The Win Odds: Statistical Inference and Regression

James Song[a*], Johan Verbeeck[b], Bo Huang[c], David C. Hoaglin[d], Margaret Gamalo-Siebers[e], Yodit Seifu[f], Duolao Wang[g], Freda Cooner[h], and Gaohong Dong[a]

Abstract: Generalized pairwise comparisons and win statistics (i.e., win ratio, win odds and net benefit) are advantageous in analyzing and interpreting a composite of multiple outcomes in clinical trials. An important limitation of these statistics is their inability to adjust for covariates other than by stratified analysis. Because the win ratio does not account for ties, the win odds, a modification that includes ties, has attracted attention. We review and combine information on the win odds to articulate the statistical inferences for the win odds. We also show alternative variance estimators based on the exact permutation and bootstrap as well as statistical inference via the probabilistic index. Finally, we extend multiple-covariate regression probabilistic index models to the win odds with a univariate outcome. As an illustration we apply the regression models to the data in the CHARM trial.

Keywords: win ratio, win odds, net benefit, win statistics, probabilistic index model, bootstrap, permutation

---

[a] BeiGene, Ridgefield Park, New Jersey, USA
[b] DSI, I-Biostat, University Hasselt, Hasselt, Belgium
[c] Pfizer Inc., Groton, Connecticut, USA
[d] Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, Massachusetts, USA
[e] Pfizer Inc., Collegeville, Pennsylvania, USA
[f] Bristol Myers Squibb, Berkeley Heights, New Jersey, USA
[g] Liverpool School of Tropical Medicine, Liverpool, UK
[h] Amgen, Thousand Oaks, California, USA
*Correspondence: James Song, BeiGene, 55 Challenger Road, Ridgefield Park, New Jersey, USA. Email: james.song@beigene.com

1. Introduction

In generalized pairwise comparisons (Buyse, 2010), each patient in the Treatment group is compared with every patient in the Control group. For a pair, if the patient in the Treatment group has a better outcome than the Control patient, it is called a ‘win’ for the Treatment group; if on the other hand, the Control patient has a better outcome, it is called a 'loss' for the Treatment group and a “win” for the Control group. If a 'win' or a 'loss' cannot be established, the result is a 'tie'. Because the win ratio (ratio of win proportions; Pocock et al., 2012), the net benefit (difference in win proportions; Buyse, 2010) and the win odds (odds of win proportions; Dong et al., 2020a) are derived from the same win proportions and can be used to test the same hypothesis of equal win probabilities in the two groups, Dong et al. (2021) proposed the unifying term *win statistics*, adopting the idea of “*win*” from Pocock et al. (2012). R packages such as the WINS package by Cui and Huang (2022) for the calculation and inference of win statistics are available.

Inclusion of ties in the win odds has some benefits compared with the win ratio since the win odds considers a tie as a half win for the Treatment group and a half win for the Control group. Peng (2020) discussed the win odds in the design of a non-inferiority (NI) trial with a composite endpoint of prioritized multiple outcomes, since ties in NI trials may reflect comparable treatment effect and the number of ties may be substantial. Brunner, Vandemeulebroecke and Mütze (2021) pointed out that a larger number of ties provide evidence of greater similarity of the two arms and should not be ignored. In addition, through several cases they demonstrated that an increase in the proportion of ties can result in an increase in the win ratio while the win odds remains unchanged. Therefore, they concluded that the win odds should be preferred over the win ratio. Brunner, Vandemeulebroecke and Mütze (2021) and Gasparyan et al (2021a) provided a way to calculate variance and confidence interval for the win odds when the response variable is univariate and can

be ordered. Further, Gasparyan et al (2021b) provided the sample size and power for the win odds with a univariate outcome. Matsouaka (2022) developed a robust statistical inference for the matched win odds and win statistics in general. Though beyond the scope of this article, methods suggested for inference with the win ratio and net benefit can be adapted to the win odds.

For a continuous non-normal or ordinal outcome, the Mann-Whitney U test (also known as the Wilcoxon-Mann-Whitney test) (Wilcoxon, 1945; Mann and Whitney, 1947) is often used to compare two treatment groups (i.e., each patient in the Treatment group is compared with every patient in the Control group). The *probabilistic index* (called the *relative effect* in Brunner et al. 2021), typically estimated as the ratio of the Mann-Whitney U to the total number of pairwise comparisons, was first formally used by Acion et al. (2006). It has been extensively studied during the past few decades in stress-strength models in reliability theory, receiver operating characteristic (ROC) curve analysis in diagnostic test accuracy, medical applications, and other areas (e.g., Church and Harris, 1970; Kotz, Lumelskii and Pensky, 2003; Pepe, 2010; Fay and Malinovsky, 2018; Verbeeck et al., 2021). In one milestone in the development of the probabilistic index, Thas et al. (2012) introduced a class of regression models, the Probabilistic Index Models (PIMs). Subsequently, PIMs have been further developed and established (e.g., De Neve and Thas, 2015; Meys et al., 2020). Because generalized pairwise comparisons (Buyse, 2010) is an extension of the Wilcoxon-Mann-Whitney test, it is related to the probabilistic index. For example, the net benefit can be shown as a linear transformation of the probabilistic index. Verbeeck et al. (2021) showed that, with a univariate outcome and no missing or censored data, the probabilistic index and net benefit are always unbiased and efficient in detecting a treatment effect in realistic clinical scenarios.

