https://doi.org/10.26219/heal.aueb.6426


## SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS
POSTGRADUATE PROGRAM

# ON THE WIN-RATIO IN CLINICAL TRIALS

**By**
**Alexandra Chotzallari**

M.Sc. Thesis
Submitted to the Department of Statistics
of the Athens University of Economics and Business
in partial fulfillment of the requirements for
the degree of Master of Science in Statistics

Athens, Greece
February 2024

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426


ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ
ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ

# ΑΝΑΦΟΡΙΚΑ ΜΕ ΤΗΝ ΑΝΑΛΟΓΙΑ ΚΕΡΔΩΝ ΣΤΙΣ ΚΛΙΝΙΚΕΣ ΔΟΚΙΜΕΣ

Αλεξάνδρα Χοτζαλλάρι

Διατριβή
Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Στατιστική

Αθήνα
Φεβρουάριος 2024

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# DEDICATION

*To my beloved parents, Antony and Zoe,*
*whose unwavering support has been my anchor*

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# ACKNOWLEDGEMENTS

I would like to express my gratitude to those people who supported me throughout my journey to completing this thesis.

Foremost, I extend my heartfelt thanks to my supervisors, Prof. Xanthi Pedeli and Prof. Karlis Dimitris. Their guidance, advice, and adept solutions to the challenges I encountered were invaluable. Their enthusiasm and dedication to their work not only emanate from a genuine love for what they do but also translate into meaningful assistance and the offering of valuable knowledge to others, thanks to their extensive expertise.

Additionally, I am deeply appreciative of my family and close friends for their unwavering support, not only in the academic realm but in every facet of life. My beloved parents, Antony and Zoe Chotzallari, along with my brother George, consistently believed in me, offering assistance to navigate and overcome the obstacles that arose. Finally, I extend my love and appreciation to my partner, Prodromos, who stood steadfastly by my side during challenging moments, skillfully making them appear more manageable.

To everyone mentioned, your support has been a cornerstone, and I am truly grateful for your contributions to my success.

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# ABSTRACT

## *On the Win-Ratio in Clinical Trials*

*by Alexandra Chotzallari*

*February 2024*

Clinical trials constitute essentially a process of evaluating innovative medical methods or treatments when implemented in human subjects. One of the principal objectives is to explore the advantages and disadvantages of the new treatment, while simultaneously patient safety must be ensured. Any new treatment proposed to the general public must undergo detailed and thorough inspection in order to discover its therapeutic attributes, as well as its safety profile. A single endpoint or composite endpoints are employed to attain the aim of clinical trials. Composite endpoints consist of multiple individual endpoints, which are known as components of the composite endpoint and represent potential medical outcomes. Various techniques have been proposed in the literature for the analysis of endpoints. The traditional methods applied, such as the Cox model or Kaplan - Meier plots, have some limited aspects. Specifically, they assume equal medical importance for all components and consider only the first event that occurs, ignoring the medical hierarchy of events. For instance, the death of a patient is considered more important than hospitalization. Hence, the win ratio method provides a solution, as it takes into account the medical hierarchy of events and the relationship between the components.

To comprehend the functionality of the win ratio method, as well as its attributes, a detailed review of existing literature regarding it is presented in this thesis. Specifically, the concept of the win ratio is described, its methodology is analyzed and its fundamental elements are reported, such as the construction of a 95% confidence interval for the specific measure or the evaluation of the statistical power of the hypothesis testing. In addition, further details about the method are provided, such as discovering the appropriate sample size for each analysis or the adaptation and flexibility of the method to different scenarios. Furthermore, the extension of the method in question to the case of stratified data, as well as to the case of clustered data, is presented.

https://doi.org/10.26219/heal.aueb.6426



A simulation study is conducted to explore the operation of the win ratio method, to comprehend it, and to completely leverage its potential in the context of clinical research. Specifically, this technique is applied to a data set derived from a breast cancer clinical trial. The different parameters employed in each case lead to different scenarios. According to the results, there exists a close relation between the win ratio measure and the hazard ratios associated with the non-fatal and fatal events under consideration. The relationship between the win ratio and sample size is also evident, as well as the connection between the win ratio and the correlation, which exhibits between the times required for the occurrence of the non-fatal event and the fatal event.

In conclusion, the win ratio method emerges as a useful asset in the framework of clinical trials, possessing particular characteristics that render it privileged over conservative methods. Although there exist limitations that may prevent its widespread utilization, these can be addressed through thorough study, analysis, and attention to the details of the method. By solving these problems, wider implementation of the method will be possible, allowing easier and more efficient analysis of research data.

https://doi.org/10.26219/heal.aueb.6426

# ΠΕΡΙΛΗΨΗ

## *Αναφορικά με την Αναλογία Κερδών στις Κλινικές Δοκιμές*

*Αλεξάνδρα Χοτζαλλάρι*

*Φεβρουάριος 2024*

Οι κλινικές δοκιμές αποτελούν ουσιαστικά μια διαδικασία αξιολόγησης καινοτόμων ιατρικών μεθόδων ή θεραπειών, όταν αυτές εφαρμόζονται σε ανθρώπους. Ένας από τους κύριους στόχους είναι να διερευνηθούν τα πλεονεκτήματα και τα μειονεκτήματα της νέας θεραπείας, ενώ ταυτόχρονα πρέπει να διασφαλιστεί η ασφάλεια των ασθενών. Κάθε νέα θεραπεία που προτείνεται στο ευρύ κοινό πρέπει να υποβληθεί σε λεπτομερή και διεξοδικό έλεγχο, προκειμένου να γνωστοποιηθούν οι θεραπευτικές της ιδιότητες, καθώς και τα στοιχεία σχετικά με την ασφάλεια των ατόμων που υφίστανται τη θεραπεία αυτή. Για την επίτευξη του στόχου των κλινικών δοκιμών χρησιμοποιείται ένα τελικό σημείο ή σύνθετα τελικά σημεία. Τα σύνθετα τελικά σημεία αποτελούνται από πολλαπλά ατομικά τελικά σημεία, τα οποία είναι γνωστά ως συστατικά του σύνθετου τελικού σημείου και αντιπροσωπεύουν τα πιθανά ιατρικά αποτελέσματα. Διάφορες τεχνικές έχουν προταθεί στην βιβλιογραφία για την ανάλυση των τελικών σημείων. Οι παραδοσιακές μέθοδοι που εφαρμόζονται, όπως το μοντέλο Cox ή τα Kaplan - Meier plots, έχουν ορισμένες περιορισμένες πτυχές. Συγκεκριμένα, θεωρούν ισάριθμη ιατρική σημασία για όλα τα συστατικά και λαμβάνουν υπόψη μόνο το πρώτο συμβάν που συμβαίνει, αγνοώντας την ιατρική ιεραρχία των γεγονότων. Για παράδειγμα, ο θάνατος ενός ασθενούς θεωρείται σημαντικότερος από μια νοσηλεία στο νοσοκομείο. Επομένως, η μέθοδος αναλογίας κερδών προσφέρει μια λύση, καθώς λαμβάνει υπόψη την ιατρική ιεραρχία των γεγονότων και τη σχέση μεταξύ των συστατικών.

Για να κατανοηθεί ο τρόπος λειτουργίας της μεθόδου της αναλογίας κερδών, καθώς και τα χαρακτηριστικά της, μια λεπτομερής ανασκόπηση της υπάρχουσας βιβλιογραφίας σχετικά με αυτή παρουσιάζεται στην παρούσα διπλωματική εργασία. Συγκεκριμένα, περιγράφεται η έννοια της αναλογίας κερδών, αναλύεται η μεθοδολογία της και γίνεται αναφορά στα θεμελιώδη στοιχεία της, όπως η κατασκευή 95% διαστήματος εμπιστο-

https://doi.org/10.26219/heal.aueb.6426



σύνης για το συγκεκριμένο μέτρο ή η αξιολόγηση της στατιστικής ισχύος των ελέγχων υποθέσεων. Επιπλέον, παρέχονται περαιτέρω λεπτομέρειες σχετικά με την μέθοδο, όπως η εύρεση του κατάλληλου μεγέθους δείγματος για την κάθε ανάλυση είτε η προσαρμογή και ευκαμψία της μεθόδου σε διαφορετικά σενάρια. Επιπλέον, παρουσιάζεται η επέκταση της εν λόγου μεθόδου στην περίπτωση στρωματοποιημένων δεδομένων, καθώς και στην περίπτωση ομαδοποιημένων δεδομένων.

Μια μελέτη προσομοίωσης διεξάγεται με στόχο την εξερεύνηση της λειτουργίας της μεθόδου αναλογίας κερδών, την κατανόηση της, καθώς και την πλήρης αξιοποιήση των δυνατότητων που προσφέρει στο πλαίσιο της κλινικής έρευνας. Συγκεκριμένα, η τεχνική αυτή εφαρμόζεται σε ένα σύνολο δεδομένων που προέρχεται από μια κλινική δοκιμή για τον καρκίνο του μαστού. Οι διάφορες παράμετροι που χρησιμοποιούνται σε κάθε περίπτωση οδηγούν σε διαφορετικά σενάρια. Σύμφωνα με τα αποτελέσματα, υφίσταται στενή σχέση μεταξύ του μέτρου της αναλογίας κερδών και των αναλογιών κινδύνου που συνδέονται με το υπό μελέτη μη θανατηφόρο και θανατηφόρο γεγονός. Εμφανής γίνεται και η σχέση μεταξύ της αναλογίας κερδών και του μεγέθους δείγματος, καθώς και ο δεσμός μεταξύ της αναλογίας κερδών και της συσχέτισης, η οποία παρουσιάζεται μεταξύ των χρόνων που απαιτούνται για την εκδήλωση του μη θανατηφόρου γεγονότος και του θανατηφόρου γεγονότος.

Συγκεντρωτικά, η μέθοδος της αναλογίας κερδών αναδεικνύεται ως ένα χρήσιμο εργαλείο στον τομέα των κλινικών δοκιμών, διαθέτοντας ιδιαίτερα χαρακτηριστικά που την καθιστούν προνομιούχα έναντι των κλασικών μεθόδων. Αν και υπάρχουν περιορισμοί που μπορεί να αποτελούν εμπόδια στην ευρεία χρήση της, αυτοί μπορούν να αντιμετωπιστούν μέσω διεξοδικής μελέτης, ανάλυσης και αφιερώνοντας προσοχή στις λεπτομέρειες της μεθόδου. Με την επίλυση αυτών των προβλημάτων, θα είναι δυνατή η ευρύτερη χρήση της μεθόδου, επιτρέποντας την ευκολότερη και πιο αποτελεσματική ανάλυση των ερευνητικών δεδομένων.

https://doi.org/10.26219/heal.aueb.6426

# Contents

https://doi.org/10.26219/heal.aueb.6426

https://doi.org/10.26219/heal.aueb.6426

# List of Tables

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# List of Figures

https://doi.org/10.26219/heal.aueb.6426





https://doi.org/10.26219/heal.aueb.6426

# Chapter 1

# Introduction

## 1.1 Motivation of the thesis

The objective of a clinical trial is to systematically assess the potential benefits and drawbacks of a novel medical strategy, technique, or essence in human subjects. By following stringent protocols and methodologies, these trials seek to ascertain whether the new intervention holds promise in outperforming existing standards or filling therapeutic voids in the field of science. They play a decisive role in shaping the future of medicine, ensuring that every new intervention introduced to the public has been comprehensively analyzed for both its therapeutic potential and safety profile (National Institutes of Health (2017); World Health Organization (2020)). Valid primary efficacy endpoint(s) are used to implement this purpose. It is crucial to begin with the careful selection of a single primary efficacy endpoint. This endpoint should provide a detailed description of the disease being studied. It should be easily and consistently diagnosed and guarantee unbiased identification and classification. It is equally important to ensure that its interpretation is clear and precise, as it will be required in drawing conclusions. Moreover, the primary endpoint should be associated with uncomplicated available statistical analysis techniques. Its purpose is to enable an optimized assessment of the treatment's effectiveness from both a clinical and statistical standpoint while offering a comprehensive and easily understood assessment of drug effect. In two-arm clinical trials, a meticulous assessment is performed, carefully comparing the observations associated with the primary endpoint in both the control and treatment groups. This careful comparison is pivotal in constructing robust scientific evidence concerning the potential efficacy of the new treatment, following a clearly defined monitoring period (see Figure 1.1) (Sankoh et al. (2014)).

However, in numerous studies, more than one outcomes are observed, which are characterized by the same degree of importance and significance, a fact that leads to their combination in one primary endpoint. It is also essential to arrive at a pure estimate of clinical treatment benefit because it maintains evidence-based medicine, patient-centered care, and informed decision-making in healthcare. It involves rigorous research methodologies and an even understanding of how treatments perform

https://doi.org/10.26219/heal.aueb.6426



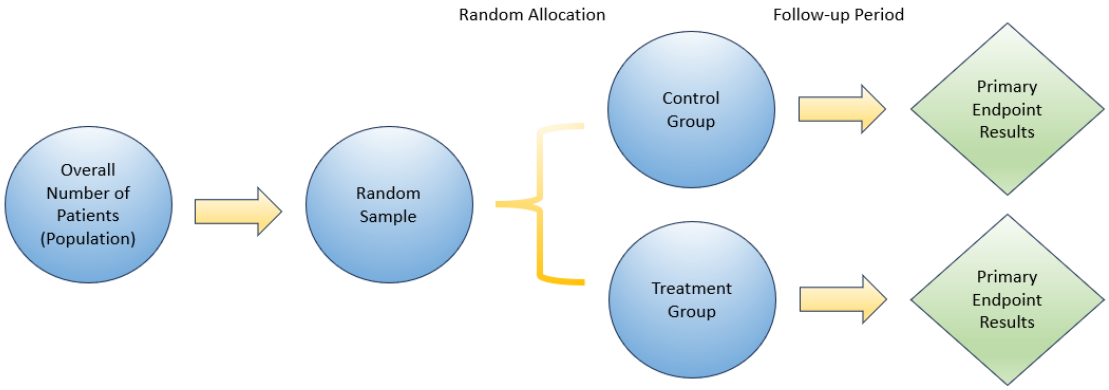



*Figure 1.1:* Structure of a randomized vertificative clinical trial (Sankoh et al. (2014)).

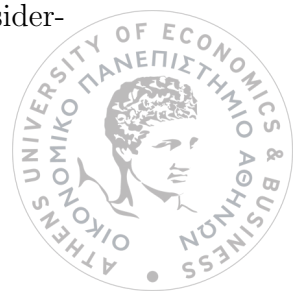
in real-world scenarios, ultimately improving patient outcomes and the overall efficiency of healthcare systems. It is noteworthy to refer that it is profitable to handle multiplicity adequately without Type I error adjustments. While Type I error adjustments play a primary role in controlling false positives, there are situations where adequately addressing multiplicity without such adjustments can be justified, especially in exploratory or preliminary research settings. At the same time, it is widely known that the problem of bias related to competing risks is of particular concern to researchers. Their aim is to eliminate bias and obtain as accurate a result as possible. Researchers are dedicated to adopting appropriate methods and strategies to achieve this goal and to make informed decisions in healthcare and research. Usually, the statistical power of a trial increases as the event rate increases. Specifically, statistical power is an underlying concept in research, particularly in hypothesis testing and the design of experiments or clinical trials. It represents the probability of correctly detecting a true effect or difference when it exists. The event rate refers to the frequency or likelihood of the event of interest occurring. This event could be a medical condition, a specific outcome, or any other defined event that researchers are investigating. When the event of interest is rare or has a low occurrence rate, it can be challenging to detect a statistically significant difference or effect because there are fewer instances of the event happening. This leads to a lower statistical power. Conversely, when the event rate increases, there are more occurrences of the event, providing a larger sample size for analysis. A larger sample size generally results in increased statistical power because there is more information available to detect differences or associations. With a higher event rate, it becomes easier to distinguish between groups or treatments, making it more likely to detect true effects. A higher event rate often leads to narrower confidence intervals and more precise estimates of the effect size. This increased precision enhances the ability to make confident conclusions about the effectiveness of an intervention or the presence of a relationship. All of the above contribute to the adoption and utilization of composite endpoint(s) in the design and evaluation of clinical trials (Gómez-Mateu (2016); McCoy C. Eric (2018)).

https://doi.org/10.26219/heal.aueb.6426



A composite endpoint, often utilized as the primary outcome measure in Randomized Clinical Trials, endues multiple individual endpoints, known as component endpoints. These component endpoints, when observed individually, may fail in the requisite sensitivity and reliability to entirely capture the desired characterization of the drug's effect. To contend the overall event rate, particularly for binary or time-to-event outcomes, or to attain a more sensitized endpoint in the case of continuous outcomes, researchers prefer the merger of these component endpoints. In essence, a composite endpoint represents the union of two or more distinct outcomes, and it is deemed achieved if any of the individual outcomes within the composite endpoint is observed. Additionally, composite endpoints may encompass total scores or health indices derived from rating scales, further broadening their applicability and relevance in clinical research. In the field of time-to-event analysis, composite endpoints operate as a precious approach to assessing clinical outcomes. These composite endpoints are constructed by measuring the time interceded from the point of randomization until the occurrence of the first component within the composite endpoint. For instance, in cancer studies, progression-free survival is a commonly used composite endpoint, implying the time it takes for either disease progression or death to be revealed. In essence, these composite endpoints offer an inclusive perspective on clinical trial outcomes, efficiently incorporating a range of critical events while determining the time it takes for the earliest of these events to occur. This approach proves particularly valuable in evaluating the efficacy of interventions across various medical domains. It is additionally worth noting that a high event rate or a great sensitivity of the endpoint reduces the required number of participants to ensure there is enough statistical power to detect a predetermined meaningful treatment benefit. Thus, employing a composite endpoint could lead to fewer individuals being subjected to unverified and potentially unsafe substances. Additionally, this approach offers economic profits, including cost savings associated with conducting trials with fewer participants and the expedited availability of promising new treatments (Sankoh et al. (2014); FDA (2022)).

Numerous techniques have been proposed for merging or integrating multiple endpoints, each offering its own unique approach and advantages depending on the context and application in which it is used. There are two ways of analyzing composite endpoints, the distinction of which is based on the characteristic of the components of the composite endpoints to be distinct events. Notably, a component of a composite endpoint is defined as a distinct event with its own clinical significance when its realization entails the occurrence of the composite endpoint. When the components represent distinct events, they are typically evaluated by measuring the time until the initial event, where the event could be any one of the components. Nevertheless, in certain scenarios, patients may encounter multiple events, in which case it is suitable to examine the cumulative count of endpoint events. In the initial scenario, the composite endpoints are assessed through a time-to-first-event analysis. This analysis utilizes statistical methods such as the Cox model, a log-rank test, and Kaplan-Meier plots to yield a hazard ratio, a confidence interval, and a p-value as outcomes. In the alternative scenario, all events are taken into consider-



https://doi.org/10.26219/heal.aueb.6426



ation, regardless of their sequence. Patients participate in event measurements for multiple components. The statistical methods that constitute this comprehensive event analysis include negative binomial regression, Andersen-Gill regression, and weighted analysis of composite endpoints. Despite this, the comparative approach is characterized as particularly complicated, and often the occurrence of an event, such as death, does not allow the realization of another event in the same patient. Thus, one group appears to have a higher probability of an event relative to other endpoints. Furthermore, time-to-first-event analysis is widely used to assess these types of outcomes. This approach, however, is determined by providing the same degree of severity to all components of the composite endpoint and taking into account the first event to occur. For example, we consider that a patient suffers from cerebrovascular disease. If the patient initially incurs a stroke and a few days later dies, only the non-fatal event will be considered in the investigation and not the death, which is of greater significance (Baracaldo-Santamaría et al. (2023); Huque et al. (2018)).

In order to overcome these limitations, the win ratio method has been introduced for the comprehensive analysis of composite endpoints, taking into account both the hierarchy of clinical significance and the temporal relationship among its constituent elements. In contrast to the conventional approach, this particular method provides the flexibility to analyze various types of endpoints. For instance, the win ratio method allows the examination of time-to-event endpoints like mortality alongside recurrent or longitudinal endpoints, such as the frequency of hospitalizations. Since its introduction, this approach has garnered significant attention, not only in its practical applications but also in its methodological development (Baracaldo-Santamaría et al. (2023)). The primary objective of this thesis is to consolidate the existing literature on the win ratio, its various extensions, inferential aspects, and trial design considerations. Furthermore, it seeks to apply these methodologies and evaluate their performance using a substantial dataset derived from a breast cancer clinical trial. This comprehensive exploration aims to enhance our understanding and harness the full potential of the win ratio method in the context of clinical research.

## 1.2 Concise outline of the thesis

Over the years, the win ratio method has gained considerable prominence, contributing to various perspectives in the literature concerning its definition, attributes, and properties, alongside its diverse applications and utilizations. This document thoroughly investigates and reports these aspects, aiming to fulfill the aforementioned objectives.

The principal part of the thesis initiates in Chapter 2, delving into the introduction of the win ratio concept and an overall analysis of its methodology. This section not only examines the fundamental characteristics of the win ratio but also studies the complications of calculating the p-value and clarifies methodologies for constructing robust 95% confidence intervals. In addition, it highlights systematic

https://doi.org/10.26219/heal.aueb.6426



procedures for assessing the power of hypothesis testing, offering an absolute understanding of the method's implementation and significance in statistical analysis.

Following a comprehensive investigation of the fundamental aspects in Chapter 2, Chapter 3 provides more details about the win ratio method. This part discusses diverse folds, encompassing the specifying of required sample sizes for each analysis and the proper handling of ties, a category emerging from the implementation of the win ratio method. Furthermore, the chapter refers to the flexibility of the method when faced with various outcome weights and offers insights into its graphical representation and its application to scenarios involving continuous non-normal data. The final sections of Chapter 3 explore the utilization of the win ratio in evaluating composite endpoints when the components exhibit opposing treatment effects. The chapter concludes with an examination of the win ratio expressed in four distinct forms, leading to an understanding of its variations and applicability across different analytical perspectives.

Moving into Chapter 4, the concept of the win ratio is extended to assess stratified data. This section offers a detailed composition of the stratified win ratio, comprising its definition and distribution. An inextricable aspect of this exploration includes the examination of the homogeneity across strata. Additionally, the chapter describes the construction of a 95% confidence interval and the computation of the power of the hypothesis test, specifically adjusted for the stratified win ratio. The chapter also reports scenarios involving diverse weights assigned to outcomes within the stratified win ratio framework. Furthermore, a discrimination of the stratified win ratio into four different categories is included, clarifying its multiple implementations within the realm of statistical analysis.

Next, Chapter 5 reinforces the application of the win ratio method to the context of clustered data. This section thoroughly introduces the definition of the win ratio for cluster-randomized data and provides insights into the distribution associated with this specific metric.

Chapter 6 involves a simulation study centering on diverse scenarios. The objective of this chapter is to implement the win ratio method on data originating from a breast cancer clinical trial. The aim is to distinguish and comprehend the agents influencing this particular analytical measure. We employ different parameters to generate various datasets, leading to the creation of distinct scenarios. Exploring the results derived from these scenarios contributes to drawing meaningful inferences regarding the effectiveness of the methodology and its applicability in real-world scenarios.

Chapter 7 provides a comprehensive summary of the pivotal marks of the thesis. It not only encompasses the substantial findings but also establishes the stage for discussion and offers feasible recommendations for future research efforts aimed at expanding the implementation of the win ratio method.

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# Chapter 2

# Win Ratio Method

## 2.1 Fundamental definitions and notation

### 2.1.1 What is the win ratio?

The win ratio is a statistical method that has gained prominence in the analysis of composite endpoints, particularly in clinical research. One notable aspect of the win ratio method is its versatility and adaptability, making it applicable in a wide range of practical scenarios and research applications. Researchers have found it to be a valuable tool for evaluating and summarizing complex clinical data. The win ratio approach has attracted significant attention in both methodology development and practical applications, as it offers a differentiated perspective on the assessment of composite endpoints. Researchers have explored various extensions and inferences associated with the win ratio, enhancing its utility and broadening its scope of application. In the context of clinical trials and medical research, the win ratio method provides a valuable framework for analyzing composite endpoints in a manner that is related to the clinical significance and temporal order of events, contributing to a more comprehensive understanding of the data. It has been particularly employed in the analysis of healthcare outcomes and patient-related studies. The method also recognizes that patients have differing risk profiles, by utilizing risk-matched pairs.

In recent years, there has been a growing and dynamic landscape in the development and diversification of methodologies related to the win ratio approach. Researchers have actively formulated and refined various strategies and techniques to enhance the utility and flexibility of this method in the analysis of composite endpoints in order to provide more robust and informative analyses, ensuring that the method remains relevant and effective in a rapidly evolving healthcare treatment. Pocock et al. (2012) contributions to the win ratio method enclose the presentation of two distinct versions, each designed to provide two different analytical scenarios. These two methodological variants offer researchers a valuable choice when employing the win ratio approach. The first version is known as the matched pairs approach, while the second constitutes the unmatched approach.

https://doi.org/10.26219/heal.aueb.6426



### 2.1.2 The matched pairs approach

In the particular strategy employed in this method, we categorize participants into two distinct sets. Typically, one group consists of individuals who are administered a placebo substance, while the other comprises those who are subjected to the intended treatment. Alternatively, it could involve a group receiving the conventional standard of care and another group being exposed to the novel treatment approach. We form matched pairs utilizing the individuals from the two groups while considering each participant's individual risk profile. Typically, the two groups do not have an equal number of subjects, resulting in the exclusion of some individuals from the larger group. The selection of these specific individuals is done randomly. If the exclusion of certain subjects from the study is assumed problematic, an alternative approach would involve conducting various analyses, each with a different set of excluded patients. Ultimately, the median win ratio across these replicated analyses could be used as an estimation.

In our analysis, we focus on two distinct events, with one carrying greater significance than the other. These events can be classified as non-fatal, such as experiencing a non-fatal medical issue, and fatal, which involves the occurrence of death. We begin the analysis by examining each pair of individuals to determine which of the patients experienced the death first. If this information is unknown, we proceed to assess the occurrence of the non-fatal event using a similar approach. Furthermore, there's a scenario where neither of the two patients encounters an event, or one patient experiences an event while the observation period for the other patient in the pair is shorter. In the latter situation, we're dealing with what is known as censored data. This process results in the creation of five (5) distinct categories:

i) The patient receiving the new treatment experiences the fatal event first.

ii) The patient receiving the conservative treatment faces the fatal event first.

iii) The patient underlying the new treatment encounters the non-fatal event first.

iv) The patient underlying the conservative treatment experiences the non-fatal event first.

v) No event occurs for either patient in the pair.

Figure 2.1 illustrates the various scenarios classifying any pair into categories (i)–(v).

The aggregation of individuals who experienced either the fatal or non-fatal event first, originating from the conservative treatment group, represents the count of winners for the new treatment. This is because the conventional treatment, employed previously, yielded unfavorable outcomes. Similarly, the sum of individuals from the new treatment group who faced either the fatal or non-fatal event first constitutes the number of losers for the new treatment, as it fails to yield promising and satisfactory results. The latter category of patients comprises individuals commonly referred to as "tied". We suppose that the number of winners is $N_W$, while the number of losers is assumed to be $N_L$. Therefore, the win ratio is defined as follows,

https://doi.org/10.26219/heal.aueb.6426




*Figure 2.1:* A visual representation illustrating potential scenarios for the win ratio method. It is supposed that 'S' denotes an individual under the standard treatment, and 'N' signifies an individual receiving the new treatment. It is mentioned that when the new treatment outperforms the conventional one, we can categorize the first three cases as wins in terms of survival (fatal event - death), while the remaining three cases represent successes relevant to the non-fatal event (Pocock et al. (2012)).

$$WR = \frac{N_W}{N_L}$$

It is worth noting to mention that the proportion can be expressed through the following relationship,

$$P_W = \frac{N_W}{N_W + N_L}$$

This allows us to determine the percentage of participants who experienced positive outcomes from the new treatment within the entire group of participants, encompassing both winners and losers. Additionally, a 95% confidence interval is established to provide a range within which we can be reasonably certain that the true value of this proportion falls.

$$P_W \pm 1.96\sqrt{\frac{P_W(1-P_W)}{N_W + N_L}} = L, U$$

where $L$ denotes the lower bound of the CI, while $U$ determines the upper limit of CI.

These two measures are linked through the following equation, which has a pivotal significance in defining the confidence interval for the win ratio.

$$WR = \frac{P_W}{1 - P_W}$$

This relationship contributes to understanding the range of values within which we can be 95% certain that the win ratio of the new treatment lies. It essentially allows

https://doi.org/10.26219/heal.aueb.6426



us to assess the degree of confidence we can have in the benefits provided by the new treatment in comparison to the conventional one. The 95% CI for the win ratio is given by,

$$\frac{L}{1-L}, \quad \frac{U}{1-U}$$

Testing the null hypothesis that there is no significant statistical difference between the two groups, which is identical to testing that the win ratio is equal to 1, can be implemented using the z-statistic, which follows the standard normal distribution under the null hypothesis.

$$z = \frac{P_W - 0.5}{\sqrt{\frac{P_W(1-P_W)}{N_W + N_L}}} \sim N(0,1)$$

We can also calculate the respective figures for the adverse outcome, specifically mortality. Assuming that among the patients who encountered death initially, $N_S$ originated the standard treatment group, while $N_N$ individuals appertained to the new treatment group, we determine the win ratio just for death as,

$$WR_D = \frac{N_S}{N_N}$$

Moreover, the proportion and its corresponding 95% confidence interval can be determined through the following equations.

$$P_{W_D} = \frac{N_S}{N_S + N_N}$$

$$P_{W_D} \pm 1.96\sqrt{\frac{P_{W_D}(1-P_{W_D})}{N_S + N_N}} = L_D, U_D$$

We also refer to the connection between these two metrics, which entails the calculation of the 95% confidence interval for the death win ratio.

$$WR_D = \frac{P_{W_D}}{1-P_{W_D}}$$

$$\frac{L_D}{1-L_D}, \quad \frac{U_D}{1-U_D}$$

Lastly, the z-statistic is transformed as outlined below.

$$z_D = \frac{P_{W_D} - 0.5}{\sqrt{\frac{P_{W_D}(1-P_{W_D})}{N_S + N_N}}} \sim N(0,1)$$

It's important to mention that the count of individuals categorized as "tied," repre-

https://doi.org/10.26219/heal.aueb.6426



sented as $N_T$. Essentially, this refers to the overall number of participants who didn't experience either of the two events, or situations where one patient experienced an event while the follow-up time for the other patient in the pair was shorter. A valuable statistic, known as the proportion of tied pairs, is determined as the count of individuals classified as tied, divided by the total number of participants included in the analysis (N).

$$P_T = \frac{N_T}{N}$$

Its 95% confidence interval employs the identical formula as used for $P_W$, with the substitution of $P_W$ with $P_T$ and $N_W + N_L$ with N.

This particular approach of the method can be further expanded to encompass the analysis of composite endpoints involving three or more components. In such cases, the methodology can be adapted and applied to comprehensively evaluate and interpret complex outcome combinations, offering a broader perspective on the study's results (Pocock et al. (2012)).

### 2.1.3 The unmatched approach

In certain clinical trial scenarios, the establishment of matched pairs for study participants is not feasible, leading to a unidirectional or unmatched approach. In this context, the unmatched approach is a one-way street, meaning that the comparisons between groups are made without the advantage of pairing similar individuals based on relevant variables. In this particular method, each patient in the new treatment group is systematically compared with every patient in the conservative treatment group. The objective is to identify which of the specific events occurs first, taking into consideration the order of clinical importance associated with these events. Assuming that $N_S$ represents the count of participants in the standard treatment group and $N_N$ symbolizes those in the new treatment group, it is obtained that the total number of subject-to-subject comparisons will be equal to $N_S \times N_N$.

Similar to the approach used with matched pairs, we are interested in determining the number of winners and losers for the new treatment. The categories that result from this comparison align with those observed in the matched pairs approach. Therefore, the winners are individuals from the conservative treatment group who experienced either a fatal or non-fatal event before their counterparts in the new treatment group. On the other hand, the losers are considered to be the total number of individuals from the new treatment group who experienced the fatal or non-fatal event before the conservative treatment group. As a result, the win ratio is represented using the formula $WR = N_W/N_L$, where $N_W$ depicts the number of winners, while $N_L$ symbolizes the number of losers.

It's requisite to highlight that the overall count of patients involved in the study, classified into the five categories following the comparison, equals $N_W \times N_L$. These unmatched pairs should not be perceived as entirely independent comparisons. In this approach, each patient from both the conservative and the new treatment groups

https://doi.org/10.26219/heal.aueb.6426



is effectively considered multiple times – $N_S$ times for those in the new treatment group and $N_N$ times for those in the conservative treatment group. This is due to the extensive cross-comparisons made between individuals in these groups, as each patient is matched with multiple counterparts from the other group to determine the order of event occurrence (Pocock et al. (2012)).

### 2.1.4 Expressing the win ratio in terms of win proportions

Dong et al. (2020) conducted comprehensive research regarding the win ratio, emphasizing introducing an alternative representation. Their focus centered on propulsion for the employing win proportions relative to the number of wins, claiming that the latter exhibits variability based on sample size and tends to yield large values, given its dependence on all potential patient pairs. Thus, they decided to formulate the win ratio as a ratio of win proportions,

$$WR = \frac{P_N}{P_S}$$

This involves the definitions,

$$P_N = \frac{N_{W_N}}{N_N \cdot N_S}$$

and

$$P_S = \frac{N_{W_S}}{N_N \cdot N_S}$$

Here, $N_{W_N}$ represents the number of wins in the new treatment group during patient comparison, $N_{W_S}$ signifies the number of wins in the standard treatment group, $N_N$ indicates the number of participants in the new treatment group, and $N_S$ denotes the number of subjects in the standard treatment group (Additional relative details regarding their analysis are available in the section 2.3.6). With this expression of the win ratio available, its interpretation comes to be straightforward. For instance, suppose WR = 1.3. According to the former clarification, this indicates that a patient in the new treatment group has a 30% higher probability of winning compared to a patient in the standard treatment group.

Another vital subject highlighted by the researchers concerns the computation of tie proportions, for which they utilized the formula,

$$P_{\text{tied}} = \frac{N_T}{N_N \cdot N_S}$$

Here, $N_T$ denotes the number of ties and can be defined employing the expression,

$$N_T = N_N \cdot N_S - N_{W_N} - N_{W_S}$$

Furthermore, in addition to the win proportion ratio, they introduced an alternative measurement, the difference of win proportions,

$$\text{WD} = P_N - P_S$$

Overall, it is assumed favorable and effective in the analysis to combine the conventional definition of the win ratio with the alternative formulations reported in this section.

https://doi.org/10.26219/heal.aueb.6426



## 2.2 Calculation of p-value in unmatched approach

Calculating the p-value for the win ratio can be a complex process. To address this challenge, Finkelstein and Schoenfeld (1999) laid out a comprehensive framework for conducting significance testing. Their work provides a structured methodology for assessing the statistical significance of win ratios, enabling researchers to make informed decisions about the efficacy and impact of different treatment approaches in clinical trials. This framework involves a rigorous analysis of the data, taking into account the win ratio as a critical indicator of treatment success. More precisely, this method involves a straightforward non-parametric test designed to assess whether a statistical difference exists between the two treatments under investigation. The null hypothesis posits that treatment has no impact on either survival or the longitudinal measure. The alternative hypothesis, on the other hand, suggests that there is an improvement in at least one, and potentially both, of these factors. It appears to integrate both a time-to-event metric and a longitudinal measure, effectively combining these elements to determine the treatment's comparative effectiveness. It's worth noting that longitudinal studies utilize continuous or repetitive data collection or measures for monitoring specific individuals across extended duration, which can endure many years or even decades (National Library of Medicine (2015)). Moreover, it's important to emphasize that this particular test leans heavily on the Wilcoxon rank test, rather than the Cox model, as a primary statistical method.

In crafting the statistical test, comprehensive comparisons are conducted among all patients. Specifically, each member from the standard treatment group, encompassing $N_S$ individuals, is systematically paired with every individual from the new treatment group, which has a total of $N_N$ members. The overall number of participants in the study equals $N = N_S + N_N$. As previously outlined in our method, we differentiate between the winners and losers of the new treatment. Let's introduce a variable, $U_{ij}$, which corresponds to the outcome of comparing patient $i$ with patient $j$. This variable assumes one of three distinct values depending on the comparison's outcome.

$$U_{ij} = \begin{cases} 1 & \text{, if patient } i \text{ is the winner} \\ -1 & \text{, if patient } i \text{ is the loser} \\ 0 & \text{, if the patients are tied} \end{cases}$$

We further define the variable $D_i$, which is assigned a value of 1 if patient $i$ originates from the new treatment group, and a value of 0 if patient $i$ is included in the conservative treatment group. Additionally, $U_i = \sum_{i \neq j} U_{ij}$ is determined for each patient, who participates in the research. We mention that the variable $U_i$ constitutes a positive integer when patient $i$ wins more often than loses compared to all other patients.

