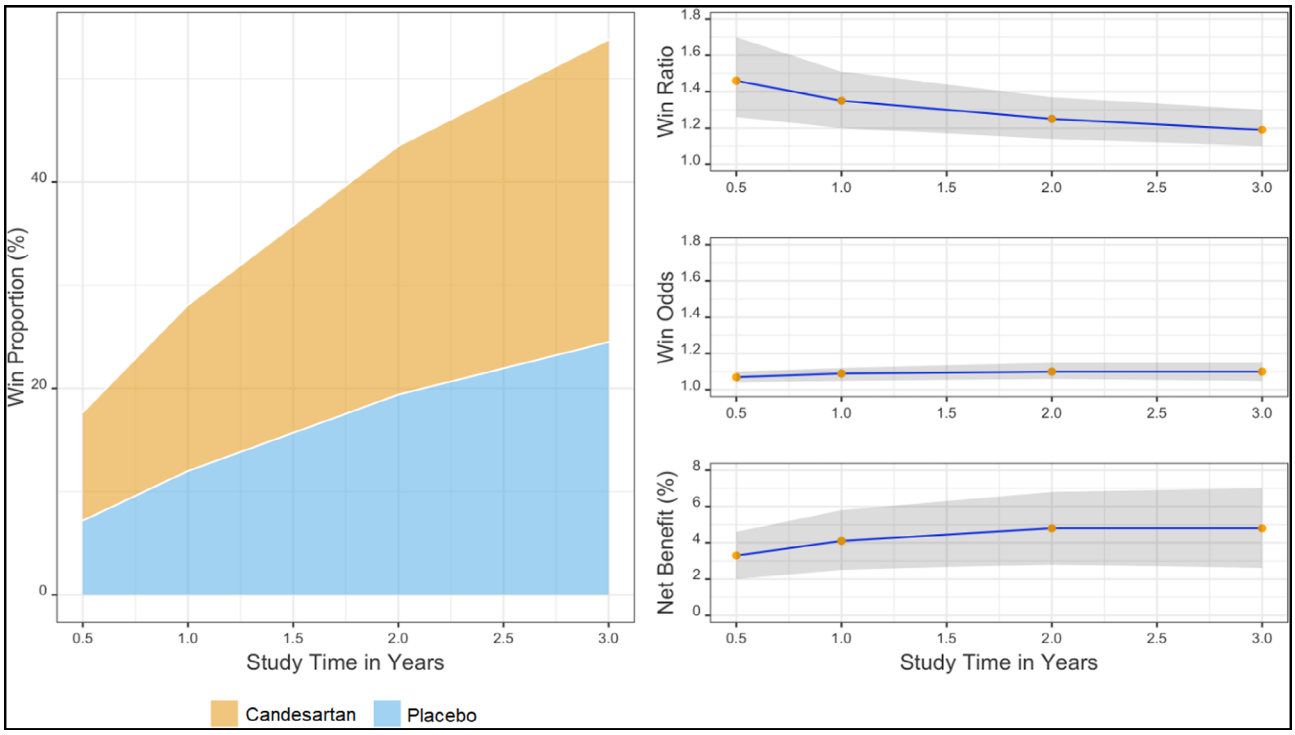

FIGURE 4 IPCW-adjusted win proportions and win statistics with 95% confidence intervals over time for the CHARM program

(b) of Simulation study 1, the width of 95% confidence intervals for the win ratio becomes narrower over time, and those for the net benefit become wider. Very interestingly, for this study as a large clinical trial, the point estimate of the win ratio declines with the follow-up time, whereas the point estimates of the win odds and the net benefit vary little with the follow-up time. Moreover, the 95% confidence interval for the win odds is very narrow; its width increases from 0.06 to 0.10, whereas the width for the win ratio decreases from 0.44 to 0.20. This may indicate that, compared with the win ratio, the win odds may have an advantage of a narrow confidence interval when the proportion of ties is not small.

6 | CONCLUSIONS AND DISCUSSION

The generalized pairwise comparisons and win statistics (in particular, win ratio and net benefit) have received increasing attention in methodological research. They also have been applied in design and analysis of Phase III clinical trials and in supporting drug approval by health authorities. However, win ratio, win odds and net benefit have been typically used separately.