Current methods for win statistics have a major limitation: when a win statistic is used to

analyze a randomized clinical trial, it is assumed that two treatment groups are balanced/comparable on patients' characteristics at baseline. When two treatment groups are not balanced, it may be desirable to take possible confounding factors into account. Therefore, this restriction has limited the application of win statistics. Dong et al. (2018) proposed the stratified win ratio. Mao and Wang (2021) introduced a regression method for the win ratio to provide flexibility in adjusting for confounders in the model. Furthermore, Gasparyan et al (2021a) provided a unified theory of win odds estimation in the presence of stratification and adjustment for one numeric variable. The probabilistic index models (Thas et al, 2012) allow adjustments for multiple covariates, but those models have not been explored for the win odds.

In this article, we focus on the (unmatched) win odds. We present statistical inferences for the win odds, including showing how to extend alternative variance estimators by the exact permutation and bootstrap to the win odds, and we demonstrate a regression for the win odds, following the probabilistic index models (Thas et al, 2012). For the matched win odds following a matched pair design (e.g., one eye vs. the other eye of the same patients) or for the paired patients matched on baseline characteristics, risk profiles and other factors, we refer to Matsouaka (2022).

2. Calculation of the win odds

Generalized pairwise comparisons (Buyse, 2010) result in three possible outcomes for comparing a patient from the Treatment group with a patient from the Control group: the Treatment patient wins, the Control patient wins, or the two patients are tied. Let π_t, π_c and π_{tie} be the probabilities of these three outcomes, for which $\pi_t + \pi_c + \pi_{tie} = 1$. The subscripts $_t$ and $_c$ denote the Treatment and Control groups, respectively. Following Dong et al. (2020a), the win odds (*WO)* is defined as follows:

$$WO = \frac{\pi_t + 0.5\pi_{tie}}{\pi_c + 0.5\pi_{tie}} = \frac{\pi_t + 0.5(1-\pi_t-\pi_c)}{\pi_c + 0.5(1-\pi_t-\pi_c)} = \frac{\pi_t + 0.5(1-\pi_t-\pi_c)}{1-[\pi_t + 0.5(1-\pi_t-\pi_c)]}. \quad (1)$$

The *WO* can be theoretically calculated as an integral or estimated using the counting approach. The win ratio (*WR*), the net benefit (*NB*) and the probabilistic index (*PI*) can also be similarly calculated as $WR = \frac{\pi_t}{\pi_c}$, $NB = \pi_t - \pi_c$ and $PI = \pi_t + 0.5\pi_{tie}$.

2.1 Theoretical calculation via integral

Let *Y* be a response variable. If the distribution of *Y* is known, one may be able to theoretically calculate the *WO* (and the other *win statistics*). As an example, suppose *Y* is a time-to-event variable that follows an exponential distribution with parameter λ, *Y~Exp(λ)*. The survival function of *Y* at time *x* is $S(x) = e^{-\lambda x}$. Following Oakes (2016), the win probabilities by time *x* can be calculated as

$$\pi_t(x) = -\int_0^x S_t(x)\, dS_c(x) = \int_0^x e^{-\lambda_t x}\, \lambda_c e^{-\lambda_c x} dx = \lambda_c \int_0^x e^{-(\lambda_t+\lambda_c)x}\, dx$$

$$= \frac{\lambda_c}{\lambda_t+\lambda_c}[1 - e^{-(\lambda_t+\lambda_c)x}],$$

$$\pi_c(x) = \frac{\lambda_t}{\lambda_t+\lambda_c}[1 - e^{-(\lambda_t+\lambda_c)x}].$$

Suppose λ_t= 0.0693 and λ_c = 0.1155 and assume that all patients are followed to 12 months. Then at 12 months, $\pi_t(12) = \frac{0.1155}{0.0693+0.1155}\left[1 - e^{-(0.0693+0.1155)*12}\right] = 0.557$, $\pi_c(12) = 0.334$ and $\pi_{tie}(12) = 1 - \pi_t(12) - \pi_c(12) = 0.109$. Therefore, *WO* = 1.57, *WR* = 1.67, *NB* = 22.3% and *PI* = 0.61 at 12 months.

