












The win ratio in cardiology trials: lessons learnt, new developments, and wise future use

Stuart J. Pocock ^{1,*}, John Gregson ¹, Timothy J. Collier ¹,
Joao Pedro Ferreira ^{2,3,4,5}, and Gregg W. Stone ⁶

¹Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK; ²Centre d'Investigations Cliniques Plurithématique 1433, INSERM, Université de Lorraine, Nancy, France; ³F-CRIN INI-CRCT (Cardiovascular and Renal Clinical Trialists), INSERM U1116, Centre Hospitalier Régional Universitaire de Nancy, Nancy, France; ⁴UnIC@RISE, Cardiovascular Research and Development Center, Department of Surgery and Physiology, Faculty of Medicine of the University of Porto, Porto, Portugal; ⁵Heart Failure Clinic, Internal Medicine Department, Unidade Local de Saúde de Gaia, Espinho, Portugal; and ⁶Department of Medicine, The Zena and Michael A Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Received 16 May 2024; revised 9 July 2024; accepted 12 September 2024

Graphical Abstract

Recommendations for future use of the win ratio in design, analysis and reporting of clinical trials	
Topic	Recommendation
 Conventional composite outcomes	Conventional composite outcomes can be misleading. They misguidedly give equal priority to all components, and time-to-first-event analysis can give less important components undue influence.
 Hierarchical composite outcomes	Hierarchical composite outcomes recognize the clinical priorities amongst components and hence better capture an overall treatment effect.
 Understanding the win ratio	The win ratio (with CI and P-value) estimates the ratio of wins to losses across all patient pairs (new vs control) based on the hierarchy of outcomes. The win ratio is the odds that the new treatment wins for any randomly chosen patient pair.
 Ties and the win ratio	It is a myth that ties are a problem in using the win ratio. The win difference (% wins - % losses) is a useful complement, quantifying the absolute benefit alongside the relative benefit of the win ratio. It also elucidates each component outcome's contribution to the overall result.
 Repeat events	The number of repeat events (e.g. hospitalizations) can be readily incorporated into the win ratio. It provides a more reliable assumption-free approach compared to other repeat events methods. But it may not enhance statistical power compared to just using the first event.
 Quantitative outcomes	Adding a quantitative outcome (e.g. quality of life score) to the hierarchy has been useful in trials that lack power for clinical events alone. Use of a margin to claim a win/loss for any patient pair is common but not statistically necessary. Use of a time-average may be advantageous.
 Misuses of the win ratio	Valid use of the win ratio requires a clinically meaningful hierarchy of outcomes, preferably with a consistency of effects across components. Misuses have involved an over-elaborate mixed hierarchy. Post hoc uses in trials with a disappointing primary result require cautious interpretation.
 Stratified win ratio	Stratifying the win ratio by key patient factors can be of value. But it will usually make little difference unless the factors are strongly related to the hierarchical outcome. The matched win ratio is not generally recommended, except for propensity matching in non-randomized studies.
 Other statistical developments	Given the win ratio is relatively novel, some appropriate methodology is still in development (e.g. covariate-adjusted win ratio, non-inferiority trials, adaptive designs). These initiatives should be encouraged.
 Statistical software	Statistical software for the win ratio is becoming widely available. An increasing diversity of software is needed to make the broad range of win ratio applications easy to implement.
 Determining trial size	Methods for determining trial size for a primary win ratio outcome often require simulations. Power calculations follow familiar principles, but require specific software to be implementable.

The graphical abstract figure explains the win ratio method, its advantages over conventional time-to-first event analysis of composite outcomes, and provides recommendations for future use of the win ratio in the design, analysis, and reporting of clinical trials. CI, confidence interval.

* Corresponding author. Tel: +44 (0)7798656561, Email: stuart.pocock@lshtm.ac.uk

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The win ratio method for analysing a composite clinical hierarchy of outcomes is growing in popularity especially in cardiovascular trials. This article gives a perspective on its use so far and the issues derived from that experience. Specifically, it focuses on the limitations of a conventional composite outcome; how does the win ratio work, what does it mean, and how to display its findings; guidance on choosing an appropriate clinical hierarchy of outcomes including clinical events, quantitative outcomes, and other options; the additional value of the win difference as a measure of absolute benefit: extension to stratified win ratio, subgroup analysis, matched win ratio, and covariate adjustment; determining trial size for a win ratio outcome; specific insights such as adaptive designs, use of repeat events, and use of margins and time averages for quantitative outcomes; a critique of potential misuses; availability of statistical software; and a statistical appendix on the methodological details. Throughout, each principle is illustrated by examples from specific cardiology trials. The article concludes with a set of recommendations for future use of the win ratio.

