Duzce Medical Journal

2026;28(1):3-10
https://doi.org/10.18678/dtfd.1778178

e-ISSN: 1307-671X
https://dergipark.org.tr/en/pub/dtfd

Research Article

The Effect of the Number of Matched Data Pairs on the Win Ratio Statistic and Confidence Intervals: A Simulation Study

İsmet Doğan*

Department of Biostatistics and Medical Informatics, Afyonkarahisar Health Sciences University, Faculty of Medicine, Afyonkarahisar, Türkiye.
ror.org/00sfg6g55

*Corresponding author:
ismet.dogan@afsu.edu.tr

ABSTRACT

Aim: The aim of this study was to demonstrate the effect of paired data numbers on the win ratio statistic and its 95% confidence intervals for two composite endpoints.

Material and Methods: This is a simulation study. Data was generated for 35 different scenarios of matched data pairs within the range $10 \leq n \leq 1000$ using the Python random library. A total of 34101 different data sets (286 for $n = 10$, 815 for $n = 15$, and 1000 data sets for each of the other 33 cases with $n \geq 20$) were used in the study.

Results: As the number of paired data increases, the win ratio increases, although not regularly. The number of cases where the value of the win ratio statistic is greater than one is not affected by the number of paired data, and the win ratio numbers considered important increase as the number of paired data increases. As the number of paired data increases, the lower and upper limits of the win ratio approach each other, although not regularly, and the differences between the upper and lower limits also decrease.

Conclusion: When evaluating the results of the study in general, it can be said that the increase in the number of paired data affects the number of significant win ratio and the maximum win ratio, and the confidence interval of the win ratio is more affected by the data structure than the size of the number of paired data.

Keywords: Treatment efficacy, composite outcomes, win ratio, matched data pairs, confidence interval.

Received: 04.09.2025
Revision: 30.01.2026
Accepted: 02.02.2026
Online: 25.04.2026

INTRODUCTION

The effectiveness of a treatment is typically assessed by the extent to which it reduces the incidence of adverse events within a specified timeframe. This adverse event may be a single outcome, such as death, or one of several possible outcomes (a composite endpoint), such as death, heart attack or stroke. More participants will experience a composite endpoint than a single component. An increase in the number of observed events enhances the statistical power of a study to detect a true difference between the treatment and control groups. Furthermore, this enables the study to recruit fewer patients and follow them for shorter periods of time (1). Composite endpoints are frequently employed in clinical trials and observational studies with the aim of increasing statistical efficiency and demonstrating overall treatment benefit. The primary advantage of using composite endpoints is that they provide greater statistical power than a study design based on a single endpoint. Nonetheless, in cases where the components of a composite endpoint differ in their clinical relevance, it is essential to incorporate the clinical weight of each component into the analysis to derive more meaningful and robust inferences (2). A composite endpoint integrates two or more individual endpoints into a single outcome variable. A patient is deemed to have experienced the clinical event if any one of the predefined component events occurs. Composite endpoints were originally developed to decrease the required sample size and duration of follow-up, while preserving sufficient statistical power to assess both the efficacy and safety of novel treatments (3). The advantages of composite endpoints have been extensively discussed in the literature. They offer a solution to the challenge of identifying a single outcome measure that comprehensively reflects the treatment effect, while also reducing or eliminating the need for multiple hypothesis testing, mitigating competing risks, and enhancing statistical power. Depending on the therapeutic area, well-established and accepted composite endpoints may already be available. In the analysis of composite endpoints, the clinical importance of each component must be carefully considered, as any composite implicitly or explicitly assigns relative weights to its components. When comparing two treatment groups, the component with the greatest clinical relevance is evaluated first to determine treatment benefit or harm; if this comparison results in a tie, subsequent components are compared in order of descending clinical importance (4). To ensure the reliability of results and prevent misleading conclusions that could overstate the efficacy of interventions, the following criteria must be satisfied (1):

a. The components must have similar clinical significance for patients.
b. The components must occur with a similar frequency within the same time period.
c. The treatment effect should be consistent across all components of the composite endpoint.