This example demonstrates a theoretical calculation for a single time-to-event endpoint without censoring. Finkelstein and Schoenfeld (2019) presented some examples with two prioritized outcomes. However, it is challenging to perform a similar calculation for prioritized multiple time-to-event endpoints. Moreover, the mathematical derivations of the variance and confidence interval are not available.

2.2 Counting approach following U-statistics

More conveniently, the *WO*, like the other *win statistics*, can be estimated using the counting approach. The main idea is to compare each patient in a treatment group with every patient in the other group. Within each pair, the comparison starts with the most important outcome, and uses lower-priority outcomes only if higher-priority outcomes have not occurred or result in a tie. When comparing patient *i* (*i* = 1, 2, …, N_t) in the Treatment group and patient *j* (*j*=1, 2, …, N_c) in the Control group, we define the kernel functions *K* and *L* as follows:

$$K_{ij} = 1,\ if\ Treatment\ patient\ i\ \boldsymbol{wins}\ over\ Control\ patient\ j$$

$$= 0, \text{otherwise.} \quad (2a)$$

$$L_{ij} = 1,\ if\ Control\ patient\ j\ \boldsymbol{wins}\ over\ Treatment\ patient\ i$$

$$= 0, \text{otherwise.} \quad (2b)$$

The numbers of wins are $n_t = \sum_{i=1}^{N_t}\sum_{j=1}^{N_c} K_{ij}$ and $n_c = \sum_{i=1}^{N_t}\sum_{j=1}^{N_c} L_{ij}$ for the Treatment and Control groups, respectively. The corresponding win proportions are $P_t = n_t/N_tN_c$ and $P_c = n_c/N_tN_c$, and the proportion of ties is $P_{tie} = 1 - P_t - P_c$. These three proportions are the estimates of the corresponding probabilities, namely, $\hat{\pi}_t = P_t$, $\hat{\pi}_c = P_c$ and $\hat{\pi}_{tie} = P_{tie}$. Therefore, the *WO* can be estimated as

$$\widehat{WO} = \frac{P_t + 0.5P_{tie}}{P_c + 0.5P_{tie}} = \frac{P_t + 0.5(1 - P_t - P_c)}{P_c + 0.5(1 - P_t - P_c)}. \quad (3)$$

Similarly, *WR*, *NB* and *PI* can be estimated as $\widehat{WR} = P_t/P_c$, $\widehat{NB} = P_t - P_c$ and $\widehat{PI} = P_t + 0.5P_{tie}$. Since *WR*, *WO* and *NB* are constructed using the same win proportions and test the same null hypothesis, H_0: $\pi_t = \pi_c$, Dong et al. (2021) proposed the unifying term *win statistics*, adopting the idea of "*win*" from Pocock et al. (2012).

3. Variance estimators for win odds

Some estimators have been proposed for the variance of the win odds. Additionally, methods for estimating the variance of the win ratio and the net benefit can be adapted to the win odds. Simulation studies (Verbeeck et al. 2020) have shown that the exact permutation method provides better Type I error rate control, coverage probability and variance estimation than the U-statistic method for the net benefit, but less so for the win ratio; and the U-statistic inference proposed by Dong et al. (2016) performs better than the U-statistic inference proposed by Bebu and Lachin (2016).

3.1 Estimators based on U-statistics

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

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

By the delta method, $log(\widehat{WO})$ is asymptotically normally distributed. The variance can be derived as (Dong et al., 2021)

$$\hat{\sigma}^2_{log(WO)} = (\hat{\sigma}_t^2 - 2\hat{\sigma}_{tc} + \hat{\sigma}_c^2)\left(\frac{1}{\hat{\gamma}} + \frac{1}{N_t N_c - \hat{\gamma}}\right)^2/4, \tag{5}$$

where $\hat{\gamma} = \hat{\theta}_t + 0.5(N_t N_c - \hat{\theta}_t - \hat{\theta}_c)$. Under the null hypothesis H_0: $\pi_t = \pi_c$, θ_t and θ_c can be estimated as

$$\hat{\theta}_t = \hat{\theta}_c = (n_t + n_c)/2, \tag{6a}$$

and the variance of $log(WO)$ can be simplified as

$$\hat{\sigma}^2_{log(WO)} = \frac{\hat{\sigma}_t^2 - 2\hat{\sigma}_{tc} + \hat{\sigma}_c^2}{(N_t N_c/2)^2}. \tag{6b}$$

Under the alternative hypothesis, their maximum likelihood estimators can be used:

$$\hat{\theta}_t = n_t, \tag{7a}$$

$$\hat{\theta}_c = n_c. \quad (7b)$$

The calculations for $\hat{\sigma}_t^2$, $\hat{\sigma}_c^2$ and $\hat{\sigma}_{tc}$ can be found in Dong et al. (2016, 2021) and Bebu and Lachin (2016).

