British Cardiovascular Society
'Promoting excellence in cardiovascular care'

Composite outcomes and making sense of the 'win ratio'

Kayla Chiew MBBS BSc MRCP

Cardiology ACF ST3

Barts Health NHS Trust

Take Home Messages

- Composite outcomes are commonly used in clinical trials to reduce sample size and increase statistical power.
- The traditional approach to evaluating composite outcomes fails to account for relative clinical priority and recurrent events.
- Hierarchical composite endpoints have been devised to overcome this, with the 'win ratio' being a method of analysis that uses pairwise comparisons.
- Understanding its advantages and limitations is important for clinicians as the adoption of the win ratio increases.

Introduction

Global improvements in healthcare provision over time has resulted in event rates in clinical trials, such as mortality and hospitalisations, to decline in frequency. Adequately powering these studies to detect meaningful differences between treatment arms (if a true difference exists) therefore mandates recruiting larger sample sizes at the expense of time, labour, and costs. Using composite as opposed to single endpoints has been popularised to overcome this barrier. The development of hierarchical endpoints and the method of the 'win ratio' has increasingly gained favour as a more sensitive way to evaluate composite outcomes. We discuss the nuances of composite endpoint analysis, how the win ratio works, and review examples of recent clinical trials using this methodology.

Composite endpoints

In cardiovascular clinical trials, the primary composite endpoint is commonly chosen to be 'major adverse cardiovascular events' (MACE), which frequently includes cardiovascular death, non-fatal stroke and non-fatal myocardial infarction. This has several advantages including increasing statistical power due to the higher total number of events, overcoming multiplicity (inflations in

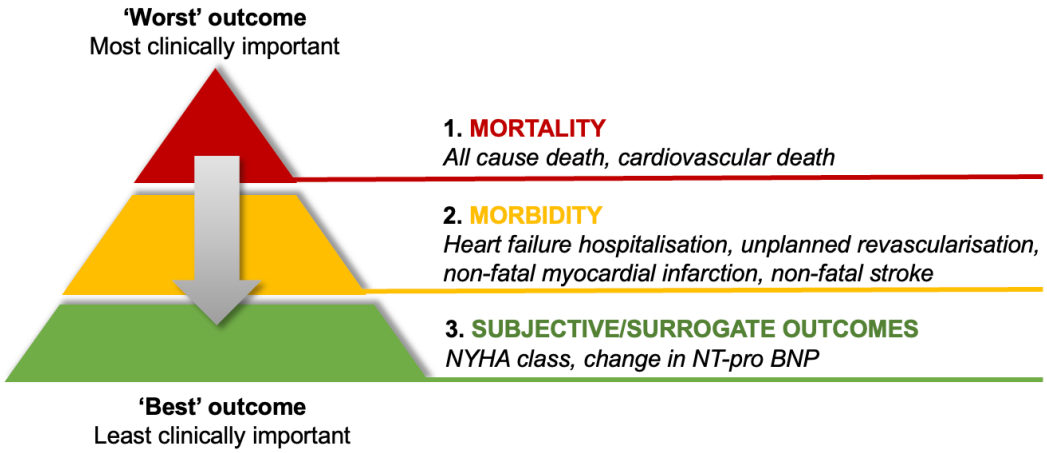
type 1 error from multiple testing), avoiding issues related to competing risks, and being able to evaluate the benefits of a treatment more holistically.

Evaluating composite outcomes however relies on a number of assumptions, such as the equal contribution of each component to the overall effect size, which may not reflect clinical importance. An individual is considered to have reached the primary outcome when any component of the composite first occurs. This traditional approach is therefore disproportionally influenced by non-fatal events which frequently occur before death, under-represents fatal events, and ignores recurrent events(1). For example, an individual who experiences a non-fatal myocardial infarction at 3 months, with subsequent death at 6 months, will be scored on their first event with the latter event ignored. Increasing awareness of these limitations has prompted the development of novel approaches to analysing composite endpoints.