A key limitation of conventional analytical methods for studies utilizing composite endpoints is the assumption of equal statistical weighting for each component, irrespective of differences in their clinical significance. The odds ratio is a ranking-based statistic that evaluates events based on their clinical significance. This is achieved by comparing each patient in the intervention group with

Cite as: Doğan İ. The Effect of the number of matched data pairs on the win ratio statistic and confidence intervals: a simulation study. Duzce Med J. 2026;28(1):3–10. doi:10.18678/dtfd.1778178.

their corresponding control counterparts. It also accounts for the importance of the composite endpoints and summarises and prioritises the important components (2,5). The fundamental principle of the win ratio involves establishing a hierarchy of clinical significance among the components of the composite outcome. All possible pairs are generated by comparing each subject in the treatment group with every subject in the control group. Each pair is then evaluated in order of decreasing importance for the occurrence of the components of the composite outcome. This process starts with the most significant outcome. If the comparison does not yield a favorable outcome for either member of the pair, the process continues down the hierarchy until either a difference is observed or all components have been evaluated and the pair is deemed tied. Win ratio uses hierarchical comparison to prioritise lethal components over non-lethal ones (6,7). This method can be readily extended to composites comprising three or more components, provided that these components can be logically ranked according to their clinical importance. For example, in a hypertension study with primary outcomes of cardiovascular death, stroke, and myocardial infarction, each pairwise comparison can be conducted sequentially in the order of cardiovascular death, stroke, and then myocardial infarction. This ordering reflects the generally greater clinical severity and long-term disability associated with stroke compared to myocardial infarction (8). This method offers several advantages, including the flexibility to incorporate non-time-to-event variables within the composite endpoint and a hierarchical framework that enables ranking components based on their clinical significance. Unlike parametric methods, it does not require strong assumptions (9). One of the major drawbacks of the approach is its novelty; some researchers may struggle to conceptualise its clinical relevance. A win ratio of 2 indicates that a patient in the treatment group has twice the likelihood of experiencing a favorable outcome compared to a patient in the control group. As such, the win ratio serves as a relative measure of treatment effect, with only informative pairs—those in which a clear winner can be determined—contributing to the estimation, while tied comparisons do not influence the result. A second limitation is the difficulty of determining sample sizes for studies using the win ratio (10). Another notable limitation of the method is its reliance on the assumption that patient characteristics are balanced between treatment groups in a randomized clinical trial. When this assumption does not hold, the presence of potential confounding factors may bias the results, necessitating their consideration in the analysis. Consequently, this limitation can restrict the method's applicability, particularly in settings where randomization does not ensure adequate balance or in observational studies (11). The statistical properties of the win ratio have been extensively investigated in the literature. Luo et al. (12), Dong et al. (13), and Oakes (14) developed inferential methods—such as variance estimators and confidence intervals—to establish a robust statistical framework for the win ratio. In parallel, Bebu and Lachin (15) formulated the win ratio as a bivariate U-statistic for independent data, facilitating large-sample inference. Luo et al. (16) proposed a weighted win ratio statistic to enhance estimation efficiency. Separately, Wang and Pocock (17) extended the application of the win ratio to non-time-to-event outcomes. However, they highlighted a key limitation: the performance of the win ratio may be compromised in the presence of a high proportion of tied comparisons. Accordingly, they recommended its use primarily in settings where the frequency of ties is minimal. Specifically:

a. In the presence of tied comparisons, the win ratio loses its interpretation as an odds measure. However, the win odds retains this interpretability, allowing for meaningful comparison of treatment effects even when ties are present.
b. In the presence of ties, the win ratio can be misleading with regard to the degree of similarity between the two distributions; this is not the case for win odds ratios.
c. In the absence of a tie, the win odds ratios are the same as the win ratio.

Win odds ratios are more reasonable, interpretable, and better aligned with established methodologies (17). Dong et al. (18) and Finkelstein and Schoenfeld (19) explored the interpretation of the win ratio, providing guidance on its clinical and statistical meaning. In parallel, Dong et al. (20) and Mao (21) extended the methodology to accommodate multivariate and stratified settings, thereby broadening its applicability across more complex analytical frameworks. This study aimed to demonstrate the effect of sample size on the win ratio statistic and 95% confidence intervals when considering two composite endpoints following the methodology outlined by Pocock et al. (8).