Alternative estimators

By considering a tie as a half win for the Treatment group and a half win for the Control group, the kernel functions *K* and *L* defined in (2a) and (2b) can be modified as

$$K'_{ij} = 1, \ if\ Treatment\ patient\ i\ \boldsymbol{wins}\ over\ Control\ patient\ j$$

$$= 0.5, \textit{ if the two patients are tied}$$

$$= 0, \text{ otherwise,} \quad (8a)$$

$$L'_{ij} = 1, \ if\ Control\ patient\ j\ \boldsymbol{wins}\ over\ Treatment\ patient\ i$$

$$= 0.5, \textit{ if the two patients are tied}$$

$$= 0, \text{ otherwise.} \quad (8b)$$

By plugging K'_{ij} and L'_{ij} into the formulas for the point estimate and variance estimate for the win ratio provided in Dong et al. (2016), one can directly obtain the point and variance estimates for the win odds.

On the other hand, Peng (2020) provided a direct way to estimate the variance of the win odds by applying K'_{ij} and L'_{ij}, based on the estimator by Bebu and Lachin (2016) for the win ratio. Peng (2020) gives the details of the calculations.

Additionally, Gasparyan et al. (2021a and 2021b) defined the individual win proportions and subsequently calculated the win odds and its variance.

3.2 Permutation and bootstrap-based estimators

Originally, Buyse (2010) and Pocock et al. (2012) suggested randomization (i.e., re-sampling permutation) and bootstrap tests for inference with the net benefit and the win ratio. However, re-

sampling is computationally intensive, and the results are less replicable (they depend on randomly generated permutations) since a precise estimate may require a large number (>700,000) of random permutations. Verbeeck et al. (2020) derived a closed-form formula for the exact permutation and bootstrap variance of *NB* and *WR*. Their approach can easily be extended to *WO*, by realizing that

$$\hat{\sigma}^2_{log(WO)} = (N_t N_c)^2 \, \hat{\sigma}^2_{NB} \left(\frac{1}{\hat{\gamma}} + \frac{1}{N_t N_c - \hat{\gamma}}\right)^2 / 4. \quad (9)$$

The variance of the exact permutation distribution requires a modification of the kernel functions *K* and *L* defined in (2a) and (2b), by allowing comparisons both between treatment arms and within treatment arms. When comparing patient *m* (*m* = 1, 2, …, *N*) with patient *n* (*n* = 1, 2, …, *N*), where $N = N_t + N_c$, the kernel functions are modified as

$$K^*_{mn} = 1, \; if \; Treatment \; patient \; m \; \boldsymbol{wins} \; over \; Control \; patient \; n$$

$$= 0, \text{ otherwise,} \quad (10a)$$

$$L^*_{mn} = 1, \; if \; Control \; patient \; n \; \boldsymbol{wins} \; over \; Treatment \; patient \; m$$

$$= 0, \text{ otherwise.} \quad (10b)$$

The two matrices K^*_{mn} and L^*_{mn} have *N* rows and *N* columns. The numbers of wins from the Treatment and Control rows are $r^*_{tm} = \sum_{n=1}^{N} K^*_{mn}$ and $r^*_{cm} = \sum_{n=1}^{N} L^*_{mn}$, respectively. The exact permutation variance for the net benefit is

$$\hat{\sigma}^2_{NB} = \frac{\sum_{m=1}^{N} (r^*_{tm} - r^*_{cm})^2}{N(N-1)N_t N_c} \quad (11)$$

Plugging (11) into (9) results in the exact permutation variance of *WO*.

The exact bootstrap variance requires only between-treatment-arm comparisons. Using the kernel functions *K* and *L* defined in (2a) and (2b), the numbers of wins from row *i* are $r_{ti} = \sum_{j=1}^{N_c} K_{ij}$ and $r_{ci} = \sum_{j=1}^{N_c} L_{ij}$, and the number of wins from column *j* are $c_{tj} = \sum_{i=1}^{N_t} K_{ij}$ and $c_{cj} = \sum_{i=1}^{N_t} L_{ij}$, respectively. The exact bootstrap variance for the net benefit is:

$$\hat{\sigma}^2_{NB} = \frac{1}{(N_t N_c)^2}\left(\frac{(N_t-1)}{N_t}\sum_{i=1}^{N_t}(r_{ti}-r_{ci})^2 + \frac{(N_c-1)}{N_c}\sum_{j=1}^{N_c}(c_{tj}-c_{cj})^2 + n_t + n_c - \frac{(N_t+N_c-1)}{N_t N_c}\sum_{i=1}^{N_t}\sum_{j=1}^{N_c}(K_{ij}-L_{ij})^2\right) \quad (12)$$

Plugging (12) into (9) results in the exact bootstrap variance of *WO*.