Therefore, the subsequent statistical function is employed as the foundation for the implementation of the significance test.

https://doi.org/10.26219/heal.aueb.6426



$$T = \sum_{i=1}^{N} U_i D_i$$

Under the assumption of the null hypothesis, the variance of the T-statistic is defined as follows,

$$Var = \frac{N_N N_S}{N(N-1)} \sum_{i=1}^{N} U_i^2$$

In this context, understanding the variance of the T-statistic is crucial because it helps to evaluate the stability and reliability of the statistical test results. This measure allows researchers to assess the degree of variability expected when no treatment effect is present and forms a critical part of the test's overall statistical framework.

Subsequently, $z = \frac{T}{\sqrt{Var}}$ represents a standardized normal deviate, providing a convenient pathway to easily obtain the p-value. This streamlined approach accommodates the quick and efficient determination of the p-value, a vital component in drawing statistical inferences. Specifically, utilizing the outcomes mentioned, when z exceeds 1.96, 2.38, and 3.29, this signifies that the associated p-values are less than 0.05, 0.01, and 0.001, respectively. These distinct z-values provide a clear indicator of the level of statistical significance, aiding researchers in estimating the strength of their results. This particular approach is broadened to encompass scenarios involving stratified data, details of which will be discussed in subsequent sections.

Dong et al. (2016) carried an independent investigation out focusing on the win ratio method and proposed an alternative approach, summarized in subsection 2.3.6. In their analysis, they innovatively integrated kernel functions and availed themselves of $U$-statistics to derive the distribution of winners within each group. They also analyzed the joint distribution of winners in the treatment and control groups. By taking a careful look at the distribution of the logarithm of the win ratio, they could conclude and offer a novel perspective on the efficacy of hypothesis testing, thereby contributing to a more ideal understanding of statistical power. Hence, for evaluation of the null hypothesis $H_0$, establishing that $\log(WR) = 0$ or equivalently $WR = 1$, in contrast to the alternative hypothesis $H_1$, inserting $\log(WR) > 0$ or equivalently $WR > 1$, they proposed a process to compute the p-value. This computation is expressed as,

$$\text{p-value} = 1 - \Phi\left(\frac{\log(WR)}{\hat{\sigma}_{\log(WR)}}\right)$$

This formula pertains to a one-sided test. However, in scenarios where a two-sided test is performed—specifically when the null hypothesis is $H_0 : \log(WR) = 0$ or equivalently $WR = 1$, against the alternative $H_1 : \log(WR) \neq 0$ or $WR \neq 1$—the p-value is given by,

$$\text{p-value} = 2 \cdot \left[1 - \Phi\left(\left|\frac{\log(WR)}{\hat{\sigma}_{\log(WR)}}\right|\right)\right]$$

https://doi.org/10.26219/heal.aueb.6426



In this context, $\Phi$ denotes the cumulative distribution function of the standard normal distribution, while the $\hat{\sigma}_{\log(WR)}$ is presented in section 2.3.6. This approach provides a comprehensive framework for hypothesis testing, comprising both one-sided and two-sided scenarios, as elucidated by Dong and colleagues.

## 2.3 Calculation of 95% CI in unmatched approach

Calculating a 95% confidence interval for the win ratio utilizing the unmatched approach presents challenges because the comparisons among patients are not truly independent. Recent research has highlighted diverse strategies for implementing the unmatched method, leading to variations in the methods employed to establish the 95% CI. This variability underscores the need for multiple procedures to determine the CI, reflecting the complexity of the deed.

### 2.3.1 Utilizing the bootstrap technique

The particular calculation's intricate nature and the absence of a straightforward analytical solution often lead us to lean on numerical and computational techniques. Widely adopted methods comprise the bootstrap, jackknife, cross-validation, and Monte Carlo simulation. One of the key benefits of these methods is that they don't require predefined assumptions about our data's distribution and they ordinarily yield results that closely align with traditional analytical methods. In our case, we'll focus on the bootstrap method, which was introduced by Efron in 1977 and has since created opportunities for statistical analysis in complex scenarios.

This method is contributory for estimating unknown parameters associated with the sampling distribution of estimators and statistical tests. Frequently, the bootstrap method comes into consideration when we need to determine standard errors for estimates, establish confidence intervals for unknown parameters, and calculate p-values for test statistics, particularly when assessing hypotheses grounded in a null hypothesis. It is essentially a resampling procedure, which involves repeatedly drawing B random samples, with replacement, from an original observed dataset to create a large number of simulated datasets. Within each of these simulated datasets, we compute the estimator for the specific quantity of interest, ultimately allowing us to build its empirical sampling distribution (Johnson (2001); Boos (2003); Fook Chong Stephanie (2011)). Specifically, for the determination of the 95% confidence interval for the win ratio, we follow the subsequent procedure:

1. We are examining two categories of individuals – one group receiving the standard treatment, referred to as "$S$," and it consists of $N_S$ subjects, while the other group is undergoing the new treatment, denoted as "$N$," and comprises $N_N$ individuals.

2. We draw a random sample $B_{S_1}$ of size $N_S$ with replacement from the initial conservative treatment group. Similarly, we formulate a random sample $B_{N_1}$ of size $N_N$ with replacement from the original new treatment group.

https://doi.org/10.26219/heal.aueb.6426



3. We implement the unmatched approach on the two generated samples, $B_{S_1}$ and $B_{N_1}$, in order to compute the win ratio.

4. We iterate through steps 2 and 3 a significant number of times, which is denoted as B and typically exceeds 1000.

5. We establish the empirical bootstrap distribution for the win ratio using the B bootstrap values.

6. We extract the $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles from this distribution, representing the estimated boundaries of the 95% confidence interval for the win ratio.

(Pocock et al. (2012)).

### 2.3.2 Utilizing logarithmic transformation

Furthermore, an alternative and simplified method for establishing a 95% CI has been discussed in the literature. In this approach, once we have computed the win ratio value, we also calculate its logarithm, denoted as $\log(WR)$, and the approximate standard error of $\log(WR)$. This standard error is determined as $se = \frac{\log(WR)}{z}$, where '$z$' represents the standardized normal deviate, as defined by the significance test mentioned earlier. Consequently, the 95% CI for $\log(WR)$ is computed as $\log(WR) \pm 1.96\, se$. In order to estimate the corresponding 95% confidence interval for the $WR$, we exponentiate the result, providing us,

$$\exp(\log(WR) \pm 1.96 \cdot se)$$

(Pocock et al. (2012)).

### 2.3.3 Utilizing the large sample distribution of certain multivariate multi-sample $U$-statistics

Bebu and Lachin (2015) conducted a study to establish a 95% confidence interval utilizing the large sample distribution of certain multivariate multi-sample $U$-statistics. Their approach involved the introduction of two distinct variables, $X = (X_1, X_2, \ldots, X_k)$ and $Y = (Y_1, Y_2, \ldots, Y_k)$, representing the $k$ potential outcomes for two individuals. One individual originated the new treatment group, while the other was from the conservative treatment group. These outcomes were categorized based on their clinical severity, starting with the most severe cases, and various measurements were taken to assess them. In this particular analysis, the assumption was made that all the results were associated with the time of occurrence of an event. The comparison of these individuals commenced with the most severe outcomes, specifically $X_1$ and $Y_1$. In instances where the available data did not provide conclusive evidence to determine a clear winner in the comparison, the analysis proceeded to the next most severe outcomes, such as $X_2$ and $Y_2$, and continued in this manner.

https://doi.org/10.26219/heal.aueb.6426



Furthermore, we import two functions, $\varphi_1(X,Y) = I_{\{X>Y\}}$ and $\varphi_2(X,Y) = I_{\{X<Y\}}$. These functions serve as binary indicators, determining whether the individual from the new treatment group or the individual from the standard treatment group is the winner in a given comparison, respectively. In this context, we determine $\tau_v = \mathbb{E}(\varphi_v(X,Y))$, where $v = 1,2$, denoting the associated probabilities. Essentially, $\tau_1$ represents the probability that the individual from the new treatment group comes out as the winner, while $\tau_2$ represents the probability that the individual from the standard treatment group emerges as the winner.

In our statistical framework, we assume that $X$ follows a distribution $F$ and $Y$ follows a distribution $G$, with $F$ and $G$ representing the joint distributions of the $k$ outcomes within the two groups. In this context, under the null hypothesis that there is no distinction between the two groups, indicating $F = G$, it follows that $\tau_1 = \tau_2 = \tau$. However, in cases where there is a possibility of no winner being definitively determined, i.e. the probability of a tie is greater than zero, $\tau$ is less than 0.5. To evaluate this null hypothesis, we employ a statistical test based on the ratio $\hat{WR}$, which has an expected value of $WR = \frac{\tau_1}{\tau_2}$, with $\tau_2 \neq 0$. $\hat{WR}$ is commonly referred to as the win ratio and is utilized to test the hypothesis of equality between the two groups, examining whether they have similar outcomes or if one group tends to win more frequently than the other.

In statistical analysis, we assume two sets of random samples, $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$. These samples are drawn from probability distributions $F$ and $G$, respectively. It is mentioned that they both have at least one element, i.e. $m, n \geq 1$, and it holds that $\forall i = 1, \ldots, m$, $X_i = (X_{i1}, \ldots, X_{ik})$ and $\forall j = 1, \ldots, n$, $Y_j = (Y_{j1}, \ldots, Y_{jk})$. The total sample size is denoted as $N$, where $N = n + m$. We are interested in estimating certain parameters, specifically $\tau_1$ and $\tau_2$. To implement this, we utilize a statistical method that involves $U$-statistic.

$$U_v = \frac{1}{n \cdot m} \sum_{i=1}^{m} \sum_{j=1}^{n} \phi_v(X_i, Y_j), \quad v = 1,2$$

It's requisite to examine the joint distribution of $U_1$ and $U_2$ since we desire to gain insights into their behavior. According to Lehmann's theorem for multivariate multi-sample $U$-statistics, the joint distribution of $U_1$ and $U_2$ approaches a normal distribution.

$$\sqrt{N} \begin{pmatrix} U_1 - \tau_1 \\ U_2 - \tau_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{0}$ is the $2 \times 1$ zero matrix and $\mathbf{\Sigma}$ is a variance-covariance matrix with elements,

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

These elements are computed as follows,

$$\sigma_{uv} = \frac{N}{m} \xi_{10}^{uv} + \frac{N}{n} \xi_{01}^{uv}, \quad u = 1,2, \quad v = 1,2$$

https://doi.org/10.26219/heal.aueb.6426



The values $\xi_{10}^{uv}$ and $\xi_{01}^{uv}$ represent the covariances between certain functions, $\varphi_u$ and $\varphi_v$, applied to the data points. Specifically,

$$\begin{aligned}\xi_{10}^{uv} &= \mathrm{Cov}(\phi_u(X_1, Y_1), \phi_v(X_1, Y_1')) \\ \xi_{01}^{uv} &= \mathrm{Cov}(\phi_u(X_1, Y_1), \phi_v(X_1', Y_1)), \quad u = 1, 2, \quad v = 1, 2\end{aligned}$$

It's important to note that $X_1$ and $X_1'$ represent values from two different subjects in the $X$ sample, and they follow the same distribution $F$. Likewise, $Y_1$ and $Y_1'$ are values from two different subjects in the $Y$ sample and follow the same distribution $G$. Importantly, all these variables are assumed to be independent. The obtained expressions can be condensed as outlined below,

$$\begin{aligned}\xi_{10}^{11} &= P(X_1 > Y_1 \ \& \ X_1 > Y_1') - [P(X_1 > Y_1)]^2 \\ \xi_{10}^{22} &= P(X_1 < Y_1 \ \& \ X_1 < Y_1') - [P(X_1 < Y_1)]^2 \\ \xi_{01}^{12} &= P(X_1 > Y_1 \ \& \ X_1 < Y_1') - P(X_1 > Y_1) \cdot P(X_1 < Y_1)\end{aligned}$$

With this provided information, it can be entailed that the logarithm of the ratio between the two statistics, $U_1$ and $U_2$, conforms to a normal distribution.

$$\log\left(\frac{U_1}{U_2}\right) \sim \mathcal{N}\left(\log(WR), \tau_1^{-2} \cdot \sigma_{11} + \tau_2^{-2} \cdot \sigma_{22} - 2 \cdot (\tau_1\tau_2)^{-1} \cdot \sigma_{12}\right)$$

Hence, statistical assessments and confidence intervals for the win ratio parameter $WR$ can be derived by employing the delta method on the logarithmic scale.

### 2.3.4 Utilizing Fieller's theorem

Expanding on the previous approach, an alternative method entails applying Fieller's theorem, which is founded on the subsequent distributional observation,

$$\sqrt{N} \cdot (U_1 - WR \cdot U_2) \sim \mathcal{N}(0, \sigma_{11} + WR^2 \cdot \sigma_{22} - 2 \cdot WR \cdot \sigma_{12})$$

Fieller's confidence interval for $WR$ is established by solving the following inequality,

$$\sqrt{N} \cdot \left| \frac{U_1 - WR \cdot U_2}{\sqrt{\sigma_{11} + WR^2 \cdot \sigma_{22} - 2 \cdot WR \cdot \sigma_{12}}} \right| \leq z_{\alpha/2}$$

The nature of this interval, whether it is finite, the complement of a finite interval, or even extending from negative infinity to positive infinity, relies on the solutions to the quadratic equation,

$$A \cdot WR^2 - 2 \cdot B \cdot WR + C = 0$$

It is noted that, $A = U_2^2 - z_{\alpha/2}^2 \cdot \frac{\sigma_{22}}{N}$, $B = U_1 \cdot U_2 - z_{\alpha/2}^2 \cdot \frac{\sigma_{12}}{N}$, and $C = U_1^2 - z_{\alpha/2}^2 \cdot \frac{\sigma_{11}}{N}$ (Bebu and Lachin (2015)).

https://doi.org/10.26219/heal.aueb.6426



### 2.3.5 Utilizing the asymptotic normality

Luo et al. (2015) invented a novel statistical method to estimate the variance of the win ratio within the context of unmatched pair analysis. Their approach was specifically applied to situations involving a two-component composite endpoint and also comprised a unique algorithm for discerning the outcomes as winners, losers, or ties. This method facilitated the construction of a 95% CI for the win ratio.

They began their study by introducing three distinct random variables: $T_{\text{N-F}}$, $T_F$, and $T_C$, which respectively denoted the time until a non-fatal event, the time until a fatal event, and the time until censoring. The variables $T_{N-F}$ and $T_F$ typically exhibit a positive correlation. Moreover, it's important to note that $T_F$ has the capability to act as a right-censoring factor for $T_{N-F}$, meaning that $T_{N-F}$ data may have incomplete information due to censoring by $T_F$. Conversely, the reverse relationship, where $T_{N-F}$ right-censors $T_F$, does not apply. In the research, patients were categorized into two groups. One group had received the new treatment, denoted as $Z = 1$, while the other group had assumed the standard treatment, signified by $Z = 0$. It is mentioned that $T_C$ is considered to be independent of both $T_{N-F}$ and $T_F$, given the condition $Z$. The observed data can be distinguished into four (4) categories, determined by the ranking of $T_{\text{N-F}}$, $T_F$, and $T_C$. These categories are represented by the variable $\delta$, which can assume values ranging from 1 to 4.

i) In cases where both $T_{\text{N-F}}$ and $T_F$ are observed, the conditions $T_{N-F} < T_F \leq T_C$ hold, and this corresponds to $\delta = 1$.

ii) When only $T_F$ is recorded, two conditions are met, $T_F \leq T_{N-F}$ and $T_F \leq T_C$, leading to $\delta = 2$.

iii) When neither $T_{\text{N-F}}$ nor $T_F$ is visible, implying that $T_C < T_{N-F}$ and $T_C < T_F$, then $\delta = 3$.

iv) If only $T_{\text{N-F}}$ is apparent and $T_{N-F} \leq T_C < T_F$ then $\delta = 4$.

Concurrently, two new variables are imported, $Y_1$ and $Y_2$. The $Y_1$ is defined as the minimum among $T_{\text{N-F}}$, $T_F$, and $T_C$, while $Y_2$ represents the minimum of $T_F$ and $T_C$. It's important to emphasize that the observed data, denoted as $(Y_{1i}, Y_{2i}, \delta_i, Z_i)$ for $i = 1, 2, \ldots, n$, represent independent and identically distributed samples of the variables $(Y_1, Y_2, \delta, Z)$. We also determine three hazard functions for different scenarios in two groups, denoted by $k = 0$ and $k = 1$. Specifically, $\lambda_{Fk}(y_2) = P(T_F = y_2 \,|\, T_F \geq y_2, Z = k)$ characterizes the hazard function of $T_F$ in group $k$, $\lambda_{N-Fk}(y_1 \,|\, y_2) = P(T_{\text{N-F}} = y_1 \,|\, T_{\text{N-F}} \geq y_1, T_F \geq y_2, Z = k)$ represents the conditional hazard of $T_{\text{N-F}}$ in group $k$, and $\lambda_{Ck}(y_2) = P(T_C = y_2 \,|\, T_C \geq y_2, Z = k)$ constitutes the hazard function of $T_C$ in group $k$. Additionally, $\lambda_{CT}(y_2)$ is the cumulative hazard function of censoring. It's calculated as the sum of $\lambda_{C0}(y_2)$ and $\lambda_{C1}(y_2)$, representing the overall hazard of censoring.

Additionally, we consider two distinct patients, denoted as $i$ and $j$. Patient $i$ belongs to the conservative treatment group, while patient $j$ is part of the new

https://doi.org/10.26219/heal.aueb.6426



treatment group. We define a set of equations using several functions. For any $y_1 \leq y_2$ and $k = 0, 1$, we have that,

$$R_{1k}(y_2) = n^{-1} \sum_{j=1}^{n} I(Y_{2j} \geq y_2, Z_j = k)$$

$$R_{2k}(y_1, y_2) = n^{-1} \sum_{j=1}^{n} I(Y_{1j} \geq y_1, Y_{2j} \geq y_2, Z_j = k)$$

$$R_{3k}(y_2) = n^{-1} \sum_{j=1}^{n} I(\delta_j = 1, 2, Y_{2j} \leq y_2, Z_j = k)$$

$$R_{4k}(y_1, y_2) = n^{-1} \sum_{j=1}^{n} I(\delta_j = 4, Y_{1j} \leq y_1, Y_{2j} \leq y_2, Z_j = k)$$

$$R_{5k}(y_1, y_2) = n^{-1} \sum_{j=1}^{n} I(\delta_j = 3, 4, Y_{1j} \geq y_1, Y_{2j} < y_2, Z_j = k)$$

$$R_{6k}(y_1, y_2) = n^{-1} \sum_{j=1}^{n} I(\delta_j = 1, 4, Y_{1j} \leq y_1, Y_{2j} > y_2, Z_j = k)$$

where $I$ constitutes the indicator function. We also specify the expected values of these $R$ functions as $r_{sk} = E(R_{sk})$ for $s = 1, \ldots, 6$. Additionally, we introduce variables $a_i$, $b_i$, $c_i$, and $d_i$, each calculated using a combination of the $R$ functions and indicator functions based on specific conditions related to $\delta_i$ and $Z_i$. In particular, we have that,

$$a_i = I(\delta_i = 1, 2, Z_i = 1) r_{10}(Y_{2i}) + I(Z_i = 0) r_{31}(Y_{2i})$$

$$b_i = I(\delta_i = 1, 2, Z_i = 0) r_{11}(Y_{2i}) + I(Z_i = 1) r_{30}(Y_{2i})$$

$$\begin{aligned} c_i = {} & I(\delta_i = 4, Z_i = 1) r_{20}(Y_{1i}, Y_{2i}) + I(Z_i = 0) r_{41}(Y_{1i}, Y_{2i}) \\ & + I(\delta_i = 1, 4, Z_i = 1) r_{50}(Y_{1i}, Y_{2i}) + I(\delta_i = 3, 4, Z_i = 0) r_{61}(Y_{1i}, Y_{2i}) \end{aligned}$$

$$\begin{aligned} d_i = {} & I(\delta_i = 4, Z_i = 0) r_{21}(Y_{1i}, Y_{2i}) + I(Z_i = 1) r_{40}(Y_{1i}, Y_{2i}) \\ & + I(\delta_i = 1, 4, Z_i = 0) r_{51}(Y_{1i}, Y_{2i}) + I(\delta_i = 3, 4, Z_i = 1) r_{60}(Y_{1i}, Y_{2i}) \end{aligned}$$

Then, we have $A_i$, $B_i$, $C_i$, and $D_i$ as their respective estimates. The primary difference is that in these estimates, we swap "$r$" with "$R$" in the previous expressions.

By implementing the exponential inequality for $U$-statistics (as presented by Giné et al. (2000), and Houdré and Reynaud-Bouret (2003)), we can establish that $n^{-1}\{N_f + E(N_f)\} - \sum_{i=1}^{n}(f_i) = o(n^{1/2})$ almost certainly, where f can be any of a, b, c, d. With these approximations in place, we calculate the estimated variance $\sigma_D^2$ of $n^{-3/2}\{WD - E(WD)\}$, where $WD$ represents the win difference between $N_W$ (number of winners) and $N_L$ (number of losers). The estimation is achieved as $\hat{\sigma}_D^2 = n^{-1} \sum_{i=1}^{n}(\varphi_i - \bar{\varphi})^2$, where, for each i ranging from 1 to n, $\varphi_i = (B_i + D_i) - (A_i + C_i)$ and

https://doi.org/10.26219/heal.aueb.6426



$\bar{\varphi}$ denotes the average calculated as $\bar{\varphi} = n^{-1} \sum_{i=1}^{n} \varphi_i$. Moreover, we can demonstrate that $WR$ (the ratio of winners to losers) converges to $wR$, which is the expected value of $N_W$ divided by the expected value of $N_L$. This signifies the true win ratio and can be computed as follows,

$$\frac{\int r_{11}(y_2)\, r_{10}(y_2)\, \lambda_{F0}(y_2)\, dy_2 + \iint_{y_1 \le y_2} r_{21}(y_1, y_2)\, r_{20}(y_1, y_2)\, \lambda_{N-F0}(y_1|y_2)\, \lambda_{CT}(y_2)\, dy_1\, dy_2}{\int r_{11}(y_2)\, r_{10}(y_2)\, \lambda_{F1}(y_2)\, dy_2 + \iint_{y_1 \le y_2} r_{21}(y_1, y_2)\, r_{20}(y_1, y_2)\, \lambda_{N-F1}(y_1|y_2)\, \lambda_{CT}(y_2)\, dy_1\, dy_2}$$

Additionally, $n^{1/2}(WR - wR)$ converges in distribution to a normal distribution with a mean of zero and a variance $\sigma_R^2$. This variance can be consistently estimated as,

$$\hat{\sigma}_R^2 = n^{-1} \sum_{i=1}^{n} u_i^2$$

where $u_i$ is calculated as,

$$u_i = \frac{(B_i + D_i) - WR(A_i + C_i)}{\frac{N_L}{n}}$$

By applying a logarithmic transformation, $n^{1/2}(\log WR - \log wR)$ converges in distribution to a normal distribution with a mean of zero and a variance of $\frac{\sigma_R^2}{wR^2}$. This variance can be reliably assessed as $\frac{\hat{\sigma}_R^2}{WR^2}$. Consequently, we can construct an approximate $(1 - \alpha)$ confidence interval for $wR$ using the following formula,

$$WR \cdot e^{\pm \frac{n^{-1/2} \cdot z_\alpha \cdot \hat{\sigma}_R}{WR}}$$

It is mentioned that $z_\alpha$ corresponds to the upper percentile of the standard normal distribution, precisely the upper $100\alpha/2$ percentile. This confidence interval allows us to evaluate the potential range of values for $wR$ while maintaining a specified level of confidence $(1 - \alpha)$.

### 2.3.6 Utilizing the asymptotic distribution of winners

Dong et al. (2016) extended the previously mentioned approach. Their method depends on the asymptotic logarithmic distribution of the win ratio and offers a more generalized technique that doesn't rely on algorithms to determine winners, losers, or ties. Particularly, their process commenced by establishing the distribution of the number of winners, followed by investigating the joint distribution of winners in both the standard treatment and new treatment groups. Ultimately, this led to the distribution of the logarithm of the win ratio and enabled the construction of a 95% CI.

Their assumptions center around two groups of participants. Let's consider $X_i$, where $i$ ranges from 1 to $N_N$, with $N_N$ denoting the total number of individuals

https://doi.org/10.26219/heal.aueb.6426



receiving the new treatment. $X_i$ incorporates the information (e.g. non-fatal events, death, etc.) concerning the $i^{\text{th}}$ participant in the new treatment cohort. On the other hand, $Y_j$, $j = 1, \ldots, N_S$, where $N_S$ signifies the total participants in the standard treatment group, captures details for the $j^{\text{th}}$ participant in the conservative group. To further their methodology, they introduce a kernel function defined as $K_{ij}$, which is based on the interaction between these two groups. Specifically, $K_{ij}$ is assumed as,

$$K_{ij} = K(X_i, Y_j) = \begin{cases} 1 & \text{, if condition A holds} \\ 0 & \text{, if condition B holds} \end{cases}$$

where:

- **Condition A:** The $i^{\text{th}}$ participant from the new treatment group wins against the $j^{\text{th}}$ participant from the conservative group.
- **Condition B:** The $i^{\text{th}}$ participant from the new treatment either loses to or ties with the $j^{\text{th}}$ participant from the standard group.

Hence, the count of winners of the new treatment group, denoted as $N_{W_N}$, can be determined utilizing the equation,

$$N_{W_N} = \sum_{i=1}^{N_N} \sum_{j=1}^{N_S} K(X_i, Y_j)$$

Essentially, this summation involves a $U$-statistic with a kernel function. To estimate $N_{W_N}$, we employ the Wei and Johnson (1985) approximation. This approximation is applied to the specific $U$-statistic mentioned but with a slight modification. We reparameterize the multiplication factor $\frac{N_N \cdot N_S}{\sqrt{N_N + N_S}}$ to the $U$-statistic as initially proposed by Wei & Johnson. As an outcome, the distribution of $N_{W_N}$ closely aligns with the asymptotic normal distribution.

$$N_{W_N} \dot{\sim} \mathcal{N}(\theta_{W_N}, \sigma^2_{W_N})$$

where,

$$\begin{aligned}
\theta_{W_N} &= N_N \cdot N_S \cdot \theta_K \\
\theta_K &= E[K(X_i, Y_j)] \\
\sigma^2_{W_N} &= \frac{1}{N_N}\sigma_1^2 + \frac{1}{N_S}\sigma_2^2 \\
\sigma_1^2 &= \frac{N_N \cdot N_S}{N_S - 1} \sum_{i=1}^{N_N} \sum_{j=1}^{N_S} \sum_{j'=1, j' \neq j}^{N_S} [K(X_i, Y_j) - \theta_K][K(X_i, Y_{j'} - \theta_K] \\
\sigma_2^2 &= \frac{N_N \cdot N_S}{N_N - 1} \sum_{j=1}^{N_S} \sum_{i=1}^{N_N} \sum_{i'=1, i' \neq i}^{N_N} [K(X_i, Y_j) - \theta_K][K(X_{i'}, Y_j) - \theta_K]
\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



When we operate under the null hypothesis, asserting that the win ratio is one, it suggests both the new treatment and standard groups demonstrate equivalent therapeutic effects. Under this assumption, it holds that $\theta_K = \theta_{K0} = \mathbb{E}_{H_0}[K(X_i, Y_j)]$. The estimate of $\theta_{K0}$ can be derived from the equation $\hat{\theta}_{K0} = \hat{\mathbb{E}}_{H_0}[K(X_i, Y_j)] = \frac{N_{W_N} + N_{W_S}}{2 \cdot N_N \cdot N_S}$, with $N_{W_S}$ representing the count of individuals who succeeded with the standard treatment (winners of the conservative treatment), and noting that this figure is essentially equivalent to the number of people who did not benefit from the new treatment (losers of the new treatment). Consequently, when considering the null hypothesis $H_0$, it can be obtained that,

$$N_{W_N} \dot{\sim} \mathcal{N}(\theta_{W_N}, \hat{\sigma}^2_{W_N 0})$$

where,

$$\begin{aligned}
\hat{\sigma}^2_{W_N 0} &= \frac{1}{N_N}\hat{\sigma}^2_{10} + \frac{1}{N_S}\hat{\sigma}^2_{20} \\
\hat{\sigma}^2_{10} &= \frac{N_N \cdot N_S}{N_S - 1} \sum_{i=1}^{N_N} \sum_{j=1}^{N_S} \sum_{j'=1, j' \neq j}^{N_S} [K(X_i, Y_j) - \hat{\theta}_{K0}][K(X_i, Y_{j'}) - \hat{\theta}_{K0}] \\
\hat{\sigma}^2_{20} &= \frac{N_N \cdot N_S}{N_N - 1} \sum_{j=1}^{N_S} \sum_{i=1}^{N_N} \sum_{i'=1, i' \neq i}^{N_N} [K(X_i, Y_j) - \hat{\theta}_{K0}][K(X_{i'}, Y_j) - \hat{\theta}_{K0}]
\end{aligned}$$

Likewise, we have the option to establish another kernel function, denoted as $L_{ij}$, to evaluate patient $j$ from the conservative treatment group in relation to patient $i$ receiving the new treatment.

$$L_{ij} = L(X_i, Y_j) = \begin{cases} 1 & \text{, if condition C holds} \\ 0 & \text{, if condition D holds} \end{cases}$$

where:

- **Condition C:** The $j^{\text{th}}$ individual from the standard treatment group wins against the $i^{\text{th}}$ individual from the new treatment group.
- **Condition D:** The $j^{\text{th}}$ individual from the conservative treatment either loses to or ties with the $i^{\text{th}}$ individual from the new treatment group.

Next, we can determine the count of winners of the standard treatment group, referred to as $N_{W_S}$, along with its corresponding asymptotic distribution under the null hypothesis $H_0$. We can compute $\hat{\sigma}^2_{W_S 0}$ in a similar manner to $\hat{\sigma}^2_{W_N 0}$, and we can define $\theta_{W_S}$ in a manner analogous to how we specified $\theta_{W_N}$.

$$N_{W_S} = \sum_{i=1}^{N_N} \sum_{j=1}^{N_S} L(X_i, Y_j)$$

$$N_{W_S} \dot{\sim} \mathcal{N}(\theta_{W_S}, \hat{\sigma}^2_{W_S 0})$$

https://doi.org/10.26219/heal.aueb.6426



In addition, the covariance between $N_{W_N}$ and $N_{W_S}$ can be derived using the following relationships,

$$\hat{\sigma}_{W_N W_S 0} = \frac{1}{N_N}\hat{\sigma}_{W_N W_S 10} + \frac{1}{N_S}\hat{\sigma}_{W_N W_S 20}$$

$$\hat{\sigma}_{W_N W_S 10} = \frac{N_N N_S}{N_S - 1}\sum_{i=1}^{N_N}\sum_{j=1}^{N_S}\sum_{j'=1, j'\neq j}^{N_S}\left[K(X_i, Y_j) - \hat{\theta}_{K0}\right]\left[L(X_i, Y_{j'}) - \hat{\theta}_{L0}\right]$$

$$\hat{\sigma}_{W_N W_S 20} = \frac{N_N N_S}{N_N - 1}\sum_{j=1}^{N_S}\sum_{i=1}^{N_N}\sum_{i'=1, i'\neq i}^{N_N}\left[K(X_i, Y_j) - \hat{\theta}_{K0}\right]\left[L(X_{i'}, Y_j) - \hat{\theta}_{L0}\right]$$

$$\hat{\theta}_{L0} = \hat{\theta}_{K0} = \frac{N_{W_N} + N_{W_S}}{2N_N N_S}$$

Consequently, when considering the null hypothesis $H_0$, the asymptotic joint distribution of $N_{W_N}$ and $N_{W_S}$ is defined as,

$$\begin{pmatrix} N_{W_N} \\ N_{W_S} \end{pmatrix} \dot{\sim} \mathcal{N}\left(\begin{bmatrix} \theta_{W_N} \\ \theta_{W_S} \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2_{W_N 0} & \hat{\sigma}_{W_N W_S 0} \\ \hat{\sigma}_{W_N W_S 0} & \hat{\sigma}^2_{W_S 0} \end{bmatrix}\right)$$

Furthermore, by applying the delta method, it is demonstrated that both the logarithm of the number of winners in the new treatment and the logarithm of the number of winners in the standard treatment exhibit an asymptotic normal distribution.

$$\log(N_{W_N}) \dot{\sim} \mathcal{N}\left(\log(\theta_{W_N}), \frac{\hat{\sigma}^2_{W_N 0}}{\hat{\theta}^2_{W_N 0}}\right)$$

$$\log(N_{W_S}) \dot{\sim} \mathcal{N}\left(\log(\theta_{W_S}), \frac{\hat{\sigma}^2_{W_S 0}}{\hat{\theta}^2_{W_S 0}}\right)$$

where,

$$\left.\begin{aligned} \hat{\theta}_{W_N 0} = N_N\, N_S\, \hat{\theta}_{K0} = \frac{N_{W_N} + N_{W_S}}{2} \\ \hat{\theta}_{W_S 0} = N_N\, N_S\, \hat{\theta}_{L0} = \frac{N_{W_N} + N_{W_S}}{2} \end{aligned}\right\} \quad \hat{\theta}_{W_N 0} = \hat{\theta}_{W_S 0} \tag{2.1}$$

The asymptotic joint distribution of $log(N_{W_N})$ and $log(N_{W_S})$ can be determined as,

$$\begin{pmatrix} \log(N_{W_N}) \\ \log(N_{W_S}) \end{pmatrix} \dot{\sim} \mathcal{N}\left(\begin{bmatrix} \log(\theta_{W_N}) \\ \log(\theta_{W_S}) \end{bmatrix}, \begin{bmatrix} \frac{\hat{\sigma}^2_{W_N 0}}{\hat{\theta}^2_{W_N 0}} & \frac{\hat{\sigma}_{W_N W_S 0}}{\hat{\theta}_{W_N 0}\hat{\theta}_{W_S 0}} \\ \frac{\hat{\sigma}_{W_N W_S 0}}{\hat{\theta}_{W_N 0}\hat{\theta}_{W_S 0}} & \frac{\hat{\sigma}^2_{W_S 0}}{\hat{\theta}^2_{W_S 0}} \end{bmatrix}\right)$$

Hence, we can deduce the asymptotic distribution of the logarithm of the win ratio, which can also be expressed as $WR = \frac{N_{W_N}}{N_{W_S}}$, as the count of losers of the new treatment is equivalent to the count of winners of the standard treatment.

$$\log(WR) \dot{\sim} \mathcal{N}\left(\log\left(\frac{\theta_{W_N 0}}{\theta_{W_S 0}}\right), \frac{\hat{\sigma}^2_{W_N 0}}{\hat{\theta}^2_{W_N 0}} + \frac{\hat{\sigma}^2_{W_S 0}}{\hat{\theta}^2_{W_S 0}} - 2\frac{\hat{\sigma}_{W_N W_S 0}}{\hat{\theta}_{W_N 0}\hat{\theta}_{W_S 0}}\right)$$

https://doi.org/10.26219/heal.aueb.6426



Utilizing equation (2.1) and considering the null hypothesis $H_0$, it can be inferred that,

$$\log(WR) \dot{\sim} \mathcal{N}(0, \hat{\sigma}^2_{\log(WR)})$$

where,

$$\hat{\sigma}^2_{\log(WR)} = \frac{\hat{\sigma}^2_{W_N 0} + \hat{\sigma}^2_{W_S 0} - 2\hat{\sigma}_{W_N W_S 0}}{\hat{\theta}_0^2}, \quad \hat{\theta}_0 = \hat{\theta}_{W_N 0} = \hat{\theta}_{W_S 0}$$

Consequently, the confidence interval for the win ratio at a confidence level of $(1-\alpha)100\%$ can be computed using the following formula, with $Z_{1-\frac{\alpha}{2}}$ representing the $\left(1-\frac{\alpha}{2}\right)$ quantile of the standard normal distribution.

$$\left(e^{\log(WR) - Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\log(WR)}},\ e^{\log(WR) + Z_{1-\frac{\alpha}{2}}\hat{\sigma}_{\log(WR)}}\right)$$

### 2.3.7 Utilizing the variance of logarithmic transformation

In prior literature, Yu and Ganju (2022) presented an outstanding perspective on $\log(WR)$ variance, which forms the basis for constructing a 95% CI for the win ratio. Their argument depends on the idea that the variance of the logarithm of the win ratio, $\log(WR)$, is contingent upon several factors, including the total sample size $(N)$, the proportion $(k)$ of patients allocated to one group (with the complementary ratio for the other group being $(1-k)$), and the proportion of tied observations, represented by $P_{\text{tied}}$. The proportion of tied observation is defined as $P_{\text{tied}} = \frac{N_T}{k(1-k)N^2}$, with $N_T$ standing for the total number of tied instances.