Keywords Clinical trial • Hierarchical composite outcome • Win ratio • Statistical methods • Presentation and interpretation • Systematic review

Introduction

The win ratio, introduced in 2012,¹ is an innovative approach to analysis of composite endpoints in randomized clinical trials (RCTs), its strength being to recognize the differing clinical importance of a composite's components, prioritizing them in a clinical hierarchy. It can also incorporate repeat events, e.g. hospitalizations,² and quantitative outcomes, e.g. quality of life scores.^{3,4}

The win ratio method's popularity has steadily grown, especially in cardiology trials so now is opportune time to review its role. In this article, we tackle several key issues:

- (1) The limitations of conventional composite outcomes
- (2) What the win ratio does and how to display and interpret it
- (3) Non-technical explanation of the estimate, confidence interval (CI), and *P*-value
- (4) Guidance on choosing an appropriate hierarchy of endpoints
- (5) Explanation of some complementary approaches, e.g. win difference
- (6) Determining trial size for a win ratio outcome
- (7) Availability of statistical software
- (8) Extensions to permit stratified, matched, or covariate-adjusted analyses
- (9) Specific insights regarding the use of repeat events and the use of margins and time averages for quantitative outcomes
- (10) Insights regarding misguided uses and misinterpretations of the win ratio
- (11) Conclusions including recommendations for future use of the win ratio

Our practical guidance is illustrated with real cardiology trial examples.

The limitations of conventional composite outcomes

Many cardiology trials evaluate treatment efficacy by its impact on fatal and non-fatal events. Hence, the primary endpoint is often a composite of such events, e.g. in heart failure (HF), cardiovascular (CV) death or HF hospitalization, and in ischaemic heart disease, CV death, stroke, myocardial infarction, or revascularization.

Analysis commonly uses a proportional hazards model⁵ for time-to-first event, with hazard ratio, its 95% CI, and logrank *P*. Examples from the EMPEROR-Preserved⁶ and CLEAR⁷ trials are in [Table 1](#). EMPEROR-Preserved is a typical pivotal trial in chronic HF. The time-to-first event

is dominated by HF hospitalizations, so CV deaths that happen subsequently do not contribute to analysis. Also, this analysis ignores repeat hospitalizations.

The CLEAR trial illustrates the challenge in interpreting a four-component primary composite of CV events. The highly positive overall result is driven by benefits in coronary revascularization and myocardial infarction, the least clinically important components, while CV death has a slight numerical excess on bempedoic acid.

Thus, conventional composite endpoints do not directly take into account the fact that the component events may well vary in their clinical importance, e.g. deaths are more important than non-fatal events. While time-to-first event analyses often work well, they sometimes may fail to adequately represent a trial's conclusions.⁸ A weighted combined effect measure could account for components of differing clinical relevance,⁹ but difficulties in agreeing a choice of weights, and consequent analytical complexities have inhibited its use. These limitations led to the development of the win ratio.

The win ratio approach to a clinical hierarchy of endpoints

The win ratio is of value because most composite outcomes have a sensible hierarchy of components that reflects their clinical priorities. At its simplest, with two time-to-event outcomes of interest, death and hospitalization, the hierarchy is (i) death and (ii) hospitalization.

Such a hierarchy can be extended to more time-to-event outcomes, e.g. (i) death, (ii) stroke, (iii) myocardial infarction, and (iv) coronary revascularization. Alternatively, for repeat outcomes, one can replace time to event by number of events, e.g. a hierarchy (i) death and (ii) number of hospitalizations. In all these scenarios, deaths could be either all-cause or cause-specific, e.g. CV. Likewise hospitalizations could be all cause, CV, or condition specific, e.g. HF. Such choices depend on what best captures potential treatment effects.

In some RCTs, there are key quantitative outcomes such as change in self-assessed health status (e.g. KCCQ score), physical function (e.g. 6-min walking distance), or biomarkers (e.g. NT-proBNP).