3.3 Rank-based estimator

Instead of using the kernel functions (2a) and (2b), the win odds can equivalently be estimated using ranks (Brunner et al. 2021; Gasparyan et al. 2021a) when the outcome is a single continuous or ordinal endpoint. Based on the ranks (mid-ranks must be used in case of ties), the probabilistic index and the win odds can be estimated by:

$$\widehat{PI} = \frac{1}{N}(\bar{R}_t - \bar{R}_c) + \frac{1}{2}, \quad (13a)$$

$$\widehat{WO} = \frac{\widehat{PI}}{1-\widehat{PI}} \quad (13b)$$

where $\bar{R}_t$ and $\bar{R}_c$ are the mean ranks in the Treatment and Control groups, respectively. By the delta method, the asymptotic variance of the logarithm of the win odds can be estimated by (Brunner et al. 2021)

$$\hat{\sigma}^2_{\log(WO)} = \frac{\hat{\sigma}^2_{PI}}{\widehat{PI}(1-\widehat{PI})}. \quad (14)$$

The estimate of the asymptotic variance for the probabilistic index, $\hat{\sigma}^2_{PI}$, can easily be obtained from statistical software such as SAS (SAS Institute Inc 2018). Brunner et al. (2021) and Gasparyan et al. (2021a and 2021b) give further details of the rank-based method when the outcome is a single continuous or ordinal endpoint.

4. Regression

4.1 Probabilistic index

Following the notation by Thas et al. (2012), let *Y* be a response variable and *X* be *p*-dimensional covariates, and (Y, X) and (Y^*, X^*) are independent and identically distributed (i.i.d.) with a joint density function. The *Probabilistic index* is defined as

$$P(Y \succcurlyeq Y^*|X, X^*) = P(Y > Y^*|X, X^*) + \frac{1}{2} P(Y = Y^*|X, X^*). \quad (15)$$

4.2 Probabilistic index models

The *Probabilistic index models* (Thas et al., 2012) are defined as

$$P(Y \succcurlyeq Y^*|X, X^*) = m(X, X^*; \beta), \quad (16)$$

Where the function $m(\cdot)$ has a range of [0, 1] and β is a p-dimensional parameter vector. This model requires semiparametric theory for inference on β because it does not make a full distributional assumption on the conditional distribution of *Y* given *X*.

Let $Z = X - X^*$ and $m(X, X^*; \beta) = g^{-1}(Z^T\beta)$, where $g(\cdot)$ is a link function (e.g., logit or probit). Then (16) has the general form of the probabilistic index model:

$$P(Y \succcurlyeq Y^*|X, X^*) = g^{-1}(Z^T\beta). \quad (17)$$

The calculations for the probabilistic index model are implemented in the R package **pim** (Meys et al., 2017).

4.3 Regression for the win odds

The regression for the win odds can use the probabilistic index model. Indeed, since WO is related to PI via $WO = \frac{PI}{1-PI}$, the natural link function is the logit:

$$logit\big(P(Y \succcurlyeq Y^*|X, X^*)\big) = Z^T\beta.$$

Since $log\left(\frac{P(Y \succcurlyeq Y^*|X, X^*)}{1-P(Y \succcurlyeq Y^*|X, X^*)}\right) = log(WO) = Z^T\beta$, we can obtain

$$WO = exp(Z^T\beta), \quad (18a)$$

$$\widehat{WO} = exp\big(Z^T\hat{\beta}\big). \quad (18b)$$

4.4 Example

As an example, consider a clinical study with a time-to-event outcome as the primary endpoint. Let *X* = 1 denote the Treatment group and *X* = 0 denote the Control group. For illustration, we

assume that the event time Y follows an exponential distribution with parameter $\lambda = 0.0693$ in the Treatment group and $\lambda = 0.1155$ in the Control group. Therefore, theoretically, the hazards are proportional with hazard ratio (HR) = 0.6 (win ratio = 1.67) and win odds = 1.57. We randomly generate 1000 samples and obtain median win odds = 1.57 and 95% percentile interval (1.25, 1.98).