In particular, utilizing the delta method, we can approximate the logarithmic transformations of the numbers of winners and losers. The logarithm of the number of winners, $\log(N_W)$, can be approximated as $\log(N_W) \approx \log(\mu_W) + \frac{N_W - \mu_W}{\mu_W}$, where $\mu_W$ represents the expected value of winners. Similarly, the logarithm of the number of losers, $\log(N_L)$, can be estimated as $\log(N_L) \approx \log(\mu_L) + \frac{N_L - \mu_L}{\mu_L}$, where $\mu_L$ signifies the expected value of losers. From the above, we can deduce the logarithm of the win ratio (WR) by taking the difference between the logarithms of the numbers of winners and losers.

$$\log(WR) = \log(N_W) - \log(N_L)$$

Substituting our approximations, we obtain that,

$$\log(WR) \approx \log\left(\frac{\mu_W}{\mu_L}\right) + \frac{N_W}{\mu_W} - \frac{N_L}{\mu_L}$$

Under the null hypothesis, when the expected values of winners, losers, and the total are all equal, which is expressed as $\mu_W = \mu_L = \mu = \frac{1}{2} \cdot k(1-k)N^2(1-P_{\text{tied}})$, we can approximate the logarithm of the win ratio as,

$$\log(WR) \approx \frac{N_W - N_L}{\mu}$$

Additionally, the variance of the logarithm of WR can be approximated as,

https://doi.org/10.26219/heal.aueb.6426



$$\text{Var}(\log(WR)) \approx \tfrac{1}{\mu^2}\text{Var}(N_W - N_L)$$

In the Finkelstein and Schoenfeld (1999) method, the difference between the numbers of winners and losers, $N_W - N_L$, can be expressed as $N_W - N_L = \sum_{i\in A} U_i$. The permutation variance of this difference can be approximated as,

$$\text{Permutation Variance} \approx k(1-k)\sum_i U_i^2$$

Thus, the variance of the logarithm of the win ratio, $Var(\log(WR))$, can be approximated as,

$$Var(\log(WR)) \approx \frac{k(1-k)}{\mu^2}\sum_i U_i^2$$

It is mentioned that $A$ and $B$ represent the two treatment groups. When ties and intransitivity are present in a ranking system, the value of $\sum_i U_i^2$ becomes dependent on the specific conformation of ties and intransitive relationships within the ranking. In the absence of ties and in-transitivity, each patient is assigned a unique rank in a sequence that ranges from $-(N-1)$ to $(N-1)$. Therefore, the summation of $U_i^2$ for all individuals can be computed as follows,

$$\begin{aligned}\sum_i U_i^2 &= (-(N-1))^2 + (-(N-3))^2 + \ldots + (N-3)^2 + (N-1)^2 \\ &= \sum_{j=1}^{N}(N+1-2j)^2 \\ &= \frac{N\cdot(N^2-1)}{3}\end{aligned}$$

Assuming there is no in-transitivity, but ties are allowed, we suppose initially dividing $N$ patients into $l = N/m$ groups, each containing $m$ individuals. This division assumes that all patients within a group have the same rank, resulting in tied ranks for each group. The proportion of Finkelstein-Schoenfeld ties, denoted as $P_{\text{FS,tied}}$, is approximated as $\frac{lm(m-1)}{N(N-1)}$, which can be further simplified to $\frac{m-1}{N-1}$. Since $P_{\text{FS,tied}}$ is approximately equal to $P_{\text{tied}}$, we can equate them, leading to the relationship $\frac{m-1}{N-1} \approx P_{\text{tied}}$, or $m \approx N_{P_{\text{tied}}}$. With this approximation, the sum of the squared $U_i$ values for all $i$ can be expressed as,

$$\begin{aligned}\sum_i U_i^2 &= m\left[(-(l-1)m)^2 + (-(l-3)m)^2 + \ldots + ((l-3)m)^2 + ((l-1)m)^2\right] \\ &= m^3\sum_{j=1}^{K}(l+1-2j)^2 \\ &= \frac{m^3 l(l^2-1)}{3} \\ &= \frac{N(N^2-m^2)}{3} \\ &\approx \frac{N^3(1-P_{\text{tied}}^2)}{3}\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



Consequently, the approximate variance of $\log(WR)$ is given by,

$$\begin{aligned}\mathrm{Var}(\log(WR)) &\approx \frac{k(1-k)}{\mu^2}\sum_i U_i^2 \\ &\approx \frac{k(1-k)}{\mu^2}\frac{N^3(1-P_{\text{tied}}^2)}{3} \\ &\approx \frac{4(1+P_{\text{tied}})}{3k(1-k)(1-P_{\text{tied}})N}.\end{aligned} \tag{2.2}$$

This expression underscores the dependency of $\log(WR)$ variance on the interplay of sample size, group allocation proportion, and tied observations, offering an alternative viewpoint on the statistical framework. Given the insights presented, we can now determine the 95% CI for the win ratio, which can be expressed as,

$$\exp(\log(\mathrm{WR}) \pm 1.96\sqrt{\mathrm{Var}})$$

In simpler terms, to evaluate the confidence interval at 95% for the win ratio, one needs to exponentiate the natural logarithm of the win ratio adjusted by $\pm 1.96$ times the square root of its variance. This interval allows us to estimate the range within which the win ratio is likely to fall with a 95% level of confidence.

## 2.4 Seeking the power within the win ratio

A subject of considerable significance for researchers specializing in statistics and analysis revolves around the efficacy of testing a statistical hypothesis. This crucial issue delves deeply into the ability of statistical tests to accurately detect true effects or differences in a given dataset, thereby influencing the reliability and robustness of scientific conclusions. In the context of employing the win ratio method, an ordinary investigation comprises testing the null hypothesis, verifying the absence of a statistically significant difference between the two groups under comparison. These groups typically consist of the control or standard treatment group and the new treatment group. This meticulous statistical testing is primary for researchers seeking to draw meaningful conclusions about the efficacy or impact of a new treatment in comparison to the established standard, ensuring a comprehensive understanding of the potential implications and significance of the study's findings. Researchers often conjure unique methods for assessing the robustness of a statistical test based on their studies and findings.

Ramchandani et al. (2016) conducted an independent study where they introduced an innovative non-parametric global rank test oriented for scenarios involving multiple outcomes and potential censored data, as detailed in section 3.1. This singular approach served as the foundation for their methodology in assessing the power of hypothesis testing. The researchers centered on a specific function, denoted as $\phi$, which briefly recorded the scores of the outcomes. They selected error probabilities I and II, represented by $\alpha$ and $\beta$, respectively. Additionally, they proceeded under the

https://doi.org/10.26219/heal.aueb.6426



assumption that the global treatment effect $\theta_\phi$, indicating the expected treatment effect, would be positive ($\theta_\phi > 0$) under the alternative hypothesis $H_1$. Thus, they established a method to compute the statistical power of hypothesis testing. The power can be calculated as below,

$$\text{Power} = 1 - \beta \approx 1 - \Phi\left(z_{1-\alpha/2} - \frac{\sqrt{N}\theta_\phi}{\sigma}\right)$$

Here, $\Phi$ indicates the standard normal cumulative distribution function. In addition, $z_{1-\alpha/2}$ declares the critical value in the upper tail for rejecting $H_0$, while $\sigma$ and $N$ correspond to the standard deviation of the $U$-statistic and the required sample size for achieving the specified power, respectively. These parameters and their relationships are further detailed in section 3.1.

The researchers in question expanded their review by investigating the potential influence of different weights assigned to various outcomes. Specifically, they utilized a vector of $U$-statistics denoted as $U = (U_1, \ldots, U_p)'$, corresponding to the outcomes. This vector, alongside its variance-covariance matrix $\Lambda = \text{cov}(U)$, the global treatment effect represented by the vector $\theta_\phi = (\theta_{\phi 1}, \ldots, \theta_{\phi p})'$, where $\theta_\phi$ denotes the expected treatment effect ($\theta_\phi = \mathbb{E}(U)$) under the alternative hypothesis $H_1$, and the introduction of a vector of weights, $w = (w_1, \ldots, w_p)'$, allowed them to compute the power of hypothesis testing. Without loss of generality, the researchers supposed that the expected treatment effect, $\theta_\phi$, would be non-negative ($\theta_\phi \geq 0$). Although they recognized the accessibility of assuming negative values for the expected treatment effect, they specifically adopted this assumption. Their syllogism was based on their desire to center on observing the positive effects of the treatment, even in the case of a single outcome. Hence, they stated that the power calculation formula is given by,

$$\text{Power} = 1 - \beta \approx 1 - \Phi\left(z_{1-\alpha/2} - \frac{\sqrt{N}\, w'\theta_\phi}{\sqrt{w'\Lambda w}}\right)$$

It is noted that in the overall analysis, a preference is placed on achieving high power, indicating the probability of rejecting the null hypothesis when the alternative is true. To augment this probability, one approach is to maximize the expression $w'\theta_\phi \, (w'\Lambda w)^{-1/2}$, concerning the weight vector $w$. In instances where a negative expected treatment effect is present ($\theta_\phi \leq 0$), the maximization of power is accomplished by minimizing the aforementioned quantity. Further details regarding their approach will be examined in section 3.1 (Ramchandani et al. (2016)).

Mao et al. (2021) recommended an alternative aspect of evaluating the effectiveness of hypothesis testing, employing a unique win ratio approach. Their novel methodology comprises defining the power of the hypothesis test through a specific formula,

$$\text{Power} = 1 - \beta = \Phi\left(\frac{W_0\theta\sqrt{q(1-q)n}}{\zeta_0} - z_{1-\frac{\alpha}{2}}\right)$$

https://doi.org/10.26219/heal.aueb.6426



In the particular expression, the descriptor $\beta$ corresponds to the type II error, while $\Phi$ denotes the standard normal cumulative distribution function. In addition, $W_0$ represents both the win and loss fraction, as the total sample size $n$ tends to infinity ($n \to \infty$) in the framework of the null hypothesis $H_0$. According to the null hypothesis, there is an absence of a statistically significant distinction between the two groups under comparison. Furthermore, $\theta$ indicates the logarithm of the win ratio, and $q$, restricted to the interval $(0, 1)$, is a parameter where the ratio of $n_1$ to $n$ converges when the minimum required sample size is defined. Here, $n_1$ constitutes the total number of individuals in the new treatment group. The variable $n$ signifies the requisite sample size to accomplish a specific power and it is expressed as per equation (3.2). The parameter $\zeta_0$, known as the standard rank deviation (SRD), is calculated according to equation (3.1). Additionally, $z_{1-\frac{\alpha}{2}}$ represents the critical value in the upper tail for rejecting the null hypothesis. Detailed insights into the method and its components can be found in section 3.1.

In their survey detailed in subsection 2.3.7, Yu and Ganju (2022) came up with a mathematical expression to measure the effectiveness of the statistical test. They claimed that the power of statistical control can be expressed by the formula,

$$\text{Power} = 1 - \Phi\left(Z_{1-\alpha} - \log(\text{WR}_{\text{true}})\sqrt{\frac{3Nk(1-k)(1-P_{\text{tied}})}{4(1+P_{\text{tied}})}}\right)$$

Here, $\Phi$ represents the cumulative distribution function of the standard normal distribution. The variables include $\alpha$ for the type I error rate, $Z$ for the quantile value from the standard normal distribution, and $WR_{\text{true}}$ for the actual value of the win ratio, with $\log(\text{WR}_{\text{true}})$ signifying its natural logarithmic transformation. Additionally, $k$ denotes the proportion of patients allocated to one group, $(1-k)$ signifies the proportion assigned to the other group, $P_{\text{tied}}$ represents the proportion of tied observations, and $N$ stands for the required sample size to attain a power of $(1-\beta)$. The determination of $N$ is clarified in the concluding paragraph of section 3.1. This formulation encapsulates a comprehensive consideration of diverse parameters, reflecting the researchers' meticulous approach to quantifying the statistical test's efficacy.

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# Chapter 3

# Elaborate Further on the Win Ratio Approach

## 3.1 Determining the required sample size

A crucial concern for researchers when analyzing data to estimate the win ratio is the sample size required to achieve their goal, which involves rejecting the null hypothesis that the win ratio equals 1. This null hypothesis implies that there is no statistically significant difference between the two treatment groups. Initially, sample size calculations involved the use of simulations, a rather complex procedure. This complexity arises from the requirement to create different scenarios with different numbers of multiple endpoints and various abandon patterns. It also involves a hierarchical comparison of patient data based on their minimum follow-up times. Additionally, within each simulation run, the data are resampled using the bootstrap method to estimate variance. Assessing multiple scenarios through simulations is intricate and demands a significant amount of time and effort.

Ramchandani et al. (2016) introduced a formula for determining sample size based on a non-parametric global rank test, specifically designed for scenarios involving multiple outcomes, which may include censored data. It is noteworthy that this approach is consistent with other research methodologies, contributing to the broader body of literature in the field. In their study, patients were categorized into two groups, one receiving conservative treatment and the other undergoing the new treatment. Each patient in one group was systematically compared with every patient in the other group across multiple outcomes. For this comparative analysis, a function denoted as $r_k$ was employed, representing the relative performance of patients concerning the $k^{\text{th}}$ outcome. The function assigned values as follows,

$$r_k = \begin{cases} -1 & \text{, if patient } i \text{ exhibited inferior outcomes compared to patient } j \\ 0 & \text{, if patient } i \text{ and patient } j \text{ achieved identical outcomes} \\ 1 & \text{, if patient } i \text{ demonstrated superior outcomes compared to patient } j \end{cases}$$

This methodology, referred to as pairwise rank, allowed for a comprehensive evaluation of treatment efficacy across various parameters. It is crucial to emphasize

https://doi.org/10.26219/heal.aueb.6426



that due to censoring and mortality, valid comparisons between patients can only be made up to their last common follow-up time. It is assumed that the variables $x_{ik}$ and $y_{jk}$ represent the observed data for individuals $i$ and $j$, where $i$ denotes individuals from the new treatment group ($i = 1, 2, \ldots, n$), $j$ denotes individuals from the conservative treatment group ($j = 1, 2, \ldots, m$), and $k$ represents the potential outcomes ($k = 1, 2, \ldots, p$). The assumption is made that the complete vector of outcomes for random variables $X_i$ and $Y_j$ are independent and identically distributed, controlled by corresponding distribution functions $F_X(x_1, x_2, \ldots, x_p)$ and $F_Y(y_1, y_2, \ldots, y_p)$. Consequently, the ranking function for an outcome is expressed as $r_k(x_{ik}, y_{jk}) = I(x_{ik} > y_{jk}) - I(x_{ik} < y_{jk})$, adopting values as previously described. Moreover, the quantity $E[r_k(x, y)] = \theta_k$ is introduced, representing a marginal treatment effect for outcome $k$, with a positive value favoring the new treatment group. The vector of scores comparing subject $i$ to subject $j$ across each of the $p$ outcomes is defined as $\boldsymbol{r}_{ij} = (r_1(x_{i1}, y_{j1}), r_2(x_{i2}, y_{j2}), \ldots, r_p(x_{ip}, y_{jp}))$. As an illustration, the vector $\boldsymbol{r}_{ij} = (1, -1, 0)$ signifies that individual $i$ performed better than individual $j$ on the first outcome, worse on the second outcome, and achieved the same or had indeterminate results on the third outcome. Given access to the score vectors, we proceeded to construct a novel function, denoted as $\phi(\boldsymbol{r}_{ij}) = \phi(r_1, \ldots, r_p)$, serving as a condensed representation of these scores. This function acts as a comprehensive measure of discrepancies in outcomes between subjects $i$ and $j$, where positive values favor individual $i$, negative values service individual $j$, and a value of 0 signifies no preference for either individual (examples of this function can be found in Appendix A). By aggregating this specific function across the two groups under analysis, a $U$-statistic is obtained and defined as,

$$U = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi(\boldsymbol{r}_{ij})$$

This statistical measure contributes to evaluating the expected treatment effect, denoted as $\theta_\phi = E[\phi(r_1(X_1, Y_1), \ldots, r_p(X_p, Y_p))]$, commonly referred to as the global treatment effect. The estimation of this parameter can be interpreted as a normalized probability of a superior outcome with the treatment. Simultaneously, it holds true that $\theta_\phi = 0$ under the null hypothesis, signifying no distinction between the two analyzed groups. In such instances, the test statistic is characterized by a null mean and serves as a basis for rejecting the null hypothesis at a given significance level. While the application of this approach necessitates strong assumptions, it is noteworthy that it can be viable even under weak conditions. Placing greater emphasis on robust assumptions, it follows that the joint distribution of outcomes is identical in both groups. Consequently, specific conditions arise for the function $\phi$,

i) $\phi(\mathbf{0}) = 0$.

ii) $\phi$ is an odd function, i.e., $\phi(\boldsymbol{r}_{ij}) = -\phi(\boldsymbol{r}_{ji})$. This implies $\phi(\boldsymbol{r}_{ij}) + \phi(\boldsymbol{r}_{ji}) = 0$.

iii) $E[\phi^2(r_1(X_1, Y_1), \ldots, r_p(X_p, Y_p))] < \infty$.

https://doi.org/10.26219/heal.aueb.6426



These conditions reflect foundational aspects of the function $\phi$ contending the statistical methodology. The inclusion of primes ensures that the distinction in scores is solely based on the sign when the arguments of the function $r_k(\cdot,\cdot)$ are inverted. Assuming the total sample size is $N = n + m$, that the third assumption holds, and that $\frac{n}{N} \to \lambda$ as $N \to \infty$, standard asymptotic theory of $U$-statistics implies that,

$$\sqrt{N}U \to N(0, \sigma^2)$$

Notably,

$$\sigma^2 = \frac{1}{\lambda}E[\phi(\boldsymbol{r}_{ij})\phi(\boldsymbol{r}_{ij'})] + \frac{1}{1-\lambda}E[\phi(\boldsymbol{r}_{ij})\phi(\boldsymbol{r}_{i'j})]$$

Given the available data, the asymptotic variance can be calculated using the following formula,

$$\hat{\sigma}^2 = \frac{N}{(nm)^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j'\neq j}^{m}\phi(\boldsymbol{r}_{ij})\phi(\boldsymbol{r}_{ij'}) + \sum_{i=1}^{n}\sum_{i'\neq i}^{n}\sum_{j=1}^{m}\phi(\boldsymbol{r}_{ij})\phi(\boldsymbol{r}_{i'j})\right]$$

Hence, consolidating the aforementioned considerations with the designation of type I and II errors, denoted as $\alpha$ and $\beta$ respectively, yields the subsequent expression for determining the sample size,

$$N = \left[\frac{\sigma(\mathrm{z}_{1-\frac{\alpha}{2}} - \mathrm{z}_{\beta})}{\theta_\phi}\right]^2$$

where $z_{1-\frac{\alpha}{2}}$ and $z_\beta$ represent the critical values corresponding to the upper $\frac{\alpha}{2}$ tail and $\beta$ tail of the standard normal distribution, respectively. Conducting simulations allows us to acquire the essential parameter values, $\theta_\phi$ and $\sigma$, required for determining the sample size. Furthermore, insights drawn from the given formula lead to the following conclusions,

$$n = \lambda N \quad \text{and} \quad m = (1-\lambda)N$$

These expressions elucidate the relationship between the number of participants in each group ($n$ and $m$) and the total sample size ($N$), where $\lambda$ serves as a constant.

The outlined methodology for determining the sample size can be extended to scenarios where distinct weights are assigned to each outcome. Specifically, let $U = (U_1, \ldots, U_p)'$ denote a vector of $U$-statistics, $\Lambda = \mathrm{cov}(U)$ represents the covariance matrix of the $U$-statistics, $\theta_\phi = (\theta_{\phi 1}, \ldots, \theta_{\phi p})' = E(U)$ signify the expected values of the $U$-statistics under the alternative hypothesis $H_1$. Here, $H_1$ assumes a statistically significant difference between the two treatment groups. Additionally, $w = (w_1, \ldots, w_p)'$ is a vector representing weights. Without loss of generality, we emphasize the alternative hypothesis that the group receiving the new treatment yields more favorable outcomes. As a consequence, we assume that $\theta_\phi \geq 0$. In the scenario where the alternative hypothesis assumes favorable results for the standard treatment group, it follows that $\theta_\phi \leq 0$. Given this context, the sample size can be computed as follows,

$$N = w'\Lambda w\left[\frac{z_{1-\alpha/2} - z_\beta}{w'\theta_\phi}\right]^2$$

https://doi.org/10.26219/heal.aueb.6426



This formulation accounts for the covariance matrix $\Lambda$, the weights vector $w$, and critical values $z_{1-\alpha/2}$ and $z_\beta$ associated with the desired significance level ($\alpha$) and power ($1-\beta$), respectively (Ramchandani et al. (2016)).

Mao et al. (2021) introduced a comprehensive method for calculating sample sizes in various win ratio analysis scenarios. This non-parametric formula centers on two key elements: the effect size, which aligns with the win ratio, and a noise parameter called the standard rank deviation (SRD). The SRD is derived from the null variance of the U-statistic. Initially, the researchers characterized the set of possible outcomes as $\mathcal{Y}$, designating $Y_\alpha \in \mathcal{Y}$ as an outcome associated with group $\alpha$. Here, $\alpha = 1$ denotes the new treatment group, while $\alpha = 0$ represents the standard treatment group. Assuming a random sample of size $n_\alpha$ with outcomes $Y_{\alpha 1}, \ldots, Y_{\alpha n_\alpha}$, and a total participant set of size $n = n_1 + n_0$, they examined the comparison between the two groups. To assess the efficacy of the new treatment, they introduced the notion of the fraction of winners, defined as,

$$\hat{W}_{1n} = \frac{1}{n_1 \cdot n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} I(Y_{1i} \succ Y_{0j})$$

and representing instances where the outcome in the new treatment group surpasses that in the standard treatment group. Similarly, the fraction of new treatment losers, $\hat{W}_{0n}$, is defined as,

$$\hat{W}_{0n} = \frac{1}{n_1 \cdot n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} I(Y_{0j} \succ Y_{1i})$$

In cases where a tie occurs in the outcome comparison, it is acknowledged that $\hat{W}_{1n} + \hat{W}_{0n} < 1$. Consequently, they introduced the win ratio,

$$WR = \frac{\hat{W}_{1n}}{\hat{W}_{0n}}$$

This metric serves as a measure of the relative success of the new treatment compared to the standard treatment. In the context of contiguous alternative hypotheses approaching the null hypothesis at a rate of $n^{-1/2}$, we define $W_{1,n}$ and $W_{0,n}$ as the population fractions of wins and losses, respectively. These fractions converge to $W_0$, i.e. $W_{a,n} \to W_0 \quad (a = 1, 0)$, where $W_0$ represents both the win and loss fraction, as $n$ tends to infinity ($n \to \infty$), under the conditions of null hypothesis $H_0$. To ensure the stability of the asymptotic process, we introduce the assumption that the logarithm of the win ratio, multiplied by $\sqrt{n}$, converges to a fixed positive real number $\mathbb{R}^+$. This is expressed as,

https://doi.org/10.26219/heal.aueb.6426



$$\sqrt{n}\log\left(\frac{W_{1,n}}{W_{0,n}}\right) \to h > 0$$

Consequently, the logarithm of the win ratio, denoted as $\theta = \log\left(\frac{W_{1,n}}{W_{0,n}}\right)$, can be approximated by $\frac{1}{\sqrt{n}} \cdot h$ in the case of large sample sizes. It's important to note that the parameter $\theta$ captures a systematic difference between the two groups. However, this difference is relative rather than absolute, emphasizing the need for a measure of power. This consideration is crucial for understanding the impact of the treatment and determining its effectiveness. This induces the introduction of a noise parameter, which is characterized by the determination of $W(y) = P(Y_0 \succ y)$ and $L(y) = P(Y_0 \prec y)$. These reflect the proportions of better and worse outcomes, respectively, within the same population. Consequently, we introduce the population-level generalized rank function $R(y) = W(y) - L(y)$, bounded within the range of $[-1, 1]$. The standard deviation of $R(Y_0)$ is then computed as,

$$\zeta_0^2 = \int R(y)^2 \nu_0(dy) \tag{3.1}$$

where $\nu_0$ represents the distribution of outcomes in the reference group. It's noteworthy that, given the asymptotic equivalence of $\nu_0$ to the null distribution under consecutive alternatives, we refrain from drawing unnecessary distinctions between the "reference" and "null" distributions. The parameter $\zeta_0$ is commonly referred to as the standard rank deviation (SRD).In the context of assessing continuous outcomes using the standard Wilcoxon-Mann-Whitney $U$-test, an interesting observation emerges. Specifically, when examining the rank probability function $R(y)$, defined as $R(y) = P(Y_0 > y) - P(Y_0 < y) = 2 \cdot P(Y_0 > y) - 1$, it is found that $R(Y_0)$ exhibits a uniform distribution within the range of $[-1, 1]$. This distribution holds true irrespective of the parameter $\nu_0$. Mathematically, $R(Y_0)$ follows a uniform distribution with $\alpha = -1$, $\beta = 1$, denoted as $R(Y_0) \sim \text{Unif}[-1, 1]$. The SRD of this distribution, denoted as $\zeta_0^2$, is calculated to be $3^{-1}$. While it may be enticing to exclusively rely on this specific distribution for win ratio analyses, it is crucial to recognize that the value of $\zeta_0^2$ can vary in different scenarios. Consequently, it is imperative not to arbitrarily select the SRD value in order to ensure the accuracy of sample size calculations. Careful consideration of the specific context and underlying assumptions is essential to obtain meaningful and reliable results. Considering the scenario of deriving a non-parametric sample size formula for a general win ratio test, the objective is on ensuring asymptotic power greater than or equal to $1 - \beta$ under the log-win ratio $\theta$, given parameters $\zeta_0$, $W_0$ and asymptotic level $\alpha$. Specifically, when the ratio $\frac{n_1}{n}$ approaches $q \in (0, 1)$, the minimum sample size $n$ required can be expressed as,

$$n = \frac{\zeta_0^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{q(1-q)W_0^2\theta^2} \tag{3.2}$$

To compute $n$, it is necessary to obtain

$$\frac{\zeta_0^2}{q(1-q)W_0^2}$$

https://doi.org/10.26219/heal.aueb.6426



as the asymptotic null variance of the test statistic $S_n$. This involves employing Hoeffding's decomposition theory (Hoeffding (1948)) for $U$-statistics, which allows breaking down the win/loss fractions $\hat{W}_{an} \quad (a = 1, 0)$ into independent and identically distributed (i.i.d.) terms. The subsequent application of the delta method to the log-ratio transformation is then employed to accomplish this task.

In the progression of their investigation, they innovatively proposed an alternative approach for determining the sample size. More precisely, upon establishing or approximating $\nu_0$, the parameters $\zeta_0$ and $W_0$ can be computed using the relationship (3.1), as they characterize the reference distribution $\nu_0$. The expressions for

$$R(y) = \int \{I(y \prec y^*) - I(y \succ y^*)\}\nu_0(dy^*)$$
$$W_0 = \int \int_{y \succ y^*} \nu_0(dy^*)\nu_0(dy)$$

comprise these computations. Depending on the intricacy of each scenario, these calculations can be executed either analytically or numerically. Conversely, the determination of the parameter $\theta$ depends on our available knowledge. Ideally, we aim to minimize the logarithm of the win ratio, seeking statistical significance. However, this effort is often intricate and not always feasible, especially in the realm of time-to-event data, where $\theta$ is associated with both the outcome and the distribution of censoring. Consequently, the methodology employed to calculate $\theta$ involves postulating an outcome distribution, denoted as $\nu_1$ for the novel treatment group, in contrast to $\nu_0$ for the standard treatment group. In this context, the log-win ratio can be defined as,

$$\theta = \log \left\{ \frac{\int_{y_1 \succ y_0} \nu_1(dy_1)\nu_0(dy_0)}{\int_{y_0 \succ y_1} \nu_1(dy_1)\nu_0(dy_0)} \right\}$$

The parameter $\theta$ is typically derived from integrals that are challenging to solve analytically, especially in the context of multiple alternatives. Analytical solutions are often elusive, except in cases involving discrete data where these integrals become more tractable. To facilitate computation, we substitute $\nu_1(\cdot)$ with $\nu(\cdot; \boldsymbol{\xi})$, a parametric model with parameter $\boldsymbol{\xi}$. Employing a first-order Taylor expansion with respect to $\boldsymbol{\xi}$, we obtain the following representations,

$$W(y; \boldsymbol{\xi}) = \int_{y^* \succ y} \nu(dy^*; \boldsymbol{\xi})$$
$$L(y; \boldsymbol{\xi}) = \int_{y^* \prec y} \nu(dy^*; \boldsymbol{\xi})$$
$$R(y; \boldsymbol{\xi}) = W(y; \boldsymbol{\xi}) - L(y; \boldsymbol{\xi})$$

Assuming $\nu_0(\cdot) = \nu(\cdot; 0)$, we have that $W(y; 0) = W(y)$, $L(y; 0) = L(y)$, and $R(y; 0) =$

https://doi.org/10.26219/heal.aueb.6426



$R(y)$. Utilizing the relationship $\log(1+x) = x + \mathcal{O}(x)$, we can express $\theta$ as,

$$\theta = \log\left\{\frac{\int W(y;\boldsymbol{\xi})\nu_0(dy)}{\int L(y;\boldsymbol{\xi})\nu_0(dy)}\right\}$$
$$= \log\left\{1 + \frac{\int R(y;\boldsymbol{\xi})\nu_0(dy)}{\int L(y;\boldsymbol{\xi})\nu_0(dy)}\right\}$$

This can be approximated as,

$$\theta \approx \frac{\int R(y;\boldsymbol{\xi})\nu_0(dy)}{\int L(y;\boldsymbol{\xi})\nu_0(dy)} \tag{3.3}$$

In essence, the parameter $\theta$ is related to the integrals of the functions $W$, $L$, and $R$, which are expressed in terms of the parametric model $\nu(\cdot;\boldsymbol{\xi})$ and the baseline model $\nu_0(\cdot)$. By utilizing the relationships,

$$\int R(y)\nu_0(dy) = 0$$

$$\int L(y)\nu_0(dy) = W_0$$

we establish that the derivative of the right-hand side of equation (3.3) at $\boldsymbol{\xi} = \mathbf{0}$ is $w_0^{-1} \cdot \boldsymbol{\delta}_0$, where

$$\boldsymbol{\delta}_0 = \int \frac{\partial R(y;\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\Big|_{\boldsymbol{\xi}=0} \nu_0(dy)$$

The integral $\delta_0$ receives an analytical solution, yet often requires numerical methods for practical computation. Combining the approximation $\theta \approx \mathbf{w}_0^{-1}\boldsymbol{\delta}_0^T\boldsymbol{\xi}$ with equation (3.2) yields an alternative formula for calculating the sample size, specifically a model-based approach. Let $\nu_1(\cdot) = \nu(\cdot;\boldsymbol{\xi})$ be a smooth parametric model parameterized by $\boldsymbol{\xi}$, with $\nu_0(\cdot) = \nu(\cdot;\mathbf{0})$. Consequently, for an asymptotic level-$\alpha$ win ratio test, the minimum sample size $n$ required to achieve an asymptotic power greater than or equal to $1-\beta$ under the effect size $\boldsymbol{\xi}$ is given by,

$$n = \frac{\zeta_0^2\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{q(1-q)(\delta_0^T\xi)^2}$$

This alternative formulation provides insight into the determination of the sample size, adopting a model-based perspective. To enhance the interpretability, $\nu_1(\cdot)$ serves as a smoothly varying parametric model, introducing flexibility to the analysis (Mao et al. (2021)).

https://doi.org/10.26219/heal.aueb.6426



Yu and Ganju (2022) have introduced a formula for calculating the sample size, which takes into account the probability that a randomly selected patient from one group performs better than a counterpart from the other group, along with the probability of a tie. What's noteworthy about this mathematical formula is that it offers the advantage of increasing statistical power even when incorporating an endpoint with low individual power into the hierarchy. Furthermore, it provides valuable insights that can guide adjustments in an ongoing blinded trial if deemed necessary. This formula streamlines the process of determining trial size with a win ratio endpoint, making it accessible to non-specialists. In particular, when considering the variance of the natural logarithm of the win ratio, as discussed in subsection 2.3.7, and aiming to attain a one-sided type I error rate of $\alpha$ along with a study power of $1-\beta$, the necessary sample size, denoted as $N$, should meet the following equation,

$$\mathrm{Var}(\log(WR))(Z_{1-\alpha}+Z_{1-\beta})^2 = \log^2(WR_{\text{true}})$$

or

$$N = \frac{4(1+P_{\text{tied}})(Z_{1-\alpha}+Z_{1-\beta})^2}{3k(1-k)(1-P_{\text{tied}})\log^2(WR_{\text{true}})}$$

where $WR_{\text{true}}$ signifies the presumed or actual value of the win ratio and $\log(WR_{\text{true}})$ denotes its natural logarithmic transformation. As is customary, $\alpha$ and $\beta$ represent the type I and type II error rates, respectively, with $Z$ indicating the quantile value from the standard normal distribution. Additionally, $P_{\text{tied}}$ stands for the proportion of tied observations, $k$ represents the proportion of patients assigned to one group, and $(1-k)$ corresponds to the proportion allocated to the other group.

## 3.2 What is the appropriate approach for addressing tied outcomes?

In our investigation of the win ratio methodology, we observe that in patient comparisons, scenarios develop where some individuals experience wins, some confront losses, and a part ends up in an indeterminate outcome referred to as a tie. Ties are frequently greater than wins or losses, due to censoring. Therefore, it becomes crucial to invent an approach for effectively handling these ties. Dong et al. (2020) imported a novel concern termed "win odds", offering an expedient to evaluate and understand the impact of ties in the analysis. This formulation employs as an extension of the Mann–Whitney odds within the context of prioritized pairwise comparisons.

Initially, it is commonly acknowledged that in formulating Mann–Whitney odds within the framework of prioritized pairwise comparisons, the employing of the following $U$-statistic is essential,

$$U = \sum_{i=1}^{N_N}\sum_{j=1}^{N_S}\left[I(X_i > Y_j) + \frac{1}{2}I(X_i \approx Y_j)\right]$$

https://doi.org/10.26219/heal.aueb.6426



In this formulation, $N_N$ corresponds to the number of participants in the new treatment group, $N_S$ denotes the number of patients in the conservative treatment group, $X_i$ represents the information regarding the priority outcome for the $i^{\text{th}}$ patient in the new treatment group, $Y_j$ signifies the information about the priority outcome for the $j^{\text{th}}$ patient in the standard treatment group, and $I$ stands for the indicator function. The symbols ">" and "≈" are utilized to signify a win and a tie, respectively. Utilizing the previous $U$-statistic permits the definition of the probability that an outcome in the new treatment group exceeds or is equivalent to the outcome in the standard treatment group. This probability is represented as $U_2$ and is computed as the ratio of $U$ to the product of $N_N$ and $N_S$,

$$U_2 = \frac{U}{N_N \cdot N_S}$$

Therefore, employing the aforementioned probability, it becomes reachable to verify the win odds, formulated as,

$$WO = \frac{U_2}{1 - U_2}$$

Thus, it is obvious that this expression constitutes an expansion of the Mann–Whitney odds. Furthermore, by utilizing an alternative expression for the $U$-statistic, an alternative equivalent form of the win odds arises. Specifically, the $U$-statistic can be written as,

$$U = N_{W_N} + 0.5N_T$$

Here, $N_{W_N}$ denotes the number of wins in the new treatment group, while $N_T$ represents the number of ties. Hence, it follows that,

$$WO = \frac{N_{W_N} + 0.5N_T}{N_{W_S} + 0.5N_T} = \frac{P_N + 0.5P_{\text{tied}}}{P_S + 0.5P_{\text{tied}}}$$

In this framework, $N_{W_S}$ signifies the number of wins in the standard treatment group, $P_N$ and $P_S$ stand for the win proportions for the new treatment and standard treatment groups, respectively, while $P_{\text{tied}}$ refers to the proportion of ties. This leads to the formulation of the interpretation of win odds, different from that of the win ratio. Unlike the latter, the win odds account for ties in its computation. For example, if $WO = 1.3$, it implies that a patient in the new treatment group is 30% more likely to win or tie compared to a patient in the standard treatment group. It is noteworthy that in scenarios distinguished by a high frequency of ties, the win odds tend to approximate the value 1, closely aligning with the null hypothesis claiming $H_0 : WO = 1$.

## 3.3 The scenario involving varied weights in the outcomes

Within the realm of clinical trials, diverse outcomes are considered, depending upon the nature of the disease under investigation. Formally, some outcomes are expressed

https://doi.org/10.26219/heal.aueb.6426



as clementer, while others come with more severe consequences, potentially leading to fatality. Accordingly, in the analysis of these outcomes and hypothesis testing, disparate weights are attributed to signify the severity of each outcome. However, relying solely on medical severity to determine these weights may not always be appropriate. This arises from the comprehension that individual patients and organisms may exhibit varying reactions to each outcome. As an alternative, Ramchandani et al. (2016) suggested a novel approach for weight assignation. Notably, they claimed that the derivation of optimal weights can be executed through the maximization of hypothesis testing power. They employed their approach to both the test introduced by O'Brien and the one presented by Finkelstein and Schoenfeld, offering a comprehensive analysis of the topic.