[Figure 1](#) illustrates the types of clinical hierarchy that have been used. The choice of primary clinical hierarchy when designing a specific trial depends on several factors: what collection of outcomes best captures patient benefit and how extensive a hierarchy provides good statistical power given constraints on trial size.

Table 1 Two examples of conventional composite outcomes with an insight into their limitations

EMPEROR-Preserved trial⁶			
	Empagliflozin N = 2997	Placebo N = 2991	HR (95% CI)
Primary composite	415	511	.79 (.69–.90), <i>P</i> < .001
Components			Difference
HFH	259	352	–93
CV death	219	244	–25
Total HFH including repeats	407	541	–134
Problems with primary analysis			
All components of composite treated equally			
Ignores 148 (32%) CV deaths that occur after HF hospitalization			
Ignores 337 repeat HF hospitalizations (36% of the total)			
CLEAR trial⁷			
	Bempedoic acid N = 6992	Placebo N = 6998	HR (95% CI)
Primary composite	819	927	.79 (.69–.90), <i>P</i> < .001
Components			Difference
CV death	269	257	+12
Stroke	135	158	–23
Myocardial infarction	261	334	–73
Revascularization	435	529	–94
Problems with primary analysis			
No benefit in the most important component, CV death			
Greatest benefit in the least important component, revascularization			
Non-fatal components occur earlier and hence dominate the analysis			

N, number; HR, hazard ratio; CI, confidence interval; CV, cardiovascular; HF, heart failure.

Hierarchical composite outcomes are increasingly popular especially in cardiology trials. New statistical methods are needed to make them usable. The first innovation was Finkelstein and Schoenfeld's¹⁰ non-parametric test that combined evidence from two or more hierarchical endpoints into a single *P*-value for a treatment difference (see [Supplementary data online, Appendix](#)). But one also requires an estimate of the magnitude of treatment effect and its 95% CI.

The win ratio was created for this purpose.¹ The principle is as follows. Consider a RCT comparing new treatment vs. control with N_T and N_C patients, respectively. Then every patient on new treatment is compared with every patient on control, that is, $N_T \times N_C$ paired comparisons. Within each pair, one evaluates the hierarchical component outcomes in descending order of importance until one of the pair shows a better outcome than the other. If the patient on new treatment does better, it is called a 'win' whereas if the control patient does better, it is a 'loss'.

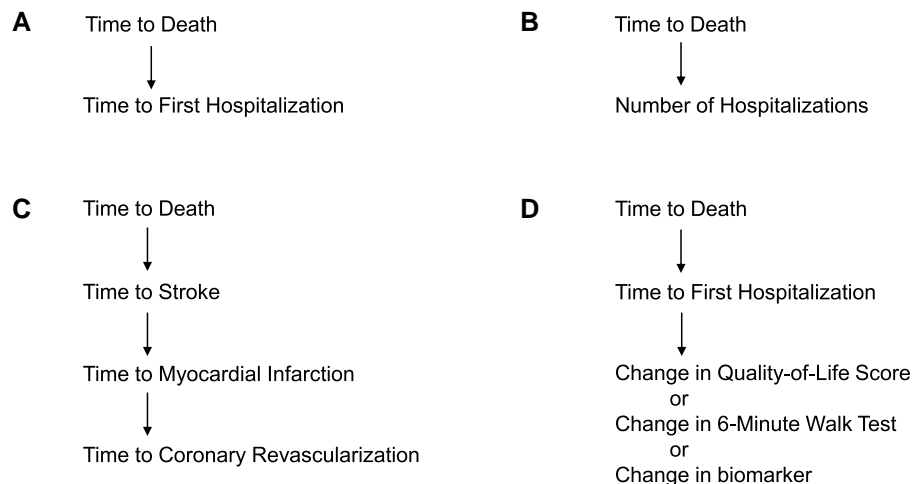
Thus, among all paired comparisons, one accumulates a total of N_{WW} wins; N_{LL} losses and the rest are ties. The win ratio is N_{WW}/N_{LL} . This is sometimes called the unmatched win ratio and is consistent with the Finkelstein–Schoenfeld test.¹⁰ Obtaining the win ratio's 95% CI involves a more complex calculation. [Supplementary data online, Appendix](#), explains statistical details.

