The stratified win statistics (win ratio, win odds, and net benefit)

Gaohong Dong[1*], David C. Hoaglin[2], Bo Huang[3], Ying Cui[4],

Duolao Wang[5], Yu Cheng[6], and Margaret Gamalo-Siebers[7]

Abstract: The win odds and the net benefit are related directly to each other and indirectly, through ties, to the win ratio. These three win statistics test the same null hypothesis of equal win probabilities between two groups. They provide similar p-values and powers, because the Z-values of their statistical tests are approximately equal. Thus, they can complement one another to show the strength of a treatment effect. In this Short Communication, we show that the estimated variances of the win statistics are also directly related regardless of ties or indirectly related through ties. Since its introduction in 2018, the stratified win ratio has been applied in designs and analyses of clinical trials, including Phase III and Phase IV studies. This Short Communication generalizes the stratified method to the win odds and the net benefit. As a result, the relations of the three win statistics and the approximate equivalence of their statistical tests also hold for the stratified win statistics.

Keywords: Mantel-Haenszel method, generalized pairwise comparisons, Finkelstein-Schoenfeld test, win ratio, win odds, net benefit

---

[1] BeiGene, Ridgefield Park, New Jersey, USA
[2] Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, Massachusetts, USA
[3] Pfizer Inc., Groton, Connecticut, USA
[4] Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA
[5] Liverpool School of Tropical Medicine, Liverpool, UK
[6] Department of Statistics, University of Pittsburg, Pennsylvania, USA
[7] Pfizer Inc., Collegeville, Pennsylvania, USA

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

1. Introduction

Many clinical trials include composite outcomes as the primary endpoint. Conventional analysis of a composite outcome uses the time to the occurrence of a first event, ignoring subsequent events and the clinical importance of each component event. To address this limitation, following the Finkelstein-Schoenfeld test[1], the generalized pairwise comparisons (GPC)[2], and the "win" concept[3], the win statistics (win ratio[3], win odds[4], and net benefit[2]) have been proposed and applied as measures of treatment effect. The Finkelstein-Schoenfeld test is equivalent to the testing of the difference in the number of wins between two groups. The generalized pairwise comparisons (GPC) is a multivariate extension of the well-known non-parametric Mann-Whitney U test. The win ratio is a ratio of win proportions, the win odds is an odds of win proportions by dividing a tie into two half wins and assigning a half win to each treatment group, and the net benefit is a difference in win proportions. These three win statistics are based on the same underlying pairwise comparison methodology, which first ranks the components in the composite endpoint by order of importance and then evaluates each possible pair (one patient from the Treatment group and the other patient from the Control group). The evaluation starts with the most important component and considers the components in order of importance until a comparison determines a win. Otherwise, the pair is considered a tie. Thus, when lower-priority outcomes occur earlier, they do not "mask" more important outcomes.

However, the two patients in a pair may not be comparable on prognostic and other factors. Often clinical trials use a stratified randomization on known factors that influence prognosis or treatment outcome, so that patient populations are relatively homogeneous within the strata. Therefore, the analysis should take those strata into account. A stratified analysis makes pairwise comparisons (and counts wins and ties) separately within each stratum. The analysis examines

those stratum-specific results and combines them to produce an overall result, and thus removes the potential confounding effect of the stratification variable(s) from the analysis. Since its introduction in 2018, the stratified win ratio[5] has been applied in designs and analyses of clinical trials, including Phase III and Phase IV studies, such as the EMPULSE study of the SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure[6] and the ACTION study of therapeutic versus prophylactic anticoagulation for patients admitted to hospital with COVID-19 and elevated D-dimer concentration[7]. In this Short Communication, we apply the stratified method developed for the win ratio (WR) to obtain the stratified win odds (WO) and the stratified net benefit (NB).