By applying the probabilistic index model with the logit link function, we have $Z_{ij} = X_i - X_j = 1$ to compare the Treatment group against the Control group:

$$logit\left(P\left(Y_i \succcurlyeq Y_j | X_i = 1, X_j = 0\right)\right) = (X_i - X_j)\beta = Z_{ij}\beta = \beta\,,$$

Therefore, WO = exp(β). From the 1000 samples, we get the median $\hat{\beta}$ = 0.452 and the 95% percentile interval = (0.221, 0.681). Hence, the median $\widehat{WO}$ = exp(0.452) = 1.57 and the 95% percentile interval = (1.25, 1.98).

4.5 Application to CHARM studies

The CHARM trial was a randomized, double-blind, controlled trial comparing candesartan with placebo in patients with chronic heart failure (Pfeffer et al., 2003). The primary endpoint was cardiovascular death or hospitalizations due to chronic heart failure. For illustration, we use the R package **pim** to explore regression analysis of the win odds in the CHARM data. Since the current **pim** package (Meys et al., 2020) considers only one single continuous or ordinal outcome, we analyze only the 1-year survival outcome, and we assume that patients who were censored were alive at the end of Year 1.

From (3) through (6b), we obtain $\widehat{WO}$ = 1.046 and $\hat{\sigma}^2_{log(WO)}$= 0.000126 (or $\hat{\sigma}_{\log(WO)}$= 0.01122), and the 95% confidence interval for WO is (1.023, 1.069). The probabilistic index models with the treatment group as a single covariate give almost the same result: $\hat{\beta}$ = 0.04499, $\widehat{WO} = \exp(\hat{\beta}) = 1.046$ and the standard error $SE_{\hat{\beta}}$ = $\hat{\sigma}_{\log(WO)}$= 0.01119.

For illustration, in a multivariate regression analysis using the probabilistic index model with the logit link function, we consider the first few most significant covariates reported in Pocock et al. (2006): age (per 10 years over age 60 years), diabetes status at baseline, and prior hospitalization. Table 1 shows the parameter estimates $\hat{\beta}$ and their standard errors (i.e., $SE_{\hat{\beta}}$). Table 2 shows the win odds and its 95% confidence interval by $\widehat{WO} = \exp(\hat{\beta})$ and (exp($\hat{\beta} - Z_{1-\alpha/2}SE_{\hat{\beta}}$), exp($\hat{\beta} + Z_{1-\alpha/2}SE_{\hat{\beta}}$)), where α = 0.05.

5. Discussion

Win statistics (win ratio, net benefit and win odds) have been increasingly used as a clinically meaningful measure to quantify treatment benefit in medical research because they consider more important outcomes first in pairwise comparisons. They compare each subject in the Treatment arm with every subject in the Control arm, starting with the most important outcome and evaluating lower-priority outcomes if and only if higher-priority outcomes cannot determine a win. Thus, lower-priority outcomes do not "mask" more important outcomes when they occur earlier. The win statistics have been used in the design and analysis of clinical trials (e.g., NCT04001504, NCT04847557 and NCT04510493 as examples of Phase III studies, registered in ClinicalTrials.gov) and in supporting drug approval by regulatory authorities (e.g., in 2019 the FDA approved tafamidis for treatment of cardiomyopathy on the basis of the ATTR-ACT trial with the win ratio as the primary analysis). The stratified win ratio (Dong et al., 2018) 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 (Voors et al., 2022) and the ACTION study of therapeutic versus prophylactic anticoagulation for patients admitted to hospital with COVID-19 and elevated D-dimer concentration (Lopes et al., 2021).

The win statistics depend on follow-up time especially when a time-to-event variable is used. To demonstrate long-term benefit from a more important but less frequent variable by the win statistics, one may need to follow patients longer (Dong et al, 2020b). In addition, censoring could cause bias in win statistics (e.g., Oakes, 2016; Mao, 2019). To correct censoring bias, Dong et al. (2020c and 2021) developed the IPCW (inverse-probability-of-censoring weighting) adjusted win statistics in the presence of independent or dependent censoring. On the other hand, win ratio, win odds and net benefit have typically been used separately. The three win statistics are based on the same win proportions, and they test the same null hypothesis of equal win probabilities in the two groups. Therefore, in general, win ratio, win odds and net benefit complement one another for assessing the magnitude and strength of the treatment effect. We will report this work separately.

In this article, we focus on the win odds, which differs from the win ratio by including ties in accounting for results from all pairwise comparisons. We also summarize the calculations and variance estimators for the win odds. Although inferential methods (U-statistics, permutation and bootstrap, rank-based) for the win statistics have been suggested, one main disadvantage is the inability to take into account covariates other than in stratified analyses. Adjustment of win statistics for important prognostic factors would yield more appropriate estimates of treatment effect, and thus aid interpretation of the results. This is the first article that explores regression analysis for the win odds using the probabilistic index model with a logit link function. We illustrate this approach using the CHARM dataset. Also, the identity link function can be used in probabilistic index models to indirectly estimate the net benefit.