The simplest hierarchy: death and a non-fatal event

We now illustrate the simplest win ratio: a hierarchy comprising time to death and time to a non-fatal event, e.g. hospitalization. For each patient pair, we evaluate who 'won' over their shared follow-up time. If one or both patients died, then the one who lived longer is the 'winner'. If neither patient died, but one or both were hospitalized, then the one who avoided hospitalization longer is the 'winner'. If both patients survived without hospitalization over their shared follow-up time, it is a 'tie'. [Figure 2](#) illustrates the possible options for any patient pair.

[Figure 3](#) shows three examples:

- (1) The PARTNER trial¹¹ randomized patients with severe aortic stenosis not suitable for surgery to transcatheter aortic valve implantation (TAVI) or standard therapy. The primary endpoint was a hierarchy of time to death and time to rehospitalization. This was the first cardiology trial adopting the Finkelstein–Schoenfeld test. There were substantially more wins than losses both for deaths and hospitalizations leading to win ratio 1.87 (95% CI 1.35–2.54; *P* < .0001). In this high-risk population, incidence of ties (11.7%) was low.

**Note:**

- 1) Deaths and hospitalizations can be all-cause or cause specific, e.g. cardiovascular deaths and heart failure hospitalizations.
- 2) For event outcomes time of follow-up may vary between patients.
- 3) Quality-of-life scores and other visit-related quantitative outcomes can be for changes from baseline at one specific time-point or a mean of several time-points.

Figure 1 Four examples of hierarchical composite outcomes reflecting clinical priorities that have been used by randomized trials using the win ratio. (A) has two levels: time to death followed by time to first hospitalization; In (B), which also has two levels, time to first hospitalization has been replaced with number of hospitalizations. (C) has four levels; time to death, followed by time to stroke; then time to myocardial infarction; and finally time to coronary revascularization. (D) has three levels where the first two are time to events and the third level is a quantitative outcome, e.g. change from baseline in a 6-min walk test distance

How to interpret this win ratio? For any randomly chosen pair of patients who are not a tie, the estimated odds that the TAVI patient wins is 1.87. For such untied pairs, one can also estimate the probability that the TAVI patient wins, which is $1.87/(1.87 + 1) = .65$.

- (2) The COAPT trial¹² randomized patients with HF and mitral regurgitation to MitraClip device ($N = 302$) or medical therapy alone ($N = 312$). The primary endpoint, all HF hospitalizations within 2 years, had hazard ratio of .53 (95% CI .40–.70). Here, we analyse the hierarchical composite of time to death and then time to first HF hospitalization. Again, there were more wins than losses for both components resulting in win ratio 1.61 (95% CI 1.29–2.04; $P < .0001$).
- (3) The EMPEROR-Preserved trial⁶ randomized patients with preserved ejection fraction HF to empagliflozin ($N = 2997$) or placebo ($N = 2991$). The primary endpoint, a conventional composite of time to CV death or HF hospitalization over median 26.2 months, yielded hazard ratio of .79 (95% CI .69–.90; $P < .001$). Here, we analyse the hierarchical composite of time to CV death and time to first HF hospitalization. Both components had more wins than losses, much more so for hospitalizations. The consequent win ratio is 1.25 (95% CI 1.09–1.43; $P = .001$).

Exploring the win ratio details, helped by the win difference

Like other outcomes, it is important to quantify absolute treatment effects, complementing the relative benefit of the win ratio. The win difference, also called net treatment benefit,¹³ is % wins – % losses. Calculating its 95% CI is more complex (see [Supplementary data online, Appendix](#)). We recommend reporting the win difference alongside the win ratio. It also helps to clarify each component's contribution to the overall win ratio.

In [Figure 3](#), the right-hand column gives win differences both overall and for each component. For PARTNER ([Figure 3A](#)), the win difference is 26.8%, indicating TAVI's marked benefit in this high-risk population. This was mostly due to deaths, win difference of 18.6%, the win ratio for death alone being 1.70 (95% CI 1.23–2.38; $P = .002$). With a hierarchical composite rather than a conventional time-to-first composite, the impact on mortality was clearer and uncorrupted by any hospitalizations prior to death.

For COAPT ([Figure 3B](#)), the win difference is 17.0%, a substantial benefit evenly split between effects on death (9.0%) and on HF hospitalizations (8.0%).