MATERIAL AND METHODS

Wang and Pocock (17) noted that a high prevalence of tied comparisons can undermine the reliability of the win ratio, and advised that the method should be applied only when the proportion of ties is negligible. This warning has been taken into account in this study, with the time condition being neglected. The Python-random library was used to generate data for 35 different paired datasets (n) in the range $10 \leq n \leq 1000$ (10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000). Sample sizes were determined by taking into account situations that are likely to be encountered in real-world studies. The goal was to generate 1000 unique tables for each sample size. This goal was achieved in most cases, but due to combinatorial limitations, only 286 unique tables were obtained for $n = 10$ and 815 for $n = 15$. The Python code used in data generation is presented in Appendix-1. The details of the simulation setup for data generation using the Python 3.9.10 programming language are as follows:

Step 1. Enter the number of paired data (n) and the number of categories required for data generation into the program.

Step 2. A random category is selected using the choice function in the Python random library. The categories are: 'Experimental group-first outcome component', 'Control group-first outcome component', 'Experimental group-second outcome component', 'Control group-second outcome component', and so on.

Step 3. Use the Choice function again to select a random integer between 0 and the desired number of

matched data pairs for the selected category. For example, if the desired number of matched data pairs is 20, an integer between 0 and 20 is selected at random.

Step 4. Steps two and three are repeated for each category. After selecting an integer for the kth category, the sum of the selected numbers is examined. If this total exceeds the desired total, the process starts again, as this indicates invalid data. The integer to be assigned to the final category is found by subtracting the current total from the desired total.

Step 5. Check whether the derived data has been derived before. If so, it is not added to the dataset.

Step 6. The programme is run repeatedly, starting from step 1, until the desired total number of data points is reached. For this study, 286 data sets were used for $n = 10$, 815 for $n = 15$ and 1000 for $n \geq 20$.

Step 7. Finally, the win ratio and confidence intervals are calculated from the derived data using a simple Excel application (Microsoft Corporation, USA).

The win ratio method was employed to estimate the treatment effect based on the framework of generalized pairwise comparisons, as originally proposed by Finkelstein and Schoenfeld (22). In generalized pairwise comparisons, each patient in the treatment group is systematically compared with every patient in the control group. For each comparison, if the patient in the treatment group has a better outcome, the treatment group is considered a 'win'; if the patient in the control group has a better outcome, the treatment group is considered a 'loss'; and if no win or loss can be determined, the outcome is considered a 'tie' (11). Win ratio analysis, which was introduced by Pocock et al. (8), consists of four steps (23):

1. Events are ranked by clinical severity: In cardiovascular disease trials, the components of the composite endpoint are typically ordered based on their perceived clinical importance—commonly including death, stroke, myocardial infarction, major bleeding, and revascularization. In trials involving drug-resistant infections, components may include outcomes such as death or renal failure. In multiple sclerosis trials, the composite endpoint often comprises death, relapse, and clinically significant disease progression.
2. Patient matching: The concept is to assign patients to different treatments based on their individual risk estimates. Pocock et al. (8) proposed calculating a composite risk score for each patient based on pre-selected baseline prognostic factors. Patients in the experimental treatment group are matched with control group patients who have comparable risk scores, assuming that their follow-up durations are sufficiently similar. If the number of patients in the two groups differs, some patients are randomly excluded to equalise the number in each group.
3. Determine a winner for each pair of patients by comparing each pair using each categorised event type (death, stroke, myocardial infarction, etc.). The events within each pair are evaluated to determine which patient has experienced the most serious event. If not, the remaining pairs are evaluated for the occurrence of the second most serious event, and so on for each subsequent ranking. If no event has occurred by the final follow-up, the pair is considered tied. The win ratio method emphasizes the more clinically severe components when comparing composite endpoints between two patient groups.
4. The win ratio is calculated by dividing the total number of favorable comparisons (wins) for the treatment group by the number of unfavorable comparisons (losses), excluding tied pairs.

The distribution of possible outcomes for the two outcome components can be summarised using a 2x2 contingency table, as shown in Table 1. The composite outcome results of the study are summarised by n_a, n_b, n_c and n_d.

Table 1. Table of distribution of observation values for the case of two outcome components.