Win statistics and generalized pairwise comparisons are particularly appealing in the setting of multiple prioritized outcomes because they consider the more important endpoints first, while also assessing the overall effect across multiple outcomes. The current probabilistic index

models (Thas et al., 2012 and Myers et al. 2020) are limited to a single continuous or ordinal outcome. To illustrate regression with multiple covariates for the win odds, we applied these models to the 1-year survival data in the CHARM studies and assumed that the patients who were censored were alive at the end of Year 1. Further research should extend the probabilistic index models for win statistics to multiple outcomes and the presence of censoring for time-to-event endpoints. Alternative regression models, such as the proportional win-fractions regression model (Mao and Wang, 2021) for composite fatal and non-fatal outcomes, can be used to account for confounding covariates and handle censoring.

References

1. Acion, L., Peterson, J., Temple, S., and Arndt, S. 2006. Probabilistic index: An intuitive nonparametric approach to measuring the size of treatment effects. *Statistics in Medicine* 25:591–602.
2. Bebu, I., and J. Lachin. 2016. Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics* 17:178–187.
3. Brunner E, Vandemeulebroecke M, Mütze T. 2021. Win odds: An adaptation of the win ratio to include ties. *Statistics in Medicine* 40(14):3367-3384.
4. Buyse M. 2010. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine* 29(30):3245-3257.
5. Church JD and Harris B. 1970. The estimation of reliability from stress-strength relationships. *Technometrics* 12: 49–54
6. Cui Y, Huang B. WINS: The R WINS Package. 2022. https://cran.r-project.org/web/packages/WINS/index.html.
7. De Neve J. and Thas O. 2015. A regression framework for rank tests based on the probabilistic

index model. *Journal of the American Statistical Association* 110(511):1276–1283.

8. Dong G, Li D, Ballerstedt S, Vandemeulebroecke M. 2016. A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components. *Pharmaceutical Statistics* 15(5):430-437.
9. Dong, G., J. Qiu, D. Wang, and M. Vandemeulebroecke. 2018. The stratified win ratio. *Journal of Biopharmaceutical Statistics* 28(4):778–796.
10. Dong G, Hoaglin DC, Qiu J, Matsouaka, RA, Chang Y, Wang J, Vandemeulebroecke M. 2020a. The win ratio: on interpretation and handling of ties. *Statistics in Biopharmaceutical Research* 12(1):99-106.
11. Dong G, Huang B, Chang Y-W, Seifu Y, Song J, Hoaglin DC. 2020b. The win ratio: Impact of censoring and follow-up time and use with nonproportional hazards. *Pharmaceutical Statistics* 19(3):168–177
12. Dong G, Mao L, Huang B, Gamalo-Siebers M, Wang J, Yu G, Hoaglin DC. 2020c. The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring. *Journal of Biopharmaceutical Statistics.* 30(5):882-899.
13. Dong G, Huang B, Wang D, Verbeeck J, Wang J, Hoaglin DC. 2021. Adjusting win statistics for dependent censoring. *Pharmaceutical Statistics* 20(3):440-450.
14. Fay MP, Malinovsky Y. 2018. Confidence intervals of the Mann-Whitney parameter that are compatible with the Wilcoxon-Mann-Whitney test. *Statistics in Medicine* 37(27):3991-4006.
15. Finkelstein DM, Schoenfeld DA. 2019. Graphing the win ratio and its components over time. *Statistics in Medicine* 38(1):53-61.
16. Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. 2021a. Adjusted win

ratio with stratification: Calculation methods and interpretation. *Statistical Methods in Medical Research*. 30(2):580-611.

17. Gasparyan SB, Kowalewski E, Folkvaljon F, Bengtsson O, Buenconsejo J, Adler J, Koch GG. 2021b. Power and sample size calculation for the win odds test: Application to an ordinal endpoint in COVID-19 trials. *Journal of Biopharmaceutical Statistics* 31(6):765-787
18. Kotz S, Lumelskii Y and Pensky M. 2003. The stress-strength model and its generalizations. Singapore: World Scientific Press,
19. Lopes RD, de Barros E Silva PGM, Furtado RHM, Macedo AVS, Bronhara B, Damiani LP, Barbosa LM, de Aveiro Morata J, Ramacciotti E, de Aquino Martins P, de Oliveira AL, Nunes VS, Ritt LEF, Rocha AT, Tramujas L, Santos SV, Diaz DRA, Viana LS, Melro LMG, de Alcântara Chaud MS, Figueiredo EL, Neuenschwander FC, Dracoulakis MDA, Lima RGSD, de Souza Dantas VC, Fernandes ACS, Gebara OCE, Hernandes ME, Queiroz DAR, Veiga VC, Canesin MF, de Faria LM, Feitosa-Filho GS, Gazzana MB, Liporace IL, de Oliveira Twardowsky A, Maia LN, Machado FR, de Matos Soeiro A, Conceição-Souza GE, Armaganijan L, Guimarães PO, Rosa RG, Azevedo LCP, Alexander JH, Avezum A, Cavalcanti AB, Berwanger O; ACTION Coalition COVID-19 Brazil IV Investigators. 2021. 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* 397(10291):2253-2263.
20. Mann, H. and Whitney, D. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18:50–60.
21. Mao, L. 2019. On the alternative hypotheses for the win ratio. *Biometrics* 75(1): 347-351.
22. Mao L, Wang T. 2021 A class of proportional win-fractions regression models for composite