The three win statistics test the same null hypothesis of equal win probabilities in the Treatment and Control groups, and they provide similar *p*-values and statistical powers because the Z-values of the corresponding statistical tests are approximately equal (Section 2.5). Therefore, in this article, we examine whether the three win statistics complement one another for analyzing the strength of the treatment effect. In the setting of time-to-event outcomes, we use simulation studies and data from a clinical trial to explain their behavior in relation to proportionality of hazards and in relation to the Mann–Whitney test. In our view, when censoring is absent or the amount of censoring is small, the three win statistics can complement one another to show the strength of the treatment effect. Specifically, the win ratio and win odds are relative quantitative measures (similar to the hazard ratio) evaluating the relative strength of one treatment group versus the other, whereas the net benefit is an absolute quantitative measure (similar to difference in response rates), bounded by -1.0 and $+1.0$, evaluating the absolute strength of one treatment group versus the other. In the presence of a positive treatment effect (i.e., win proportion for the Treatment group is higher than that for the Control group), the win ratio is always greater than the win odds (they are equal in the absence of ties).

For continuous, ordinal and binary outcomes, the win odds may be preferable to the win ratio to handle the ties more appropriately,[20] because a tie implies that the two patients in a pair had the same value of an outcome.

For time-to-event outcomes, however, censoring-induced ties do not necessarily mean that the two patients in a pair have the same value of the outcomes (details in Section 3). As noted in Oakes[4] and Dong et al.,[16] unless the proportional hazards assumption holds, the win ratio can be affected by censoring and follow-up time. The same issue arises for the win odds and the net benefit, which are sensitive to censoring-induced ties. Handling of ties caused by censoring is complex because of various censoring mechanisms (noninformative administrative censoring, informative censoring owing to drop-out or confounding intercurrent event). For example, when the censoring is primarily administrative, the win ratio may have a more meaningful clinical interpretation than the win odds. This is because the win ratio can be viewed as a special version of the win odds, by imputing $100*\pi_t/(\pi_t+\pi_c)\%$ and $100*\pi_c/(\pi_t+\pi_c)\%$ of the ties as win proportions for the Treatment and Control groups, respectively, instead of 50% of ties for each group in the win odds. This imputation approach assumes missing-at-random and is analogous to the conditional power approach based on observed data at the interim analysis. However, when censoring is primarily due to dropout or informative intercurrent events, without an adjustment for censoring, the win odds and the net benefit may have an advantage for interpreting treatment effect, as this type of censoring may force the observed win ratio substantially away from the true value in either a positive or a negative direction; with an adjustment (e.g., IPCW adjustment) for censoring, the three win statistics can complement one another to show the strength of the treatment effect.

In general, for time-to-event outcomes, comparisons among the three win statistics are subtle. Nevertheless, the Z-values of the three statistical tests are approximately equal and the tests provide similar *p*-values. Therefore, the three win statistics can complement one another to show the strength of the treatment effect; and presenting win proportions, win ratio, win odds, and net benefit together can give a more detailed picture of an analysis. One may use and present only one statistical measure (win ratio, win odds, or net benefit) for clinical trial design and analysis because the three win statistics are complementary. It is also advisable to graphically display the win statistics over follow-up time, following Finkelstein and Schoenfeld[15], to assess the variability and robustness of the win statistics, because censoring-induced ties decrease over time.

ACKNOWLEDGMENT

The authors thank two anonymous reviewers for their insightful and constructive comments and suggestions.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The simulated data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Gaohong Dong https://orcid.org/0000-0002-9848-1329
Bo Huang https://orcid.org/0000-0002-3088-9328
Johan Verbeeck https://orcid.org/0000-0002-4923-1032
Margaret Gamalo-Siebers https://orcid.org/0000-0003-2428-3298
David C. Hoaglin https://orcid.org/0000-0003-1336-181X
Tobias Mütze https://orcid.org/0000-0002-4111-1941
John Kolassa https://orcid.org/0000-0002-8246-4276

REFERENCES

1. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med*. 2010;29(30):3245-3257.
2. 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. *Eur Heart J*. 2012;33(2):176-182.
3. Dong G, Hoaglin DC, Qiu J, et al. The win ratio: on interpretation and handling of ties. *Stat Biopharm Res*. 2020;12(1):99-106.
4. Oakes D. On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika*. 2016;103(3):742-745.
5. Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharm Stat*. 2016;15(3):238-245.
6. 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.