	Treatment	Control
Composite Outcome I	n_a	n_b
Composite Outcome II	n_c	n_d
	$n_{Loss} = n_a + n_c$	$n_{Loss} = n_b + n_d$
	$n_{Win} = n_b + n_d$	$n_{Win} = n_a + n_c$

The number of winners, the number of losers, and the win ratio of the treatment group are calculated using Equation 1, Equation 2, and Equation 3, respectively.

$$n_w = n_b + n_d \quad [1]$$

n_w, is the number of winners for the treatment group, i.e. those matched pairs where control group fared worse.

$$n_L = n_a + n_c \quad [2]$$

n_L is the number of losers for the treatment group

$$R_w = \frac{n_w}{n_L} \quad [3]$$

R_w is the win ratio. To find the confidence interval for the win ratio, it was recommended to use the proportion given in Equation 4.

$$p_w = \frac{n_w}{n_w + n_L} \quad [4]$$

p_w is a proportion and the lower and upper bound values of the 95% confidence interval for p_w are calculated using Equation 5.

$$p_w \pm 1.96 * \sqrt{\frac{p_w(1-p_w)}{(n_w+n_L)}} \quad [5]$$

When the p_w proportion is used, the win ratio is calculated as given in Equation 6.

$$R_w = \frac{p_w}{(1-p_w)} \quad [6]$$

Using the lower and upper bound values obtained from Equation 5, the lower and upper bound values of the 95% confidence interval for the win ratio are calculated as follows:

$$Lower\ bound = \frac{p_{Lower\ bound}}{(1-p_{Lower\ bound})}$$

$$Upper\ bound = \frac{p_{Upper\ bound}}{(1-p_{Upper\ bound})}$$

The value obtained from Equation 7 for the significance test has an asymptotic standard normal distribution under the null hypothesis that there is no difference between the treatment and control groups (8).

$$z = \frac{p_w - 0{,}5}{\sqrt{\frac{p_w(1-p_w)}{(n_w+n_L)}}} \quad [7]$$

Values outside the range $-1.96 \leq z \leq 1.96$ obtained using Equation 7 are considered significant. A win ratio greater than 1 indicates evidence of a treatment effect in favour of the treatment group (24).

RESULTS

Table 2 provides basic statistics for the win ratio (R_w) based on paired data numbers.

Table 2. Basic statistics for win ratio (R_w) based on paired data counts (n).

n	R_w	Min R_w	Max R_w	Important R_w	$R_w > 1$
10	1.81±2.31	0	9	35.6%	41.5%
15	1.97±2.85	0	14	46.9%	49.1%
20	2.05±3.24	0	19	41.2%	46.7%
25	2.13±3.61	0	24	47.2%	49.6%
30	2.19±3.92	0	29	48.6%	48.0%
35	2.10±3.77	0	34	52.3%	50.7%
40	1.91±3.69	0	39	60.9%	45.9%
45	2.03±3.80	0	44	64.0%	49.4%
50	1.91±3.58	0.0204	49	61.6%	48.7%
55	2.04±3.82	0	54	61.2%	49.5%
60	2.08±3.74	0	59	66.1%	49.8%
65	2.16±4.94	0	64	66.1%	50.4%
70	2.10±4.89	0.0145	69	69.2%	47.9%
75	1.99±3.93	0.0135	74	67.7%	51.1%
80	2.06±5.07	0.0127	79	67.4%	50.3%
85	2.33±5.57	0.0241	84	70.0%	51.1%
90	2.32±6.05	0.0112	89	67.7%	51.8%
95	2.05±4.16	0	46.5	72.0%	48.7%
100	1.93±2.91	0	32.3	72.7%	50.5%
125	2.17±4.92	0	61.5	74.0%	50.6%
150	1.81±2.48	0.0135	24	74.7%	51.6%
175	2.00±3.71	0.0174	57.3	78.9%	50.7%
200	2.02±3.65	0.0309	65.7	80.2%	50.9%
225	2.31±6.57	0.0135	111.5	78.4%	50.2%
250	2.07±5.39	0.0246	124	81.1%	50.0%
300	1.75±2.97	0.0204	41.9	83.3%	46.6%
350	1.99±3.54	0.0116	49	85.5%	50.3%
400	1.91±2.83	0.0127	27.6	84.9%	50.1%
450	2.32±7.32	0.0181	149	86.5%	51.0%
500	2.24±3.90	0.0142	44.5	87.6%	51.0%
600	1.90±2.98	0.0187	30.6	88.4%	50.7%
700	1.86±3.65	0.0057	76.8	90.0%	48.1%
800	2.04±5.88	0.0088	132.3	91.6%	50.4%
900	2.11±8.19	0.0112	224	90.2%	49.9%
1000	1.91±3.25	0.0204	40.7	91.1%	52.8%
Total	2.05±4.45	0	224	72.0%	49.8%

As shown in Table 2, the maximum win ratio tends to increase with the number of paired data points, although this increase is not uniform. The frequency of cases with $R_w > 1$ remains unaffected by the sample size, whereas the number of statistically significant win ratios increases as the number of paired observations grows. The statistical significance of this trend was evaluated using the Pearson chi-square test, revealing a significant association ($p < 0{,}001$). Table 3 summarizes the lower and upper bounds of the 95% confidence intervals for the win ratio (R_w), along with the corresponding interval widths.