Dong et al. [8] showed that the three win statistics can complement one another to show the strength of the treatment effect, because (1) their definitions use the same win proportions based on generalized pairwise comparisons (GPC); (2) they test the same null hypothesis of equal win probabilities between two groups; (3) for point estimates, the win odds and the net benefit are related directly to each other and are related indirectly to the win ratio through ties; (4) the estimated variances of the three win statistics are also related directly regardless of ties or indirectly related through ties (Section 3.2 gives the details); and (5) Z-values of statistical tests of the three win statistics are approximately equal (i.e., the three win statistics provide similar p-values and powers). In this Short Communication, we show that the relations among the three win statistics and the approximate equivalence of their statistical tests also apply to the stratified win statistics. For calculations, we use the R package WINS by Cui and Huang[9].

2. The stratified win statistics

Consider a clinical trial with patients randomized into two groups within *M* strata. Let $N_t^{(m)}$ and $N_c^{(m)}$ denote the number of patients in the Treatment and Control groups, respectively, and

$N^{(m)} = N_t{}^{(m)} + N_c{}^{(m)}$ the total sample size, in the m^{th} stratum (m = 1, 2, …, M). The m^{th} stratum has $N_t{}^{(m)} N_c{}^{(m)}$ comparisons between a patient in the Treatment group and a patient in the Control group. In the m^{th} stratum, let the kernel function $K_{ij}^{(m)} = 1$ if the i^{th} patient from the Treatment group wins against the j^{th} patient from the Control group, and = 0 otherwise. Similarly, let the kernel function $L_{ij}^{(m)} = 1$, if the j^{th} patient from the Control group wins against the i^{th} patient from the Treatment group, and = 0 otherwise. As long as $K_{ij}^{(m)}$ and $L_{ij}^{(m)}$ are constructed in this way (i.e., = 1 for a win, = 0 otherwise), the results in this Short Communication hold, regardless of the types of outcomes.

2.1 Stratified win ratio

The stratified win ratio[5] is defined as

$$WR = \frac{\sum_{m=1}^{M} w^{(m)} n_t^{(m)}}{\sum_{m=1}^{M} w^{(m)} n_c^{(m)}}, \qquad (1)$$

where $n_t{}^{(m)} = \sum_{i=1}^{N_t{}^{(m)}} \sum_{j=1}^{N_c{}^{(m)}} K_{ij}^{(m)}$ and $n_c{}^{(m)} = \sum_{i=1}^{N_t{}^{(m)}} \sum_{j=1}^{N_c{}^{(m)}} L_{ij}^{(m)}$ are the numbers of wins in the Treatment and Control groups, respectively, in the m^{th} stratum (m = 1, 2, …, M), and $w^{(m)}$ is the weight for the m^{th} stratum. Dong et al.[5] evaluated Mantel-Haenszel-type weights ($w^{(m)} = 1/N^{(m)}$), equal weights, and inverse-variance weights. They recommended the Mantel-Haenszel-type weights, particularly for sparse data, because the stratified win ratio defined in (1) reduces to the Mantel-Haenszel stratified odds ratio when the outcome is a single binary endpoint. A few years later Hermans et al.[10] studied stratified odds ratios and pointed out that the stratified odds ratio with the Mantel–Haenszel weights does not follow from optimality considerations, but nevertheless has properties similar to and often better than the optimal estimator.

Although not formally described in Dong et al.[5], the stratified win proportions, P_t and P_c,

for the Treatment and Control groups can be calculated as follows:

$$P_t = \frac{\sum_{m=1}^{M} w^{(m)} n_t^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)}}, \quad (2a)$$

$$P_c = \frac{\sum_{m=1}^{M} w^{(m)} n_c^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)}}. \quad (2b)$$

Therefore, the stratified proportion of ties is

$$P_{tie} = \frac{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)} - \sum_{m=1}^{M} w^{(m)} n_t^{(m)} - \sum_{m=1}^{M} w^{(m)} n_c^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)}} = 1 - P_t - P_c. \quad (2c)$$