Initially, within the context of O'Brien's test, the brief presentation function $\phi$ characterizing patient scores on outcomes is denoted as,

$$\phi(r_1, \ldots, r_p) = r_1 + r_2 + \ldots + r_p$$

with each $r_k$ corresponding the relative performance of patients on the $k^{\text{th}}$ outcome (as represented in section 3.1). The computation involves summing this function for both the treatment and control group, obtaining the $U$-statistic for the $k^{\text{th}}$ outcome as follows,

$$U_k = \frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} r_k(X_{ik}, Y_{jk})$$

Therefore, the O'Brien statistic, including outcome weights, is expressed as the product of the weight vector $w$ and the sum of $U$-statistics for each outcome, $w'U$. Within the framework of statistical analysis, based on knowledge from existing literature, particularly the relationship $\sqrt{N}U_k \sim \mathcal{N}(0, \sigma_k^2)$ where $N$ represents the total sample size and $\sigma_k^2$ signifies the variance for the $k^{\text{th}}$ outcome, leads to the conclusion that $\sqrt{N}\sum_k U_k \sim \mathcal{N}(0, \Lambda)$ with $\Lambda$ denotes the covariance matrix of $U$. When different weights are assigned to distinct outcomes, the distribution of $\sqrt{N}w'U$ follows a normal pattern with a mean of 0 and a variance of $w'\Lambda w$. Notably, it is given that,

$$\sqrt{N}w'U \sim \mathcal{N}(0, w'\Lambda w)$$

In the broader analytical context with a focus on maximizing statistical power, defined as the probability of rejecting the null hypothesis when the alternative is true, an optimization criterion comprises maximizing the expression $|w'\theta|(w'\Lambda w)^{-1/2}$, where $\theta = (\theta_1, \ldots, \theta_p)' = (E[U_1], \ldots, E[U_p])'$. Upon aiming this maximization, it becomes obvious that the optimal weight vector $(w)$ is proportional to the inverse of the covariance matrix $\Lambda$ multiplied by the vector of expected values $(\theta)$. Therefore, the optimal weight vector is given by $w = \Lambda^{-1}\theta$. Due to this circumstance, it becomes required to preselect a value for $\theta$ based on the specific aspects researchers aim to explore. For instance, the option for a positive $\theta$ signifies an interest in observing the positive effects of the treatment, even in the case of a single outcome. In general, each outcome allows for the evaluation of a treatment effect with either a positive or

https://doi.org/10.26219/heal.aueb.6426



negative value, denoted as $\theta_k > 0$ or $\theta_k < 0$. Respectively, the choice of the covariance matrix $\Lambda$ relies on prior knowledge, drawn from available data or hypotheses about treatment effects. When dealing with missing data, both selection procedures demand making assumptions about the data distribution and implementing analytical methods or simulations. In the case of O'Brien's test, the covariance matrix $\Lambda$ encompasses elements $\sigma_{k,l} = \text{cov}(U_k, U_l)$, where $k, l = 1, 2, \ldots, p$. These elements can be estimated as follows,

$$\hat{\sigma}_{k,l} = \frac{N}{(nm)^2}\left[\sum_{i=1}^{n}\sum_{i\neq i'}^{n}\sum_{j=1}^{m} r_k(X_{ik}, Y_{jk}) r_l(X_{i'l}, Y_{jl}) + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j\neq j'}^{m} r_k(X_{ik}, Y_{jk}) r_l(X_{il}, Y_{j'l})\right]$$

In the context of the Finkelstein–Schoenfeld (FS) test, the method encompasses utilizing independent $U$-statistics, and the test statistic is expressed as the sum of these statistics. The primary step involves the classification of participants into two groups: the control group, represented as Group 0, and the treatment group, denoted as Group 1. The analysis centers on comparing participants, with particular emphasis on the moment when an outcome becomes apparent. Let the variables $X_{ik}$ and $Y_{jk}$ correspond to the follow-up times of individuals $i$ and $j$, respectively, where one individual originates from the control group (Group 0) and the other from the treatment group (Group 1). For each outcome $k = 1, 2, \ldots, p$, indicator functions $\delta_{ik}$ and $\delta_{jk}$ are inserted to signify whether an outcome was observed in subject $i$ and subject $j$, respectively. As a result, a pairwise rank is established for subject $i$ versus subject $j$ concerning outcome $k$,

$$r_{ijk} = I(X_{ik} > Y_{jk}) - I(X_{ik} < Y_{jk})$$

This formulation depicts the comparison of follow-up times between subjects from distinct groups, offering a basis for assessing the development of an outcome in the analysis. It can be further distinguished into a specific expression known as the pairwise Gehan rank for the $k^{th}$ outcome,

$$r_{ijk} = I(X_{ik} > Y_{jk})\delta_{jk} - I(X_{ik} < Y_{jk})\delta_{ik}$$

This formulation combines the time until an event occurs with the censoring time for distinct $\delta$ values. However, for analysis purposes, the general form is employed. In addition, the import of the function $e_{ijk}$ is determined as follows,

$$e_{ijk} = \begin{cases} 1 & \text{, if } k = 1 \\ I(r_{ij1} = 0, r_{ij2} = 0, \ldots, r_{ij,k-1} = 0) & \text{, if } k \geq 2 \end{cases}$$

Here, $k = 1$ is assumed the fatal event that can occur. Consequently, the test statistic can be denoted as,

$$U = \sum_k U_k = \sum_k \left[\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{ijk} r_{ijk}\right]$$

https://doi.org/10.26219/heal.aueb.6426



Similar to the former test, it can be deduced that $\sqrt{N}w'U \sim \mathcal{N}(0, w'\Lambda w)$. Consequently, the optimal weights are specified by solving the equation $w = \Lambda^{-1}\theta$, where $\theta = (\theta_1, \ldots, \theta_p)' = (E[U_1], \ldots, E[U_p])'$ and $\Lambda = \text{cov}(U)$. These parameters need to be predetermined. To evaluate the elements of the covariance matrix $\Lambda$, the following formula can be employed,

$$\hat{\sigma}_{k,l} = \frac{N}{(nm)^2}\left[\sum_{i=1}^{n}\sum_{i\neq i'}^{n}\sum_{j=1}^{m} e_{ijk}r_{ijk}e_{i'jl}r_{i'jl} + \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{j\neq j'}^{m} e_{ijk}r_{ijk}e_{ij'l}r_{ij'l}\right]$$

Hence, in the aforementioned two test cases, the acquisition of optimal weights is accomplished. However, this approach's applicability that expands beyond these scenarios is suppressed. In clinical trials, for instance, an ordinary practice comprises employing both positive and negative weights, depending on the considered outcome. Additionally, specific weight values are often defined as maximum or minimum. The previously discussed method proves less efficient in such cases, requiring optimization with constraints. The programming language R contributes to encountering these challenges, as it allows the implementation of constrained optimization. This involves maximizing the quantity $|w'\theta_\phi|\,(w'\Lambda w)^{-\frac{1}{2}}$ with respect to the vector $w$, exploiting the **'optim'** function and box-constraints. Box-constraints determine the upper and lower bounds for each element of the vector, allowing for flexibility. The core of the analysis is detected in comprehending the characteristics of the outcomes under investigation. This understanding leads to the identification of the appropriate and optimal weights that effectively describe them.

## 3.4 Depiction of the win ratio and its features via graphical representation

Our survey denotes that the win ratio method possesses substantial prominence in the existing literature. With its flexible features, it is obvious that this approach holds considerable importance and promises greater advances in future analyses. Nevertheless, a notable downside concerns its sensitivity to the impact of follow-up time. Encountering this aspect, Finkelstein and Schoenfeld (2018) have made efforts to come up with a method for evaluating the win ratio in the presence of follow-up time, along with clarifying the confluences of individual components within composite measures through graphical representations.

Initially, considering the temporal factor in the analysis, it became crucial to illustrate the win ratio as a time-dependent function. Hence, the researchers assorted all patients into two groups: the new treatment group and the control group, and monitored their progression up to time "$c$". Highlighting the independence of the two analyzed groups, they centered on two distinct outcomes, one fatal, which was prioritized, and the other non-fatal. To characterize these events, the researchers employed distributions denoted as $F_0$ and $F_1$ for the fatal event in group 0 (control group) and group 1 (treatment group), respectively. Additionally, $G_0$ and $G_1$ signified the distributions for the non-fatal event in group 0 and group 1, respectively.

https://doi.org/10.26219/heal.aueb.6426



The upcoming stage involved defining the proportion of patients accomplishing win in either the treatment or control group, contributing to the calculation of the win ratio. Therefore, four integrals were expressed, comprising the marginal distribution of the fatal event and the joint distribution of both the fatal and non-fatal events.

As regards the primary integral, it assesses the probability of a patient from the treatment group exceeding the lifetime of a patient from the control group, with the latter experiencing death before time "$c$". In this case, a win is attributed to the treatment group, formulated as,

$$\int_0^c f_0(t)\bar{F}_1(t)\,dt$$

where $f_j$ signifies the density and $\bar{F}_j$ indicates the survivorship function corresponding to $F_j$, for $j = 0, 1$. The second integral refers to the likelihood of a patient from the control group surviving longer than a patient from the treatment group, while the latter passed away before time "$c$". In this scenario, the control group ensures a win, stated as,

$$\int_0^c f_1(t)\bar{F}_0(t)\,dt$$

The third integral is obtained when both patients from the treatment and control groups do not experience the fatal event, contributing to a comparison based on the non-fatal event. Thus, the probability of the non-fatal event occurring later in the treatment group patient compared to the control group patient, with both survival times exceeding $c$, is calculated as,

$$\bar{F}_1(c)\bar{F}_0(c)\int_0^c g_0(x|c)\bar{G}_1(x|c)\,dx$$

Here, $g_j(x|c)$ indicates the density for the non-fatal event given survival beyond $c$, and $\bar{G}_j(x|c)$ is the corresponding survivorship function, where $j = 0, 1$. Likewise, the fourth integral is determined as the probability of the non-fatal event occurring later in the control group patient relative to the treatment group patient, both experiencing survival times greater than $c$,

$$\bar{F}_0(c)\bar{F}_1(c)\int_0^c g_1(x|c)\bar{G}_0(x|c)\,dx$$

By combining the above-mentioned four integrals, it becomes reachable to establish the probability of the treatment group exceeding in both fatal and non-fatal events, as well as the likelihood of the control group developing superior outcomes in both events. Consequently, the first probability is specified as,

$$\int_0^c f_0(t)\bar{F}_1(t)\,dt + \bar{F}_1(c)\bar{F}_0(c)\int_0^c g_0(x|c)\bar{G}_1(x|c)\,dx$$

Parallelly, the second probability is stated by the formula,

$$\int_0^c f_1(t)\bar{F}_0(t)\,dt + \bar{F}_0(c)\bar{F}_1(c)\int_0^c g_1(x|c)\bar{G}_0(x|c)\,dx$$

https://doi.org/10.26219/heal.aueb.6426



Hence, the win ratio can be expressed through the equation,

$$\mathrm{WR} = \frac{\int_0^c f_0(t)\bar{F}_1(t)\,dt + \bar{F}_1(c)\bar{F}_0(c)\int_0^c g_0(x|c)\bar{G}_1(x|c)\,dx}{\int_0^c f_1(t)\bar{F}_0(t)\,dt + \bar{F}_0(c)\bar{F}_1(c)\int_0^c g_1(x|c)\bar{G}_0(x|c)\,dx}$$

This formulation of the win ratio highlights its dependency on the follow-up time $c$. This property offers a means to distinguish how the monitoring time and the components constituting composite measures play a decisive role in assessing the win ratio. The unique attributes and measures utilized vary across different clinical trials, leading to diverse graphical representations of the win ratio and its quiet features. For example, the implementation of this approach diverges between cancer and heart disease studies. Overall, the computation of the four integrals accounts for a critical step in determining the win ratio, subsequently enabling its graphical representation.

## 3.5 Win ratio for continuous non-normal outcomes

Within the domain of clinical trials, a variety of data is observed, carefully collected, and submitted for analysis to deduce several inferences. Regularly, encountered data diverges from a normal distribution, importing the challenge of operating non-normally distributed outcomes. Furthermore, there exists a class of data described by continuous attributes, such as the duration of a patient's hospital stay or the persistence of a virus or microbe within an individual's organism. Frequently, these categories of outcomes come out simultaneously, contributing to the investigation of continuous non-normal data. The win ratio method is applicable in this scenario as well. Wang and Pocock (2016) dealt with this situation, outlining the win ratio in a manner suitable for the analysis of continuous non-normal data. This approach is not significantly distinct from the formulation introduced by Pocock et al. (2012).

Initially, the researchers developed their survey with the assumption that continuous non-normal outcomes were available. Pursuing the conventional context of the win ratio, the study participants were categorized into two different groups: the treatment group, represented as group $A$, and the control group, indicated as group $B$. Group $A$ consisted of $N_A$ individuals, while Group $B$ encompassed $N_B$ subjects. A total of $N_A \times N_B$ comparisons were conducted among the patients, contributing to the identification of winners, losers, and ties according to the following standard criterion:

*Suppose patient $i$ originates from the treatment group and patient $j$ derives from the control group. If patient $i$ manifests more advantageous outcomes than patient $j$, it is denoted as a winner. In contrast, if patient $i$ demonstrates less favorable outcomes than patient $j$, it is characterized as a loser. In scenarios where patient $i$ and patient $j$ experience identical outcomes, a tie arises.*

https://doi.org/10.26219/heal.aueb.6426



Thus, the total of winners, losers, and indeterminate outcomes, denoted as $N_W$, $N_L$, and $N_T$ respectively, is obtained. It is noteworthy to mention that the relationship $N_W + N_L + N_T = N_A \times N_B$ holds. Consequently, the win ratio ($WR$) is formulated as,

$$WR = \frac{N_W}{N_L}$$

In addition, in their investigation of non-normal outcomes, the researchers expanded their survey to comprise the definition of the win difference. Specifically, they stated that the win difference could be expressed as the difference between the proportions of winners and losers. Indicated by the symbol $\Delta$, they clarified that it assumes values of 1, $-1$, and 0 under circumstances where the treatment group surpasses the control group, the control group exceeds the treatment group, and there is equality between the two groups, respectively. In the initial scenario, all members of the treatment group are identified as winners, while in the second, all members are regarded as losers in the treatment group.

## 3.6 Employing the win ratio approach to assess composite time-to-event endpoints when the components exhibit contrary treatment effects

Over the years, the employ of the win ratio method can be attributed to its unique and advantageous properties, as outlined in the particular study. Another pivotal aspect within the context of the preceding discussions involves the consideration of composite time-to-event endpoints, encompassing components with contrary effects deriving from treatment implementation. Take, for instance, the case of breast cancer, where treatment may positively affect the mortality event, extending the lifetime of a patient compared to those not subjected to the procedure. Additionally, this treatment might manifest favorable outcomes in terms of intended events, potentially reducing the time until recovery. Liao et al. (2022) dealt with this particular data structure and its analysis, utilizing an alternative version of the win ratio that takes account of the opposite effects of treatment on the components of composite endpoints. Their approach relied on the ranking of these components and the application of the Clayton copula, offering a robust framework for their research.

### 3.6.1 Win ratio approach, introduced by Liao et al. (2022)

Liao et al. (2022) recommended a remarkable perspective within the context of the win ratio method, outlining a methodology that entails the classification of participants into two different groups: the treatment group ($Z_i = 1$) and the control group ($Z_i = 0$). They imported the variable $X_i = (X_{i1}, \ldots, X_{ip})$ to denote the various potential outcomes, which were prioritized based on their clinical significance. They considered a joint distribution $F$, (i.e. $X_i \sim F$) for the control group and an alternative joint distribution $G$, (i.e. $X_i \sim G$) for the treatment group. The number of

https://doi.org/10.26219/heal.aueb.6426



individuals in the control group was represented as $n_1$, while the treatment group consisted of $n_2$ subjects, resulting in a total participant count of $n = n_1 + n_2$. They supposed that the outcomes under investigation comprised both event times and event times in absorption scenarios.

Supplementary, they implemented pairwise comparisons between patients in the two groups, allowing to specify winners, losers, and ties in each instance. They stated that in assessing two patients, $i$ and $j$, the expression $X_i \succ X_j$ represented that $X_i$ arose as the winner and $X_j$ as the loser, $X_i \prec X_j$ indicated $X_j$ as the winner and $X_i$ as the loser, while $X_i \succ\prec X_j$ signified an undetermined outcome, known as a tie. Moreover, they inserted the functions,

$$\psi_1(X_i, Z_i, X_j, Z_j) = I(X_i \succ X_j, Z_i = 0, Z_j = 1)$$

and

$$\psi_2(X_i, Z_i, X_j, Z_j) = I(X_i \prec X_j, Z_i = 0, Z_j = 1)$$

where $I$ corresponds to the indicator function. These relationships contributed to the definition of the probability that a patient in the control group would bring out as the winner, denoted as,

$$\tau_1 = E(\psi_1(X_i, Z_i, X_j, Z_j))$$

as well as the probability that a patient in the control group would emerge as the loser, expressed as,

$$\tau_2 = E(\psi_2(X_i, Z_i, X_j, Z_j))$$

These computations were conducted under the conditions of the joint distributions $F$ and $G$. It's worthwhile to emphasize that when there is no statistically significant difference between the two groups, indicating the validity of the null hypothesis, it yields that,

$$F = G \qquad \Rightarrow \qquad \tau_1 = \tau_2 = \tau$$

Here, $\tau < 0.5$ if $P((X_i, Z_i = 0) \succ\prec (X_j, Z_j = 1)) > 0$, a circumstance usually observed due to censoring.

Consequently, in their survey, they successfully developed the formulations for both the win ratio and win difference. In particular, they stated that the win ratio (WR) is determined by the ratio of $\tau_1$ to $\tau_2$, denoted as,

$$WR = \frac{\tau_1}{\tau_2}$$

Simultaneously, they claimed that the win difference (WD) can be signified as,

$$WD = \tau_1 - \tau_2$$

https://doi.org/10.26219/heal.aueb.6426



### 3.6.2 Win ratio approach within the components of composite endpoints exhibiting equivalent effects

In the domain of clinical trials, researchers are often occupied with the analysis of outcomes encompassing components that are distinguished from an alike direction of impact of the administered treatment. In their integrated study, Liao et al. (2022) clarified this unique aspect of clinical trial features. More precisely, they centered on one of the $p$ outcomes, represented as $X_i$, which they considered to consist of two components, expressed as $X_i = (X_{i1}, X_{i2})$. The research progressed with the assumption that the initial component corresponds to the time taken for recovery, while the latter component signifies the duration until the first remarkable improvement. These two components exhibit a certain degree of correlation. In addition, they took into consideration the censoring time, denoted as $c_i$, and highlighted its independence with regard to $(X_{i1}, X_{i2})|Z_i$. They also emphasized that the time taken for recovery was prioritized, and highlighted that this span can right-censor the time until the primary evident improvement, while the latter duration does not dispose of the identical property to imitate so as the time to recovery.

Afterward, the observed data was classified into four different groups according to the ranking of two events and the timing of censoring. The parameter $\delta$ was utilized to denote each allegation, contributing to four values for this parameter $(\delta = 1, 2, 3, 4)$. The categories are determined as follows,

i) In cases where both $X_{i1}$ and $X_{i2}$ are observed, the conditions $X_{i2} < X_{i1} \leq c_i$ hold, and this corresponds to $\delta = 1$.

ii) When only $X_{i1}$ is recorded, two conditions are met, $X_{i1} \leq X_{i2}$ and $X_{i1} \leq c_i$, leading to $\delta = 2$.

iii) When neither $X_{i2}$ nor $X_{i1}$ is visible, then $c_i < X_{i2}$ and $c_i < X_{i1}$, implying that $\delta = 3$.

iv) If only $X_{i2}$ is apparent, the relationship $X_{i2} \leq c_i < X_{i1}$ is established, resulting in $\delta = 4$.

The upcoming stage comprised the comparison of individuals within one group to each subject in the other group, resulting in conclusions about wins, losses, and indeterminate outcomes. This comparison yielded five different instances,

i) The patient receiving the new treatment experiences recovery first. It is posited as a win for the new treatment group when the time until the recovery of patient $i$ is captured, and concurrently, the time until the recovery of patient $j$ or its censoring time is shorter. This can be formulated as,

$$(X_j \succ X_i, Z_i = 0, Z_j = 1) = I(\{X_{i1} < \min(c_j, X_{j1})\}, Z_i = 0, Z_j = 1, \delta = 1, 2)$$

ii) The patient underlying the conservative treatment or placebo encounters recovery first. It is assumed a win for the control group when the time until

https://doi.org/10.26219/heal.aueb.6426



the recovery of patient $j$ is recorded, and simultaneously, the time until the recovery of patient $i$ or its censoring time is shorter. This can be stated as,

$$(X_i \prec X_j, Z_i = 0, Z_j = 1) = I(\{X_{j1} < \min(X_{i1}, c_i)\}, Z_i = 0, Z_j = 1, \delta = 1, 2)$$

iii) The patient undergoing the new treatment exhibits substantial improvement first. The new treatment group emerges as the winner. Mathematically, this is equivalent to,

$$(X_i \prec X_j, Z_i = 0, Z_j = 1) = I(\{X_{i2} < \min(X_{j2}, c_j)\} \cap \{X_{i1} \not< \min(c_j, X_{j1})\}, Z_i = 0, Z_j = 1, \delta = 1, 4)$$

iv) The patient receiving the conservative treatment or the placebo performs remarkable improvement first. The control group is denoted as the winner. This can be written as,

$$(X_i \prec X_j, Z_i = 0, Z_j = 1) = I(\{X_{j2} < \min(X_{i2}, c_i)\} \cap \{X_{j1} \not< \min(c_i, X_{i1})\}, Z_i = 0, Z_j = 1, \delta = 1, 4)$$

v) No event occurs for either patient in the pair. The outcome is regarded as indeterminate. Equivalently,

$$(X_i \succ\prec X_j, Z_i = 0, Z_j = 1) = I(\{c < \min(X_{i1}, X_{i2}, X_{j1}, X_{j2})\}, Z_i = 0, Z_j = 1, \delta = 3)$$

In light of these five distinct scenarios, the win, loss, and tie functions for the treatment group can be derived according to the outcomes of patients. In particular, the number of winners ($N_W$) for the new treatment group is computed by aggregating the instances where patients in the new treatment group encountered recovery or initial evident improvement first. Thus, the win function can be formulated as follows,

$$\begin{aligned}\psi_1(X_i, Z_i, c_i, X_j, Z_j, c_j) &= I(X_i \succ X_j, Z_i = 0, Z_j = 1) \\ &= I(\{X_{i1} < \min(c_j, X_{j1})\}, Z_i = 0, Z_j = 1, \delta = 1, 2) \\ &\quad + I(\{X_{i2} < \min(X_{j2}, c_j)\} \cap \{X_{i1} \not< \min(c_j, X_{j1})\}, \\ &\quad Z_i = 0, Z_j = 1, \delta = 1, 4)\end{aligned}$$

Similarly, the set of subjects within the control group, who experience recovery or initial remarkable improvement first, comprises the total of patients not favoring the new treatment, represented as $N_L$. Therefore, the loss function can be expressed as,

$$\begin{aligned}\psi_2(X_i, Z_i, c_i, X_j, Z_j, c_j) &= I(X_i \prec X_j, Z_i = 0, Z_j = 1) \\ &= I(\{X_{j1} < \min(X_{i1}, c_i)\}, Z_i = 0, Z_j = 1, \delta = 1, 2) \\ &\quad + I(\{X_{j2} < \min(X_{i2}, c_i)\} \cap \{X_{j1} \not< \min(c_i, X_{i1})\}, \\ &\quad Z_i = 0, Z_j = 1, \delta = 1, 4)\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



Meanwhile, the number of ties is signified by the individuals encompassed in the latter scenario, where no event occurred for any patient within the pair. Hence, the tie function can be stated as,

$$\begin{aligned}\psi_3(X_i, Z_i, c_i, X_j, Z_j, c_j) &= I(X_i \succ\prec X_j, Z_i = 0, Z_j = 1) \\ &= I(\{c < \min(X_{i1}, X_{i2}, X_{j1}, X_{j2})\}, Z_i = 0, Z_j = 1, \delta = 3)\end{aligned}$$

Consequently, the researchers exhibited the win ratio, calculated as the quotient of the total of winners to the total of losers, utilizing the formula,

$$\text{WR} = \frac{N_W}{N_L}$$

Concurrently, they determined the win difference as the difference between the count of winners and the count of losers. As a result, the expression for the win difference is clarified as,

$$\text{WD} = N_W - N_L$$

Overall, they referred to the relationship between the number of winners and losers through both a ratio and a difference metric.

### 3.6.3 Win ratio approach within the components of composite endpoints exhibiting contrary effects

Clinical trials often investigate composite endpoints, which encompass multiple components introducing a unique property. More precisely, these components exhibit opposite effects under the impact of the treatment. Essentially, when a patient underlies the new treatment, it may both expedite the development of one event and hold the realization of another over. Liao et al. (2022) centered on a particular outcome, represented as $X_i$, incorporating two components ($X_{i1}$, $X_{i2}$). For the initial component, exposure to the new treatment is expected to diminish the time until the examined event. For instance, the administration of the new treatment is possible to contribute to a decrease of the time until recovery or improvement. Concurrently, the latter component is anticipated to perform a contrary trend, with the receiving of the new treatment leading to an increase in the time until the realization of another event under examination, such as the time until death. In assessing such endpoints, a challenge was developed pertaining to the potential cancellation of one event by another. This subject arose because the success of one component implied the failure of the other, and vice versa. Selecting the exclusive study of one of the components was considered incommodious, as it would yield inaccurate inferences. Hence, the simultaneous investigation of both components, as well as the singular focus on just one, was identified as compromising the power of the hypothesis test—an undesirable outcome. Researchers attempted to confront this issue throughout their analysis.

Initially, they entered their research by implementing it in the framework of COVID-19. They supposed an outcome involved two components, represented as

https://doi.org/10.26219/heal.aueb.6426



$X_i = (X_{i1}, X_{i2})$, as well the censoring time $c_i$. The researchers considered the independence between $(X_{i1}, X_{i2})|Z_i$ and $c_i$. The outcomes under examination comprised the time until recovery, the time until intubation, and the time until fatality. To proceed with their analysis, certain situations and assumptions were employed as preliminary elements. These encompassed,

1. The initial component indicated the patient's enhancement, while the secondary outcome signified a deterioration.

2. Recovery constituted the ultimate phase for the patient, denoting that accomplished recovery contributed to the release from the hospital in good health, preventing entry into the intubation procedure.

3. Patient health deterioration could come before recovery, resulting in the necessity for intubation.

4. The realization of death excludes the opportunity for patient recovery. Moreover, scenarios were analyzed where a patient recuperates, discharges, and afterward dies, compelling improbable the inclusion of the death event in the research. Therefore, the event of death right censors the event of recovery and vice versa.

5. Death might take place after a patient underlies the intubation stage.

Given these considerations, for the construction of an elementary yet balanced model, the researchers elided the event of death. Meanwhile, they utilized initially the event of recovery and subsequently the event of intubation for the comparison of patients in each pair.

Thus, the comparison of individuals within one group to each subject in the other group results in conclusions about wins, losses, and indeterminate outcomes. This comparison yielded five different instances,

i) The patient receiving the new treatment experiences recovery first. The new treatment group is posited as the winner. This scenario implies that the time until the recovery of the patient $j$ is captured, and concurrently, the time until the recovery of the patient $i$ or its censoring time is longer. This can be formulated as,

$$(X_j \succ X_i, Z_i = 0, Z_j = 1) = I(\{X_{j1} < \min(c_j, X_{i1})\}, Z_i = 0, Z_j = 1, \delta = 1, 2)$$

ii) The patient underlying the conservative treatment or placebo encounters recovery first. The control group is assumed the winner. This entails that the time until the recovery of patient $i$ is recorded, and simultaneously, the recovery time for patient $j$ or its censoring time is longer. This can be stated as,

$$(X_j \prec X_i, Z_i = 0, Z_j = 1) = I(\{X_{i1} < \min(X_{j1}, c_j)\}, Z_i = 0, Z_j = 1, \delta = 1, 2)$$

https://doi.org/10.26219/heal.aueb.6426



iii) The patient undergoing the conservative treatment or placebo exhibits intubation first. The new treatment group is deemed the winner. Mathematically, this is equivalent to,

$$(X_i \prec X_j, Z_i = 0, Z_j = 1) = I(\{X_{i2} < \min(X_{j2}, c_j)\} \cap \{X_{i1} \not\preceq \min(c_j, X_{j1})\}, \\ Z_i = 0, Z_j = 1)$$

iv) The patient receiving the new treatment faces intubation first. The control group emerges as the winner. This can be written as,

$$(X_i \prec X_j, Z_i = 0, Z_j = 1) = I(\{X_{j2} < \min(X_{i2}, c_i)\} \cap \{X_{j1} \not\preceq \min(c_i, X_{i1})\}, \\ Z_i = 0, Z_j = 1, \delta = 1, 4)$$

v) No event occurs for either patient in the pair. The outcome is regarded as indeterminate. Equivalently,

$$(X_i \succ\prec X_j, Z_i = 0, Z_j = 1) = I(\{c < \min(X_{i1}, X_{i2}, X_{j1}, X_{j2})\}, \\ Z_i = 0, Z_j = 1, \delta = 3)$$

In the context of these five different scenarios, the number of winners ($N_W$) for the new treatment is computed by aggregating instances where patients in the new treatment group recover before their counterparts in the control group, and instances where patients in the control group encounter intubation before those in the new treatment group. Similarly, the total of losers ($N_L$) for the new treatment is defined by summing cases where patients in the control group experience recovery first and cases where patients in the new treatment group face intubation before their counterparts. Consequently, the win ratio (WR) is expressed through the relationship,

$$WR = \frac{N_W}{N_L}$$

## 3.7 Various win ratio categories

In our study, we have underlined the main role played by the win ratio method in the context of clinical trials, highlighting its extraordinary benefits over conventional techniques for analyzing composite endpoints. Notably, this approach stands out for its unique capability to take the clinical significance of each event into account. However, it is essential to acknowledge a limitation associated with recurrent events, where non-fatal incidents like microbial infections or hospitalizations may occur multiple times within an individual. The conventional win ratio method, as originally formulated by Pocock et al. (2012) and as discussed thus far, focuses exclusively on the initial occurrence of a non-fatal event. Recognizing the need to enhance the statistical efficiency of clinical trials, Mao et al. (2022) have introduced a comprehensive formulation for the win ratio. This formulation generates four distinct variations of

https://doi.org/10.26219/heal.aueb.6426



the win ratio, including the traditional Pocock et al. (2012) model. This innovative approach effectively addresses the challenges obtained by recurrent events, offering a more diverse and adaptable framework for the analysis of clinical trial data.

Mao et al. (2022) invented a regular approach to deduce the general form of the win ratio and subsequently its four variations. They initiated their procedure by distinguishing between censored and uncensored events, specifying the initial as observed data and the latter as complete or full data. Within the complete data context, they introduced the variable $T_F$ to denote the time until the occurrence of a fatal event and $T_k$, where $k = 1, 2, \ldots,$ to represent the time until the $k^{\text{th}}$ non-fatal event. They assumed a continuous distribution for these time variables and established the relationships,

$$N_{T_F}(t) = I(T_F \leq t) \quad \text{and} \quad N_{T_{N-F}}(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$$

Here, $I$ poses the indicator function. Given that $T_F$ signifies a terminal event, $N_{T_{N-F}}(.)$ should be zero in the interval $[T_F, \infty)$. To assemble all events experienced by an individual up to time $t$, they introduced the variable $Y(t)$,

$$Y(t) = \{N_{T_F}(u), N_{T_{N-F}}(u) : 0 \leq u \leq t\}$$

This formulation allowed them to express the full data as $Y := Y(\tau)$, where $\tau$ denotes the maximum follow-up time for an individual. They also referred to the censoring time, denoting it as $T_C$. This is obtained because of the completion of the study or the participant's departure from the follow-up period due to various reasons. To characterize the observed data, they introduced the variable $X$, defined as $X = T_C \wedge T_F$, where $T_b \wedge T_c = \min\{T_b, T_c\}$. Consequently, they successfully signified the observed data as $\{Y(X), X\}$. Furthermore, the participants were categorized into two groups: the first included individuals who received the new treatment, referred to as the treatment group, and the second encompassed those who put themselves through either a placebo or conservative treatment, determined as the control group. They utilize the variable $\alpha$ to express the group to which each individual belonged, with $\alpha$ taking the value 0 for the control group and 1 for the treatment group. The number of subjects in the control group was denoted as $n_0$, while the number of patients in the treatment group was declared as $n_1$. The total number of participants in both groups was represented as $N = n_0 + n_1$. Consequently, the observed data could be expressed in the form,

$$\{Y_a(X_a), X_a\} \text{ where } a = 1, 0$$

To be more precise,

$$\{Y_{ai}(X_{ai}), X_{ai}\} \text{ where } i = 1, \ldots, n_a \text{ and } a = 1, 2$$

Subsequently, the researchers employed a general comparison rule to assess individuals in two groups, determining total wins and losses. This regulation, denoted as,

$$\mathcal{W}(.,.)(.) : \mathcal{Y} \times \mathcal{Y} \times [0, \tau] \to \{0, 1\}$$

takes place on the sample space $\mathcal{Y}$ for outcomes $Y$. According to this rule:

https://doi.org/10.26219/heal.aueb.6426



$$\mathcal{W}(Y_a, Y_{1-a})(t) = 1 \text{ if } Y_a \text{ wins over } Y_{1-a} \text{ at time } t$$
$$\mathcal{W}(Y_{1-a}, Y_a)(t) = 1 \text{ if } Y_{1-a} \text{ wins over } Y_a \text{ at time } t$$

Assumptions were crucial and had to be taken into account for implementing this rule. Firstly, they clarified that $\mathcal{W}(Y_a, Y_{1-a})(t)$ is dependent exclusively on $Y_a(t)$ and $Y_{1-a}(t)$, ensuring the realization of the comparison between an individual from the treatment group and one from the control group. It is noted that the comparison was valid only if both subjects were uncensored simultaneously; otherwise, it was undefined. Secondly, they argued that,

$$\mathcal{W}(Y_a, Y_{1-a})(t) + \mathcal{W}(Y_{1-a}, Y_a)(t) \in \{1, 0\}$$

indicating that $\mathcal{W}(Y_a, Y_{1-a})(t)$ and $\mathcal{W}(Y_{1-a}, Y_a)(t)$ cannot be valid at the same time. That is, they could not both be 1 simultaneously. When one equaled 1, the other equaled 0. In the instance that both were 0, the comparison was undetermined. Thirdly, they assumed that,

$$\mathcal{W}(Y_a, Y_{1-a})(t) = \mathcal{W}(Y_a, Y_{1-a})(T_{F_a} \wedge T_{F_{1-a}} \wedge t)$$

in order to obtain a constant win rate even if patients died, especially in cases prioritizing fatal events. This led to the statement,

$$\mathcal{W}(Y_a, Y_{1-a})(X_1 \wedge X_0) = \mathcal{W}(Y_a, Y_{1-a})(T_{C_1} \wedge T_{C_0})$$

Hence, given a win function $\mathcal{W}$ that satisfies the above-mentioned assumptions, they provided the definition for the general form of the win ratio as follows,

$$\hat{\mathcal{E}}_{n_1,n_0}(\mathcal{W}) = \frac{\frac{1}{n_1 n_0}\sum_{i=1}^{n_1}\sum_{j=1}^{n_0}\mathcal{W}(Y_{1i}, Y_{0j})(X_{1i} \wedge X_{0j})}{\frac{1}{n_1 n_0}\sum_{i=1}^{n_1}\sum_{j=1}^{n_0}\mathcal{W}(Y_{0j}, Y_{1i})(X_{1i} \wedge X_{0j})}$$

Here, similar to the conventional win ratio definition, the numerator represents the count of wins, and the denominator corresponds to the count of losses. This formulation captures the relative difference between the treatment group and the control group while accounting for the specified conditions.

When employing various functions for the win ratio calculation, the researchers discovered four distinct win ratios. Initially, they examined a function denoted as,

$$\mathcal{W}_S(Y_a, Y_{1-a})(t) = I\{T_{F_{1-a}} < T_{F_a} \wedge t^+\} + I\{T_{F_a} \wedge T_{F_{1-a}} > t, T_{1,1-a} < T_{1,a} \wedge t^+\}$$

Here, $T_b < T_c \wedge T_f^+$ signifies that $T_b < T_c$ and $T_b \leq T_f$. The resulting win ratio coincides with the conventional form known as the standard win ratio. Notably, the acronym SWR is utilized when referring to this particular formulation. It's essential to highlight that, in this context, the analysis assumes an absence of tied outcomes.

An alternative expression for the $\mathcal{W}$ function is given by,

$$\begin{aligned}\mathcal{W}_N(Y_a, Y_{1-a})(t) = {} & I\{T_{F_{1-a}} < T_{F_a} \wedge t^+\} \\ & + I\{T_{F_a} \wedge T_{F_{1-a}} > t, N_{T_{N-F},a}(t) < N_{T_{N-F},1-a}(t)\}\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



In this formulation, the cumulative frequency of non-fatal events, denoted by $N_{T_{\text{N-F}}}$, is encompassed. When this function is applied to the general win ratio formula, the well-known naive win ratio (NWR) arises. It is notable that NWR includes a larger range of events compared to the standard win ratio (SWR), as it assumes the cumulative frequency of non-fatal events rather than just the occurrence of the first non-fatal event, which is considered by the SWR. However, a limitation of NWR is its failure to take the timing of these events into consideration, a factor bargained by SWR. Consequently, NWR is often not the preferred choice in certain analyses.