Table 3. 95% Confidence intervals for win ratio (R_w) based on paired data counts (n).

n	R_w	*Lower Bound*	*Upper Bound*	|Upper Bound − Lower Bound|
10	1.81±2.31	0.48±0.69	6.43±21.80	12.69±18.82
15	1.97±2.85	0.64±0.91	-25.01±102.87	32.99±101.02
20	2.05±3.24	0.76±1.09	-2.90±33.01	12.58±31.17
25	2.13±3.61	0.86±1.26	-0.67±29.34	11.32±27.61
30	2.19±3.92	0.93±1.40	2.85±25.41	9.66±23.90
35	2.10±3.77	0.96±1.40	3.20±26.65	9.13±25.38
40	1.91±3.69	0.91±1.38	2.61±23.39	6.83±22.73
45	2.03±3.80	1.00±1.49	3.80±23.17	6.83±22.44
50	1.91±3.58	0.99±1.44	4.09±26.50	6.73±25.79
55	2.04±3.82	1.07±1.59	5.18±30.88	7.60±30.02
60	2.08±3.74	1.13±1.58	5.05±29.56	6.90±28.81
65	2.16±4.94	1.15±1.92	4.31±34.65	7.33±34.01
70	2.10±4.89	1.14±1.93	3.40±26.63	5.57±26.33
75	1.99±3.93	1.13±1.71	5.69±36.57	6.45±35.83
80	2.06±5.07	1.16±1.97	4.88±34.42	5.66±33.88
85	2.33±5.57	1.30±2.30	5.38±47.69	8.83±46.84
90	2.32±6.05	1.31±2.37	5.48±41.77	6.89±41.12
95	2.05±4.16	1.22±1.97	4.70±47.05	7.32±46.35
100	1.93±2.91	1.20±1.58	6.03±41.51	5.44±40.66
125	2.17±4.92	1.34±2.34	4.14±46.64	6.20±46.15
150	1.81±2.48	1.23±1.53	3.21±7.17	1.97±5.83
175	2.00±3.71	1.36±2.04	2.42±24.05	3.08±24.38
200	2.02±3.65	1.41±2.10	2.93±20.81	2.67±20.74
225	2.31±6.57	1.54±3.26	6.56±94.36	8.27±93.25
250	2.07±5.39	1.45±2.71	6.93±100.79	6.24±99.55
300	1.75±2.97	1.31±1.94	2.67±7.69	1.36±6.03
350	1.99±3.54	1.50±2.31	3.05±9.67	1.55±7.74
400	1.91±2.83	1.48±2.01	2.62±4.86	1.14±2.94
450	2.32±7.32	1.71±3.91	3.55±53.72	4.32±53.21
500	2.24±3.90	1.75±2.72	3.14±7.26	1.39±4.73
600	1.90±2.98	1.54±2.20	2.46±4.68	0.92±2.56
700	1.86±3.65	1.52±2.51	2.46±8.17	0.95±6.00
800	2.04±5.88	1.64±3.60	3.22±23.43	1.58±20.34
900	2.11±8.19	1.69±4.49	12.65±320.89	10.96±317.30
1000	1.91±3.25	1.62±2.49	2.35±4.77	0.73±2.38
Total	2.05±4.45	1.26±2.26	3.21±68.06	6.28±67.11

As can be seen in Table 3, as the number of paired data increases, the upper and lower bounds of the maximum win ratio approach each other, albeit irregularly, and the difference between them decreases. This is also clearly evident from the locally estimated scatterplot smoothing (LOESS) graph in Figure 1.

Figure 1 illustrates that the width of the confidence interval is both wide and highly variable at low sample sizes (n), but decreases substantially as n increases. The Kruskal-Wallis H test was employed to evaluate differences in win ratios across varying sample sizes, with no statistically significant differences observed ($p > 0{,}05$). This is also evident in Figure 2.