7. 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.
8. 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. *Pharm Stat*. 2016;15(5):430-437.
9. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *J Biopharm Stat*. 2018;28(4):778-796.
10. Luo X, Qiu J, Bai S, Tian H. Weighted win loss approach for analyzing prioritized outcomes. *Stat Med*. 2017;36(15):2452-2465.
11. Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. Adjusted win ratio with stratification: calculation methods and interpretation. *Stat Methods Med Res*. 2021;30(2):580-611.
12. Verbeeck J, Ozenne B, Anderson WN. Evaluation of inferential methods for the net benefit and win ratio statistics. *J Biopharm Stat*. 2020;30(5):765-782.
13. Péron J, Buyse M, Ozenne B, Roche L, Roy P. An extension of generalized pairwise comparisons for prioritized outcomes in the presence of censoring. *Stat Methods Med Res*. 2018;27(4):1230-1239.
14. Verbeeck J, Spitzer E, de Vries T, et al. Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints. *Stat Med*. 2019;38(30):5641-5656.
15. Finkelstein DM, Schoenfeld DA. Graphing the win ratio and its components over time. *Stat Med*. 2019;38(1):53-61.
16. Dong G, Huang B, Chang Y-W, Seifu Y, Song J, Hoaglin DC. The win ratio: impact of censoring and follow-up time and use with non-proportional hazards. *Pharm Stat*. 2020;19(3):168-177.
17. Dong G, Mao L, Huang B, et al. The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring. *J Biopharm Stat*. 2020;30(5):882-899.
18. Dong G, Huang B, Wang D, Verbeeck J, Wang J, Hoaglin DC. Adjusting win statistics for dependent censoring. *Pharm Stat*. 2021;20(3):440-450.
19. Peng L. The use of the win odds in the design of non-inferiority clinical trials. *J Biopharm Stat*. 2020;30(5):941-946.
20. Brunner E, Vandemeulebroecke M, Mütze T. Win odds: an adaptation of the win ratio to include ties. *Stat Med*. 2021;40(14):3367-3384.
21. Gasparyan SB, Kowalewski EK, Folkvaljon F, et al. Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials. *J Biopharm Stat*. 2021;31(6):765-787.
22. Song J, Verbeeck J, Huang B, et al. The win odds: statistical inference and regression. *J Biopharm Stat*. 2022; (in press).
23. Mao L, Wang T. A class of proportional win-fractions regression models for composite outcomes. *Biometrics*. 2021;77(4):1265-1275.
24. Mao L, Kim K, Miao X. Sample size formula for general win ratio analysis. *Biometrics*. 2021. doi:10.1111/biom.13501
25. Mao L, Kim K. Statistical models for composite endpoints of death and nonfatal events: a review. *Stat Biopharm Res*. 2021;13(3):260-269.
26. Yang S, Troendle J, Pak D, Leifer E. Event-specific win ratios for inference with terminal and non-terminal events. *Stat Med*. 2022;41(7):1225-1241.
27. Matsouaka RA. Robust statistical inference for matched win statistics. *Stat Methods Med Res*. 2022. doi:10.1177/09622802221090761
28. Mao L, Kim K, Li Y. On recurrent-event win ratio. *Stat Methods Med Res*. 2022. doi:10.1177/09622802221084134
29. Voors AA, Angermann CE, Teerlink JR, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. *Nat Med*. 2022;28(3):568-574.
30. Lopes RD, de Barros E, Silva PGM, et al. ACTION coalition COVID-19 Brazil IV investigators. Therapeutic versus prophylactic anticoagulation for patients admitted to hospital with COVID-19 and elevated D-dimer concentration (ACTION): an open-label, multicentre, randomised, controlled trial. *Lancet*. 2021;397(10291):2253-2263.
31. Pfeffer MA, Swedberg K, Granger CB, et al. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-overall programme. *Lancet*. 2003;362:759-766.
32. Cui Y, Huang B. WINS: The R WINS Package, 2022. https://cran.r-project.org/web/packages/WINS/index.html.
33. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18(1):50-60.
34. Lehmann EL. *Elements of Large-Sample Theory*. Springer; 1999.
35. Verbeeck J, Deltuvaite-Thomas V, Berckmoes B, et al. Unbiasedness and efficiency of non-parametric and UMVUE estimators of the probabilistic index and related statistics. *Stat Methods Med Res*. 2021;30(3):747-768.
36. 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. *Pharm Stat*. 2018;17(3):202-213.

How to cite this article: Dong G, Huang B, Verbeeck J, et al. Win statistics (win ratio, win odds, and net benefit) can complement one another to show the strength of the treatment effect on time-to-event outcomes. *Pharmaceutical Statistics*. 2023;22(1):20-33. doi:10.1002/pst.2251