The stratified win ratio can be expressed as $WR = \frac{P_t}{P_c}$, in the same form as the unstratified win ratio. The logarithm of the stratified win ratio is asymptotically normally distributed with the following variance under the null hypothesis of equal win probabilities in the two groups:

$$\hat{\sigma}^2_{log(WR)} = \frac{\sum_{m=1}^{M} (w^{(m)})^2 \hat{\sigma}_t^{2\,(m)} + \sum_{m=1}^{M} (w^{(m)})^2 \hat{\sigma}_c^{2\,(m)} - 2\sum_{m=1}^{M} (w^{(m)})^2 \hat{\sigma}_{tc}{}^{(m)}}{(\hat{\lambda})^2}, \quad (3)$$

where $\hat{\lambda} = \frac{1}{2}\left(\sum_{m=1}^{M} w^{(m)} n_t^{(m)} + \sum_{m=1}^{M} w^{(m)} n_c^{(m)}\right)$; $\hat{\sigma}_t^{2\,(m)}$ and $\hat{\sigma}_c^{2\,(m)}$ are the estimated variances for $n_t^{(m)}$ and $n_c^{(m)}$, respectively, and $\hat{\sigma}_{tc}{}^{(m)}$ is their estimated covariance (for details see Dong et al.[5]).

2.2 Stratified win odds and stratified net benefit

To define the stratified win odds and the stratified net benefit, we apply the stratified method (Section 2.1) to the unstratified versions:

$$WO = \frac{P_t + 0.5(1 - P_t - P_c)}{P_c + 0.5(1 - P_t - P_c)} = \frac{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)} + \sum_{m=1}^{M} w^{(m)} n_t^{(m)} - \sum_{m=1}^{M} w^{(m)} n_c^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)} - \sum_{m=1}^{M} w^{(m)} n_t^{(m)} + \sum_{m=1}^{M} w^{(m)} n_c^{(m)}}. \quad (4a)$$

$$NB = P_t - P_c = \frac{\sum_{m=1}^{M} w^{(m)} n_t^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)}} - \frac{\sum_{m=1}^{M} w^{(m)} n_c^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_t^{(m)} N_c^{(m)}} \quad (4b)$$

By the delta method, *log(WO)* and *NB* are also asymptotically normally distributed. The variances of *log(WO)* and *NB* under the null hypothesis can be estimated by

$$\hat{\sigma}^2_{log(WO)} = \frac{\sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_t^{2\,(m)} + \sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_c^{2\,(m)} - 2\sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_{tc}{}^{(m)}}{\left(\frac{\sum_{m=1}^{M} w^{(m)}N_t^{(m)}N_c^{(m)}}{2}\right)^2}, \quad (5a)$$

$$\hat{\sigma}^2_{NB} = \frac{\sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_t^{2\,(m)} + \sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_c^{2\,(m)} - 2\sum_{m=1}^{M}(w^{(m)})^2\hat{\sigma}_{tc}{}^{(m)}}{\left(\sum_{m=1}^{M} w^{(m)}N_t^{(m)}N_c^{(m)}\right)^2}. \quad (5b)$$

The stratified win statistics defined in (1), (4a) and (4b) are in a general form. One can plug in fixed weights ($w^{(m)}, m = 1, 2, \ldots, M$) and variances and covariances of the numbers of wins for each stratum ($\hat{\sigma}_t^{2\,(m)}$, $\hat{\sigma}_c^{2\,(m)}$, and $\hat{\sigma}_{tc}{}^{(m)}$). The variances of the three (stratified) win statistics are constructed under the null hypothesis of equal (stratified) win probabilities in the two groups. Dong et al.[11] discussed variance estimators under the alternative hypothesis. For clinical trial practice (e.g., the win ratio is far above 0.25 and below 4.0), the variance estimators under the null and alternative hypotheses do not differ to a meaningful extent.

2.3 Hypothesis test of the stratified win statistics and homogeneity test of win statistics across strata

The stratified win statistics test the null hypothesis that the stratified win probabilities are equal in the two groups (i.e., the stratified win ratio = 1, the stratified win odds = 1, or the stratified net benefit = 0) vs. the alternative hypothesis that the stratified win probability in the Treatment group is greater than the stratified win probability in the Control group.