While conducting the analysis, it is regarded desirable to consider both the number of events and their occurrence time. Hence, two variations of the $\mathcal{W}$ function were employed, giving rise to divergent win ratios. One formulation supposes an equal frequency of non-fatal events in the two compared groups and depends on the primary event's demonstration.

$$\begin{aligned}\mathcal{W}_F(Y_a, Y_{1-a})(t) = & I\{T_{F_{1-a}} < T_{F_a} \wedge t^+\} \\ & + I\{T_{F_a} \wedge T_{F_{1-a}} > t, N_{T_{N-F},a}(t) < N_{T_{N-F},1-a}(t)\} \\ & + I\{T_{F_a} \wedge T_{F_{1-a}} > t, N_{T_{N-F},a}(t) = N_{T_{N-F},1-a}(t), T_{1,1-a} < T_{1,a} \wedge t^+\}\end{aligned}$$

This formulation yields what is termed the first-event-assisted win ratio (FWR).

Likewise, the latter expression of the $\mathcal{W}$ function is derived, relying on an equivalent number of non-fatal events in the two groups, similar to the prior one. However, the outcome is determined based on the occurrence of the last event.

$$\begin{aligned}\mathcal{W}_L(Y_a, Y_{1-a})(t) = & I\{T_{F_{1-a}} < T_{F_a} \wedge t^+\} \\ & + I\{T_{F_a} \wedge T_{F_{1-a}} > t, N_{T_{N-F},a}(t) < N_{T_{N-F},1-a}(t)\} \\ & + I\{T_{F_a} \wedge T_{F_{1-a}} > t, N_{T_{N-F},a}(t) = N_{T_{N-F},1-a}(t) = k \geq 1, T_{k,1-a} < T_{k,a} \wedge t^+\}\end{aligned}$$

By applying this specific function to the general form of the win ratio, we arrive at the well-known last-event-assisted win ratio (LWR). Figure 3.1 visually represents the process of determining winners, losers, or ties when comparing two patients. Each variant of the win ratio is employed to derive the outcome of the comparison.

The various expressions derived from the general win ratio form display interconnected relations. Specifically, when recurrent non-fatal events do not exist, the FWR and the LWR agree with the SWR. We can write that,

$$N_{N-F}(\tau) \leq 1 \implies \text{FWR} = \text{LWR} = \text{SWR}$$

Additionally, the sets of indeterminate pairs resulting from the comparisons are identical across SWR, FWR, and LWR. These pairs satisfy the conditions,

$$T_{F_a} \wedge T_{F_{1-a}} > T_{C_a} \wedge T_{C_{1-a}} \text{ and } N_{T_{N-F},a}(T_{C_a} \wedge T_{C_{1-a}}) = N_{T_{N-F},1-a}(T_{C_a} \wedge T_{C_{1-a}}) = 0$$

Additionally, for the indeterminate pairs under the NWR, it yields that,

$$T_{F_a} \wedge T_{F_{1-a}} > T_{C_a} \wedge T_{C_{1-a}} \text{ and } N_{T_{N-F},a}(T_{C_a} \wedge T_{C_{1-a}}) = N_{T_{N-F},1-a}(T_{C_a} \wedge T_{C_{1-a}})$$

https://doi.org/10.26219/heal.aueb.6426



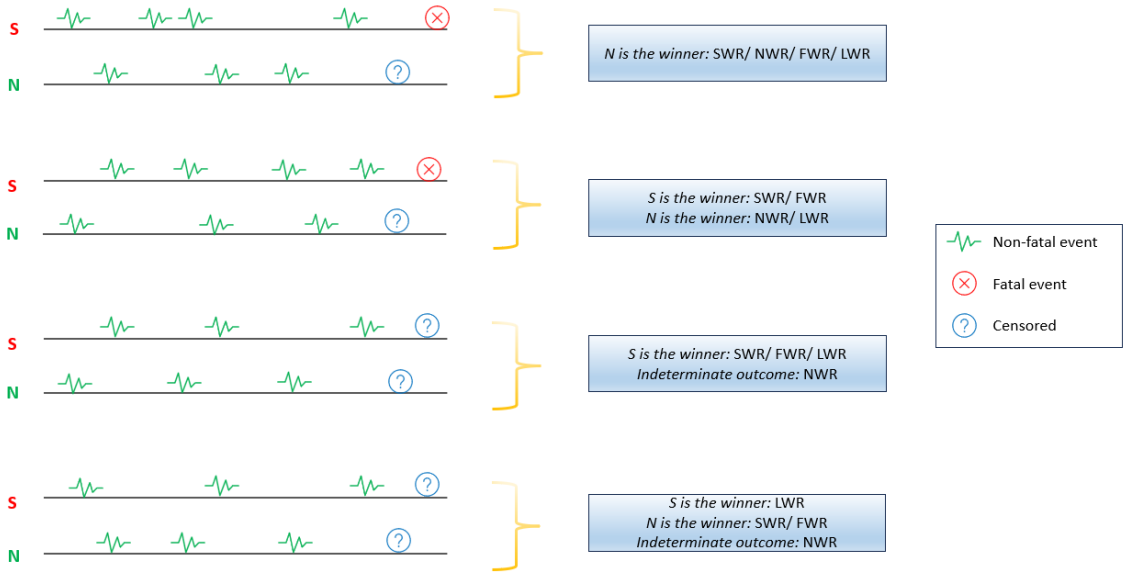



*Figure 3.1:* A visual representation of the process of determining winners, losers, or ties when comparing two patients. It is supposed that 'S' denotes an individual under the control group or standard treatment group, while 'N' signifies an individual receiving the new treatment. Each variant of the win ratio is employed to derive the outcome of the comparison (Pocock et al. (2012), Mao et al. (2022)).

Considering these observations, the preference leans towards employing FWR and LWR over NWR. FWR and LWR, by simultaneously assuming the number of non-fatal events and the timing of their occurrences, exceed NWR, which lacks the ability to recognize the temporal aspect of events. Furthermore, LWR overcomes FWR by deploying the latter event, which is typically distinguished by greater clinical significance, and by assigning the latest win-loss status in the case of disparate frequencies. A summary of the various win ratio forms and their distinctive features is presented in Table 3.1.

https://doi.org/10.26219/heal.aueb.6426



*Table 3.1:* Various representations of the win ratio and their corresponding features (Mao et al. (2022)).

| Various Win Ratio Categories | | | | |
|---|---|---|---|---|
| | **SWR** | **NWR** | **FWR** | **LWR** |
| **Foundational element** | Occurrence of the first event | Cumulative frequency of non-fatal events | ▪ Equal number of non-fatal events in the two comparison groups<br>▪ Occurrence of the first event | ▪ Equal number of non-fatal events in the two comparison groups<br>▪ Occurrence of the last event |
| **Preferential ranking** | SWR is favored over NWR because the latter neglects the temporal aspect of event occurrences | NWR considers a larger number of events than SWR but disregards the timing of their occurrences | FWR is favored over NWR and SWR because it considers both the quantity and timing of non-fatal events | LWR is favored over FWR because it considers the last fatal event, which typically holds greater medical significance, and assigns the latest win-loss status when the pair has disparate frequencies |

https://doi.org/10.26219/heal.aueb.6426

# Chapter 4

# Stratified Win Ratio

## 4.1 Definition of the stratified win ratio

In the realm of clinical trials, it is imperative to consider various factors that may influence the outcome of a treatment, ensuring the accuracy of the final research results. Commonly encountered variables, such as age and gender, require a particular approach to data analysis. One prevalent method involves stratifying the data based on these variables and subsequently conducting a stratified analysis. The win ratio method is extended in the context of stratified data. Referred to as the "stratified win ratio," this methodology builds upon the Mantel-Haenszel stratified odds ratio. Its development hinges on the strategic mobilization of data within strata, offering a nuanced perspective that enhances the reliability and accuracy of the analysis.

Consider a sample of patients, randomly assigned to two groups across $M$ strata in a clinical study. One group undergoes a novel treatment, while the other receives either a placebo or a conservative treatment. Let $N_N^{(m)}$ signify the number of participants in the new treatment group, and $N_S^{(m)}$ denotes the count in the standard treatment group, also referred to as the control group, within the $m^{th}$ stratum. The total sample size in the $m^{th}$ stratum is $N^{(m)} = N_N^{(m)} + N_S^{(m)}$, where $m = 1, 2, \ldots, M$. For each stratum $(m)$, the information about patient outcomes is captured through $X_i^{(m)}$ for those in the new treatment group and $Y_j^{(m)}$ for those in the control group $(i = 1, 2, \ldots, N_N^{(m)}; j = 1, 2, \ldots, N_S^{(m)}; m = 1, 2, \ldots, M)$. Two kernel functions, $K$ and $L$, are defined as follows,

$$K(X_i^{(m)}, Y_j^{(m)}) = \begin{cases} 1 & \text{, if } X_i^{(m)} \text{ "wins" over } Y_j^{(m)} \\ 0 & \text{, if } X_i^{(m)} \text{ "loses" over } Y_j^{(m)} \end{cases} = I(X_i^{(m)} > Y_j^{(m)})$$

$$L(X_i^{(m)}, Y_j^{(m)}) = \begin{cases} 1 & \text{, if } Y_j^{(m)} \text{ "wins" over } X_i^{(m)} \\ 0 & \text{, if } Y_j^{(m)} \text{ "loses" over } X_i^{(m)} \end{cases} = I(X_i^{(m)} < Y_j^{(m)})$$

In essence, $K$ yields 1 if the outcome $X_i^{(m)}$ surpasses $Y_j^{(m)}$, and $L$ yields 1 if $Y_j^{(m)}$ exceeds $X_i^{(m)}$. Both functions result in 0 in other cases. The sum of these functions

https://doi.org/10.26219/heal.aueb.6426



can only be 1 or 0, indicating a decisive outcome or an indeterminate result, respectively. The sum cannot reach 2. In the context of the $m^{th}$ stratum, we determine the overall number of wins for both the new treatment and standard treatment groups through the following computations,

$$N_{W_N}^{(m)} = \sum_{i=1}^{N_N^{(m)}} \sum_{j=1}^{N_S^{(m)}} K(X_i^{(m)}, Y_j^{(m)})$$

$$N_{W_S}^{(m)} = \sum_{i=1}^{N_N^{(m)}} \sum_{j=1}^{N_S^{(m)}} L(X_i^{(m)}, Y_j^{(m)})$$

The computation of wins is conducted independently for each stratum, and these results are amalgamated to derive the estimate for the stratified win ratio. Consequently, the weighted stratified win ratio ($\text{WR}_{\text{stratified}}$) is formulated as follows,

$$\text{WR}_{\text{stratified}} = \frac{\sum_{m=1}^{M} w^{(m)} N_{W_N}^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_{W_S}^{(m)}} \tag{4.1}$$

In this equation, $w^{(m)}$ represents the weight assigned to the $m^{th}$ stratum. Notably, for the scenario where $M = 1$, the formula reverts to the original definition of the win ratio as initially defined by Pocock et al. (2012). It is essential to underscore that this approach effectively eliminates the potential confounding influence of prognostic variables, which serve as the basis for strata creation, thereby enhancing the analytical precision.

A critical aspect of the mathematical formula under consideration involves the determination of weights. The simplest approach involves assigning a weight of 1 to $w^{(m)}$ for all $M$ layers. This implies an equal weighting across all strata, resulting in the well-known unweighted stratified win ratio. An alternative strategy is to set $w^{(m)} = \frac{1}{N^{(m)}}$, where the inverse of the stratum size is considered. Notably, this choice aligns with the Mantel-Haenszel stratified odds ratio estimator. It is essential to recognize that various weight assignment options exist, contingent upon the specific analytical goals and preferences. The selection of weights plays a crucial role in influencing the overall impact and interpretation of the stratified win ratio (Dong et al. (2018)).

## 4.2 Dispersal patterns of the stratified win ratio

Within the realm of research and analysis, a central focus turns around understanding the distributional characteristics of the statistical measures and formulas employed. In the case of the stratified win ratio, our attention is directed towards elucidating its distribution. Initially, we propose $M$ strata, each navigated by uniform weights $w^{(m)}$ for every stratum $m = 1, 2, \ldots, M$. Building upon the arguments

https://doi.org/10.26219/heal.aueb.6426



presented in section 2.3.6 and asserting that the counts of wins for both the treatment and control groups exhibit an asymptotic normal distribution, we enunciate the distribution of $N_{W_N}^{(m)}$ and $N_{W_S}^{(m)}$ within stratum $m$ as follows,

$$\begin{pmatrix} N_{W_N}^{(m)} \\ N_{W_S}^{(m)} \end{pmatrix} \dot{\sim} \mathcal{N}\left( \begin{bmatrix} \theta_{W_N}^{(m)} \\ \theta_{W_S}^{(m)} \end{bmatrix}, \begin{bmatrix} \sigma_{W_N}^{2(m)} & \sigma_{W_N W_S}^{(m)} \\ \sigma_{W_N W_S}^{(m)} & \sigma_{W_S}^{2(m)} \end{bmatrix} \right)$$

Specifically, applying the approach developed by Dong et al. (2016) produces the following results,

$$N_{W_N}^{(m)} \dot{\sim} \mathcal{N}(\theta_{W_N}^{(m)}, \sigma_{W_N}^{2(m)})$$

where,

$$\theta_{W_N}^{(m)} = N_N^{(m)} \cdot N_S^{(m)} \cdot \theta_K^{(m)}$$

$$\theta_K^{(m)} = E[K(X_i^{(m)}, Y_j^{(m)})]$$

$$\sigma_{W_N}^{2(m)} = \frac{1}{N_N^{(m)}} \sigma_1^{2(m)} + \frac{1}{N_S^{(m)}} \sigma_2^{2(m)}$$

$$\sigma_1^{2(m)} = \frac{N_N^{(m)} \cdot N_S^{(m)}}{N_S^{(m)} - 1} \sum_{i=1}^{N_N^{(m)}} \sum_{j=1}^{N_S^{(m)}} \sum_{j'=1, j' \neq j}^{N_S^{(m)}} [K(X_i^{(m)}, Y_j^{(m)}) - \theta_K^{(m)}][K(X_i^{(m)}, Y_{j'}^{(m)} - \theta_K^{(m)}]$$

$$\sigma_2^{2(m)} = \frac{N_N^{(m)} \cdot N_S^{(m)}}{N_N^{(m)} - 1} \sum_{j=1}^{N_S^{(m)}} \sum_{i=1}^{N_N^{(m)}} \sum_{i'=1, i' \neq i}^{N_N^{(m)}} [K(X_i^{(m)}, Y_j^{(m)}) - \theta_K^{(m)}][K(X_{i'}^{(m)}, Y_j^{(m)}) - \theta_K^{(m)}]$$

Likewise, we derive the results for the variable corresponding to the number of wins in the standard treatment.

$$N_{W_S}^{(m)} \dot{\sim} \mathcal{N}(\theta_{W_S}^{(m)}, \sigma_{W_S}^{2(m)})$$

The covariance between the variables $N_{W_N}^{(m)}$ and $N_{W_S}^{(m)}$ is computed as follows,

$$\sigma_{W_N W_S}^{(m)} = \frac{1}{N_N^{(m)}} \sigma_{W_N W_S 1}^{(m)} + \frac{1}{N_S^{(m)}} \sigma_{W_N W_S 2}^{(m)}$$

with,

$$\sigma_{W_N W_S 1}^{(m)} = \frac{N_N^{(m)} N_S^{(m)}}{N_S^{(m)} - 1} \sum_{i=1}^{N_N^{(m)}} \sum_{j=1}^{N_S^{(m)}} \sum_{j'=1, j' \neq j}^{N_S^{(m)}} \left[ K(X_i^{(m)}, Y_j^{(m)}) - \theta_K^{(m)} \right] \left[ L(X_i^{(m)}, Y_{j'}^{(m)}) - \theta_L^{(m)} \right]$$

$$\sigma_{W_N W_S 2}^{(m)} = \frac{N_N^{(m)} N_S^{(m)}}{N_N^{(m)} - 1} \sum_{j=1}^{N_S^{(m)}} \sum_{i=1}^{N_N^{(m)}} \sum_{i'=1, i' \neq i}^{N_N^{(m)}} \left[ K(X_i^{(m)}, Y_j^{(m)}) - \theta_K^{(m)} \right] \left[ L(X_{i'}^{(m)}, Y_j^{(m)}) - \theta_L^{(m)} \right]$$

In the next crucial stage, the evaluation of $\theta_K^{(m)}$ and $\theta_L^{(m)}$ takes precedence. Under the null hypothesis $H_0$, this estimation is accomplished utilizing the mathematical expression,

$$\hat{\theta}_K^{(m)} = \hat{\theta}_L^{(m)} = \frac{N_{W_N}^{(m)} + N_{W_S}^{(m)}}{2 \cdot N_N^{(m)} \cdot N_S^{(m)}}$$

https://doi.org/10.26219/heal.aueb.6426



By substituting these estimated values into the previously mentioned formulas for variance and covariance under the null hypothesis, we can derive the estimations for $\hat{\sigma}^{2(m)}_{W_N}$, $\hat{\sigma}^{2(m)}_{W_S}$ and $\hat{\sigma}^{(m)}_{W_N W_S}$. It is recognized that the win ratio for the $m^{th}$ stratum is expressed as,

$$WR^{(m)} = \frac{N^{(m)}_{W_N}}{N^{(m)}_{W_S}}$$

Utilizing the delta method, we derive that the logarithm of the win ratio follows an asymptotically normal distribution,

$$\log(WR^{(m)}) \dot{\sim} \mathcal{N}(\theta_{\log(WR^{(m)})}, (\sigma^2_{\log(WR^{(m)})})$$

where,

$$\sigma^2_{\log(WR^{(m)})} = \frac{\sigma^{2(m)}_{W_N}}{(\theta^{(m)}_{W_N})^2} + \frac{\sigma^{2(m)}_{W_S}}{(\theta^{(m)}_{W_S})^2} - \frac{2\sigma^{(m)}_{W_N W_S}}{\theta^{(m)}_{W_N} \cdot \theta^{(m)}_{W_S}}$$

To estimate the variation of the logarithm of the win ratio for the $m^{th}$ stratum, $\hat{\sigma}^2_{\log(WR^{(m)})}$, we need to employ estimators for the elements in the mathematical formula. With estimates already available for the numerators of the three fractions from the previous step, it is adequate to assess $\hat{\theta}^{(m)}_{W_N}$ and $\hat{\theta}^{(m)}_{W_S}$. Under the null hypothesis, the specific calculation is as follows,

$$\hat{\theta}^{(m)}_{W_N} = \hat{\theta}^{(m)}_{W_S} = \frac{N^{(m)}_{W_N} + N^{(m)}_{W_S}}{2}$$

With these at our disposal, we can assert that both the numerator and denominator of the win ratio in equation (4.1) exhibit an asymptotically normal distribution,

$$\begin{aligned}\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_N} &\dot{\sim} \mathcal{N}\left(\sum_{m=1}^{M} w^{(m)}\theta^{(m)}_{W_N}, \operatorname{Var}\left(\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_N}\right)\right) \\ &= \mathcal{N}\left(\sum_{m=1}^{M} w^{(m)}\theta^{(m)}_{W_N}, \sum_{m=1}^{M} (w^{(m)})^2 \sigma^{2(m)}_{W_N}\right)\end{aligned}$$

and

$$\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_S} \dot{\sim} \mathcal{N}\left(\sum_{m=1}^{M} w^{(m)}\theta^{(m)}_{W_S}, \sum_{m=1}^{M} (w^{(m)})^2 \sigma^{2(m)}_{W_S}\right)$$

These two quantities exhibit covariance, a computation method for which can be

https://doi.org/10.26219/heal.aueb.6426



described as follows,

$$
\begin{aligned}
&\mathrm{Cov}\left(\sum_{m=1}^{M} w^{(m)} N_{W_N}^{(m)}, \sum_{m=1}^{M} w^{(m)} N_{W_S}^{(m)}\right) \\
&= \mathrm{Cov}(w^{(1)} N_{W_N}^{(1)}, w^{(1)} N_{W_S}^{(1)}) + \mathrm{Cov}(w^{(1)} N_{W_N}^{(1)}, w^{(2)} N_{W_S}^{(2)}) + \ldots + \mathrm{Cov}(w^{(1)} N_{W_N}^{(1)}, w^{(M)} N_{W_S}^{(M)}) \\
&+ \mathrm{Cov}(w^{(2)} N_{W_N}^{(2)}, w^{(1)} N_{W_S}^{(1)}) + \mathrm{Cov}(w^{(2)} N_{W_N}^{(2)}, w^{(2)} N_{W_S}^{(2)}) + \ldots + \mathrm{Cov}(w^{(2)} N_{W_N}^{(2)}, w^{(M)} N_{W_S}^{(M)}) \\
&+ \ldots \\
&+ \mathrm{Cov}(w^{(M)} N_{W_N}^{(M)}, w^{(1)} N_{W_S}^{(1)}) + \mathrm{Cov}(w^{(M)} N_{W_N}^{(M)}, w^{(2)} N_{W_S}^{(2)}) + \ldots + \mathrm{Cov}(w^{(M)} N_{W_N}^{(M)}, w^{(M)} N_{W_S}^{(M)}) \\
&= \mathrm{Cov}(w^{(1)} N_{W_N}^{(1)}, w^{(1)} N_{W_S}^{(1)}) + \mathrm{Cov}(w^{(2)} N_{W_N}^{(2)}, w^{(2)} N_{W_S}^{(2)}) + \mathrm{Cov}(w^{(3)} N_{W_N}^{(3)}, w^{(3)} N_{W_S}^{(3)}) \\
&+ \ldots + \mathrm{Cov}(w^{(M)} N_{W_N}^{(M)}, w^{(M)} N_{W_S}^{(M)})
\end{aligned}
$$

We assume that the strata are mutually independent. This implies that, for any two distinct strata, denoted as $m$ and $n$ (where $m, n = 1, 2, \ldots, M$ and $m \neq n$), the count of wins for the new treatment in stratum $m$, $N_{W_N}^{(m)}$, is independent of the count of wins for the new treatment in stratum $n$, $N_{W_N}^{(n)}$. Similarly, the count of wins for the standard treatment in stratum $m$, $N_{W_S}^{(m)}$, is independent of the count of wins for the standard treatment in stratum $n$, $N_{W_S}^{(n)}$. Additionally, the count of wins for the new treatment in stratum $m$, $N_{W_N}^{(m)}$, is independent of the count of wins for the standard treatment in stratum $n$, $N_{W_S}^{(n)}$. Hence, we deduce that the covariance between the numerator and denominator of the win ratio is equivalent to,

$$
\begin{aligned}
\mathrm{Cov}\left(\sum_{m=1}^{M} w^{(m)} N_{W_N}^{(m)}, \sum_{m=1}^{M} w^{(m)} N_{W_S}^{(m)}\right) &= \sum_{m=1}^{M} (w^{(m)})^2 \mathrm{Cov}(N_{W_N}^{(m)}, N_{W_S}^{(m)}) \\
&= \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)}
\end{aligned}
$$

Therefore, the joint distribution of the numerator and denominator of the win ratio is established in the following manner,

$$
\begin{pmatrix} \sum_{m=1}^{M} w^{(m)} N_{W_N}^{(m)} \\ \sum_{m=1}^{M} w^{(m)} N_{W_S}^{(m)} \end{pmatrix} \dot{\sim} \mathcal{N}\left( \begin{bmatrix} \sum_{m=1}^{M} w^{(m)} \theta_{W_N}^{(m)} \\ \sum_{m=1}^{M} w^{(m)} \theta_{W_S}^{(m)} \end{bmatrix}, \begin{bmatrix} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N}^{2(m)} & \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)} \\ \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)} & \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_S}^{2(m)} \end{bmatrix} \right)
$$

Now, we implement the delta method to derive the joint distribution of the loga-

https://doi.org/10.26219/heal.aueb.6426



rithms of the aforementioned quantities.

$$\begin{pmatrix} \log\left(\sum_{m=1}^{M} w^{(m)} N_{W_N}^{(m)}\right) \\ \log\left(\sum_{m=1}^{M} w^{(m)} N_{W_S}^{(m)}\right) \end{pmatrix}$$
$$\dot{\sim} \mathcal{N}\left(\begin{bmatrix} \log\left(\sum_{m=1}^{M} w^{(m)} \theta_{W_N}^{(m)}\right) \\ \log\left(\sum_{m=1}^{M} w^{(m)} \theta_{W_S}^{(m)}\right) \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} \frac{1}{(\lambda_{W_N})^2} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N}^{2(m)} & \frac{1}{\lambda_{W_N} \cdot \lambda_{W_S}} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)} \\ \frac{1}{\lambda_{W_N} \cdot \lambda_{W_S}} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)} & \frac{1}{(\lambda_{W_S})^2} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_S}^{2(m)} \end{bmatrix} \right)$$

It is noted that $\lambda_{W_N}$ and $\lambda_{W_S}$ are specified as,

$$\lambda_{W_N} = \sum_{m=1}^{M} w^{(m)} \theta_{W_N}^{(m)} \quad \text{and} \quad \lambda_{W_S} = \sum_{m=1}^{M} w^{(m)} \theta_{W_S}^{(m)}$$

All of these factors collectively contribute to establishing the distribution of the logarithm of the stratified win ratio, ultimately resulting in an asymptotically normal distribution.

$$\log(\mathrm{WR}_{\mathrm{stratified}}) \dot{\sim} \mathcal{N}(\theta_{\log(\mathrm{WR}_{\mathrm{stratified}})}, \sigma^2_{\log(\mathrm{WR}_{\mathrm{stratified}})})$$

where,

$$\theta_{\log(\mathrm{WR}_{\mathrm{stratified}})} = \log\left(\frac{\sum_{m=1}^{M} w^{(m)} \theta_{W_N}^{(m)}}{\sum_{m=1}^{M} w^{(m)} \theta_{W_S}^{(m)}}\right)$$

and

$$\sigma^2_{\log(\mathrm{WR}_{\mathrm{stratified}})} = \frac{1}{(\lambda_{W_N})^2} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N}^{2(m)} + \frac{1}{(\lambda_{W_S})^2} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_S}^{2(m)} - \frac{2}{\lambda_{W_N} \lambda_{W_S}} \sum_{m=1}^{M} (w^{(m)})^2 \sigma_{W_N W_S}^{(m)}$$

Additionally, we provide the variances of the logarithm of the stratified win ratio for two frequently used weight values $w^{(m)}$. Specifically, when $w^{(m)} = \frac{1}{N^{(m)}}$, the expression is given by,

$$\sigma^2_{\log(\mathrm{WR}_{\mathrm{stratified}})} = \frac{1}{(\lambda_{W_N})^2} \sum_{m=1}^{M} \frac{\sigma_{W_N}^{2(m)}}{(N^{(m)})^2} + \frac{1}{(\lambda_{W_S})^2} \sum_{m=1}^{M} \frac{\sigma_{W_S}^{2(m)}}{(N^{(m)})^2} - \frac{2}{\lambda_{W_N} \lambda_{W_S}} \sum_{m=1}^{M} \frac{\sigma_{W_N W_S}^{(m)}}{(N^{(m)})^2}$$

While for $w^{(m)} = 1$, the expression simplifies to,

https://doi.org/10.26219/heal.aueb.6426



$$\sigma^2_{\log(\mathrm{WR}_{\mathrm{stratified}})} = \frac{1}{(\lambda_{W_N})^2}\sum_{m=1}^{M}\sigma^{2(m)}_{W_N} + \frac{1}{(\lambda_{W_S})^2}\sum_{m=1}^{M}\sigma^{2(m)}_{W_S} - \frac{2}{\lambda_{W_N}\lambda_{W_S}}\sum_{m=1}^{M}\sigma^{(m)}_{W_N W_S}$$

It is essential to note that the estimation of the logarithm of the stratified win ratio is determined as follows,

$$\hat{\theta}_{\log(\mathrm{WR}_{\mathrm{stratified}})} = \log\left(\frac{\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_N}}{\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_S}}\right)$$

Simultaneously, for evaluating the variance of the logarithm of the stratified win ratio, denoted as $\hat{\sigma}^2_{\log(\mathrm{WR}_{\mathrm{stratified}})}$, we employ the values $\hat{\sigma}^{2(m)}_{W_N}$, $\hat{\sigma}^{2(m)}_{W_S}$ and $\hat{\sigma}^{(m)}_{W_N W_S}$ as computed using the approach proposed by Dong et al. (2016). Additionally, we utilize the estimated values of $\lambda_{W_N}$ and $\lambda_{W_S}$, obtained as follows,

$$\hat{\lambda}_{W_N} = \hat{\lambda}_{W_S} = \frac{1}{2}\left(\sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_N} + \sum_{m=1}^{M} w^{(m)} N^{(m)}_{W_S}\right)$$

It is emphasized that these estimates are derived under the null hypothesis, assuming no statistically significant difference between the two treatment groups (Dong et al. (2018)).

## 4.3 Examination about homogeneity

A pivotal concern captivating the attention of researchers involves the investigation of the homogeneity of the win ratio across diverse strata. This crucial issue holds significant relevance in current research, where a rigorous examination is undertaken to ascertain whether there exists uniformity in the win ratio patterns among various strata. Specifically, this examination hangs on to the principles outlined by Cochran (1954), employing the inverse-variance weighted approach. The foundation of this methodology lies in a key statistic, pivotal for determining the presence or absence of homogeneity in the win ratio among strata. The said statistic is defined as follows,

$$Q = \sum_{m=1}^{M} w^{(m)}_{\mathrm{IV}}[\log(WR^{(m)}) - \log(WR_{\mathrm{IV}})]^2$$

Here, $\log(WR_{\mathrm{IV}})$ represents the inverse-variance weighted (log) win ratio, calculated as,

$$\log(WR_{\mathrm{IV}}) = \frac{\sum_{m=1}^{M} w^{(m)}_{\mathrm{IV}} \log(WR^{(m)})}{\sum_{m=1}^{M} w^{(m)}_{\mathrm{IV}}}$$

Additionally,

$$w^{(m)}_{\mathrm{IV}} = \frac{1}{\sigma^2_{\log(WR^{(m)})}}$$

https://doi.org/10.26219/heal.aueb.6426



denotes the inverse-variance weight assigned to the $m^{\text{th}}$ stratum. Moreover, it is noteworthy to highlight that the particular statistic exhibits a specific distribution. Especially, when operating under the null hypothesis, considering equality of win ratios across all strata, and benefiting from large sample sizes within each stratum, we can assert that the $Q$ statistic follows, in the asymptotic sense, a chi-squared ($\chi^2$) distribution with $M-1$ degrees of freedom.

$$Q \dot{\sim} \chi^2_{M-1}$$

Despite its capability to assess the homogeneity of win ratios across strata, this test is not notably distinguished for its statistical power; it tends to exhibit low power. Nevertheless, even in cases where the null hypothesis can be rejected, indicating some dissimilarity in the win ratios between strata, the stratified win ratio remains a vital summary measure for the stratified data. This inherent quality also holds true for the Mantel-Haenszel stratified odds ratio. Conclusively, although the homogeneity test offers valuable insights into the consistency of win ratios across strata, researchers should exercise caution in its application (Dong et al. (2018)).

## 4.4 Calculation of 95% CI for the stratified win ratio

In the computation of the 95% confidence interval for the stratified win ratio, Finkelstein and Schoenfeld (1999) introduced an extension of their method adjusted to the stratified environment. Consider the scenario where individuals are divided into $m$ strata ($m = 1, 2, \ldots$) and $A_m$ represents the set of indices of $N^{(m)}$ patients in the $m^{\text{th}}$ stratum. Within each stratum, the variable $U_i$ is computed for each patient, i.e., for $i \in A_m$ holds that $U_i = \sum_{j \in A_m} u_{ij}$. Let's recollect that $U_{ij}$ takes values of 1, $-1$, and 0, corresponding to the patient $i$ being the winner, the loser, or encountering a tie, respectively. Recalling also that $D_i$ is a binary variable (1 if the patient comes from the new treatment group, 0 if the individual originates from the conservative treatment group), the test statistic $T$ is expressed as,

$$T = \sum_m \sum_{i \in A_m} D_i U_i$$

In the context of calculating the mean and variance of this specific statistic, our analysis relies on fixed quantities denoted as $U_1, U_2, \ldots$ and random variables denoted as $D_1, D_2, \ldots$. We consider that the number of patients in the $m^{th}$ stratum from the new treatment group is denoted as $n^{(m)}$. Thus, in the $m^{th}$ stratum, we establish the following relationships,

$$E(D_i) = \frac{n^{(m)}}{N^{(m)}}$$

and

$$\text{Cov}(D_i, D_j) = \frac{n^{(m)}(N^{(m)} - n^{(m)})}{(N^{(m)} - 1)(N^{(m)})^2}(\delta_{ij} N^{(m)} - 1)$$

https://doi.org/10.26219/heal.aueb.6426



where $\delta_{ij}$ is the Kronecker delta defined as,

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Given that $\sum_{i \in A_m} U_i = 0$, we deduce that the mean value of the $T$ statistic is zero. Furthermore, the variance is expressed as,

$$\text{Var} = \sum_m \frac{n^{(m)}(N^{(m)} - n^{(m)})}{N^{(m)}(N^{(m)} - 1)} \left( \sum_{i \in A_m} U_i^2 \right)$$

With the variance of the $T$ statistic at our disposal, we proceeded to compute the standardized normal deviate, denoted as $z$. This crucial quantity is defined as,

$$z = \frac{T}{\sqrt{\text{Var}}}$$

and assumes a pivotal role in determining the standard error ($se$) of the logarithm of the win ratio. This relationship is expressed by,

$$se = \frac{\log(\text{WR}_{\text{stratified}})}{z}$$

Simultaneously, we have calculated the logarithm of the win ratio, denoted as $\log(\text{WR}_{\text{stratified}})$. Consequently, a 95% confidence interval for $\log(\text{WR}_{\text{stratified}})$ is formulated as,

$$\log(\text{WR}_{\text{stratified}}) \pm 1.96 \times se$$

Upon exponentiation, this yields the 95% confidence interval for the win ratio,

$$\exp\left(\log(\text{WR}_{\text{stratified}}) \pm 1.96 \times se\right)$$

Abdalla et al. (2016) carried a separate survey out on the stratified win ratio, inventing a method to set out a 95% confidence interval for this ratio. In their research, participants were randomly categorized into two groups—namely, the treatment and placebo groups—with the assumption of $M$ strata. Utilizing the win ratio procedure, the researchers computed the number of winners and losers within each stratum. These counts were added across all strata, contributing to the determination of the overall stratified win ratio, stated as the total number of winners divided by the total number of losers. The specific formula employed for this calculation was,

$$\text{WR}_{\text{stratified}} = \frac{\sum_{m=1}^{M} N_W^{(m)}}{\sum_{m=1}^{M} N_L^{(m)}}$$

Meanwhile, the investigators analyzed the notion of win difference, considering that it can be computed as,

$$\text{WD}_{\text{stratified}} = \sum_{m=1}^{M} N_W^{(m)} - \sum_{m=1}^{M} N_L^{(m)}$$

https://doi.org/10.26219/heal.aueb.6426



With the objective of specifying the variance of the win ratio, the researchers leveraged the variance of the win difference. Initially, they considered a stratum $m$, where $m = 1, 2, \ldots, M$, involving a sample size $N^{(m)} = N_N^{(m)} + N_S^{(m)}$. Here, $N_N^{(m)}$ denoted the number of patients in the treatment group and $N_S^{(m)}$ indicated the number of individuals in the placebo group within the given stratum. They continued in pairwise comparisons, evaluating each patient $i$ from one group with each patient $j$ from the other group. This procedure contributed to the generation of the $U$-statistic,

$$U_i^{(m)} = \sum_{i \neq j} u_{ij}^{(m)}$$

where

$$u_{ij}^{(m)} = \begin{cases} -1, & \text{if patient } i \text{ experiences earlier the considered outcome} \\ & \text{in relation to patient } j \\ +1, & \text{if individual } i \text{ does not experience earlier the assumed outcome} \\ & \text{in relation to patient } j \\ 0, & \text{if the outcome is undetermined} \end{cases}$$

Hence, the formula employed to calculate the variance within each stratum $m$ is expressed as,

$$\text{Var}^{(m)} = \frac{N_N^{(m)} N_S^{(m)}}{N^{(m)}(N^{(m)} - 1)} \sum_{i=1}^{N^{(m)}} (U_i^{(m)})^2$$