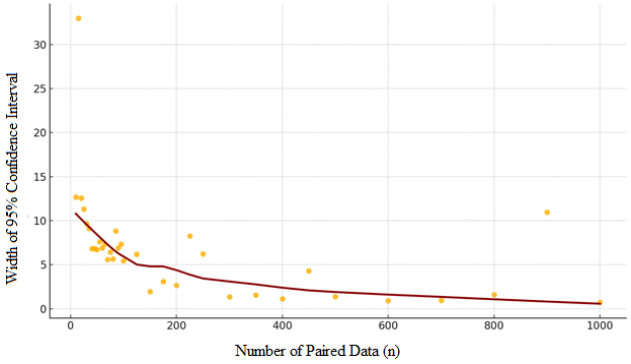

Figure 1. LOESS curve width of 95% confidence interval according to the number of paired data.

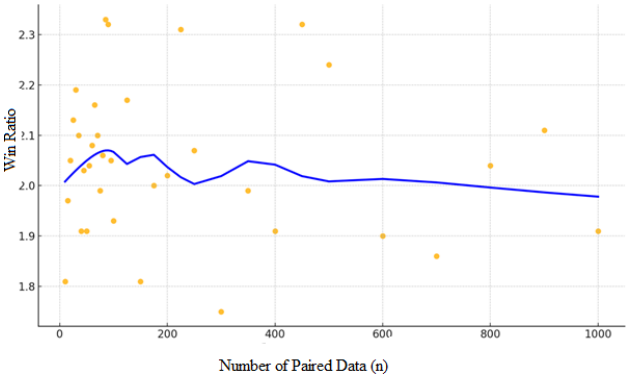

Figure 2. Win Ratio LOESS curve based on number of paired data.

As can be seen from the LOESS graph in Figure 2, the win ratio generally increases as the number of paired data increases, with small fluctuations at the beginning, but then becomes stationary after n ≈ 150.

DISCUSSION

Before an intervention can be widely adopted in clinical practice, it must demonstrate that it achieves its intended therapeutic effects, satisfying important legal and ethical standards. Clinicians should not rely solely on the win ratio statistic when interpreting study results; instead, they must consider the overall treatment effect and ensure sufficient information is available to estimate clinically meaningful measures, such as the number needed to treat. Consequently, a comprehensive assessment of the overall clinical benefit is essential to translate findings from studies reporting win ratios. This evaluation should encompass not only the magnitude of the effect but also the temporal dynamics, durability of the response, and ultimately the overall clinical value. Additionally, the drivers of the overall win ratio should be clearly identified, including any potential placebo effects. In instances where the win ratio is unlikely to substantially influence the analytical outcome, traditional analytical methods may be preferable. Importantly, consensus standards should be established for the hierarchical composite endpoints employed when the win ratio is used as the primary analytic approach, to ensure that clinically meaningful endpoints are consistently generated for research purposes. Summary statistics should be reported alongside win statistics—including wins, losses, and draws—for both the composite outcome and each component of the hierarchical composite endpoint. The final interpretation of the win ratio must clearly acknowledge the drivers of the outcome. In the absence of consistent wins across components, or for measures sensitive to placebo effects, clinical significance should not be inferred unless the totality of the evidence, including physiological and biological outcomes, supports an overall clinical effect. When a treatment's win or loss is determined by a time-dependent component, the durability of the effect size should be reported separately from the win ratio (7). When interpreting win ratio statistics inferentially, it should be considered that wide confidence intervals may lead to unbiased but non-significant results. The matching approach optimally balances confounding variables, resulting in matched win ratios that are more likely to be covered. However, matching also results in greater patient loss, necessitating cautious interpretation of win ratio results (25). A small sample size may result in non-significant findings accompanied by wide confidence intervals. Therefore, when employing the win-loss statistic, it is crucial to ensure that the sample size is sufficiently large to achieve adequate statistical power to reject the null hypothesis (H_0) in the presence of a true treatment effect. There is a risk of inconsistent results when the total sample size falls significantly below 100. Nevertheless, the proposed method appears to be a viable option for sample sizes of $n = 100$ and above in clinical trials (25,26). The study also identified this instability, encountering negative upper bound values for small sample sizes. Furthermore, irrespective of the degree of imbalance, matching within a simulated environment generally yielded higher coverage probabilities. However, simulations are based on specific distributions, proportions and scenarios, so they may not accurately reflect the diversity, heterogeneity and complexity of real-world patients. While real patient data depends on external factors such as time, context, and the healthcare system, these factors may not be incorporated into simulations. Furthermore, although simulations can examine the statistical power of measures, they may be of limited relevance for clinical decision-making.