The overall effect size of the stratified win statistics depends on the weights assigned to the strata. In a non-stratified analysis, the impact of multiple outcomes on the win statistics can be complex. Wang et al.[12] focused on the win ratio and pointed out that the first-priority component (e.g., death) plays a dominant role, especially when that component has a large treatment effect and a high event rate; when adding a component to the composite endpoint, the performance of the win ratio ~~[win statistics]~~ depends on the treatment effect, the event rate, and the position of the

component in the importance order. This finding on the win ratio generally applies to the other win statistics. In a stratified analysis, the impact of strata on the win statistics can be more complex. For example, Mao[13] showed that the stratified win ratio test is guaranteed to be consistent only if the treatment effects in all strata point in the same direction; otherwise, the stratum-wise contributions to the noncentrality parameter may cancel. It is important to assess homogeneity of win statistics across strata (i.e., H_0: all stratum-specific win statistics are equal). Dong et al.[5] discussed a homogeneity test modeled on the inverse-variance-weighted approach of Cochran[14]. Further research is warranted to develop a better homogeneity test for win statistics across strata. Nevertheless, in general, it is helpful to present the stratified win statistics together with the stratum-specific win statistics, regardless of whether those win statistics are homogeneous across strata. If the win statistics are heterogeneous across strata, stratum-specific win statistics are more interpretable than their overall counterparts.

3 Relations among the stratified win statistics

3.1 Relations of point estimates

The stratified win statistics defined in Section 2 have the same relations as the unstratified versions described in Dong et al[8]. As indicated in Dong et al.[8] and shown in Figure 1 and Figure 2, the win odds increases (or decreases) as the net benefit increases (or decreases) regardless of ties. However, the relation between the win odds and the win ratio depends on the proportion of ties (and similarly for the net benefit):

$$NB = \frac{WR-1}{WR+1}(1 - P_{tie}) \quad (6a)$$

$$NB = \frac{WO-1}{WO+1} \quad (6b)$$

$$WO = \frac{1+NB}{1-NB} \quad (6c)$$

$$WO = \frac{WR - 0.5P_{tie}(WR-1)}{1+0.5P_{tie}(WR-1)}. \quad (6d)$$

Moreover, Figure 2 shows the win ratio can take any positive value regardless of the proportion of ties, whereas the win odds approaches 1 and the net benefit approaches 0 as the proportion of ties increases. Therefore, when the proportion of ties is large, the win odds and the net benefit are more interpretable, and they may be preferred as a measure of treatment effect. This is how the win odds was motivated. This is consistent with previous findings[3,15,16,17]. On the other hand, the Z-values of the three win statistics are approximately equal (i.e., they provide similar p-values and powers, as detailed in Section 4).

3.2 Relations of estimated variances

The variances of the three stratified win statistics are also related. Because the relations derived in this Section also apply to non-stratified win statistics, we use the notations *WO*, *WR*, and *NB* for both stratified and non-stratified win statistics. From (5a) and (5b), we obtain the relation of the estimated variances for *log(WO)* and *NB* as

$$\hat{\sigma}^2_{log(WO)} = 4\hat{\sigma}^2_{NB}. \quad (7a)$$

From (3), (5b) and (7a), we show that the estimated variances for *NB* and *log(WO)* are indirectly related to the estimated variance of *log(WR)* through ties:

$$\hat{\sigma}^2_{NB} = \frac{1}{4}\,\hat{\sigma}^2_{log(WR)}(1 - P_{tie})^2, \quad (7b)$$

$$\hat{\sigma}^2_{log(WO)} = \hat{\sigma}^2_{log(WR)}(1 - P_{tie})^2. \quad (7c)$$

Figure 3 and Figure 4 show the relations of these estimated variances for the three win statistics. We use the ranges of the estimated variances from the simulation study of Cui et al.[18] and the CHARM data[19].

For point estimates, in the presence of a large proportion of ties, the win odds is close to one, but the win ratio can be much larger (Figure 2). When the proportion of ties is large, the estimated variance of log(*WO*) greatly decreases, but the estimated variance of log(*WR*) can be much larger (Figure 4). Consequently, the Z-values of statistical tests for the win odds and the win ratio are still approximately equal, regardless of ties, as shown by (8a) and (8b) in Section 4. On the other hand, since the variance of log(*WO*) can be smaller and the point estimate of the win odds can be closer to 1.0 when the proportion of ties is large, the interpretation of the win odds vs the win ratio by Brunner et al.[16] is intuitive and can aid communication of the trial results.