Hierarchical endpoints

Hierarchical endpoints assign an order of importance to the components of the composite. Braunwald and colleagues were early adopters of this approach, combining mortality with other non-fatal complications and assigning arbitrary weights to each event during evaluation of fibrinolysis for the treatment of acute myocardial infarctions(2). Choosing the hierarchical order of importance is guided by clinical reasoning, factoring in event frequency, severity, and time-to-event. Commonly, mortality is favoured as the most severe and important, followed by "softer" clinical endpoints such as heart failure hospitalisations, and lastly subjective endpoints such as patient-reported symptoms (**Figure 1**).

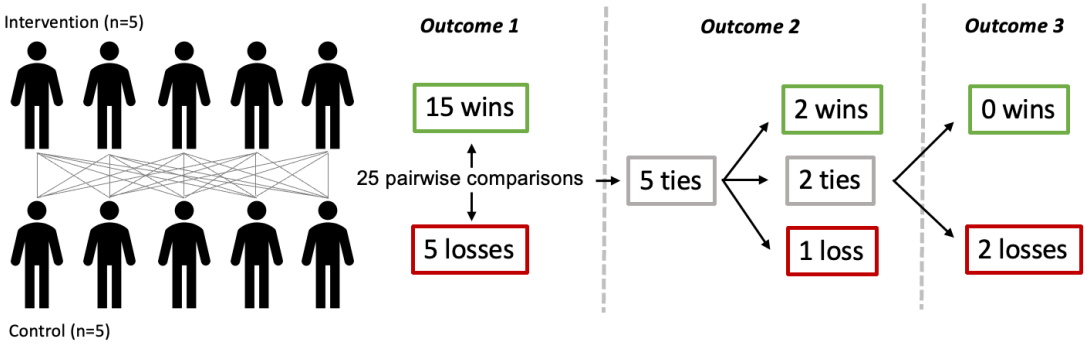
Figure 1. Commonly used hierarchy for the components of a clinical composite endpoint. *Original figure produced by K Chiew, Feb 2023.*



The win ratio

The win ratio is a method of analysing the hierarchical composite endpoint using pairwise comparisons. Introduced in 2012 into the arena of cardiovascular clinical trials by Pocock et al(3), it takes into account relative priorities of the components. Starting with the outcome at the top of the hierarchy, every individual in one arm is directly compared with every individual in the second arm. This generates a certain number of 'wins' if the adverse event is avoided, 'ties' if both individuals experience the same event, or 'losses' if the adverse event is achieved. All individuals who received a 'tie' will then move on to have pairwise comparisons for the 2nd outcome in the hierarchical chain. The win ratio is therefore calculated as the total number of wins after performing all pairwise comparisons down the hierarchy, divided by the total number of losses (**Figure 2**). It is therefore a relative measure, and represents the *odds* of doing better in the intervention group compared to the control. A win ratio of 1.4 means that when comparing an individual from the intervention to the control group, the odds that the individual in the intervention group fares better is 1.4. The estimated *probability* for this win is thus 1.4/(1+1.4) = 0.58, or 58%. In contrast, a win ratio less than one means that the outcome is less favourable for the intervention compared to control group.

Figure 2. Hypothetical example of the win ratio calculation involving 25 pairwise comparisons. *Original figure produced by K Chiew, Feb 2023.*



$$\text{Win ratio} = \frac{\text{No. of wins}}{\text{No. of losses}} = \frac{17}{8} = 2.125$$ (*Odds* of intervention group faring better)

$$\text{Probability} = \frac{2.125}{1 + 2.125} = 0.68 \text{ or } 68\%$$ (*Probability* of intervention group faring better)