CONCLUSION

Overall, the study results indicate that an increase in the number of paired data points influences both the number of significant win ratios and the maximum win ratio observed. However, the width and position of the confidence interval for the win ratio are more strongly impacted by the underlying data structure than by the sample size. Similarly, when $n \leq 25$, the confidence interval may exhibit a positive lower bound and a negative upper bound, a phenomenon attributable once again to the characteristics of the data structure.

DECLARATIONS

Acknowledgements: I would like to thank Asst. Prof. Gülcan GENCER for his help in drawing the figures in this study.

Author Contributions: All parts of the study were conducted by the author.

Conflict of Interest: The author declares no conflict of interest.

Supporting Institutions: This research received no external funding.

Ethical Approval: This study does not involve human or animal participants. All procedures followed scientific and ethical principles, and all referenced studies are appropriately cited.

Plagiarism Statement: This article has been evaluated for plagiarism, and no instances of plagiarism were detected.

Use of Artificial Intelligence (AI) Tools: The author declares that no AI tools were used in the creation of this article.

REFERENCES

1. Palileo-Villanueva LM, Dans AL. Composite endpoints. J Clin Epidemiol. 2020;128:157-8. doi:10.1016/j.jclinepi.2020.07.017.
2. Huque MF, Alosh M, Bhore R. Addressing multiplicity issues of a composite endpoint and its components in clinical trials. J Biopharm Stat. 2011;21(4):610-34. doi:10.1080/10543406.2011.551327.
3. Baracaldo-Santamaría D, Feliciano-Alfonso JE, Ramirez-Grueso R, Rojas-Rodríguez LC, Dominguez-Dominguez CA, Calderon-Ospina CA. Making sense of composite endpoints in clinical research. J Clin Med. 2023;12(13):4371. doi:10.3390/jcm12134371.
4. Peng L. The use of the win odds in the design of non-inferiority clinical trials. J Biopharm Stat. 2020;30(5): 941-6. doi:10.1080/10543406.2020.1757690.
5. Monzo L, Levy B, Duarte K, Baudry G, Combes A, Ouattara A, et al. Use of the win ratio analysis in critical care trials. Am J Respir Crit Care Med. 2024;209(7):798-804. doi:10.1164/rccm.202309-1644CP.
6. Mao L, Kim K, Li Y. On recurrent-event win ratio. Stat Methods Med Res. 2022;31(6):1120-34. doi:10.1177/09622802221084134.
7. Ajufo E, Nayak A, Mehra MR. Fallacies of using the win ratio in cardiovascular trials: challenges and solutions. JACC Basic Transl Sci. 2023;8(6):720-7. doi:10.1016/j.jacbts.2023.05.004.
8. 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):176-82. doi:10.1093/eurheartj/ehr352.
9. Wang B, Zhou D, Zhang J, Kim Y, Chen LW, Dunnmon P, et al. Statistical power considerations in the use of win ratio in cardiovascular outcome trials. Contemp Clin Trials. 2023;124:107040. doi:10.1016/j.cct.2022.107040.
10. Redfors B, Gregson J, Crowley A, McAndrew T, Ben-Yehuda O, Stone GW, et al. The win ratio approach for composite endpoints: practical guidance based on previous experience. Eur Heart J. 2020;41(46):4391-9. doi:10.1093/eurheartj/ehaa665.
11. Song J, Verbeeck J, Huang B, Hoaglin DC, Gamalo-Siebers M, Seifu Y, et al. The win odds: statistical inference and regression. J Biopharm Stat. 2023;33(2):140-50. doi:10.1080/10543406.2022.2089156.
12. Luo X, Tian H, Mohanty S, Tsai WY. An alternative approach to confidence interval estimation for the win ratio statistic. Biometrics. 2015;71(1):139-45. doi:10.1111/biom.12225.
13. Dong G, Li D, Ballerstedt S, Vandemeulebroecke M. A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components. Pharm Stat. 2016;15(5):430-7. doi:10.1002/pst.1763.
14. Oakes D. On the win-ratio statistic in clinical trials with multiple types of event. Biometrika. 2016;103(3):742-5. doi:10.1093/biomet/asw026.
15. Bebu I, Lachin JM. Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. Biostatistics. 2016;17(1):178-87. doi:10.1093/biostatistics/kxv032.
16. Luo X, Qiu J, Baid S, Tian H. Weighted win loss approach for analyzing prioritized outcomes. Stat Med. 2017;36(15):2452-65. doi:10.1002/sim.7284.
17. Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharm Stat. 2016;15(3):238-45. doi:10.1002/pst.1743.
18. Dong G, Hoaglin DC, Qiu J, Matsouaka RA, Chang YW, Wang J, et al. The win ratio: on interpretation and handling of ties. Stat Biopharm Res. 2020;12(1):99-106. doi:10.1080/19466315.2019.1575279.
19. Finkelstein DM, Schoenfeld DA. Graphing the win ratio and its components over time. Stat Med. 2019;38(1):53-61. doi:10.1002/sim.7895.
20. Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. J Biopharm Stat. 2018;28(4):778-96. doi:10.1080/10543406.2017.1397007.
21. Mao L. On the alternative hypotheses for the win ratio. Biometrics. 2019;75(1):347-51. doi:10.1111/biom.12954.
22. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. Stat Med. 1999;18(11):1341-54. doi:10.1002/(sici)1097-0258(19990615)18:11<1341::aid-sim129>3.0.co;2-7.
23. Hara H, van Klaveren D, Kogame N, Chichareon P, Modolo R, Tomaniak M, et al. Statistical methods for composite endpoints. EuroIntervention. 2021;16(18):e1484-95. doi:10.4244/EIJ-D-19-00953.
24. Wang H, Peng J, Zheng JZ, Wang B, Lu X, Chen C, et al. Win ratio -an intuitive and easy-to-interpret composite outcome in medical studies. Shanghai Arch Psychiatry. 2017;29(1):55-60. doi:10.11919/j.issn.1002-0829.217011.
25. Chiaruttini MV, Lorenzoni G, Spolverato G, Gregori D. Win statistics in observational cancer research: integrating clinical and quality-of-life outcomes. J Clin Med. 2024;13(11):3272. doi:10.3390/jcm13113272.
26. Brunner E, Vandemeulebroecke M, Mütze T. Win odds: an adaptation of the win ratio to include ties. Stat Med. 2021;40(14):3367-84. doi:10.1002/sim.8967.