4 Approximate equivalence of statistical tests for the stratified win statistics

Dong et al.[8] showed that the Z-values of the statistical tests for the three unstratified win statistics are approximately equal. Using the approximate equality of $\log(x)$ and $\frac{2(x-1)}{x+1}$ for $x > 0$, this approximate equivalence of statistical tests also applies to the stratified win statistics:

$$Z_{NB} = \frac{\widehat{NB}}{\hat{\sigma}_{NB}} = \frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{n_t+n_c}{N_tN_c}\frac{1}{\hat{\sigma}_{\log(WR)}}\frac{2N_tN_c}{n_t+n_c} = \frac{\widehat{WR}-1}{\widehat{WR}+1}\frac{2}{\hat{\sigma}_{\log(WR)}} \approx \frac{\log(\widehat{WR})}{\hat{\sigma}_{\log(WR)}} = Z_{\log(WR)}, \quad (8a)$$

$$Z_{NB} = \frac{\widehat{NB}}{\hat{\sigma}_{NB}} = \frac{\widehat{WO}-1}{\widehat{WO}+1}\frac{1}{\hat{\sigma}_{NB}} = \frac{\widehat{WO}-1}{\widehat{WO}+1}\frac{2}{\hat{\sigma}_{\log(WO)}} \approx \frac{\log(\widehat{WO})}{\hat{\sigma}_{\log(WO)}} = Z_{\log(WO)}. \quad (8b)$$

The approximate equality of the Z-values makes sense; the null hypotheses for the three (stratified) win statistics are equivalent to testing the null hypothesis of equal (stratified) win probabilities in the Treatment and Control groups. Therefore, the three (stratified) win statistics should provide similar p-values and statistical powers.

5. Application to CHARM studies

The CHARM program[19] was designed as three separate randomized, double-blind, placebo-controlled trials (CHARM-Added, CHARM-Alternative, and CHARM-Preserved) comparing candesartan with placebo in patients with chronic heart failure. The primary endpoint was a

composite of cardiovascular death or hospitalization for chronic heart failure. A total of 7599 patients were randomized to the two groups. For illustration, we use the studies as strata (N = 2548, 2028, and 3023). Table 1 presents win statistics for each stratum (study). Table 2 shows the stratified win statistics with Mantel-Haenszel-type weights and p-values. As expected, the p-values for the three stratified win statistics are almost identical.

6. Summary and discussion

In this Short Communication, we apply the stratified method developed for the stratified win ratio to derive the stratified win odds and the stratified net benefit. The three stratified win statistics are defined in a general form. For the *M* strata, one can plug in fixed weights ($w^{(m)}$) and estimated variances and covariances of the numbers of wins for each stratum ($\hat{\sigma}_t^{2\,(m)}$, $\hat{\sigma}_c^{2\,(m)}$, and $\hat{\sigma}_{tc}{}^{(m)}$). We show that the relations among the three win statistics and the approximate equality of their statistical tests, as presented in Dong et al.[8], also apply to the stratified win statistics. Therefore, the three stratified win statistics provide similar p-values and statistical powers.

For the stratified win statistics, defined in (1), (4a) and (4b), the weights are applied to the numbers of wins ($n_t^{(m)}$ and $n_c^{(m)}$) in a similar way as in the Mantel-Haenszel stratified odds ratio, risk ratio, and risk difference, which apply the weights to the cell counts of a stratified 2×2 table. As Dong et al.[5] proposed, use of Mantel-Haenszel-type weights ($w^{(m)} = \frac{1}{N^{(m)}}$) is more robust, particularly for sparse data. Equal weights can be reasonable if the data are not sparse. Buyse[1] suggested equal weights for the net benefit.