Aggregating the variances across the $M$ strata arises the variance of the win difference, which is calculated as the sum of the variances within each stratum,

$$\text{Var(WD)} = \sum_{m=1}^{M} \text{Var}^{(m)}$$

Additionally, the researchers instituted a relation linking the win ratio (WR) to the win difference (WD). Particularly, the win ratio can be obtained utilizing the win difference through the formulation,

$$\text{WR} = \frac{1}{\sum_{m=1}^{M} N_L^{(m)}} \cdot \text{WD} + 1$$

https://doi.org/10.26219/heal.aueb.6426



Afterward, the delta method was employed to verify the variance of the win ratio (Var(WR)), and its calculation is depicted as follows,

$$\begin{aligned}\text{Var(WR)} &= \left(\frac{\Delta\text{WR}}{\Delta\text{WD}}\right)^2 \cdot \text{Var(WD)} \\ &= \left(\frac{1}{\sum_{m=1}^{M} N_L^{(m)}}\right)^2 \cdot \text{Var(WD)} \\ &= \frac{\text{Var(WD)}}{\left(\sum_{m=1}^{M} N_L^{(m)}\right)^2}\end{aligned}$$

Consequently, armed with this available knowledge, one can simply compute the 95% CI for the win ratio by implementing the formula,

$$\text{WR} \pm 1.96 \cdot \sqrt{\text{Var(WR)}}$$

Alternatively, Yu and Ganju (2022) expanded their approach for assessing the variance of the win ratio, extending it to account for the stratified win ratio. Thus, they contributed to the computation of the 95% confidence interval for the stratified win ratio. In their approach, the stratified win ratio considers the total number of winners $N_W$ across all strata ($m = 1, 2, \ldots, M$) to the total number of losers $N_L$ across the same strata, incorporating corresponding weights $w^{(m)}$. The formulation for the stratified win ratio is expressed as follows,

$$\text{WR}_{\text{stratified}} = \frac{\sum_{m=1}^{M} w^{(m)} N_W^{(m)}}{\sum_{m=1}^{M} w^{(m)} N_L^{(m)}}$$

Here, $M$ represents the total number of strata, while $N_W^{(m)}$ and $N_L^{(m)}$ denote the number of winners and losers in the $m^{th}$ stratum, respectively. The weights $w^{(m)}$ are taken into account for each stratum in the calculation. Utilizing the delta method and implementing the natural logarithm transformation, we can express the logarithm of the sum of winners and losers within strata as follows,

$$\begin{aligned}\log\left(\sum_{m=1}^{M} N_W^{(m)}\right) &\approx \log(\mu_W) + \frac{\sum_{m=1}^{M} w^{(m)} N_W^{(m)} - \mu_W}{\mu_W} \\ \log\left(\sum_{m=1}^{M} N_L^{(m)}\right) &\approx \log(\mu_L) + \frac{\sum_{m=1}^{M} w^{(m)} N_L^{(m)} - \mu_L}{\mu_L}\end{aligned}$$

where,

$$\begin{aligned}\mu_W &= E\left(\sum_{m=1}^{M} w^{(m)} N_W^{(m)}\right) \\ \mu_L &= E\left(\sum_{m=1}^{M} w^{(m)} N_L^{(m)}\right)\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



Under the null hypothesis asserting no statistically significant difference between the two groups ($\mu_W = \mu_L = \mu$), it follows that,

$$\log(\mathrm{WR}_{\text{stratified}}) \approx \frac{\sum_{m=1}^{M} w^{(m)}(N_W^{(m)} - N_L^{(m)})}{\mu}.$$

Consequently,

$$\begin{aligned}\mathrm{Var}(\log(\mathrm{WR}_{\text{stratified}})) &\approx \frac{1}{\mu^2}\mathrm{Var}\left(\sum_{m=1}^{M} w^{(m)}(N_W^{(m)} - N_L^{(m)})\right) \\ &= \frac{1}{\mu^2}\sum_{m=1}^{M}(w^{(m)})^2\mathrm{Var}(N_W^{(m)} - N_L^{(m)}).\end{aligned}$$

This approximation holds as we consider the independence of $(N_W^{(m)} - N_L^{(m)})$ and $(N_W^{(n)} - N_L^{(n)})$ for two distinct strata $m \neq n$. By extending the same argument as in equation (2.2) to each stratum, assuming homogeneous tie proportions across strata, we obtain,

$$\mathrm{Var}(\log(\mathrm{WR}_{\text{stratified}})) \approx \frac{1}{\mu^2}\sum_{m=1}^{M}\frac{(w^{(m)})^2 k(1-k)(N^{(m)})^3(1-P_{\text{tied}}^2)}{3}$$

where $k$ represents the proportion of individuals allocated to one group, $(1-k)$ denotes the proportion allocated to the other group, $N^{(m)}$ is the number of patients within the $m^{\text{th}}$ stratum ($m = 1, 2, \ldots, M$), and $P_{\text{tied}}$ constitutes the proportion of ties. Additionally, we utilize the information that,

$$\mu = \frac{1}{2}\sum_{m=1}^{M} w^{(m)} k(1-k)(N^{(m)})^2(1-P_{\text{tied}})$$

Hence, the variance of the natural logarithm of the stratified win ratio, $\mathrm{Var}(\log(\mathrm{WR}_{\text{stratified}}))$, can be expressed as,

$$\mathrm{Var}(\log(\mathrm{WR}_{\text{stratified}})) \approx \frac{4(1+P_{\text{tied}})}{3k(1-k)(1-P_{\text{tied}})} \cdot \frac{\sum_{m=1}^{M}(w^{(m)})^2(N^{(m)})^3}{\left(\sum_{m=1}^{M} w^{(m)}(N^{(m)})^2\right)^2}$$

In the scenario where all weights are equal to 1 ($w^{(m)} = 1$, $m = 1, 2, \ldots, M$), the variance of the natural logarithm of the stratified win ratio simplifies as,

$$\mathrm{Var}(\log(\mathrm{WR}_{\text{stratified}})) \approx \frac{4(1+P_{\text{tied}})}{3k(1-k)(1-P_{\text{tied}})} \cdot \frac{\sum_{m=1}^{M}(N^{(m)})^3}{\left(\sum_{m=1}^{M}(N^{(m)})^2\right)^2}$$

This expression represents the fluctuation of the logarithm of the unstratified win ratio. Armed with these insights, the construction of a 95% confidence interval for the logarithm of the stratified win ratio can be formulated as,

$$\exp\left(\log(\mathrm{WR}_{\text{stratified}}) \pm 1.96\sqrt{\mathrm{Var}}\right)$$

https://doi.org/10.26219/heal.aueb.6426



## 4.5 Statistical power within the stratified win ratio

Finkelstein and Schoenfeld (1999) extended their methodology to compute a 95% confidence interval for the logarithm of the stratified win ratio (section 4.4). This expansion enabled them to draw pieces of knowledge into the statistical power of hypothesis testing. They promoted that to assess the test's power, it is substantial to determine the probability that a patient from the new treatment group ($D_i = 1$) exhibits better outcomes than a patient from the standard treatment group ($D_i = 0$) within the minimum follow-up time. This probability, denoted as $p_m$ for subjects in strata $m$, avails as a key factor in calculating the mean value of the distribution of the utilized statistic under the alternative hypothesis. This mean, in turn, is decisive in determining the power of the statistical test. In situations where ties are absent, the researchers established that the mean of the numerator of the statistic is computed as,

$$\sum_m 2n^{(m)}(N^{(m)} - n^{(m)}) \left(p_m - \frac{1}{2}\right)$$

Additionally, the mean of the denominator is approximately determined,

$$\left\{\sum_m \frac{n^{(m)}(N^{(m)} - n^{(m)})N^{(m)}}{3}\right\}^{\frac{1}{2}}$$

Following this, they recommended the configuration of a model depicting both mortality and longitudinal measures. Emphasizing the crucial role of model selection, they underlined that this decision critically influences the subsequent analysis. Ultimately, they successfully computed the probability $p_m$ and derived the statistical power of the hypothesis test, employing either analytical methods or simulations as considered appropriate.

## 4.6 The approach of stratified win ratio concerning varied weights

Within the realm of clinical trials, diverse outcomes are considered, depending upon the specific characteristics of the disease under investigation. Typically, variations in the severity of outcomes, fluctuating from gentler exhibitions to those with potentially fatal consequences, require the attributing of different weights in the analysis and hypothesis testing procedure. Ramchandani et al. (2016) proposed an approach, outlined in section 3.3, resulting in the determining of these weights for each distinct outcome. Alongside the consideration of varying weights for individual outcomes, the scenario of disparate weights within each stratum often emerges when the data are classified for stratification. Dealing with this particular issue, the same researchers expressed their aspect on adaptive weighting in 2016. In the framework of

https://doi.org/10.26219/heal.aueb.6426



a certain stratum, the investigators declared the determination of optimal weights by specifying the parameters $\theta$ and $\Lambda$. These parameters, denoting the expected treatment effect and the covariance matrix, respectively, played a decisive role in accomplishing power maximization. The a priori delineation of these parameters would enable the consequent solution for the optimal weights represented by '$w$'. Nevertheless, in real-world scenarios, the advanced setting up of these parameters is often complicated, establishing a potential obstacle to obtaining accurate inferences. Identifying this challenge, the researchers introduced a practical and flexible approach, particularly applicable in the presence of strata. Their methodology relied on availing data from preceding strata to evaluate and derive weights for upcoming strata, providing a solution to the intricacy related to parameter definition.

Initially, they assumed $p$ potential outcomes and coordinated participants into $S$ strata, ordering them as $1, 2, \ldots, S$. This layout could be relied on the consecutive admission of individuals in the survey or specified randomly. The notation $U_{sk}$ represented the $U$-statistic for the $k^{\text{th}}$ outcome in the $s^{\text{th}}$ stratum. Their methodology comprised the calculation of the $U$-statistic in the primary stratum, $U_{1k}$, where $k$ ranged from 1 to $p$. This computation derived an estimate $\theta_k = E[U_{sk}]$ for the following strata, with $U_1 = (U_{11}, \ldots, U_{1p})'$ employing as an estimator for $\theta = (\theta_1, \ldots, \theta_p)'$. Afterward the calculation of $U_1$, the researchers determined the covariance matrix $\hat{\Lambda}_1$ for the first stratum. Utilizing these results, they then sought optimal weights in the second stratum by solving the equation $w_2 = \hat{\Lambda}_1^{-1} U_1$. To avoid undesirable negative results, constraints were inflicted during the power maximization procedure. It is worthwhile to underscore that the aggregate of weights within a stratum equaled 1, expressed as $\sum_{k=1}^{p} w_{2k} = 1$. Thus, for the second stratum, the statistic is encompassed by a numerator $w_2' U_2$ and variance $\sigma_2^2 = w_2' \Lambda_2 w_2$.

To evolve through upcoming strata, a weighted average approach was utilized for the required parameters. This comprised availing knowledge acquired from the study of preceding strata, with weights defined by the total pairwise comparisons within each stratum. Hence, the parameter estimates for $\theta$ and $\Lambda$ in the $s^{\text{th}}$ stratum can be expressed as follows,

$$\hat{\theta}_s = \frac{1}{\sum_{j=1}^{s-1} n_j m_j} \sum_{j=1}^{s-1} n_j m_j U_j$$

and

$$\hat{\Sigma}_s = \frac{1}{\sum_{j=1}^{s-1} n_j m_j} \sum_{j=1}^{s-1} n_j m_j \hat{\Lambda}_j$$

respectively. Here, $n$ represents the number of individuals in the new treatment group, while $m$ corresponds to the total of subjects in the standard treatment group.

Consequently, the definition of optimal weights for the $s^{\text{th}}$ stratum comprises the relationship,

$$w_s = \hat{\Sigma}_s^{-1} \hat{\theta}_s$$

The utilization of $\hat{\Sigma}$ instead of $\hat{\Lambda}$ is carried out to prevent frustration between the within-stratum covariance estimates ($\hat{\Lambda}$) and the average of those covariances across

https://doi.org/10.26219/heal.aueb.6426



strata ($\hat{\Sigma}$). Ultimately, it is noteworthy to refer that opting for the proper statistic relies on the hypothesis testing conducted, safeguarding that the procedure of exploring optimal weights can progress within the scenario of the stratified data. Further details about the formulations employed in this methodology can be derived in sections 3.1 and 3.3 (Ramchandani et al. (2016)).

## 4.7 Various stratified win ratio categories

Mao et al. (2022) introduced a comprehensive formulation of the win ratio for the assessing of recurrent events. This expression generates four distinct variations of the win ratio, as outlined in the precedent chapter in section 3.7. The researchers expanded their survey to involve the instance of the existence of strata, resulting in a general form of the stratified win ratio, which in turn will produce the four varieties of the stratified win ratio. This perceptive investigation underlines the flexibility of their approach in handling various complexities in the estimating of recurrent events, illuminating diversified perspectives within this research domain.

In particular, the researchers considered the availability of $M$ strata. For the $m^{\text{th}}$ stratum ($m = 1, 2, \ldots, M$), they articulated the observed data as follows,

$$\{Y_{a1}^{(m)}(X_{a1}^{(m)}), X_{a1}^{(m)}\}, \ldots, \{Y_{an_{a^{(m)}}}^{(m)}(X_{an_{a^{(m)}}}^{(m)}), X_{an_{a^{(m)}}}^{(m)}\}$$

Here, the parameter $\alpha$ receives the values 0 and 1, indicating the control group and the treatment group, respectively. Additionally, a specific stratum $m$ encompasses $N^{(m)} = n_1^{(m)} + n_0^{(m)}$ individuals. Hence, the total number of patients in each group can be calculated through the equation,

$$n_a = \sum_{m=1}^{M} n_a^{(m)}$$

contributing to an overall count of survey participants $N = n_1 + n_0$.

By opting for a particular win function $\mathcal{W}$, characterized by the three properties examined in section 3.7, they demonstrated the formulation for the general structure of the stratified win ratio, expressed as,

$$\hat{\mathcal{E}}_{n_1,n_0}^{(M)}(\mathcal{W}) = \frac{\sum_{m=1}^{M} \omega^{(m)} (n_1^{(m)} n_0^{(m)})^{-1} \sum_{i=1}^{n_1^{(m)}} \sum_{j=1}^{n_0^{(m)}} \mathcal{W}(Y_{1i}^{(m)}, Y_{0j}^{(m)})(X_{1i}^{(m)} \wedge X_{0j}^{(m)})}{\sum_{m=1}^{M} \omega^{(m)} (n_1^{(m)} n_0^{(m)})^{-1} \sum_{i=1}^{n_1^{(m)}} \sum_{j=1}^{n_0^{(m)}} \mathcal{W}(Y_{0j}^{(m)}, Y_{1i}^{(m)})(X_{1i}^{(m)} \wedge X_{0j}^{(m)})}$$

In the framework of the provided information, the $\omega^{(m)}$ is utilized to represent the diverse weights related to each stratum. Typically, the numerator denotes the count of wins, while the denominator corresponds to the count of losses. This formulation captures the relative difference between the treatment group and the control group while accounting for the specified conditions. Subsequently, employing various functions results in the generation of four distinct stratified win ratios. The process reflects that of a scenario involving a single stratum, denoted by $M = 1$ (Section 3.7) (Mao et al. (2022)).

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# Chapter 5

# Win Ratio for Clustered Data

## 5.1 Definition of the win ratio for cluster randomized data

Clinical trial analyses often leverage the widely recognized approach of utilizing and implementing clusters. This method involves a randomized procedure, ensuring that the results obtained after the research are not compromised. The prevalence of cluster-based approaches has increased in recent years, primarily due to their associated benefits. One crucial advantage lies in their capability to moderate experimental contamination that may occur during the investigation of a specific subject. Additionally, the use of random clusters tends to yield more precise conclusions, enhancing the overall effectiveness of the research and increasing the likelihood of obedience to standards. In this context, the data employed to generate random clusters is referred to as semi-competing risk data. This dataset, employed throughout our study, is distinguished by the characteristic that a terminal-fatal event can censor a non-terminal-fatal event, but the reverse scenario is not possible. Expanding on the win ratio method, we apply it to amalgamate this data, resulting in the widely recognized win ratio for cluster-randomized semi-competing risks data. This ratio relies on bivariate clustered U-statistics. It's crucial to highlight that the validity of the research is contingent upon acknowledging the correlation among subjects within a cluster.

Initially, as highlighted repeatedly, the research involves categorizing patients into two distinct groups. The first group, known as the treatment group, comprises individuals receiving the new treatment, while the second group, referred to as the control group, encompasses those delivered a placebo or the conventional treatment. The variable $z$ denotes the patient's group belonging, with $z = 1$ indicating membership in the treatment group and $z = 0$ signifying inclusion in the control group. Furthermore, a composite endpoint with two events, namely a fatal event, and a non-fatal event, is considered. The prioritization of the fatal event is based on its medical significance. The time to reach the composite endpoint is defined as $\min\{T_{N-F_0}, T_{F_0}, T_{C_0}\}$ for subjects in the control group and $\min\{T_{N-F_1}, T_{F_1}, T_{C_1}\}$ for those in the treatment group. Here, $T_{N-F}$, $T_F$, and $T_C$ represent the time to

https://doi.org/10.26219/heal.aueb.6426



non-fatal event, fatal event, and censoring, respectively. It is important to note that a patient's status is considered censored when they may be lost during the follow-up before experiencing another event, i.e., $T_{C_z} < \min\{T_{N-F_z}, T_{F_z}\}$. Another crucial aspect that requires consideration is addressing the cluster effect, entailing the introduction of the variable $\gamma$. This variable is further outlined as $\gamma_1$ for individuals in the treatment group and $\gamma_0$ for those in the control group, and it is referred to as cluster-specific frailty. Our hypothesis posits that patients within the same cluster may exhibit greater similarities than those from distinct clusters. Furthermore, in the context of randomly generated clusters with independent comparison groups, we assume that $\gamma_1$ and $\gamma_0$ are independent of each other and share identical distributions. It is crucial to take into account two additional assumptions concerning the distributions of individuals involved in the study. Firstly, individuals within the same cluster are considered exchangeable, implying that they can be reallocated from one comparison group to another while conserving their distribution, as long as they remain within the same cluster. Secondly, the marginal distributions of participants within clusters are identical within the same comparison group, although they may vary in the other comparison group. Simultaneously, we establish the following distribution functions for individuals in group $z$:

i) Survival function of the time to a fatal event:

$$F_z(y_2|\gamma_z) = P(T_{F_z} \geq y_2|\gamma_z)$$

where $y_2$ represents the realization of the time-to-fatal event variable $T_{F_z}$.

ii) Joint survival function of the time to a fatal event and the time to a non-fatal event:

$$G_z(y_1, y_2|\gamma_z) = P(T_{N-F_z} \geq y_1, T_{F_z} \geq y_2|\gamma_z)$$

where $y_1$ denotes the realization of the time-to-non-fatal event variable $T_{N-F_z}$.

iii) Hazard function of the time to a fatal event:

$$\lambda_{F_z}(y_2|\gamma_z) = P(T_{F_z} = y_2|T_{F_z} \geq y_2, \gamma_z)$$

iv) Hazard function of the time to a non-fatal event conditional on the time to a fatal event:

$$\lambda_{N-F|F_z}(y_1|y_2, \gamma_z) = P(T_{N-F_z} = y_1|T_{N-F_z} \geq y_1, T_{F_z} \geq y_2, \gamma_z)$$

v) Survival function of the censoring distribution:

$$Q_z(y_2) = P(T_{C_z} \geq y_2)$$

vi) Hazard function of the censoring distribution:

$$\lambda_{C_z}(y_2) = P(T_{C_z} = y_2|T_{C_z} \geq y_2)$$

https://doi.org/10.26219/heal.aueb.6426



It is important to note that the censoring approach is presumed to be independent of the cluster effect, as well as the time to fatal and non-fatal events.

With these components at our disposal, the definition of the clustered win ratio is established as follows,

$$\mathrm{WR}_{\text{clustered}} = \frac{E(E(N_W|\gamma_0,\gamma_1))}{E(E(N_L|\gamma_0,\gamma_1))}$$

Here, $N_W$ represents the number of winners, and $N_L$ denotes the number of losers. This relationship is further extended using the methodology proposed by Luo et al. (2015), as detailed in section 2.3.5.

$$\mathrm{WR}_{\text{clustered}} = \frac{\iint \{A+B\}\, r(\gamma_0)r(\gamma_1)\, d\gamma_0\, d\gamma_1}{\iint \{C+D\}\, r(\gamma_0)r(\gamma_1)\, d\gamma_0\, d\gamma_1}$$

It is noted that $r(\cdot)$ represents the probability density function of the common frailty distribution for the placebo and treatment groups, and the terms $A$, $B$, $C$, and $D$ are determined as follows,

$$A = \int F_1(y_2|\gamma_1)Q_1(y_2)F_0(y_2|\gamma_0)Q_0(y_2)\lambda_{F_0}(y_2|\gamma_0)\, dy_2$$

$$B = \iint_{y_1\le y_2} G_1(y_1,y_2|\gamma_1)Q_1(y_2)G_0(y_1,y_2|\gamma_0)Q_0(y_2)\lambda_{N-F|F_0}(y_1|y_2,\gamma_0) \times \{\lambda_{C_0}(y_2)+\lambda_{C_1}(y_2)\}\, dy_1\, dy_2$$

$$C = \int F_1(y_2|\gamma_1)Q_1(y_2)F_0(y_2|\gamma_0)Q_0(y_2)\lambda_{F_1}(y_2|\gamma_1)\, dy_2$$

$$D = \iint_{y_1\le y_2} G_1(y_1,y_2|\gamma_1)Q_1(y_2)G_0(y_1,y_2|\gamma_0)Q_0(y_2)\lambda_{N-F|F_1}(y_1|y_2,\gamma_1) \times \{\lambda_{C_0}(y_2)+\lambda_{C_1}(y_2)\}\, dy_1\, dy_2$$

In essence, the clustered win ratio is influenced by the distributions of the fatal event, joint distribution of the fatal and non-fatal event, censoring distributions, and the distribution of the cluster effect (Zhang and Jeong (2021)).

## 5.2 Distribution of win ratio in a clustered setting

Throughout our investigation, we have frequently referred to the notion of $U$-statistics, a concept that finds application in scenarios involving clustered data. In essence, under the assumption of cluster independence and treating patients as independent within a given cluster, we aggregate the comparison outcomes for each

https://doi.org/10.26219/heal.aueb.6426



pair of treatment and control clusters. Subsequently, we compute the average across all cluster pairs to formulate the statistic at the cluster level.

At the outset, in the treatment group, $m$ clusters are formed, while in the control group, $n$ clusters are created. To denote the outcomes within these clusters, we introduce variables $X_i^{(k)}$ and $Y_j^{(l)}$, representing the results of subjects $i^{\text{th}}$ and $j^{\text{th}}$ in the $k^{\text{th}}$ treatment and $l^{\text{th}}$ control clusters, respectively. Here, $i = 1, 2, \ldots, I_k$ and $j = 1, 2, \ldots, J_l$. Additionally, we define two kernel functions, namely $\phi_1(X_i^{(k)}, Y_j^{(l)})$ and $\phi_2(X_i^{(k)}, Y_j^{(l)})$, formulated as follows,

$$\begin{aligned}\phi_1(X_i^{(k)}, Y_j^{(l)}) &= I\{X_i^{(k)} > Y_j^{(l)}\} \\ &= \begin{cases} 1, & \text{if } X_i^{(k)} \text{ wins} \\ 0, & \text{otherwise} \end{cases}\end{aligned} \tag{5.1}$$

$$\begin{aligned}\phi_2(X_i^{(k)}, Y_j^{(l)}) &= I\{X_i^{(k)} < Y_j^{(l)}\} \\ &= \begin{cases} 1, & \text{if } Y_j^{(l)} \text{ wins} \\ 0, & \text{otherwise} \end{cases}\end{aligned} \tag{5.2}$$

Here, $I$ denotes the indicator function. In essence, $\phi_1$ takes the value of 1 when the subject from the treatment group emerges as the winner, $\phi_2$ is assigned 1 when the subject from the control group arises as the winner, and both functions assume the value of 0 when the comparison yields an indeterminate outcome. This framework operates under the assumption that the comparison is initially based on fatal events, and if no decisive result is obtained, the comparison proceeds based on non-fatal events. In instances where a definitive result cannot be ascertained, a tie is declared. The subsequent stage involves amalgamating these functions to formulate the statistics $U_1$ and $U_2$. These statistics denote the average wins and losses, respectively, within the treatment group, encompassing all possible pairings of treatment and control clusters.

$$U_1 = \frac{1}{mn} \sum_{k=1}^{m} \sum_{l=1}^{n} \sum_{i=1}^{I_k} \sum_{j=1}^{J_l} \phi_1(X_i^{(k)}, Y_j^{(l)})$$

and

$$U_2 = \frac{1}{mn} \sum_{k=1}^{m} \sum_{l=1}^{n} \sum_{i=1}^{I_k} \sum_{j=1}^{J_l} \phi_2(X_i^{(k)}, Y_j^{(l)})$$

Hence, an estimator for the clustered win ratio can be expressed as the ratio of the above statistics,

$$WR_{\text{clustered}} = \frac{U_1}{U_2}$$

To make inferences about this clustered win ratio, it becomes urgent to ascertain its distribution. However, this necessitates the exploration of the joint distribution of the statistics $U_1$ and $U_2$. Lee and Dehling (2005) established that both $U_1$ and $U_2$

https://doi.org/10.26219/heal.aueb.6426



exhibit an asymptotically normal distribution. Reclaiming this understanding, we can deduce the joint distribution of these two statistics. Specifically, we establish the following at a fundamental stage,

$$\phi_1(X_i^{(k)}) = E_Y(\phi_1(X_i^{(k)}, Y_j^{(l)})) - \theta_1 = P_Y(X_i^{(k)} > Y_j^{(l)}) - \theta_1$$

$$\phi_1(Y_j^{(l)}) = E_X(\phi_1(X_i^{(k)}, Y_j^{(l)})) - \theta_1 = P_X(X_i^{(k)} > Y_j^{(l)}) - \theta_1$$

$$\phi_2(X_i^{(k)}) = E_Y(\phi_2(X_i^{(k)}, Y_j^{(l)})) - \theta_2 = P_Y(X_i^{(k)} < Y_j^{(l)}) - \theta_2$$

$$\phi_2(Y_j^{(l)}) = E_X(\phi_2(X_i^{(k)}, Y_j^{(l)})) - \theta_2 = P_X(X_i^{(k)} < Y_j^{(l)}) - \theta_2$$

It is noted that,

$$\theta_1 = E_{X,Y}(\phi_1(X_i^{(k)}, Y_j^{(l)})) = P_{X,Y}(X_i^{(k)} > Y_j^{(l)})$$

$$\theta_2 = E_{X,Y}(\phi_2(X_i^{(k)}, Y_j^{(l)})) = P_{X,Y}(X_i^{(k)} < Y_j^{(l)})$$

In this context, the notations $E_X(\cdot)$ and $E_Y(\cdot)$ refer to expectations underlied on random variables $X$ and $Y$, respectively. Correspondingly, $P_X(\cdot)$ and $P_Y(\cdot)$ denote probabilities contingent on $X$ and $Y$, respectively. The parameter $\theta_1$ signifies the likelihood of patients undergoing treatment winning over those in the control group, while $\theta_2$ represents the likelihood of control group patients prevailing over those undergoing treatment. All these elements collectively contribute to the establishment of the following theorem.

Let,

$$A_1^{(k)} = \sum_{i=1}^{I_k} \phi_1(X_i^{(k)})$$

$$B_1^{(l)} = \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)})$$

$$A_2^{(k)} = \sum_{i=1}^{I_k} \phi_2(X_i^{(k)})$$

$$B_2^{(l)} = \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)})$$

Additionally, we determine,

$$\bar{I}_m = \frac{1}{m} \sum_{k=1}^{m} I_k$$

https://doi.org/10.26219/heal.aueb.6426



$$\bar{J}_n = \frac{1}{n}\sum_{l=1}^{n} J_l$$

where $m$ and $n$ represent the numbers of clusters in the two comparison groups. Considering the assumption of independence between comparison groups and clusters, along with the exchangeability of subjects, when $\min(m,n) \to \infty$, the $U$-statistics, namely $U_1 - \bar{I}_m\bar{J}_n\theta_1$ and $U_2 - \bar{I}_m\bar{J}_n\theta_2$, asymptotically follow a bivariate normal distribution with mean 0 and variance,

$$\Sigma = \frac{\bar{J}_n^2}{m}\Sigma_1 + \frac{\bar{I}_m^2}{n}\Sigma_2$$

where $\Sigma_1$ and $\Sigma_2$ are matrices encompassing variances and covariances of the defined summation variables,

$$\Sigma_1 = \begin{bmatrix} \mathrm{Var}(A_1^{(k)}) & \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) \\ \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) & \mathrm{Var}(A_2^{(k)}) \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} \mathrm{Var}(B_1^{(l)}) & \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) \\ \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) & \mathrm{Var}(B_2^{(l)}) \end{bmatrix}$$

By implementing the delta method and employing the logarithmic transformation, it can be inferred that the logarithm of the ratio of the two statistics, $\log\left(\frac{U_1}{U_2}\right)$, approaches an asymptotically normal distribution. This occurs as the minimum number of clusters between $m$ and $n$ tends to infinity. The distribution has a mean of $\log\left(\frac{\mu_1}{\mu_2}\right)$ and a variance given by,

$$\frac{\mathrm{Var}(U_1)}{\mu_1^2} - \frac{2\mathrm{Cov}(U_1,U_2)}{\mu_1\mu_2} + \frac{\mathrm{Var}(U_2)}{\mu_2^2}$$

Here, $\mu_1$ and $\mu_2$ represent the true mean values of $U_1$ and $U_2$, respectively. Consequently, the true value of the win ratio can be defined as $\frac{\mu_1}{\mu_2}$, a quantity that can be evaluated through the $\frac{U_1}{U_2}$. Additionally, the variances and covariances are expressed as,

$$\begin{aligned}\mathrm{Var}(U_1) =& \frac{\bar{J}_n^2}{m}\left\{\bar{I}_m\mathrm{Var}(\phi_1(X_i^{(k)})) + \bar{I}_m^{(2)}\mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_1(X_{i'}^{(k)}))\right\} \\ &+ \frac{\bar{I}_m^2}{n}\left\{\bar{J}_n\mathrm{Var}(\phi_1(Y_j^{(l)})) + \bar{J}_n^{(2)}\mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_1(Y_{j'}^{(l)}))\right\}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(U_2) =& \frac{\bar{J}_n^2}{m}\left(\bar{I}_m\mathrm{Var}(\phi_2(X_i^{(k)})) + \bar{I}_m^{(2)}\mathrm{Cov}(\phi_2(X_i^{(k)}), \phi_2(X_{i'}^{(k)}))\right) \\ &+ \frac{\bar{I}_m^2}{n}\left(\bar{J}_n\mathrm{Var}(\phi_2(Y_j^{(l)})) + \bar{J}_n^{(2)}\mathrm{Cov}(\phi_2(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)}))\right)\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



$$\begin{aligned}\mathrm{Cov}(U_1, U_2) =& \frac{\bar{J}_n^2}{m}\left(\bar{I}_m \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_i^{(k)})) + \bar{I}_m^{(2)} \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_{i'}^{(k)}))\right) \\ &+ \frac{\bar{I}_m^2}{n}\left(\bar{J}_n \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_j^{(l)})) + \bar{J}_n^{(2)} \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)}))\right)\end{aligned}$$

where,

$$\bar{I}_m^{(2)} = \frac{1}{m}\sum_{k=1}^{m} I_k(I_k - 1)$$

and

$$\bar{J}_n^{(2)} = \frac{1}{n}\sum_{l=1}^{n} J_l(J_l - 1)$$

It should be noticed that the variables denoted by $X_i^{(k)}$ and $X_{i'}^{(k)}$ represent distinct individuals within the same treatment cluster denoted as $k$. Correspondingly, the variables $Y_j^{(l)}$ and $Y_{j'}^{(l)}$ signify different individuals within the identical control cluster identified as $l$. We observe that the variance in the $U1$ and $U2$ statistics is associated not only with the individual subject's variability but also with the variation between two distinct individuals within the same cluster, distinguished by the same kernel function. Regarding the covariance between the two statistics, it encompasses the covariance of a subject for the two different kernel functions, as well as the covariance between two disparate subjects within the same cluster, each characterized by a different kernel function (Zhang and Jeong (2021)).

https://doi.org/10.26219/heal.aueb.6426

https://doi.org/10.26219/heal.aueb.6426

# Chapter 6

# Simulation Study: Implementing the Win Ratio Method to a Breast Cancer Clinical Trial Dataset

## 6.1 A brief discussion on breast cancer

Numerous diseases continually afflict the global population, giving rise to various health challenges and expediting substantial impacts. Among these trials and tribulations, cancer has developed as an almighty adversary, remaining a complicated ailment that researchers have not yet entirely seduced across all its stages and manifestations. On a daily basis, we come across a large amount of instances of individuals facing diverse forms of this formidable malady. Among these, breast cancer emerges as a widespread subtype, particularly influencing women, with a higher incidence observed among the elderly demographic.

Breast cancer specifically occurs from the uncontrolled multiplication of cancerous cells within the breast, leading to the configuration of tumors. Principally, this situation comes out as an invasive form, indicating that the tumor possesses the potential to metastasize and spread beyond the verges of the breast to other areas of the body. While the majority of cases involve women, it is noteworthy that a small percentage of instances also occur in men, emphasizing the indiscreet nature of this disease. Typically associated with women, breast cancer mainly influences individuals aged 50 and above. However, it is crucial to acknowledge the existence of cases where younger women experience this ailment (CDC (2023)).

The accurate causes underscoring the transformation of normal breast cells into cancerous ones, leading to the onset of breast cancer, remain unclear. This intricacy redounds to the challenge of discovering a conclusive therapy for the ailment. Nevertheless, comparative research has shed light on specified factors that increase the probability of appearance breast cancer. Among these factors, age distinguishes, with advancing years relating to a higher sensibility to the disease. Furthermore, gender is a substantial determinant, with the comminuted majority of cases occurring in women. Family history emerges as another principal element, as individuals

https://doi.org/10.26219/heal.aueb.6426



with close relatives affected by breast cancer face a heightened risk themselves. Genetic tendency further entangles the background, with inherited mutations being responsible for a notable proportion of breast cancer diagnoses. Lifestyle choices exert deeply influence as well, with smoking, extravagant alcohol consumption, obesity, and extended sun exposure all contributing to premium risk levels. Additionally, individuals who have undergone hormone replacement therapy may find themselves at greater risk of developing the condition.

The comprehension of the diverse stages of breast cancer is vital for determining appropriate treatment strategies and predicting outcomes. A structured staging system contributes to classifying the progression of the disease based on factors like tumor size and expansion. Breast cancer is typically categorized into five stages for this purpose. The primary stage, referred to as stage 0, represents a localized presence of cancer cells within the breast ducts without spreading to surrounding tissues. Proceeding to stage I, cancer cells begin to penetrate nearby breast tissue. Stage II underlies a further progression, where carcinoma cells develop into noticeable tumors ranging from 2 to 5 centimeters in diameter. These tumors may or may not influence the lymph nodes. Advancing to stage III, the cancer expands into nearby tissues and lymph nodes, denoting locally advanced breast cancer and requiring comprehensive treatment approaches. The final stage, stage IV, indicates the most critical phase, as cancer cells have metastasized from the breast to distant organs such as bones, lungs, or liver, posing significant challenges for treatment and prognosis (Cleveland Clinic (2023)).

Ultimately, breast cancer remains a complex and almighty disease affecting individuals worldwide, with a particularly significant impact on women, especially those aged 50 and above. Figure 6.1 illustrates the anatomical structure of the breast alongside the development of breast cancer. Efforts in early detection, research, and comprehensive treatment approaches are essential in combating this relentless adversary and improving outcomes for those affected by breast cancer.

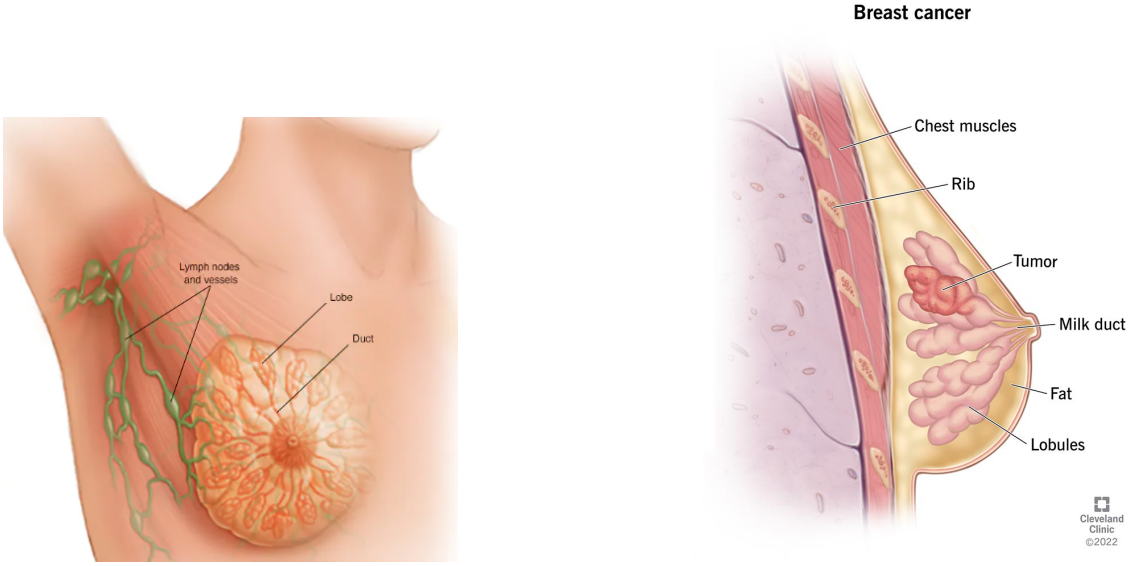
## 6.2 Data simulation: diverse scenarios

Various analyses can be conducted, and diverse scenarios can be investigated concerning the utilization and implementation of the win ratio method. In our thesis, we decided to apply the win ratio approach in deliberating a dataset from a clinical trial focused on breast cancer. We aim to counter the effectiveness of two treatments for breast cancer: Trastuzumab, a conventional treatment, and Anti-HER2, a novel approach. Both treatments target the reduction of Her2 protein levels, a pivotal factor in the multiplication of cancer cells, particularly increased on their surfaces (Hygeia (2020)). Our research centers on two primary composite endpoints: disease recurrence, a nonfatal occurrence, and death, a critical endpoint due to its implications in medical outcomes. Through the utilization of simulations, we have advanced our analytical approach, allowing us to evaluate the efficacy of the novel treatment in combating breast cancer.