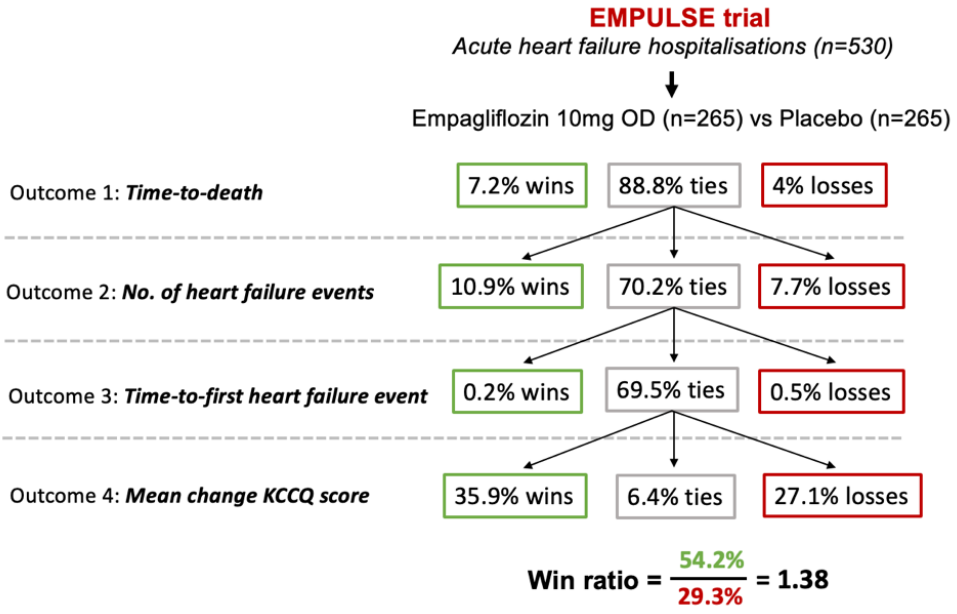
It is important to note that assessing the components of the win ratio at any other level besides the first is no longer comparing all individuals in that trial. The win ratio in fact ignores all those who 'tied', allowing the ratio to be dictated by a small number of clinical events should there be a very large proportion of ties in the trial. It is for this reason that the win ratio should not be mistakenly interpreted as a hazard ratio, and cannot be used to depict overall treatment effect.

The win ratio in practice: an analysis of the EMPULSE trial

530 individuals with an acute heart failure hospitalisation were randomised to receive 10mg empagliflozin or placebo in EMPULSE(4). The primary outcome was a hierarchical composite of time-to-death, number of heart failure events, time-to-first heart failure event, and mean change in Kansas City Cardiomyopathy Questionnaire (KCCQ) score at 90 days (**Figure 3**). The third step in this win ratio analysis, time-to-heart failure event, is arguably superfluous as it actually adds very little (0.7%) to the overall proportion of wins and losses. This is because this step is dependent on patient pairs who had the same number of heart failure events (i.e. all those who 'tied' in step 2), the majority of whom did not have a heart failure event at all. The 70.2% ties in step 2 only marginally reduces to 69.5% in step 3. In fact, the most influential step swaying the

win ratio was the lowest ranked outcome (step 4) of change in KCCQ score which had 35.9% wins, compared to 27.1% for mortality and heart failure events (steps 1-3) combined. The win ratio for mortality and heart failure events, excluding KCCQ score, is 1.50 (95% CI 0.99-2.26), demonstrating consistency with the overall win ratio of 1.38. It is however underpowered to reach a definitive conclusion as the confidence intervals cross unity(5). Using a hierarchical composite outcome that has multiple levels has thus helped to increase the statistical power of the trial overall. Careful consideration must be taken when interpreting the results however, as statistical significance may be "driven by" the least clinically important event. A number of other recent trials utilising the win ratio are summarised in **Table 1.**

Figure 3. Hierarchical composite endpoint analysis with win ratio calculation for the EMPULSE trial. *Data from Voors et al(4), figure adapted from Pocock et al(5).*



Conclusion

The development of a hierarchy has allowed for a more refined method of analysing composite outcomes, considering clinical priorities, time-to-event, and recurrent events. The win ratio method of pairwise comparisons has seen a rapid increase in popularity in the recent years,

providing a means of combining disparate outcomes into a single treatment measure, and can be used as a tool to reduce sample size compared to event-driven trials.