Alternatively, one may construct stratified win statistics by applying weights directly to stratum-specific win statistics such as $WR = \sum_{m=1}^{M} w^{(m)}\, WR^{(m)}$, $WO = \sum_{m=1}^{M} w^{(m)}\, WO^{(m)}$, and $NB = \sum_{m=1}^{M} w^{(m)}\, NB^{(m)}$. Possible choices include: (1) Weights based on the number of subjects

$w^{(m)} = \frac{N^{(m)}}{\sum_{m=1}^{M} N^{(m)}}$, (2) Weights based on the number of events (for time-to-event endpoints) $w^{(m)} = \frac{N_{event}^{(m)}}{\sum_{m=1}^{M} N_{event}^{(m)}}$, and (3) Weights $w^{(m)} = 1/M$ to average stratum-specific win statistics across strata. The resulting stratified win statistics are asymptotically normally distributed, and their variances under the null hypothesis can be estimated by $\hat{\sigma}^2_{log(WR)} \approx \frac{\hat{\sigma}^2_{WR}}{\widehat{WR}^2} = \frac{\sum_{m=1}^{M} w^{(m)2} \hat{\sigma}^2_{WR(m)}}{\widehat{WR}^2}$, $\hat{\sigma}^2_{log(WO)} \approx \frac{\hat{\sigma}^2_{WO}}{\widehat{WO}^2} = \frac{\sum_{m=1}^{M} w^{(m)2} \hat{\sigma}^2_{WO(m)}}{\widehat{WO}^2}$, and $\hat{\sigma}^2_{NB} = \sum_{m=1}^{M} w^{(m)2} \hat{\sigma}^2_{NB(m)}$. The WINS package[9] in R implements all Mantel-Haenszel-type and stratum-specific-type win statistics. A referee suggested weights from rank-based analyses (e.g., Mehrotra et al.[20]), because the Mann-Whitney test and the win statistics are related. Such weights would be worth exploring in the setting of the win statistics, together with other stratified analyses such as Gasparyan et al.[21] and Wang and Mao[22]. On the other hand, Hermans et al.[10] pointed out that, although the stratified odds ratio with the Mantel–Haenszel weights does not follow from optimality considerations, it nevertheless has properties similar to and often better than the optimal estimator.

Although stratified analyses can reduce the impact of confounding and potentially improve statistical efficiency, stratified win statistics (similar to the unstratified win statistics) still depend on the censoring distribution for time-to-event outcomes. To address the effects of independent and dependent censoring, Dong et al.[23,24] provided an IPCW (inverse-probability-of-censoring weighting) adjustment approach for the estimate of the effect size.

The win ratio has been mostly applied in designs and analyses of cardiovascular and COVID-19 trials. The win odds has also been applied in designs and analyses for clinical trials such as the DARE-19 Phase III study[17]. We encourage statisticians to explore the win statistics for their clinical trial designs and analyses. We welcome discussions and sharing of experience with

their pros and cons. The more experience we gain with them, the better we will understand when they perform well and when they are not useful.

Acknowledgement

The authors would like to thank the Associate Editor and two anonymous referees for their very helpful and constructive comments, which greatly improved this article. They also thank Johan Verbeeck, George Chu, Hong Tian, Dali Zhou, James Song, and Ran Liao for discussions, reviews, and comments.

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed.

References

1. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine*. 1999;18:1341-1354.
2. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Statistics in Medicine.* 2010;29(30):3245-3257.
3. 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.

4. 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.
5. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *Journal of Biopharmaceutical Statistics.* 2018;28(4):778-796.
6. Voors AA, Angermann CE, Teerlink JR, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. *Nature Medicine*. 2022;28(3):568-574.
7. 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.
8. Dong G, Huang B, Verbeeck J, Cui Y, Song J, Gamalo-Siebers M, Wang D, Hoaglin DC, Seifu Y, Mütze T, Kolassa J. 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.
9. Cui Y, Huang B. WINS: The R WINS Package. 2022. https://cran.r-project.org/web/packages/WINS/index.html.
10. Hermans L, Molenberghs G, Verbeke G, Kenward MG, Mamouris P, Vaes B. Optimal weighted estimation versus Cochran–Mantel–Haenszel. *Communications in Statistics - Simulation and Computation*. 2022;51(7): 3645-3659.
11. Dong G, Guo M, Hoaglin DC, Vandemeulebroecke M. 2016. Is the asymptotic variance estimator of the logarithm of the win ratio under the null hypothesis sound? – A note on the

generalized analytic solution to the win ratio by Dong et al. (2016)". Supporting information #7 (a post-publishing note) to Dong et al. (2016) (http://onlinelibrary.wiley.com/store/10.1002/pst.1763/asset/supinfo/pst1763-sup-0007-supplementary.pdf?v=1&s=cd757e23cb40556b3a1511b53c73747fe06fae89). *Pharmaceutical Statistics* 15(5):430–437.