For EMPEROR-Preserved ([Figure 3C](#)), the win difference of 1.9% is much smaller, as expected given the lower overall event rate: there were 75.6% ties, because most patients in both groups survived without HF hospitalization.

It is also useful to look at more conventional analyses of the composite and its components in a non-hierarchical fashion (see [Table 2](#)). For PARTNER, the reduction in risks of death and rehospitalization are clear and comparable. The conventional composite provides a highly significant combined effect, but fails to capture that 61 deaths (19 TAVI, 42 controls) occurred after a hospitalization. Only by doing a hierarchical composite do we capture all deaths in the win ratio analysis.

For COAPT, a similar pattern emerges of strong device effects on both death and HF hospitalization. Again, the conventional composite ignores 124 deaths (43 MitraClip, 81 controls) that occur after HF hospitalization.

For EMPEROR-Preserved, the conventional composite and its components reveal what was also found with the win difference in [Figure 3](#): that is, the treatment effect is driven by HF hospitalization with little effect on CV mortality. Note that 138 CV deaths (53 empagliflozin, 85 placebo) occurred after HF hospitalization and so are excluded from the non-hierarchical composite.

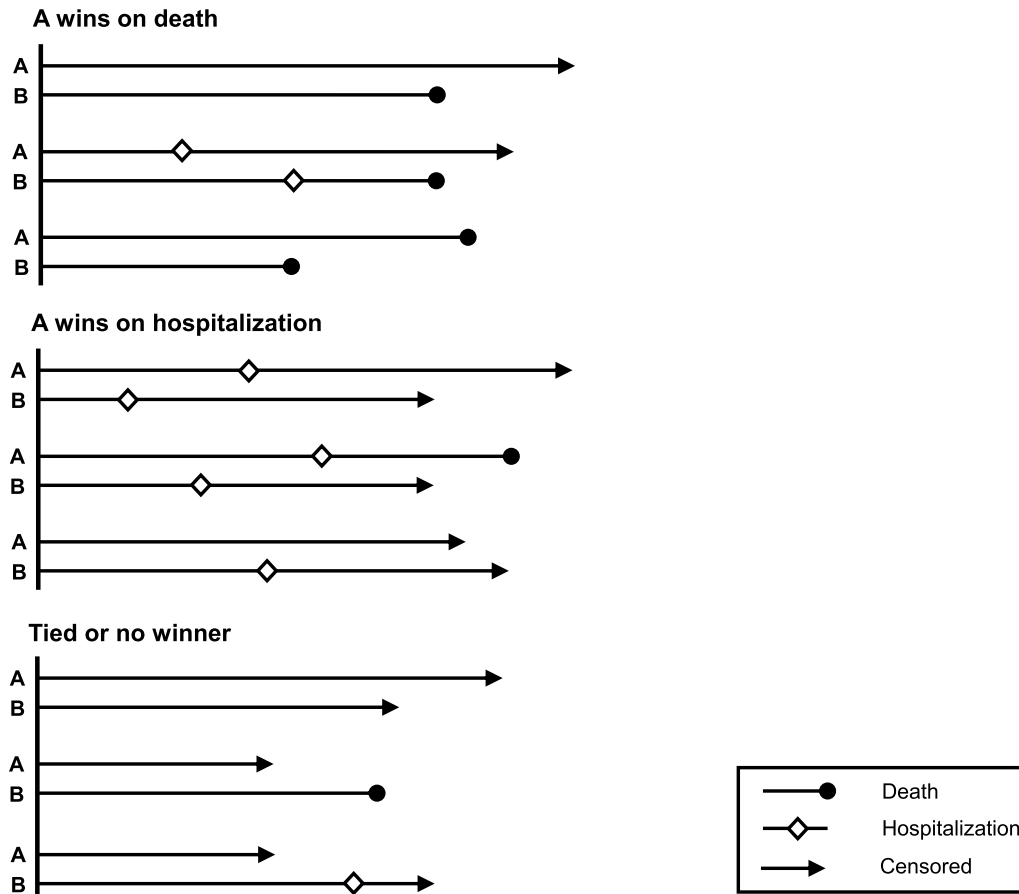


Figure 2 A schema showing nine patient pair (A, B) scenarios illustrating all the options that can arise for the simplest win ratio of the composite hierarchy of time to death and time to hospitalization. In the first set of three scenarios, A wins on the first level of the hierarchy, time to death, since B dies earlier and within the shared follow-up time. In the second set of three scenarios, A wins on the second level of the hierarchy, time to hospitalization. Note that although in the second of this set of scenarios A dies whereas B does not, the death occurs outside the shared follow-up time and is therefore ignored for this patient pair comparison. In the final set of scenarios, neither patient wins on either death or hospitalization and therefore they are tied