Appendix-1

```
import csv
from random import choice, randint
def generate(n, sum_constraint, categories):
    generated = []
    generated_set = set()
    while len(generated) != n:
        sample = generate_sample(categories, sum_constraint)
        frozen = frozenset(sample.items())
        if frozen not in generated_set:
            generated_set.add(frozen)
            generated.append(sample)
    return generated
def generate_sample(categories, sum_constraint):
    current_constraint = sum_constraint
    already_used_categories = set()
    sample = {}
    while len(already_used_categories) != len(categories):
        category = choice(categories)
        if category not in already_used_categories:
            already_used_categories.add(category)
            value = choice(list(range(current_constraint)))
            sample[category] = value
            # print(sample)
            # You may try uncommenting the following line if sample generation takes too long
            # current_constraint = current_constraint - value
            if len(sample.keys()) == (len(categories) - 1):
                total = sum(sample.values())
                if total > sum_constraint:
                    print("Could not generate, sum is: ", total)
                    current_constraint = sum_constraint
                    already_used_categories = set()
                    sample = {}
                else:
                    unused_set = set(categories).difference(already_used_categories)
                    unused = unused_set.pop()
                    sample[unused] = sum_constraint - total
                    print("Generated this sample: ", sample)
                    return sample
def export_as_csv(data, filepath):
    print(f"Exporting to: {filepath}..")
    with open(filepath, mode='w') as out_file:
        csv_writer = csv.writer(out_file, delimiter=',')
        for sample in data:
            sample = dict(sorted(sample.items()))
            csv_writer.writerow(sample.values())
    print("Export complete.")
if __name__ == "__main__":
    n = 1000
    sum_constraint = 30
    start = 100
    end = 200
    step = 100
    for i in range(start, end, step):
        categories = ['A', 'B', 'C', 'D']
        output_filename = f"toplam{i}-{n}.csv"
        export_path = "./" + output_filename
        data = generate(n, i, categories)
        print("DATA: ", data)
        export_as_csv(data, export_path)
```