12. Wang B, Zhou D, Zhang J, Kim Y, Chen LW, Dunnmon P, Bai S, Liu Q, Ishida E. Statistical power considerations in the use of win ratio in cardiovascular outcome trials. *Contemporary Clinical Trials*. 2023;124:107040.
13. Mao L. On the alternative hypotheses for the win ratio. *Biometrics*. 2019;75(1):347-351.
14. Cochran WG. The combination of estimates from different experiments. *Biometrics.* 1954;10:101–129.
15. Peng L. The use of the win odds in the design of non-inferiority clinical trials. *Journal of Biopharmaceutical Statistics.* 2020; 30(5):941–946.
16. Brunner, E., Vandemeulebroecke, M., & Mütze, T. Win odds: An adaptation of the win ratio to include ties. *Statistics in Medicine.* 2021; 40(14), 3367-3384.
17. Gasparyan SB, Buenconsejo J, Kowalewski EK, Oscarsson J, Bengtsson OF, Esterline R, Koch GG, Berwanger O, Kosiborod MN. Design and Analysis of Studies Based on Hierarchical Composite Endpoints: Insights from the DARE-19 Trial. *Therapeutic Innovation & Regulatory Science.* 2022;56(5):785-794.
18. Cui Y, Dong G, Kuan PF, Huang B. Evidence synthesis analysis with prioritized benefit outcomes in oncology clinical trials. *Journal of Biopharmaceutical Statistics*. 2022. doi: 10.1080/10543406.2022.2141769.
19. 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.

20. Mehrotra DV, Lu X, Li X. Rank-based analyses of stratified experiments: alternatives to the van Elteren test. *The American Statistician.* 2010; *64*(2):121-130.
21. Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. Adjusted win ratio with stratification: calculation methods and interpretation. *Statistical Methods in Medical Research.* 2021;30(2): 580-611.
22. Wang T, Mao L. Stratified proportional win-fractions regression analysis. *Statistics in Medicine.* 2022. 10.1002/sim.9570.
23. Dong G, Mao L, Huang B, Gamalo-Siebers M, Wang J, Yu G, Hoaglin DC. The inverse-probability-of-censoring weighting (IPCW) adjusted win ratio statistic: an unbiased estimator in the presence of independent censoring. *J Biopharm Stat*. 2020;30(5):882-899.
24. Dong G, Huang B, Wang D, Verbeeck J, Wang J, Hoaglin DC. Adjustingwin statistics for dependent censoring. *Pharm Stat*. 2021;20(3):440-450.

Table 1 Win statistics by stratum (study) for the CHARM program

Stratum (study)	Win proportion (%)		Proportion of ties (%)	Win ratio (95% CI)	Win odds (95% CI)	Net benefit (%) (95% CI)
	Candesartan	Placebo				
CHARM-Added	33.9	28.5	37.6	1.19 (1.05, 1.35)	1.11 (1.03, 1.21)	5.4 (1.6, 9.3)
CHARM-Alternative	31.6	24.6	43.8	1.27 (1.10, 1.48)	1.15 (1.06, 1.25)	7.0 (2.8, 11.1)
CHARM-Preserved	20.6	18.5	60.9	1.11 (0.96, 1.30)	1.04 (0.98, 1.11)	2.2 (-0.8, 5.1)

Table 2 Stratified win statistics with Mantel-Haenszel-type weights and p-values for the CHARM program

Stratified win proportion (%)		Stratified win ratio		Stratified win odds		Stratified net benefit (%)	
Candesartan	Placebo	Stratified win ratio (95% CI)	p-value	Stratified win odds (95% CI)	p-value	Stratified net benefit (95% CI)	p-value
28.0	23.5	1.19 (1.10, 1.29)	0.000016	1.10 (1.05, 1.14)	0.000017	4.5 (2.5, 6.6)	0.000017