Disclosures

No relevant disclosures.

References

1. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite Outcomes in Randomized Trials Greater Precision But With Greater Uncertainty? JAMA. 2003 May 21;289(19):2554–9.

2. Braunwald E, Cannon CP, McCabe CH. Use of composite endpoints in thrombolysis trials of acute myocardial infarction. American Journal of Cardiology. 1993 Dec 16;72(19):G3–12.

3. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. European Heart Journal. 2012 Jan 1;33(2):176–82.

4. Voors AA, Angermann CE, Teerlink JR, Collins SP, Kosiborod M, Biegus J, et al. The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial. Nat Med. 2022 Mar;28(3):568–74.

5. Pocock SJ, Ferreira JP, Collier TJ, Angermann CE, Biegus J, Collins SP, et al. The win ratio method in heart failure trials: lessons learnt from EMPULSE. European Journal of Heart Failure. 2023;25(5):632–41.

Table 1. Recent cardiovascular clinical trials utilising the win ratio.

Trial name	Population	Intervention vs control	Hierarchical composite endpoint	Win ratio
Primary outcome analysis				
ATTRibute-CM	Transthyretin amyloid cardiomyopathy	Acoramidis 800mg BD vs placebo	1. All-cause death 2. No. of CV hospitalisations 3. Change in NT-proBNP 4. Change in 6MWT	1.77 95% CI 1.42-2.22 p<0.0001
DAPA MI	Acute myocardial infarction, LVEF<40%, no diabetes	Dapagliflozin 10mg OD vs placebo	1. All-cause death 2. Heart failure hospitalisation 3. Non-fatal MI 4. Atrial fibrillation/flutter 5. Type 2 diabetes 6. NYHA class 7. Body weight reduction of 5%	1.34 95% CI 1.20-1.50 p<0.001
HEART FID	LVEF<40%, iron deficiency	IV ferric carboxymaltose every 6 months vs placebo	1. All-cause death 2. No. of heart failure hospitalisations 3. Change in 6MWT	1.10 95% CI 0.99-1.23 p=0.02
PARTNER 3 (5-year follow up)	Severe symptomatic aortic stenosis, low surgical risk	TAVR vs SAVR	1. All-cause death 2. Non-fatal stroke 3. Hospitalisations related to the valve, procedure, or heart failure	1.17 95% CI 0.90-1.51 p=0.25
TRILUMINATE	Severe symptomatic tricuspid regurgitation	Tricuspid TEER vs medical therapy	1. All-cause death or tricuspid valve surgery 2. Heart failure hospitalisation 3. Change in KCCQ score	1.48 95% CI 1.06-2.13 p=0.02
Non-primary outcome analysis				
STEP-HFpEF	HFpEF, BMI>30	Semaglutide 2.4mg once a week vs placebo	*Secondary outcome:* 1. All-cause death 2. No. of heart failure events 3. Time-to-heart failure event 4. Change in KCCQ score 5. Change in 6MWT	1.72 95% CI 1.37-2.15 p<0.001
PARADISE-MI	Acute myocardial infarction, LVEF<40% or pulmonary congestion	Salcubitril-valsartan (97/103mg BD) vs Ramipril 5mg BD	*Secondary analysis:* 1. Cardiovascular death 2. No. of heart failure events 3. Time-to-heart failure hospitalisation 4. Time-to-outpatient heart failure event	1.17 95% CI 1.03-1.33 p=0.015

6MWT = 6-minute walk test, CV = cardiovascular, HFpEF = heart failure with preserved ejection fraction, IV = intravenous, KCCQ = Kansas City Cardiomyopathy Questionnaire, MI = myocardial infarction, SAVR = surgical aortic valve replacement, TAVR = transcatheter aortic valve replacement, TEER = transcatheter edge-to-edge repair