One expressed concern^{14,15} is that ties do not contribute to the win ratio. However, this issue is overcome by also presenting the win difference. That is, any measure of relative effect (win ratio, hazard ratio, or relative risk) should be accompanied by a measure of absolute effect.¹⁶

An alternative is to estimate the win odds¹⁷ = $(\% \text{ wins} + \frac{1}{2} \text{ ties}) \div (\% \text{ losses} + \frac{1}{2} \text{ ties})$. This is nearer unity, more so as the % ties gets bigger. But we feel it lacks insight and so is not recommended.

Adapting the win ratio to include repeat events

In HF trials, there is interest in incorporating repeat hospitalizations into analysis either as a primary composite endpoint, e.g. PARAGON-HF¹⁸ or COAPT,¹² or as a key secondary endpoint, e.g. EMPEROR^{6,19} or DELIVER.²⁰ Sometimes death counts as an extra event, whereas others use a joint frailty model to adjust for its competing risk.² In ischaemic heart disease, repeat events such as myocardial infarction and revascularization can be incorporated into secondary analyses, e.g. REDUCE-IT.²¹ Such analyses are complex and make strong assumptions and may not enhance statistical power.^{2,22}

The win ratio can incorporate repeat events into the hierarchical composite endpoint. ATTR-ACT^{23,24} was first to do this. In evaluating

tafamidis vs. placebo in 441 patients with transthyretin amyloid cardiomyopathy, the primary hierarchical composites were time to death and number of CV hospitalizations. The win ratio was 1.70 (95% CI 1.26–2.29; $P < .0001$), a more impressive result than achieved by alternative analyses. Note that this approach is assumption free.

In other trials, the use of repeat events in a win ratio may not enhance the treatment effect, e.g. Table 3 for the EMPEROR-Reduced and EMPEROR-Preserved trials.^{6,19} For each trial, we converted the original primary endpoint into a hierarchical composite of time to CV death and time to first HF hospitalization. Both yielded highly significant win ratios of 1.340 and 1.251, respectively. Since CV deaths are prioritized in the win ratio and are less influenced by treatment than HF hospitalizations, z-scores are somewhat smaller, e.g. in EMPEROR-Reduced, $z = 3.99$ for win ratio compared with $z = 4.41$ in the Cox model.

Changing the second level of the hierarchy to be number of HF hospitalizations yields similar results, though for both trials, the consequent win ratio and its z-statistic become slightly smaller. Replacing the second level by number of days in hospital due to HF further reduces the win ratio estimates and z-scores.

This loss of statistical power when using repeat hospitalizations in HF trials² appears surprising: e.g. in EMPEROR-Reduced and EMPEROR-Preserved, there were respectively 69 and 41 fewer repeat