Initially, according to Zhang's and Jeong's investigation and theory, outlined in

https://doi.org/10.26219/heal.aueb.6426






(a) (b)

*Figure 6.1:* (a) The structure and composition of the breast, often referred to as breast anatomy (Mayo Clinic (2022)). (b) Development of breast cancer due to the abnormal growth and mutation of cells within the breast tissue, leading to the formation of tumors (Cleveland Clinic (2023)).

the previous chapter (Zhang and Jeong (2021)), we assume disposing of a single cluster. Regarding the intra-cluster correlation, the variable $\gamma$ is utilized, which corresponds to cluster-specific frailty. This $\gamma$ variable is supposed to be equal to 1 in both comparison groups since we are referring to a single cluster. Furthermore, we adopt proportional hazard models to implement the analysis. In this context, for the $i^{\text{th}}$ patient in the new treatment group, the hazard function for time to recurrence, can be formulated as,

$$\lambda_{R_{i1}} = \gamma \lambda_R \exp(-\eta_R)$$

Here, $\gamma$ denotes the frailty specific to our single cluster, $\lambda_R$ signifies the baseline hazard function for the recurrence time distribution, and $\eta_R$ represents the non-fatal event rate per unit time. Similarly, for the $j^{\text{th}}$ patient in the conservative treatment group, their hazard function is expressed as,

$$\lambda_{R_{j0}} = \gamma \lambda_R$$

For our convenience, we determine the hazard function until the time of recurrence as $\lambda_{R_z}$, where $z$ corresponds to the patient's treatment group, either Trastuzumab ($z = 0$) or Anti-HER2 ($z = 1$). Additionally, in the same manner, we define the hazard function for the time to death as $\lambda_{D_z}$. We utilize a bivariate exponential distribution with the Gumbel-Hougaard copula to model the bivariate distribution of time to recurrence $T_{R_z}$ and time to death $T_{D_z}$. The copula provides the generation of the joint survival function, which is articulated as,

$$P(T_{R_z} > y_1, T_{D_z} > y_2 | z, \gamma) = \exp\left\{ - \left[ (\lambda_{R_z} y_1)^\rho + (\lambda_{D_z} y_2)^\rho \right]^{1/\rho} \right\}$$

https://doi.org/10.26219/heal.aueb.6426



Here, $\rho \geq 1$ indicates the association parameter between $T_{R_z}$ and $T_{D_z}$. It is noted that the process is initiated by generating bivariate uniform random variables for the marginal survival functions in the copula, leveraging stable distribution. Regarding the censoring time $T_{C_z}$, it is supposed to be independent of the times until the nonfatal event and fatal event, as well as the cluster-specific frailty $\gamma$. It is worthwhile to mention that $T_{C_z}$ follows an exponential distribution with the hazard function,

$$\lambda_{C_z} = \lambda_C$$

To conduct the simulations and generate our dataset, we employ various parameter values. By amalgamating these values, we obtain diverse scenarios for analysis. Specifically, we select two values for the parameter $\rho$ ($\rho = 2, 5$), allowing us to observe how the value of the win ratio, our in-consideration measure, alters with a smaller and larger correlation between time to nonfatal and fatal events. In addition, we investigate three hazard ratio values for recurrence and death: 1.0, 1.2, and 1.5. These values indicate no survival difference, a statistically significant difference, and a notably significant difference between the treatment groups, respectively. The hazard ratio is employed as an evaluation of relative risk, denoting the probability of an event or ailment exhibition concerning exposure. Relative risk compares the probability of the event occurring among subjects exposed to a specific factor with those who are not exposed, typically forming a ratio between the two groups (William and Nicolas (2010)). Consequently, the logarithms of these hazard ratios, denoted by $\eta_R$ and $\eta_D$, take values of 0, 0.18, and 0.41, respectively. By combining these parameters, we generate data for a total of 9 different scenarios when $\rho$ equals 2, and another 9 scenarios when $\rho$ equals 5. Moreover, we explore two different total sample sizes: $N = 200$ and $N = 1000$. For the former, we set $\lambda_R$ and $\lambda_D$ to 0.1 and 0.08, respectively, while for the latter, we use $\lambda_R = 0.066$ and $\lambda_D = 0.077$. The aforementioned cases are examined for both sample sizes. We suppose that the constant value of $\lambda_C$ remains 0.01 across all cases. Each scenario undergoes 1000 repetitions, where for each repetition, we compute the win ratio, a 95% confidence interval for this metric, and the power of the hypothesis testing.

For the computation of the win ratio, we employ the unmatched approach, introduced by Pocock et al. (2012). Initially, after generating a random sample encompassing 200 individuals (1000 individuals utilizing the second sample size), we divide it into two groups: one comprising 100 subjects (500 subjects in the alternative case) receiving the standard treatment, Trastuzumab, and the other comprising 100 subjects (500 subjects in the alternative case) undergoing the experimental treatment, Anti-HER2. Afterward, we simulate the times until two critical events occur: non-fatal recurrence and death. We assume that the participants are followed up for 21 months, and we proceed with pairwise comparisons. Specifically, each patient from the Trastuzumab group is compared against each patient from the Anti-HER2 group, resulting in a total of $100 \times 100 = 10000$ patient pair comparisons ($500 \times 500 = 250000$ patient pair comparisons in the alternative case). For each patient pair, we conduct individual studies. Initially, our comparison prioritizes mortality, given its highest medical significance. Specifically, the objective is

https://doi.org/10.26219/heal.aueb.6426



to determine which patient among the pair encounters the death first. In instances where this information is unavailable, our comparison shifts to recurrence, identifying the patient who experiences the non-fatal event first. If neither of the two patients encounters an event, or if one patient experiences an event while the observation period for the other patient in the pair is shorter, we classify the scenario as censored data. This process results in the creation of five categories:

i) The patient receiving Anti-HER2 treatment experiences death first.

ii) The patient undergoing Trastuzumab treatment faces death first.

iii) The patient receiving Anti-HER2 therapy encounters the recurrence first.

iv) The patient undergoing Trastuzumab treatment experiences the recurrence first.

v) No event occurs for either patient in the pair.

It is evident that the aggregation of patients falling into the second and fourth categories represents the number of winners for Anti-HER2 therapy. This result arises due to the unfavorable outcomes associated with the administration of the conventional treatment, Trastuzumab. Similarly, the sum of the first and third categories denotes the number of losers for Anti-HER2 therapy. The last category comprises individuals commonly referred to as tied, signifying a draw between the treatments.

In each iteration, where the win ratio is computed as per the aforementioned procedure, the subsequent step involves calculating the 95% confidence interval for this metric. To accomplish this objective, we utilize the variance of the logarithm transformation, as described in subsection 2.3.7. According to the methodology introduced by Yu and Ganju (2022), we calculate the logarithm of the win ratio and depend on the variance of log(WR). The same researchers offered the formula for determining this variance.

In surveys, a crucial aspect comprises evaluating the power of hypothesis testing. In our scenario, we are involved in assessing the following hypothesis testing:

$$\begin{cases} H_0 : \text{There is no difference between the two compared groups} \\ H_1 : \text{There is difference between the two compared groups} \end{cases}$$

This hypothesis testing setup can be denoted as follows:

$$\begin{cases} H_0 : \text{WR} = 1 \\ H_1 : \text{WR} \neq 1 \end{cases}$$

https://doi.org/10.26219/heal.aueb.6426



Our objective in statistical analysis constitutes rejecting the null hypothesis. This underlies the importance of comprehending the power of hypothesis testing. This concept describes the probability of correctly rejecting a false null hypothesis, illustrating the test's ability to identify a real effect or difference if present. A higher power ensures greater confidence in correctly rejecting the false null hypothesis. Since calculating the win ratio in each repetition, the way forward is to verify the power for hypothesis testing. Thus, we examine whether the unit falls within the calculated confidence interval. In cases where the unit belongs to the confidence interval, we retain the null hypothesis. However, if the unit lies outside of this interval, specifically when it does not include 1, we reject the null hypothesis. Following this procedure, we compute the value of the power of hypothesis testing, by employing the proportion of time out of 1000 repetitions that we rejected the null hypothesis.

## 6.3 Results

After carrying the aforementioned procedures out, results for each scenario are obtained, which are then discussed within this specific subsection. It's worthwhile to underline that the outcomes derived from each repetition of every case - comprising the win ratio value, the 95% confidence interval, and the power - are assembled over the 1000 repetitions of each case. This aggregation contributes to the determination of an average win ratio, an average 95% confidence interval, an average power, and the standard error per case.

Table 6.1 displays the results for the 9 scenarios with a sample size of 200, an $\rho$ value of 2, and different combinations of hazard ratios for the non-fatal and fatal events. We observe that when the hazard ratios for both events are set to 1.0 or the hazard ratio for death is 1.0 while the hazard ratio for the non-fatal event alters, the average win ratio tends to fluctuate around unity. This indicates a lack of statistically significant difference between the two comparison groups. Similarly, examination of the 95% confidence intervals in these instances uncovers that they encompass the value 1, forbidding us to reject the null hypothesis. Hence, the average power exhibits a very small value in these scenarios. Regarding the standard error, it ranges around 0.005, signifying the small deviation expected in the sample estimate of win ratio relative to the true population value. It's noteworthy that the win ratio values increase for the rest of the compounds of hazard ratio values. Particularly increased values are observed in scenarios where the hazard ratio for death is 1.5 while the hazard ratio for recurrence varies. The highest average win ratio is accomplished when both event hazard ratios are 1.5, reaching 1.4424, implying the new treatment's superiority over the standard treatment. The 95% confidence intervals include the unit in cases where the hazard ratio for death is 1.2, while they do not encompass the value 1 when the hazard ratio for death is 1.5 with the latter circumstance leading to the rejection of the null hypothesis. Notably, the hazard ratio for recurrence alters in both scenarios. As the average win ratio increases, the average power increases, reinforcing our confidence in correctly rejecting the null hypothesis and detecting statistical differences between the groups. A small increase

https://doi.org/10.26219/heal.aueb.6426



is also observed in the standard error, which now fluctuates around 0.008.

Table 6.2 presents the outcomes across 9 scenarios, each with a sample size of 200 and diverse combinations of hazard ratios for both events, with an $\rho$ value of 5 this time. Exploring this table contributes to similar conclusions as when $\rho$ was set to 2. Nevertheless, an intriguing remark arises: with a higher correlation between event times ($\rho = 5$), we observe slightly lower average power compared to scenarios with a lower correlation. On the contrary, it can be noted slightly higher average win ratios. Additionally, there's a petty rise in standard error values under the $\rho$ 5 condition compared to $\rho$ 2. This implies that the degree of correlation between event times can affect the effectiveness of our measures, with a stronger correlation slightly diminishing our power and introducing a slight increase in observed win ratios and measurement uncertainty.

In the consequent exploration, we investigate 9 distinct scenarios created by selecting a larger sample size of 1000 and preserving an $\rho$ value of 2, while altering the hazard ratios for both non-fatal and fatal events. The discoveries from this probe are summarized in Table 6.3. Despite the modification in sample size, the relations linking the average win ratio, the average 95% confidence interval, the average power, and the standard error with diverse combinations of hazard ratios remain consistent with our earlier observations when the sample size was 200. However, it's important to emphasize that with the larger sample size, the typical 95% confidence intervals display a particular pattern. Specifically, they include the unit when the hazard ratio for death is 1.0, yet they do not comprise it for hazard ratio values of 1.2 and 1.5. Notably, the hazard ratio for recurrence alters across all three scenarios. Hence, we avoid rejecting the null hypothesis when the hazard ratio for death is 1.0, despite changes in the hazard ratio for recurrence. In contrast, in the other two instances, we decide to reject the null hypothesis. Furthermore, it is worthwhile to highlight the remarkable enhancement in results with the larger sample size. With more participants comprised in the study, we observe a substantial rise in the average power, which is particularly favorable in statistical analyses. Simultaneously, there's a marked decrease in the standard error, conforming with our anticipations for improved accuracy in measurements with a larger sample size. These findings underline the advantages of increasing the sample size in research efforts, highlighting how it can contribute to more precise and reliable outcomes during statistical analyses.

Proceeding our investigation, we study 9 scenarios produced with a sample size of 1000, an $\rho$ value set at 5, and assorted combinations of hazard ratios for non-fatal and fatal events. The findings conform to our earlier analyses with a sample size of 200, revealing similar inferences. The trend observed in the 95% confidence intervals, as outlined for a sample size of 1000 and an $\rho$ value of 2, remains apparent in the given scenario. Once more, we observe a trend towards more accurate and credible outcomes, underlying the benefits of a larger sample size in improving the precision and reliability of our analyses. Table 6.4 displays these specific results. Moreover, Figures 6.2 and 6.3 provide visual depictions of these findings. Figure 6.2 exhibits the average win ratio across the diverse scenarios, while Figure 6.3 illustrates the

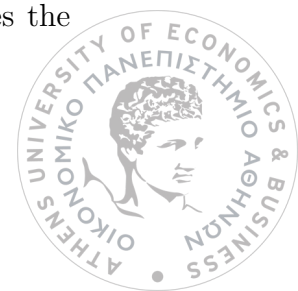


*Table 6.1:* Presenting the findings for the 9 scenarios involving a sample size of 200, an $\rho$ value set at 2, and different combinations of hazard ratios for the non-fatal and fatal events. The results encompass the average win ratio, its average 95% confidence interval, the average power of the hypothesis testing, and the standard error across 1000 iterations for each case.

| Sufficient Insights from the Simulations | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Total Sample Size N = 200** | | | | | | | | | | | | |
| | **ρ = 2** | | | | | | | | | | | |
| **HR (Recurrence)** | **1.0** | | | | **1.2** | | | | **1.5** | | | |
| **HR (Death)** | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* |
| **1.0** | 1.0054 | (0.7293, 1.3860) | 0.075 | 0.0055 | 1.0113 | (0.7334, 1.3945) | 0.070 | 0.0055 | 1.0180 | (0.7380, 1.4043) | 0.067 | 0.0056 |
| **1.2** | 1.1707 | (0.8490, 1.6141) | 0.156 | 0.0064 | 1.1801 | (0.8556, 1.6277) | 0.169 | 0.0065 | 1.1912 | (0.8632, 1.6438) | 0.182 | 0.0066 |
| **1.5** | 1.4070 | (1.0202, 1.9403) | 0.507 | 0.0078 | 1.4230 | (1.0314, 1.9634) | 0.533 | 0.0079 | 1.4424 | (1.0448, 1.9914) | 0.557 | 0.0081 |



https://doi.org/10.26219/heal.aueb.6426

*Table 6.2:* Displaying the findings for the 9 scenarios involving a sample size of 200, an $\rho$ value set at 5, and different combinations of hazard ratios for the non-fatal and fatal events. The results encompass the average win ratio, its average 95% confidence interval, the average power of the hypothesis testing, and the standard error across 1000 iterations for each case.

| Sufficient Insights from the Simulations | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Total Sample Size N = 200** | | | | | | | | | | | | |
| | **ρ = 5** | | | | | | | | | | | |
| **HR (Recurrence)** | **1.0** | | | | **1.2** | | | | **1.5** | | | |
| **HR (Death)** | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* |
| **1.0** | 1.0105 | (0.7319, 1.3952) | 0.062 | 0.0056 | 1.0188 | (0.7374, 1.4076) | 0.061 | 0.0057 | 1.0257 | (0.7419, 1.4180) | 0.066 | 0.0057 |
| **1.2** | 1.1715 | (0.8483, 1.6177) | 0.146 | 0.0066 | 1.1860 | (0.8580, 1.6393) | 0.171 | 0.0067 | 1.2008 | (0.8679, 1.6615) | 0.184 | 0.0068 |
| **1.5** | 1.3942 | (1.0095, 1.9255) | 0.482 | 0.0079 | 1.4181 | (1.0257, 1.9605) | 0.516 | 0.0081 | 1.4476 | (1.0456, 2.0042) | 0.570 | 0.0083 |





https://doi.org/10.26219/heal.aueb.6426

*Table 6.3:* Presenting the findings for the 9 scenarios involving a sample size of 1000, an $\rho$ value set at 2, and different combinations of hazard ratios for the non-fatal and fatal events. The results encompass the average win ratio, its average 95% confidence interval, the average power of the hypothesis testing, and the standard error across 1000 iterations for each case.

| Sufficient Insights from the Simulations | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Total Sample Size N = 1000** | | | | | | | | | | | | |
| | **ρ = 2** | | | | | | | | | | | |
| **HR (Recurrence)** | **1.0** | | | | **1.2** | | | | **1.5** | | | |
| **HR (Death)** | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* |
| **1.0** | 1.0026 | (0.8677, 1.1584) | 0.065 | 0.0024 | 1.0075 | (0.8718, 1.1644) | 0.066 | 0.0024 | 1.0126 | (0.8760, 1.1705) | 0.062 | 0.0024 |
| **1.2** | 1.1662 | (1.0090, 1.3478) | 0.535 | 0.0028 | 1.1746 | (1.0160, 1.3580) | 0.575 | 0.0028 | 1.1836 | (1.0235, 1.3687) | 0.599 | 0.0029 |
| **1.5** | 1.3987 | (1.2098, 1.6171) | 0.995 | 0.0034 | 1.4139 | (1.2225, 1.6353) | 0.996 | 0.0035 | 1.4306 | (1.2364, 1.6553) | 0.996 | 0.0035 |



https://doi.org/10.26219/heal.aueb.6426

https://doi.org/10.26219/heal.aueb.6426



average power values for each case.

Ultimately, we can comprehend that several factors influence the win ratio. The values we receive for the win ratio are promptly related to the hazard ratio values assigned to the two events. When both hazard ratios are equal to 1.0, or at least the hazard ratio for death is, the win ratio ranges to unity. Nevertheless, as the hazard ratios diverge from 1.0, the win ratio also inclines from unity. The opting of sample size arises as a crucial agent. A larger sample size improves the power of the hypothesis test. Consequently, greater confidence in our ability to correctly reject the null hypothesis is ensured. In addition, the correlation between the event times emerges as another substantial factor impacting the win ratio, as designated by its influence on the value of the win ratio in our simulations.

In the clinical trials datasets on breast cancer, it is acknowledged the remarkably increased hazard ratio values for both recurrence and mortality, alongside a notable correlation between them. This correlation is substantial, as in many real-life instances, death often follows recurrence. Based on the outcomes generated from our simulations, we can assuredly state that the novel Anti-HER2 treatment for breast cancer exhibits several benefits over the conventional Trastuzumab therapy. These findings signify that Anti-HER2 holds considerable promise as a prospective treatment option for breast cancer patients, offering more favorable results than its conservative counterpart.

*Table 6.4:* Displaying the findings for the 9 scenarios involving a sample size of 1000, an $\rho$ value set at 5, and different combinations of hazard ratios for the non-fatal and fatal events. The results encompass the average win ratio, its average 95% confidence interval, the average power of the hypothesis testing, and the standard error across 1000 iterations for each case.

| Sufficient Insights from the Simulations | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Total Sample Size N = 1000** | | | | | | | | | | | | |
| | **ρ = 5** | | | | | | | | | | | |
| **HR (Recurrence)** | **1.0** | | | | **1.2** | | | | **1.5** | | | |
| **HR (Death)** | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* | *WR* | *95% CI* | *Power* | *S.E* |
| **1.0** | 1.0034 | (0.8671, 1.1612) | 0.061 | 0.0025 | 1.0069 | (0.8699, 1.1654) | 0.065 | 0.0025 | 1.0088 | (0.8714, 1.1678) | 0.063 | 0.0025 |
| **1.2** | 1.1664 | (1.0074, 1.3506) | 0.514 | 0.0029 | 1.1750 | (1.0144, 1.3611) | 0.560 | 0.0029 | 1.1808 | (1.0191, 1.3682) | 0.575 | 0.0029 |
| **1.5** | 1.3926 | (1.2022, 1.6132) | 0.993 | 0.0035 | 1.4126 | (1.2186, 1.6375) | 0.995 | 0.0036 | 1.4300 | (1.2328, 1.6588) | 0.998 | 0.0036 |



https://doi.org/10.26219/heal.aueb.6426

https://doi.org/10.26219/heal.aueb.6426



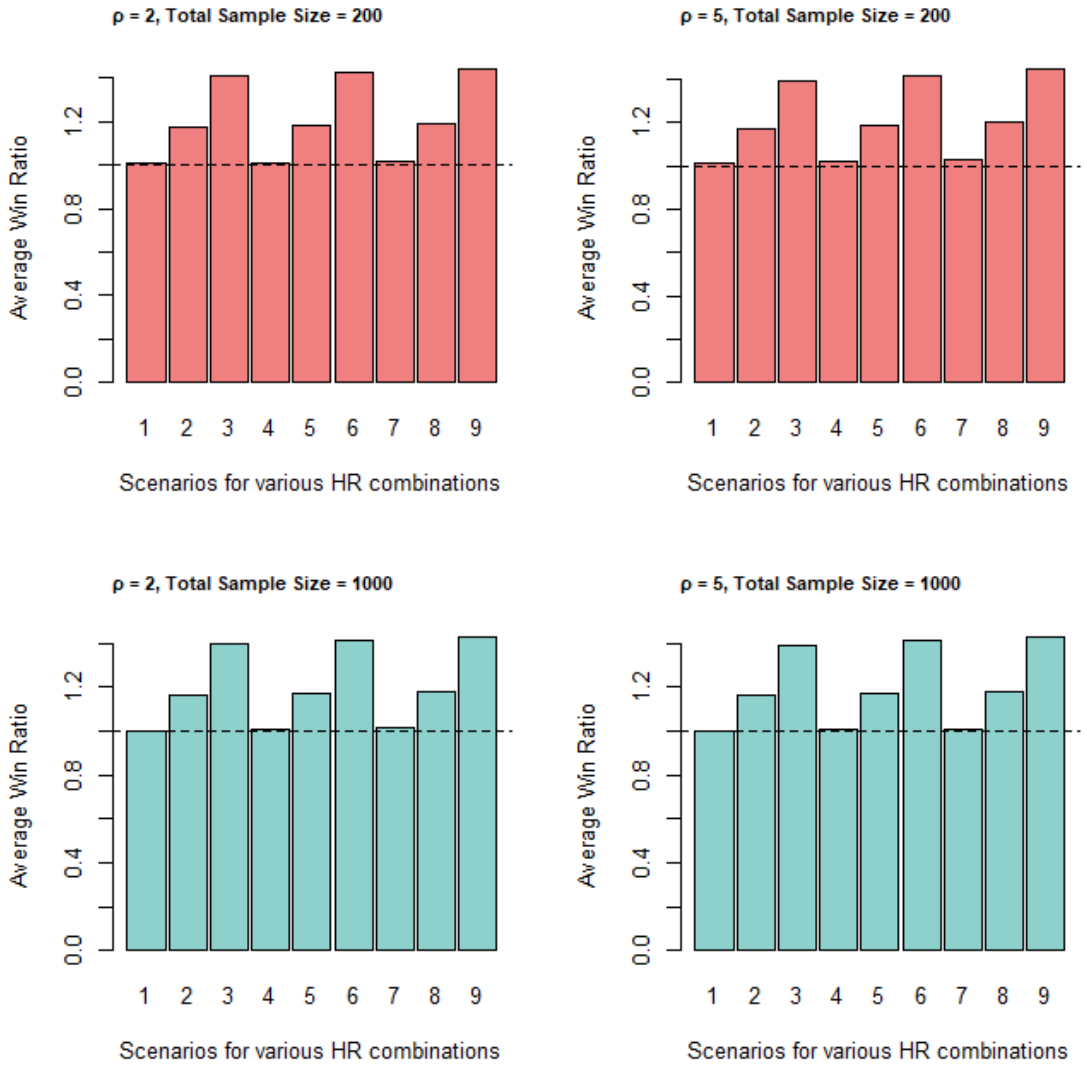



*Figure 6.2:* Illustration of the fluctuation of the average win ratio across the different scenarios. Cases 1 through 9 correspond to the following combinations of hazard ratios for non-fatal and fatal events: 1) $HR_R = 1.0$, $HR_D = 1.0$, 2) $HR_R = 1.0$, $HR_D = 1.2$, 3) $HR_R = 1.0$, $HR_D = 1.5$, 4) $HR_R = 1.2$, $HR_D = 1.0$, 5) $HR_R = 1.2$, $HR_D = 1.2$, 6) $HR_R = 1.2$, $HR_D = 1.5$, 7) $HR_R = 1.5$, $HR_D = 1.0$, 8) $HR_R = 1.5$, $HR_D = 1.2$, 9) $HR_R = 1.5$, $HR_D = 1.5$. The dashed line corresponds to the average win ratio equal to 1.

https://doi.org/10.26219/heal.aueb.6426



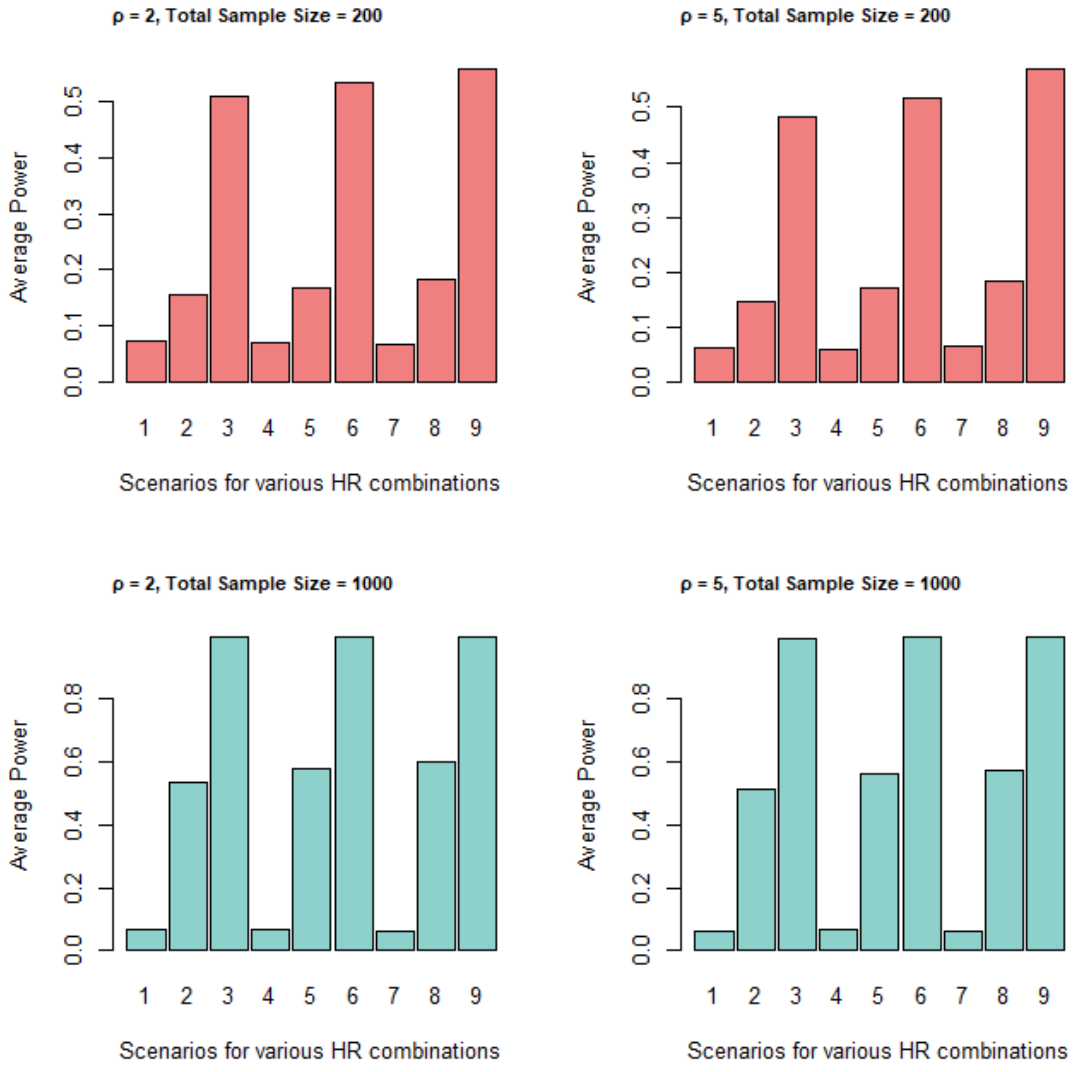



*Figure 6.3:* Illustration of the fluctuation of the average power of the hypothesis testing across the different scenarios. Cases 1 through 9 correspond to the following combinations of hazard ratios for non-fatal and fatal events: 1) $HR_R = 1.0$, $HR_D = 1.0$, 2) $HR_R = 1.0$, $HR_D = 1.2$, 3) $HR_R = 1.0$, $HR_D = 1.5$, 4) $HR_R = 1.2$, $HR_D = 1.0$, 5) $HR_R = 1.2$, $HR_D = 1.2$, 6) $HR_R = 1.2$, $HR_D = 1.5$, 7) $HR_R = 1.5$, $HR_D = 1.0$, 8) $HR_R = 1.5$, $HR_D = 1.2$, 9) $HR_R = 1.5$, $HR_D = 1.5$.

https://doi.org/10.26219/heal.aueb.6426

# Chapter 7

# Discussion and Avenues for Future Research

The win ratio method emerged as a useful asset within the realm of clinical trials, possessing particular characteristics that render it privileged over conservative methods. Its individualistic property to take into account the medical hierarchy of events and the interaction between the various components has greatly improved researchers' analyses. This methodology permits studies to be conducted in a manner that reflects real-world scenarios more closely, offering a more comprehensive understanding of the outcomes.

In addition, the penetration into the definition and methodology of the win ratio measure allows the construction of a 95% confidence interval for this specific metric, alongside the evaluation of statistical power for hypothesis testing. Literature provides knowledge into specifying the optimal sample size to achieve desired levels of power in analysis. Moreover, the win ratio method is distinguished for its flexibility and adaptability across diverse scenarios. Its expansion to stratified and clustered data is particularly noteworthy, displaying critical advantages in these contexts.

The simulation study conducted in this specific thesis investigates various scenarios to comprehend the functionality of the win ratio method. As outlined in Chapter 6, the results reveal a strong relation between the win ratio measure and the hazard ratios associated with both non-fatal and fatal events under consideration. Furthermore, the analysis underscores the interaction between the win ratio and sample size, as well as the relationship between the win ratio and the correlation, which exhibits between the times required for the occurrence of the non-fatal event and the fatal event.

While the win ratio method is widely employed, certain drawbacks are associated with its implementation. One notable limitation is the relative unfamiliarity of researchers with this particular approach. Despite its introduction into the literature in 2012 by Pocock et al., there persist some challenges in its adoption that require time to address. Additionally, it's worth noting that this method does not rely on precise timing data related to the period between randomization and event occurrence. Instead, it simply assesses whether an event occurred earlier in one

https://doi.org/10.26219/heal.aueb.6426



patient compared to his corresponding counterpart during the comparison process. However, it's important to note that this aspect does not result in a reduction of statistical power when compared to conventional methods (Pocock et al. (2012)).

Ultimately, while there exist constraints that may impede its extensive utilization, these can be overcome through comprehensive study, analysis, and thorough attention to the specifications of the method. By solving these challenges, broader implementation of the method can be accomplished, contributing to smoother and more efficient analysis of research data.

https://doi.org/10.26219/heal.aueb.6426

# Appendix A

## Several instances for the function $\phi$ within the Ramchandani et al. (2016) methodology

In the existing body of literature, diverse non-parametric tests have been suggested, with researchers employing varied formulations of the function $\phi$. This variability in the application of $\phi$ underscores the range of approaches taken by scholars in exploring and implementing non-parametric testing methodologies.

Initially, a non-parametric test was introduced for the analysis of composite endpoints by O'Brien in 1984. The fundamental concept underlying this approach involves assigning an overall rank to each participant in the research study. This notion is implemented by summing individual ranks for each outcome and utilizing either a cumulative rank statistic or an ANOVA test. In this specific scenario, the function $\phi$ can be apparent in two distinct forms,

$$\phi(r_1, \ldots, r_p) = r_1 + r_2 + \ldots + r_p$$

or

$$\phi(r_1, \ldots, r_p) = w_1 r_1 + w_2 r_2 + \ldots + w_p r_p$$

where each weight $(w_k)$ is non-negative for all $k$. The selection of different weights for each outcome hangs on its clinical severity, reflecting the distinct considerations in the design of the test methodology.

In 1999, the Finkelstein–Schoenfeld test was introduced to assess fatal and longitudinal events, incorporating their clinical hierarchy (Finkelstein and Schoenfeld (1999)). Initially, study participants were compared based on the more medically serious event, that is, the fatal event. In cases where the comparison yielded indeterminate results, the evaluation shifted to the longitudinal event. Each of these events is associated with a scoring function: $r_1$ refers to the lethal event, while $r_2$ corresponds to the longitudinal event. Subsequently, the composite function $\phi$ is defined as follows,

$$\phi(r_1, r_2) = r_1 + I(r_1 = 0) r_2$$

where $I$ represents the indicator function. For a general scenario involving $p$ outcomes in hierarchies, the function $\phi$ is extended as,

$$\phi(r_1, r_2, \ldots, r_p) = r_1 + I(r_1 = 0) r_2 + \ldots + I(r_1 = \ldots = r_{p-1} = 0) r_p$$

https://doi.org/10.26219/heal.aueb.6426



Similar to the preceding test, it is allowed for the incorporation of different weights for each outcome, implying that,

$$\phi(r_1, r_2, \ldots, r_p) = w_1 r_1 + I(r_1 = 0) w_2 r_2 + \ldots + I(r_1 = \ldots = r_{p-1} = 0) w_p r_p$$

where $\forall k \, (w_k \geq 0)$. This formulation provides flexibility in accommodating varied weights for the respective results, contributing to a comprehensive assessment within the clinical hierarchy.

An alternative non-parametric test introduced by Wittkowski et al. (2004) involves the comparison of patients based on various ordinal measures. When evaluating the performance of two patients, $i$ and $j$, a score of 1 is assigned if at least one outcome favors patient $i$, while a score of $-1$ is assigned if patient $j$ performs better. In cases where outcomes are mixed, resulting in both favorable and unfavorable results, the score is set to 0. In this specific approach, the scoring function $\phi$, is defined as follows,

$$\phi(r_1, \ldots, r_p) = I\left(\max_k\{r_k : k = 1, \ldots, p\} > 0\right) - I\left(\min_k\{r_k : k = 1, \ldots, p\} < 0\right)$$

Here, $I$ represents the indicator function. The function can be alternatively expressed as,

$$\phi(r_1, \ldots, r_p) = I\left(\sum_k r_k > 0\right) - I\left(\sum_k r_k < 0\right)$$

This modification allows scoring of 1 when one individual exhibits more favorable outcomes than the other. For scenarios involving different weights assigned to each result, denoted as $w_k > 0$ for all $k$, the function transforms,

$$\phi(r_1, \ldots, r_m) = I\left(\sum_k w_k r_k > 0\right) - I\left(\sum_k w_k r_k < 0\right)$$

This formulation accommodates the consideration of varied weights for each outcome.