outcomes. *Biometrics* 77(4):1265-1275.

23. Matsouaka RA. 2022. Robust statistical inference for matched win statistics. *Statistical Methods in Medical Research*. doi:10.1177/09622802221090761.
24. Meys, J., De Neve, J., Sabbe, N., and Guimaraes de Castro Amorim, G. 2020. pim: Fit Probabilistic Index Models. R package version 2.0.2. Available at https://cran.r-project.org/web/packages/pim.
25. Oakes, D. 2016. On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika* 103:742-745.
26. Pfeffer MA, Swedberg K, Granger CB, et al. 2003. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 362:759-766.
27. Peng L. 2020. The use of the win odds in the design of non-inferiority clinical trials. *Journal of Biopharmaceutical Statistics* 30(5):941-946.
28. Pepe MS. 2003. The statistical evaluation of medical tests for classification and prediction. Oxford, UK: Oxford University Press.
29. Pocock SJ, Ariti CA, Collier TJ, Wang D. 2012. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal* 33(2):176-182.
30. Pocock SJ, Wang D, Pfeffer MA, et al. 2006. Predictors of mortality and morbidity in patients with chronic heart failure. *European Heart Journal* 27:65-75.
31. SAS Institute Inc. 2018. *SAS/STAT®: 15.1 user's guide*. Cary, NC: SAS Institute Inc.
32. Thas, O., De Neve, J., Clement, L., and Ottoy, J.P. 2012. Probabilistic index models (with discussion). *Journal of the Royal Statistical Society - Series B* 74:623–671.

33. Verbeeck J, Deltuvaite-Thomas V, Berckmoes B, et al. 2021. Unbiasedness and efficiency of non-parametric and UMVUE estimators of the probabilistic index and related statistics. *Statistical Methods in Medical Research* 30(3):747-768.
34. Verbeeck J, Ozenne B, Anderson WN. 2020. Evaluation of inferential methods for the net benefit and win ratio statistics. *Journal of Biopharmaceutical Statistics* 30(5):765-782.
35. Voors AA, Angermann CE, Teerlink JR, Collins SP, Kosiborod M, Biegus J, Ferreira JP, Nassif ME, Psotka MA, Tromp J, Borleffs CJW, Ma C, Comin-Colet J, Fu M, Janssens SP, Kiss RG, Mentz RJ, Sakata Y, Schirmer H, Schou M, Schulze PC, Spinarova L, Volterrani M, Wranicz JK, Zeymer U, Zieroth S, Brueckmann M, Blatchford JP, Salsali A, Ponikowski P. 2022. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. *Nature Medicine* 228(3):568-574.
36. Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics Bulletin* 1:80–83.

Table 1 Parameter estimate and inference from the probabilistic index model for the 1-year survival outcome in the CHARM trial

Covariate	Estimate ($\hat{\beta}$)	Standard error ($SE_{\hat{\beta}}$)	Z value	Prob(>|z|)
Group (Treatment vs Control)	0.0450	0.0111	4.040	<0.001
Age (per 10 years over age 60)	-0.0494	0.0079	-6.272	<0.001
Diabetes at baseline (Yes vs No)	-0.0513	0.0134	-3.830	<0.001
Prior hospitalization (Yes vs No)	-0.0536	0.0111	-4.805	<0.001

Table 2 Win odds and 95% confidence interval for the 1-year survival outcome in the CHARM trial

Covariate	Win odds	95% confidence interval	p-value
Group (Treatment vs Control)	1.046	1.023, 1.069	<0.001
Age (per 10 years over age 60)	0.952	0.937, 0.967	<0.001
Diabetes at baseline (Yes vs No)	0.950	0.925, 0.975	<0.001
Prior hospitalization (Yes vs No)	0.948	0.927, 0.969	<0.001