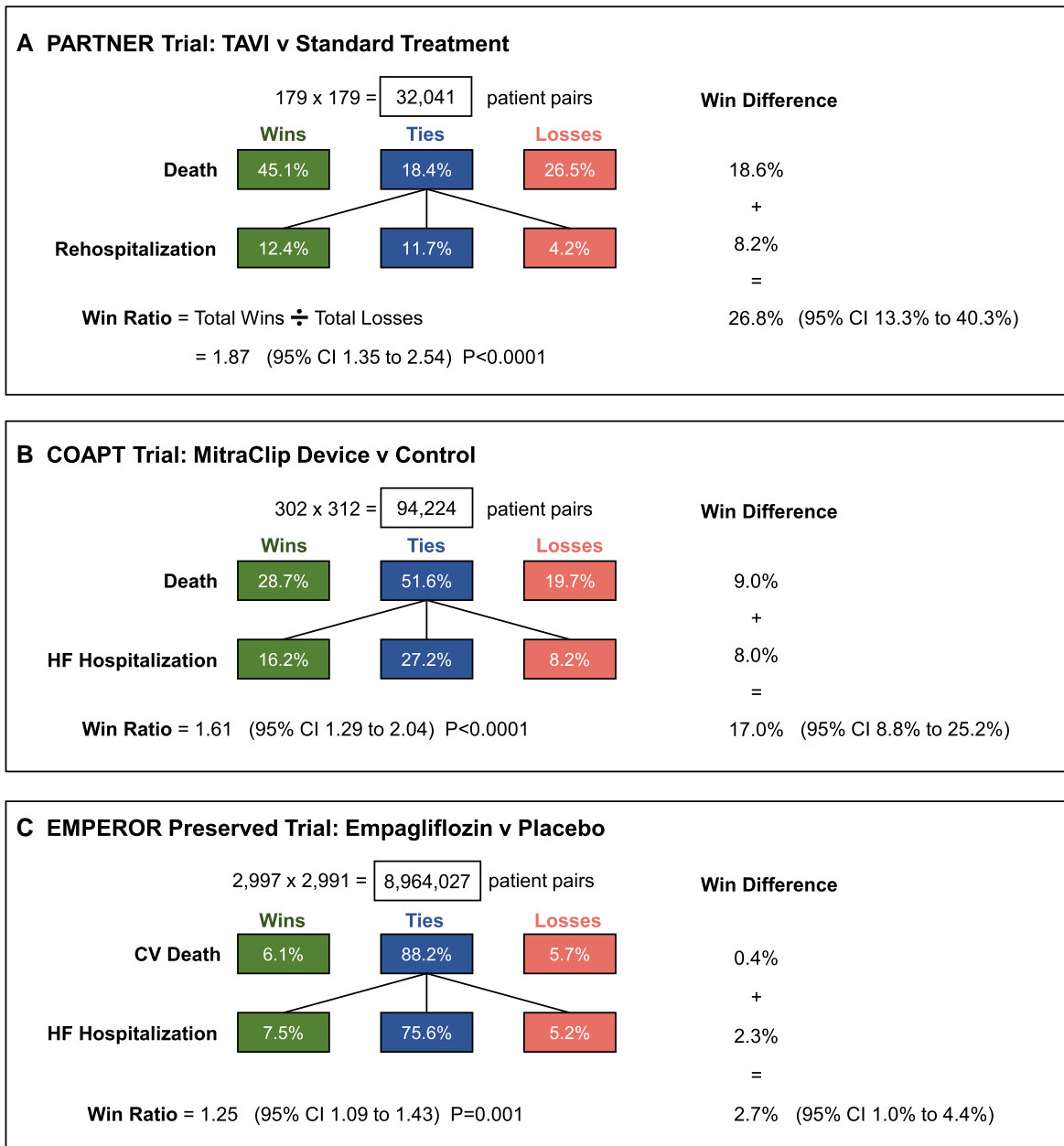


Figure 3 Three examples ((A) PARTNER,¹¹ (B) COAPT,¹² and (C) EMPEROR-Preserved⁶) of a two-level hierarchy win ratio analysis for a hierarchical composite of time to death and then time to a non-fatal event, hospitalization. Each panel shows the number of patients in each treatment group and the total number of patient pair comparisons. The green and red boxes show the percentage of all patient pair comparisons that were wins or losses, respectively, for the active arm. The blue boxes show the percentage of ties. The right-hand column gives win differences both overall and for each component. TAVI, transcatheter aortic valve implantation; CI, confidence interval

HF hospitalizations on empagliflozin compared with placebo on top of 104 and 93 fewer first HF hospitalizations respectively. Possible explanations are as follows: treatment effects may be more pronounced early on when first events dominate; repeat events have a highly skewed distribution with a few patients having many such events; and treatment switches may occur after a first event.²² However, there are trials, e.g. RELIEVE-HF,²⁵ where recurrent hospitalizations are common and repeat event analyses enhance power.

Some trials, e.g. EMPULSE,^{26,27} have a hierarchical composite outcome containing both number of HF events and time to first event.

But such an approach appears unnecessary: the win ratio finding is not enhanced and the result is harder to interpret.