An alternative strategy widely acknowledged in the literature involves the amalgamation of multiple tests. A practical instance is the combination of the O'Brien and Finkelstein–Schoenfeld tests, where the formulation of the function $\phi$ is given by,

$$\phi(r_1, \ldots, r_p) = r_1 + I(r_1 = 0) \frac{1}{p-1} \sum_{k=2}^{p} r_k$$

It is observed that this formulation derives the score from the initial outcome, utilizing a hierarchical basis. In instances where outcomes are indeterminate, the score is evidently an average computed across all other outcomes (Ramchandani et al. (2016)).

https://doi.org/10.26219/heal.aueb.6426



# Proof of the theorem outlined in section 3.1

**Theorem 1.** *In a test assessing win ratios with an asymptotic level $\alpha$, the minimum sample size $n$ required to attain asymptotic power equal to or exceeding $1-\beta$, considering the logarithm of the win ratio, denoted by $\theta$, is determined as,*

$$n = \frac{\zeta_0^2 \left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{q(1-q)W_0^2\theta^2}$$

*Proof.* Initially, we investigate the asymptotic variance of the quantity

$$\sqrt{n}\log\left(\frac{\hat{W}_{1n}}{\hat{W}_{0n}}\right)$$

under the null hypothesis, which indicates no statistically significant difference between the two groups. This variance can be denoted as $\sigma_0^2$. Therefore, the test statistic can be formulated as follows,

$$\begin{aligned} S_n = \hat{\sigma}_n^{-1}\sqrt{n}\left\{\log\left(\frac{\hat{W}_{1n}}{\hat{W}_{0n}}\right) - \log\left(\frac{W_{1,n}}{W_{0,n}}\right)\right\} \\ + \sigma_0^{-1}\sqrt{n}\log\left(\frac{W_{1,n}}{W_{0,n}}\right) \\ + \sqrt{n}\log\left(\frac{W_{1,n}}{W_{0,n}}\right)\left(\hat{\sigma}_n^{-1} - \sigma_0^{-1}\right) \end{aligned}$$

Under relatively gentle circumstances of normality concerning $\hat{W}_{1n}$, $\hat{W}_{0n}$, and $\hat{\sigma}_n$, it can be obtained that,

$$\hat{\sigma}_n^{-1}\sqrt{n}\left\{\log\left(\frac{\hat{W}_{1n}}{\hat{W}_{0n}}\right) - \log\left(\frac{W_{1,n}}{W_{0,n}}\right)\right\} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\sigma_0^{-1}\sqrt{n}\log\left(\frac{W_{1,n}}{W_{0,n}}\right) \xrightarrow{p} \sigma_0^{-1}h$$

$$\sqrt{n}\log\left(\frac{W_{1,n}}{W_{0,n}}\right)\left(\hat{\sigma}_n^{-1} - \sigma_0^{-1}\right) \xrightarrow{p} 0$$

These conditions remain valid as the sample size $n$ approaches infinity ($n \to \infty$) under the hypothesis $H_{A,n}$. Therefore, we observe that,

$$S_n \xrightarrow{d} \mathcal{N}(\sigma_0^{-1}h, 1) \quad \text{under } H_{A,n}$$

Moreover, concerning the power of $1-\beta$ for quite large sample sizes, we can state that,

$$\begin{aligned} 1-\beta &= P(|S_n| > z_{1-\frac{\alpha}{2}}) \\ &\approx P(S_n > z_{1-\frac{\alpha}{2}}) \\ &\approx \Phi(\sigma_0^{-1}h - z_{1-\frac{\alpha}{2}}) \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



It is underlined that the initial approximation is accounted for the condition $h > 0$. Furthermore, the parameter $h$ assumes the value approximately expressed by,

$$h \approx n^{-\frac{1}{2}}\theta$$

Combining these relationships and solving for $n$ contributes to the conclusion that,

$$n = \frac{\sigma_0^2(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2}{\theta^2}$$

Now, it is sufficient to establish that,

$$\sigma_0^2 = W_0^{-2}\zeta_0^2\{q(1-q)\}^{-1}$$

By utilizing the delta method, it emerges that,

$$\sigma_0^2 = W_0^{-2}\tilde{\sigma}_0^2$$

In this context, $\tilde{\sigma}_0^2$ represents the asymptotic variance of $\sqrt{n}(\hat{W}_{1n} - \hat{W}_{0n})$. Ultimately, by employing the decomposition theory of the two-sample $U$-statistic, we derive that,

$$\begin{aligned}\sqrt{n}(\hat{W}_{1n} - \hat{W}_{0n}) &= -\frac{\sqrt{n}}{n_1}\sum_{i=1}^{n_1} R(Y_{1i}) + \frac{\sqrt{n}}{n_0}\sum_{j=1}^{n_0} R(Y_{0j}) + o_p(1) \\ &\xrightarrow{d} \mathcal{N}\{0, q^{-1}(1-q)^{-1}\zeta_0^2\} \quad \text{under } H_0\end{aligned}$$

Consequently, we reach the intended result, and the demonstration is concluded (Mao et al. (2021)). □

# Proof of the theorem outlined in section 5.2

**Theorem 2.** *Let,*

$$A_1^{(k)} = \sum_{i=1}^{I_k} \phi_1(X_i^{(k)})$$

$$B_1^{(l)} = \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)})$$

$$A_2^{(k)} = \sum_{i=1}^{I_k} \phi_2(X_i^{(k)})$$

$$B_2^{(l)} = \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)})$$

*Additionally, we determine,*

https://doi.org/10.26219/heal.aueb.6426



$$\bar{I}_m = \frac{1}{m}\sum_{k=1}^{m} I_k$$

$$\bar{J}_n = \frac{1}{n}\sum_{l=1}^{n} J_l$$

*where $m$ and $n$ represent the numbers of clusters in the two comparison groups. Considering the assumption of independence between comparison groups and clusters, along with the exchangeability of subjects, when $\min(m,n) \to \infty$, the $U$-statistics, namely $U_1 - \bar{I}_m\bar{J}_n\theta_1$ and $U_2 - \bar{I}_m\bar{J}_n\theta_2$, asymptotically follow a bivariate normal distribution with mean 0 and variance,*

$$\Sigma = \frac{\bar{J}_n^2}{m}\Sigma_1 + \frac{\bar{I}_m^2}{n}\Sigma_2$$

*where $\Sigma_1$ and $\Sigma_2$ are matrices encompassing variances and covariances of the defined summation variables,*

$$\Sigma_1 = \begin{bmatrix} Var(A_1^{(k)}) & Cov(A_1^{(k)}, A_2^{(k)}) \\ Cov(A_1^{(k)}, A_2^{(k)}) & Var(A_2^{(k)}) \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} Var(B_1^{(l)}) & Cov(B_1^{(l)}, B_2^{(l)}) \\ Cov(B_1^{(l)}, B_2^{(l)}) & Var(B_2^{(l)}) \end{bmatrix}$$

*Proof.* Initially, we exploit pieces of knowledge from Lee and Dehling (2005), who established that as the minimum of the cluster counts $m$ and $n$ approaches infinity, the following holds,

$$U_1 - \bar{I}_m\bar{J}_n\theta_1 \approx \bar{J}_n\frac{1}{m}\sum_{k=1}^{m}\sum_{i=1}^{I_k}\phi_1(X_i^{(k)}) + \bar{I}_m\frac{1}{n}\sum_{l=1}^{n}\sum_{j=1}^{J_l}\phi_1(Y_j^{(l)})$$

Moreover, Lee and Dehling claimed that the prime term in the above expression asymptotically follows a normal distribution with mean 0 and variance,

$$\bar{J}_n^2\frac{1}{m}\mathrm{Var}\left(\sum_{i=1}^{I_k}\phi_1(X_i^{(k)})\right)$$

Similarly, the second term asymptotically follows a normal distribution with the same mean (0) but another variance,

$$\bar{I}_m^2\frac{1}{n}\mathrm{Var}\left(\sum_{j=1}^{J_l}\phi_1(Y_j^{(l)})\right)$$

https://doi.org/10.26219/heal.aueb.6426



Comparable conclusions apply to $U_2 - \bar{I}_m \bar{J}_n \theta_2$. Specifically,

$$U_2 - \bar{I}_m \bar{J}_n \theta_2 \approx \bar{J}_n \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) + \bar{I}_m \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)})$$

For the initial term,

$$\bar{J}_n \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) \dot{\sim} \mathcal{N}\left(0, \bar{J}_n^2 \frac{1}{m} \mathrm{Var}\left(\sum_{i=1}^{I_k} \phi_2(X_i^{(k)})\right)\right)$$

And for the secondary term,

$$\bar{I}_m \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}) \dot{\sim} \mathcal{N}\left(0, \bar{I}_m^2 \frac{1}{n} \mathrm{Var}\left(\sum_{j=1}^{J_l} \phi_2(Y_j^{(l)})\right)\right)$$

It is commonly known that the clusters utilized in this investigation develop independence, and within each group, participants share an equivalent distribution. Consequently, we can infer that the variables,

$$\sum_{i=1}^{I_k} \phi_1(X_i^{(k)}), \quad k = 1, 2, \ldots, m$$

$$\sum_{j=1}^{J_l} \phi_1(Y_j^{(l)}), \quad l = 1, 2, \ldots, n$$

$$\sum_{i=1}^{I_k} \phi_2(X_i^{(k)}), \quad k = 1, 2, \ldots, m$$

$$\sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}), \quad l = 1, 2, \ldots, n$$

are independent and identically distributed. Continuing the theorem's proof, it is noteworthy to underline that the expected values of the kernel functions $\phi_1$ and $\phi_2$ are zero for individuals in both the treatment and control groups,

$$E(\phi_1(X_i^{(k)})) = E(\phi_1(Y_j^{(l)})) = 0$$

$$E(\phi_2(X_i^{(k)})) = E(\phi_2(Y_j^{(l)})) = 0$$

Additionally, we deem that the vectors $(A_1^{(k)}, A_2^{(k)})$, $k = 1, 2, \ldots, m$, and $(B_1^{(l)}, B_2^{(l)})$, $l = 1, 2, \ldots, n$, follow a bivariate distribution possessing the same mean vector $\big(0, 0\big)^T$ and variance,

$$\begin{bmatrix} \mathrm{Var}(A_1^{(k)}) & \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) \\ \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) & \mathrm{Var}(A_2^{(k)}) \end{bmatrix}$$

https://doi.org/10.26219/heal.aueb.6426



and

$$\begin{bmatrix} \mathrm{Var}(B_1^{(l)}) & \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) \\ \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) & \mathrm{Var}(B_2^{(l)}) \end{bmatrix}$$

respectively. Thus, utilizing this piece of information, we can conclude that,

$$\sqrt{m}\bar{J}_n(\bar{A}_1 + \bar{A}_2 - 0) \dot{\sim} \mathcal{N}\left(0, \bar{J}_n^2\left(\mathrm{Var}(A_1^{(k)}) + \mathrm{Var}(A_2^{(k)}) + 2\mathrm{Cov}(A_1^{(k)}, A_2^{(k)})\right)\right)$$

and

$$\sqrt{n}\bar{I}_m(\bar{B}_1 + \bar{B}_2 - 0) \dot{\sim} \mathcal{N}\left(0, \bar{I}_m^2\left(\mathrm{Var}(B_1^{(l)}) + \mathrm{Var}(B_2^{(l)}) + 2\mathrm{Cov}(B_1^{(l)}, B_2^{(l)})\right)\right)$$

We remark that,

$$\bar{A}_1 = \frac{1}{m}\sum_{k=1}^{m} A_1^{(k)}$$

$$\bar{A}_2 = \frac{1}{m}\sum_{k=1}^{m} A_2^{(k)}$$

$$\bar{B}_1 = \frac{1}{n}\sum_{l=1}^{n} B_1^{(l)}$$

$$\bar{B}_2 = \frac{1}{n}\sum_{l=1}^{n} B_2^{(l)}$$

Carrying out the Cramer–Wold theorem, we can infer that,

$$\sqrt{m}\bar{J}_n(\bar{A}_1 - 0, \bar{A}_2 - 0) \dot{\sim} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \bar{J}_n^2\begin{bmatrix} \mathrm{Var}(A_1^{(k)}) & \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) \\ \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) & \mathrm{Var}(A_2^{(k)}) \end{bmatrix}\right)$$

and

$$\sqrt{n}\bar{I}_m(\bar{B}_1 - 0, \bar{B}_2 - 0) \dot{\sim} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \bar{I}_m^2\begin{bmatrix} \mathrm{Var}(B_1^{(l)}) & \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) \\ \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) & \mathrm{Var}(B_2^{(l)}) \end{bmatrix}\right)$$

The independence among clusters becomes advantageous once more, allowing us to assert that the covariance between the mean terms $\bar{A}_1, \bar{A}_2$ and $\bar{B}_1, \bar{B}_2$ is zero in all combinations. Specifically,

$$\mathrm{Cov}(\bar{A}_r, \bar{B}_r) = 0 \quad \text{with} \quad r = 1, 2$$

Consequently, we arrive at the desired conclusion,

$$(\mathbf{U}_1 - \bar{I}_m\bar{J}_n\boldsymbol{\theta}_1, \mathbf{U}_2 - \bar{I}_m\bar{J}_n\boldsymbol{\theta}_2)^T \dot{\sim} \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \frac{\bar{J}_n^2}{m}\Sigma_1 + \frac{\bar{I}_m^2}{n}\Sigma_2\right)$$

https://doi.org/10.26219/heal.aueb.6426



where,

$$\Sigma_1 = \begin{bmatrix} \mathrm{Var}(A_1^{(k)}) & \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) \\ \mathrm{Cov}(A_1^{(k)}, A_2^{(k)}) & \mathrm{Var}(A_2^{(k)}) \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} \mathrm{Var}(B_1^{(l)}) & \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) \\ \mathrm{Cov}(B_1^{(l)}, B_2^{(l)}) & \mathrm{Var}(B_2^{(l)}) \end{bmatrix}$$

(Zhang and Jeong (2021)). □

## Establishing the covariance between $U_1$ and $U_2$ in section 5.2

Our objective is to demonstrate the formulation for the covariance between $U_1$ and $U_2$, as detailed in section 5.2. To accomplish this, we will avail the determinations of the kernel functions $\phi_1(X_i^{(k)}, Y_j^{(l)})$ and $\phi_2(X_i^{(k)}, Y_j^{(l)})$, expressed in equations (5.1) and (5.2) respectively. Employing these definitions, we can state the covariance between $U_1$ and $U_2$ as follows,

$$\begin{aligned} \mathrm{Cov}(U_1, U_2) = \mathrm{Cov}\Bigg\{ & \bar{J}_n \cdot \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_1(X_i^{(k)}) + \bar{I}_m \cdot \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)}), \\ & \bar{J}_n \cdot \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) + \bar{I}_m \cdot \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}) \Bigg\} \end{aligned}$$

For our convenience, we carry out the following assumptions,

$$\bar{J}_n \cdot \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_1(X_i^{(k)}) = \alpha$$

$$\bar{I}_m \cdot \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)}) = \beta$$

$$\bar{J}_n \cdot \frac{1}{m} \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) = \gamma$$

$$\bar{I}_m \cdot \frac{1}{n} \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}) = \delta$$

Thus, the covariance of $U_1$ and $U_2$ can be articulated as,

$$\begin{aligned} \mathrm{Cov}(U_1, U_2) &= \mathrm{Cov}(\alpha + \beta, \gamma + \delta) \\ &= \mathrm{Cov}(\alpha, \gamma) + \mathrm{Cov}(\beta, \delta) + \mathrm{Cov}(\alpha, \delta) + \mathrm{Cov}(\beta, \gamma) \end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



Among the clusters, independence dominates. Hence, it can be claimed that,

$$\mathrm{Cov}(\alpha, \delta) = \mathrm{Cov}(\beta, \gamma) = 0$$

Regarding the remaining two covariances, it yields that,

$$\begin{aligned}
\mathrm{Cov}(\alpha, \gamma) &= \frac{\bar{J}_n^2}{m^2} \mathrm{Cov}\left( \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_1(X_i^{(k)}), \sum_{k=1}^{m} \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) \right) \\
&\overset{\text{"independent clusters"}}{=} \frac{\bar{J}_n^2}{m^2} \sum_{k=1}^{m} \mathrm{Cov}\left( \sum_{i=1}^{I_k} \phi_1(X_i^{(k)}), \sum_{i=1}^{I_k} \phi_2(X_i^{(k)}) \right) \\
&= \frac{\bar{J}_n^2}{m^2} \sum_{k=1}^{m} \left\{ I_k \, \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_i^{(k)})) + I_k(I_k - 1) \, \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_{i'}^{(k)})) \right\} \\
&= \frac{\bar{J}_n^2}{m} \left\{ \bar{I}_m \, \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_i^{(k)})) + \bar{I}_m^{(2)} \, \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_{i'}^{(k)})) \right\}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(\beta, \delta) &= \frac{\bar{I}_m^2}{n^2} \mathrm{Cov}\left( \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)}), \sum_{l=1}^{n} \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}) \right) \\
&\overset{\text{"independent clusters"}}{=} \frac{\bar{I}_m^2}{n^2} \sum_{l=1}^{n} \mathrm{Cov}\left( \sum_{j=1}^{J_l} \phi_1(Y_j^{(l)}), \sum_{j=1}^{J_l} \phi_2(Y_j^{(l)}) \right) \\
&= \frac{\bar{I}_m^2}{n^2} \sum_{l=1}^{n} \left\{ J_l \, \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_j^{(l)})) + J_l(J_l - 1) \, \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)})) \right\} \\
&= \frac{\bar{I}_m^2}{n} \left\{ \bar{J}_n \, \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_j^{(l)})) + \bar{J}_n^{(2)} \, \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)})) \right\}
\end{aligned}$$

Ultimately, as the quantities $m$ and $n$ approach infinity, signified as $m, n \to \infty$, the asymptotic covariance of $U_1$ and $U_2$ is obtained.

$$\begin{aligned}
\mathrm{Cov}(U_1, U_2) = &\frac{\bar{J}_n^2}{m} \left( \bar{I}_m \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_i^{(k)})) + \bar{I}_m^{(2)} \mathrm{Cov}(\phi_1(X_i^{(k)}), \phi_2(X_{i'}^{(k)})) \right) \\
&+ \frac{\bar{I}_m^2}{n} \left( \bar{J}_n \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_j^{(l)})) + \bar{J}_n^{(2)} \mathrm{Cov}(\phi_1(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)})) \right)
\end{aligned}$$

where,

$$\bar{I}_m^{(2)} = \frac{1}{m} \sum_{k=1}^{m} I_k(I_k - 1)$$

and

$$\bar{J}_n^{(2)} = \frac{1}{n} \sum_{l=1}^{n} J_l(J_l - 1)$$

(Zhang and Jeong (2021)).

https://doi.org/10.26219/heal.aueb.6426



# Investigation of the variances and covariances of the kernel functions in section 5.2

Our purpose is to investigate and clarify the variances and covariances related to the kernel functions outlined in section 5.2. Initially, we initiated the research with the delineation,

$$\hat{\theta}_1 = \frac{1}{(\sum_{k=1}^{m} I_k)(\sum_{l=1}^{n} J_l)} \sum_{k=1}^{m} \sum_{l=1}^{n} \sum_{i=1}^{I_k} \sum_{j=1}^{J_l} I\{X_i^{(k)} > Y_j^{(l)}\}$$

Afterward, we proceed to determine the elements incorporated in the variance of $U_1$, denoted as $\mathrm{Var}(U_1)$. Specifically, it holds that,

$$\begin{aligned}
\widehat{\mathrm{Var}}(\phi_1(X_i^{(k)})) &= \widehat{\mathrm{Var}}(P_Y(X_i^{(k)} > Y_j^{(l)})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} > Y_{j'}^{(l')}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_i^{(k)} > Y_{j'}^{(l')}) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} > Y_{j'}^{(l')}) - [P(X_i^{(k)} > Y_j^{(l)})]^2 \\
&= \frac{\sum_{k=1}^{m} \sum_{i=1}^{I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} > Y_{j'}^{(l')}\}}{(\sum_{k=1}^{m} I_k)\,(\sum_{1 \le l \ne l' \le n} J_l J'_l)} - (\hat{\theta}_1)^2 \\
&= \frac{\sum_{k=1}^{m} \sum_{i=1}^{I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} > Y_{j'}^{(l')}\}}{m\bar{I}_m \left(n^2(\bar{J}_n)^2 - n\bar{J}_n^{(2)} - n\bar{J}_n\right)} - (\hat{\theta}_1)^2
\end{aligned}$$

Here,

$$\begin{aligned}
1 \le k \le m \\
1 \le l \ne l' \le n \\
1 \le i \le I_k \\
1 \le j \le J_l \\
1 \le j' \le J'_l
\end{aligned}$$

$$\begin{aligned}
\widehat{\mathrm{Var}}(\phi_1(Y_j^{(l)})) &= \widehat{\mathrm{Var}}(P_X(X_i^{(k)} > Y_j^{(l)})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_j^{(l)}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k')} > Y_j^{(l)}) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_j^{(l)}) - [P(X_i^{(k)} > Y_j^{(l)})]^2 \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \le k \ne k' \le m} \sum_{j=1}^{J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_j^{(l)}\}}{(\sum_{l=1}^{n} J_l)\,(\sum_{1 \le k \ne k' \le m} I_k I'_k)} - (\hat{\theta}_1)^2 \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \le k \ne k' \le m} \sum_{j=1}^{J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_j^{(l)}\}}{n\bar{J}_n \left(m^2(\bar{I}_m)^2 - m\bar{I}_m^{(2)} - m\bar{I}_m\right)} - (\hat{\theta}_1)^2
\end{aligned}$$

https://doi.org/10.26219/heal.aueb.6426



where,

$$\begin{gathered} 1 \leq k \neq k' \leq m \\ 1 \leq l \leq n \\ 1 \leq i \leq I_k \\ 1 \leq i' \leq I'_k \\ 1 \leq j \leq J_l \end{gathered}$$

$$\begin{aligned} &\widehat{\mathrm{Cov}}(\phi_1(X_i^{(k)}), \phi_1(X_{i'}^{(k)})) \\ &= \widehat{\mathrm{Cov}}(P_Y(X_i^{(k)} > Y_j^{(l)}), P_Y(X_{i'}^{(k)} > Y_{j'}^{(l')})) \\ &= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} > Y_{j'}^{(l')}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k)} > Y_{j'}^{(l')}) \\ &= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} > Y_{j'}^{(l')}) - [P(X_i^{(k)} > Y_j^{(l)})]^2 \\ &= \frac{\sum_{k=1}^{m} \sum_{1 \leq i \neq i' \leq I_k} \sum_{1 \leq l \neq l' \leq n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} > Y_{j'}^{(l')}\}}{\left(\sum_{k=1}^{m} I_k(I_k - 1)\right) \left(\sum_{1 \leq l \neq l' \leq n} J_l J'_l\right)} - (\hat{\theta}_1)^2 \\ &= \frac{\sum_{k=1}^{m} \sum_{1 \leq i \neq i' \leq I_k} \sum_{1 \leq l \neq l' \leq n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} > Y_{j'}^{(l')}\}}{m \bar{I}_m^{(2)} \left(n^2 (\bar{J}_n)^2 - n \bar{J}_n^{(2)} - n \bar{J}_n\right)} - (\hat{\theta}_1)^2 \end{aligned}$$

with $1 \leq k \leq m$; $1 \leq l \neq l' \leq n$; $1 \leq i \neq i' \leq I_k$; $1 \leq j \leq J_l$; $1 \leq j' \leq J'_l$.

$$\begin{aligned} &\widehat{\mathrm{Cov}}(\phi_1(Y_j^{(l)}), \phi_1(Y_{j'}^{(l)})) \\ &= \widehat{\mathrm{Cov}}(P_X(X_i^{(k)} > Y_j^{(l)}), P_X(X_{i'}^{(k')} > Y_{j'}^{(l)})) \\ &= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_{j'}^{(l)}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k')} > Y_{j'}^{(l)}) \\ &= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_{j'}^{(l)}) - [P(X_i^{(k)} > Y_j^{(l)})]^2 \\ &= \frac{\sum_{l=1}^{n} \sum_{1 \leq k \neq k' \leq m} \sum_{1 \leq j \neq j' \leq J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_{j'}^{(l)}\}}{\left(\sum_{l=1}^{n} J_l(J_l - 1)\right) \left(\sum_{1 \leq k \neq k' \leq m} I_k I'_k\right)} - (\hat{\theta}_1)^2 \\ &= \frac{\sum_{l=1}^{n} \sum_{1 \leq k \neq k' \leq m} \sum_{1 \leq j \neq j' \leq J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} > Y_{j'}^{(l)}\}}{n \bar{J}_n^{(2)} \left(m^2 (\bar{I}_m)^2 - m \bar{I}_m^{(2)} - m \bar{I}_m\right)} - (\hat{\theta}_1)^2 \end{aligned}$$

It is worthwhile to note that $1 \leq k \neq k' \leq m$; $1 \leq l \leq n$; $1 \leq i \leq I_k$; $1 \leq i' \leq I'_k$; $1 \leq j \neq j' \leq J_l$. Similar outcomes are accomplished for the components involved in the variance of $U_2$, represented as $\mathrm{Var}(U_2)$.

Concerning the elements comprised in the expression of the covariance between

https://doi.org/10.26219/heal.aueb.6426



$U_1$ and $U_2$, indicated as $\mathrm{Cov}(U_1, U_2)$, the following formulations are applicable,

$$
\begin{aligned}
&\widehat{\mathrm{Cov}}(\phi_1(X_i^{(k)}), \phi_2(X_i^{(k)})) \\
&= \widehat{\mathrm{Cov}}(P_Y(X_i^{(k)} > Y_j^{(l)}), P_Y(X_i^{(k)} < Y_{j'}^{(l')})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} < Y_{j'}^{(l')}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_i^{(k)} < Y_{j'}^{(l')}) \\
&= \frac{\sum_{k=1}^{m} \sum_{i=1}^{I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} < Y_{j'}^{(l')}\}}{\left(\sum_{k=1}^{m} I_k\right) \left(\sum_{1 \le l \ne l' \le n} J_l J'_l\right)} - (\hat{\theta}_1)(\hat{\theta}_2) \\
&= \frac{\sum_{k=1}^{m} \sum_{i=1}^{I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_i^{(k)} < Y_{j'}^{(l')}\}}{m \bar{I}_m \left(n^2 (\bar{J}_n)^2 - n \bar{J}_n^{(2)} - n \bar{J}_n\right)} - (\hat{\theta}_1)(\hat{\theta}_2)
\end{aligned}
$$

It is given that $1 \le k \le m; 1 \le l \ne l' \le n; 1 \le i \le I_k; 1 \le j \le J_l; 1 \le j' \le J'_l$.

$$
\begin{aligned}
&\widehat{\mathrm{Cov}}(\phi_1(Y_j^{(l)}), \phi_2(Y_j^{(l)})) \\
&= \widehat{\mathrm{Cov}}(P_X(X_i^{(k)} > Y_j^{(l)}), P_X(X_{i'}^{(k')} < Y_j^{(l)})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_j^{(l)}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k')} < Y_j^{(l)}) \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \le k \ne k' \le m} \sum_{j=1}^{J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_j^{(l)}\}}{\left(\sum_{l=1}^{n} J_l\right) \left(\sum_{1 \le k \ne k' \le m} I_k I'_k\right)} - (\hat{\theta}_1)(\hat{\theta}_2) \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \le k \ne k' \le m} \sum_{j=1}^{J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_j^{(l)}\}}{n \bar{J}_n \left(m^2 (\bar{I}_m)^2 - m \bar{I}_m^{(2)} - m \bar{I}_m\right)} - (\hat{\theta}_1)(\hat{\theta}_2)
\end{aligned}
$$

where $1 \le k \ne k' \le m$; $1 \le l \le n$; $1 \le i \le I_k$; $1 \le i' \le I'_k$; $1 \le j \le J_l$.

$$
\begin{aligned}
&\widehat{\mathrm{Cov}}(\phi_1(X_i^{(k)}), \phi_2(X_{i'}^{(k)})) \\
&= \widehat{\mathrm{Cov}}(P_Y(X_i^{(k)} > Y_j^{(l)}), P_Y(X_{i'}^{(k)} < Y_{j'}^{(l')})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} < Y_{j'}^{(l')}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k)} < Y_{j'}^{(l')}) \\
&= \frac{\sum_{k=1}^{m} \sum_{1 \le i \ne i' \le I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} < Y_{j'}^{(l')}\}}{\left(\sum_{k=1}^{m} I_k (I_k - 1)\right) \left(\sum_{1 \le l \ne l' \le n} J_l J'_l\right)} - (\hat{\theta}_1)(\hat{\theta}_2) \\
&= \frac{\sum_{k=1}^{m} \sum_{1 \le i \ne i' \le I_k} \sum_{1 \le l \ne l' \le n} \sum_{j=1}^{J_l} \sum_{j'=1}^{J'_l} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k)} < Y_{j'}^{(l')}\}}{m \bar{I}_m^{(2)} \left(n^2 (\bar{J}_n)^2 - n \bar{J}_n^{(2)} - n \bar{J}_n\right)} - (\hat{\theta}_1)(\hat{\theta}_2)
\end{aligned}
$$

https://doi.org/10.26219/heal.aueb.6426



with $1 \leq k \leq m$; $1 \leq l \neq l' \leq n$; $1 \leq i \neq i' \leq I_k$; $1 \leq j \leq J_l$; $1 \leq j' \leq J'_l$.

$$
\begin{aligned}
&\widehat{\mathrm{Cov}}(\phi_1(Y_j^{(l)}), \phi_2(Y_{j'}^{(l)})) \\
&= \widehat{\mathrm{Cov}}(P_X(X_i^{(k)} > Y_j^{(l)}), P_X(X_{i'}^{(k')} < Y_{j'}^{(l)})) \\
&= P(X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_{j'}^{(l)}) - P(X_i^{(k)} > Y_j^{(l)}) P(X_{i'}^{(k')} < Y_{j'}^{(l)}) \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \leq k \neq k' \leq m} \sum_{1 \leq j \neq j' \leq J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_{j'}^{(l)}\}}{\left(\sum_{l=1}^{n} J_l(J_l - 1)\right) \left(\sum_{1 \leq k \neq k' \leq m} I_k I'_k\right)} - (\hat{\theta}_1)(\hat{\theta}_2) \\
&= \frac{\sum_{l=1}^{n} \sum_{1 \leq k \neq k' \leq m} \sum_{1 \leq j \neq j' \leq J_l} \sum_{i=1}^{I_k} \sum_{i'=1}^{I'_k} I\{X_i^{(k)} > Y_j^{(l)}, X_{i'}^{(k')} < Y_{j'}^{(l)}\}}{n \bar{J}_n^{(2)} \left(m^2 (\bar{I}_m)^2 - m \bar{I}_m^{(2)} - m \bar{I}_m\right)} - (\hat{\theta}_1)(\hat{\theta}_2)
\end{aligned}
$$

It is highlighted that $1 \leq k \neq k' \leq m$; $1 \leq l \leq n$; $1 \leq i \leq I_k$; $1 \leq i' \leq I'_k$; $1 \leq j \neq j' \leq J_l$ (Zhang and Jeong (2021)).

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# Bibliography

Abdalla, S., Montez-Rath, M. E., Parfrey, P. S., and Chertow, G. M. (2016). The win ratio approach to analyzing composite outcomes: An application to the evolve trial. *Contemporary Clinical Trials*, 48:119–124.

Baracaldo-Santamaría, D., Feliciano-Alfonso, J. E., Ramirez-Grueso, R., Rojas-Rodríguez, L. C., Dominguez-Dominguez, C. A., and Calderon-Ospina, C. A. (2023). Making sense of composite endpoints in clinical research. *Journal of Clinical Medicine*, 12(13):4371.

Bebu, I. and Lachin, J. M. (2015). Large sample inference for a win ratio analysis of a composite outcome based on prioritized components. *Biostatistics*, 17(1):178–187.

Boos, D. D. (2003). Introduction to the bootstrap world. *Statistical Science*, 18(2):168–174.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10:101–129.

Dong, G., Hoaglin, D. C., Qiu, J., Matsouaka, R. A., Chang, Y.-W., Wang, J., and Vandemeulebroecke, M. (2020). The win ratio: On interpretation and handling of ties. *Statistics in Biopharmaceutical Research*, 12(1):99–106.

Dong, G., Li, D., Ballerstedt, S., and 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*.

Dong, G., Qiu, J., Wang, D., and Vandemeulebroecke, M. (2018). The stratified win ratio. *Journal of Biopharmaceutical Statistics*, 28(4):778–796.

Finkelstein, D. M. and Schoenfeld, D. A. (1999). Combining mortality and longitudinal measures in clinical trials. *Statistics in Medicine*, 18(11):1341–1354.

Finkelstein, D. M. and Schoenfeld, D. A. (2018). Graphing the win ratio and its components over time. *Statistics in Medicine*.

Fook Chong Stephanie, C. R. (2011). Introduction to bootstrap. *Proceedings of Singapore Healthcare*, 20(3):2011.

https://doi.org/10.26219/heal.aueb.6426



Giné, E., Latala, R., and Zinn, J. (2000). Exponential and moment inequalities for u-statistics. Working Paper 47, Progress in Probability.

Gómez-Mateu, M. (2016). *Composite Endpoints in Clinical Trials: Computational Tools, Practical Guidelines, and Methodological Extensions*. PhD thesis, Universitat Politècnica de Catalunya, Departament d'Estadística i Investigació Operativa.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325.

Houdré, C. and Reynaud-Bouret, P. (2003). Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic Inequalities and Applications*, volume 56 of *Progress in Probability*, pages 55–69.

Huque, M. F., Alosh, M., and Guerra, M. (2018). Consistency-ensured parametric tests for critical events of composite endpoints. *Journal of Biopharmaceutical Statistics*, 28(1):82–98.

Johnson, R. W. (2001). An introduction to the bootstrap. *Teaching Statistics*, 23(2):49–54.

Lee, M.-L. T. and Dehling, H. G. (2005). Generalized two-sample u-statistics for clustered data. *Statistica Neerlandica*, 59(3):313–323.

Liao, R., Chakladar, S., and Gamalo, M. (2022). Win ratio approach for analyzing composite time-to-event endpoint with opposite treatment effects in its components. *Pharmaceutical Statistics*, 21(6):1342–1356.

Luo, X., Tian, H., Mohanty, S., and Tsai, W. Y. (2015). An alternative approach to confidence interval estimation for the win ratio statistic. *Biometrics*, 71(1):139–145.

Mao, L., Kim, K., and Li, Y. (2022). On recurrent-event win ratio. *Statistical Methods in Medical Research*, 31(6):1120–1134.

Mao, L., Kim, K., and Miao, X. (2021). Sample size formula for general win ratio analysis. *Biometrics*, 78(3):1257–1268.

McCoy C. Eric, MD, M. (2018). Understanding the use of composite endpoints in clinical trials. *Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health*, 19(4).

Pocock, S. J., Ariti, C. A., Collier, T. J., and 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*, 33:176–182.

Ramchandani, R., Schoenfeld, D. A., and Finkelstein, D. M. (2016). Global rank tests for multiple, possibly censored, outcomes. *Biometrics*, 72(3):926–935.

https://doi.org/10.26219/heal.aueb.6426



Sankoh, A. J., Li, H., and D'Agostino, Ralph B., S. (2014). Use of composite endpoints in clinical trials. *Statistics in Medicine*, 33(27):4709–4714.

Wang, D. and Pocock, S. (2016). A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharmaceutical Statistics*, 15(3):238–245.

Wei, L. J. and Johnson, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika*, 72(2):359–364.

William, P. and Nicolas, G. (2010). Clinical applications of kinase inhibitors in solid tumors. In *Handbook of Cell Signaling*, chapter 82, pages 615–631. Second edition.

Wittkowski, K. M., Lee, E., Nussbaum, R., Chamian, F. N., and Krueger, J. G. (2004). Combining several ordinal measures in clinical studies. *Statistics in Medicine*, 23:1579–1592.

Yu, R. X. and Ganju, J. (2022). Sample size formula for a win ratio endpoint. *Statistics in Medicine*, 41(6):950–963.

Zhang, D. and Jeong, J.-H. (2021). Inference on win ratio for cluster-randomized semi-competing risk data. *Japanese Journal of Statistics and Data Science*, 4(2):1263–1292.

https://doi.org/10.26219/heal.aueb.6426



https://doi.org/10.26219/heal.aueb.6426

# Web References

Centers for Disease Control and Prevention. (2023). Breast Cancer: What Is Breast Cancer? Retrieved from https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm

Cleveland Clinic. (2023). Breast Cancer. Retrieved from https://my.clevelandclinic.org/health/diseases/3986-breast-cancer

Hygeia. (2020). Νεότερες θεραπείες του καρκίνου του μαστού: Ποιες δυνατότητες υπάρχουν και σε ποιες περιπτώσεις βοηθούν. Retrieved from https://www.hygeia.gr/neoteres-therapeies-toy-karkinoy-toy-mastoy-poies-dynatotites-yparxoyn-kai-se-poies-periptoseis-voithoyn/

Mayo Clinic. (2022). Breast Cancer: Symptoms & Causes. Retrieved from https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470

National Institutes of Health. (2017). NIH Clinical Trials Definition. Retrieved from https://grants.nih.gov/policy/clinical-trials/definition.htm

National Library of Medicine. (2015). Longitudinal studies. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4669300/

U.S. Food and Drug Administration. (2022). Multiple Endpoints in Clinical Trials Guidance for Industry. Retrieved from https://www.fda.gov/media/162416/download

World Health Organization. (2020). WHO Questions and Answers on Clinical Trials. Retrieved from https://www.who.int/news-room/questions-and-answers/item/clinical-trials