Extending the win ratio to quantitative outcomes

Another option is to extend the win ratio to incorporate outcomes other than clinical events, especially quantitative outcomes recorded at specific times.^{3,4} The clinical hierarchy includes not only deaths and non-fatal events but also other measures of patient well-being, measures of physical

Table 2 Results for conventional composite endpoint and its components to relate to the hierarchical composite win ratio analyses of the three trials in Figure 3

PARTNER trial¹¹	TAVI	Control	HR (95% CI)
No. of patients	179	179	
Death	55 (30.7%)	89 (50.7%)	.55 (.40, .74)
Rehospitalization	40	79	
Composite	76 (42.5%)	126 (71.6%)	.46 (.35, .59)
COAPT trial¹²	MitraClip	Control	
No. of patients	302	312	
Death	80 (29.1%)	121 (46.1%)	.62 (.48, .82)
HF hospitalization	92 (35.7%)	151 (56.7%)	.52 (.40, .67)
Composite	129 (45.7%)	191 (67.9%)	.57 (.45, .71)
EMPEROR-Preserved⁶	Empagliflozin	Placebo	
No. of patients	2997	2991	
CV death	219	244	.91 (.76, 1.09)
HF hospitalization	259	352	.71 (.60, .83)
Composite	415	511	.79 (.69, .90)

HR, hazard ratio; CI, confidence interval; TAVI, transcatheter aortic valve implantation; No., number; HF, heart failure.

Table 3 Two trial examples comparing three alternative ways of handling hospitalizations in the second level of a hierarchical composite: (i) time to first, (ii) no. of hospitalizations, or (iii) total days in hospital

Hierarchy	Win ratio (95% CI)	z-value^a
EMPEROR-Reduced trial¹⁹		
CV death then first HFH	1.340 (1.160, 1.547)	3.99
CV death then no. of HFHs	1.335 (1.156, 1.541)	3.94
CV death then days in hospital due to HFH	1.330 (1.153, 1.134)	3.91
EMPEROR-Preserved trial⁶		
CV death then first HFH	1.251 (1.094, 1.430)	3.27
CV death then no. of HFHs	1.248 (1.091, 1.427)	3.24
CV death then days in hospital due to HFH	1.238 (1.083, 1.416)	3.14

CI, confidence interval; CV, cardiovascular; HFH, heart failure hospitalization; no., number.

^aThe z-value is the standardized normal deviate corresponding to the P-value. For instance, z = 1.96, 2.58, 3.29, and 3.88 correspond to P = .05, .01, .001, and .0001, respectively.

functioning, and surrogate biomarkers. There are two goals here: to achieve a more rounded view of treatment efficacy and to enhance statistical power. But such extended hierarchical outcomes need to be clinically meaningful and capture treatment benefits convincingly. There are also methodological challenges, illustrated by three examples (Figure 4).

The EMPULSE trial^{26,27} randomized 530 acute HF patients after initial stabilization to empagliflozin or placebo. The primary endpoint over 90 days was a hierarchical composite of time to death, number of HF events, time to first HF event, and change in KCCQ-Total Symptom Score (TSS) at 90 days. The consequent win ratio analysis is in Figure 4A. Note that 69.5% of patient pairs were tied after the first three hierarchy levels, since most patients on both treatments were alive and HF event free. Hence, the fourth level KCCQ-TSS change at 90 days contributed most to the positive result, win ratio of 1.38 (95% CI 1.11–1.71; P = .0036). Regarding absolute benefit, the overall win difference 14.9% was largely due to KCCQ-TSS change (win difference 8.8%) leaving only 6.4% tied.

There are two specific issues here: what were the rules defining win, tie, or loss for this quantitative outcome and how were missing data handled.

In EMPULSE, for any patient pair, a win (or loss) required the 90-day change in KCCQ-TSS to be ≥ 5 points better (or worse) for the empagliflozin patient compared with the placebo patient. It was a tie if the difference was < 5 points. This fixed margin (5 points) is open to debate, and so sensitivity analyses used alternative criteria for a win: any difference, ≥ 2 points, ≥ 10 points, and ≥ 15 points. The estimated win ratio got slightly smaller as the margin got less, but so did the CI width, meaning the P-value scarcely changed.

It is debatable whether margins are necessary for valid use of the win ratio. The statistical argument against margins is that the inference is like other non-parametric tests in essentially ranking patients from worst outcome (first death after randomization) to best outcome (event free with the largest quantitative improvement), with the added complication of differing follow-up times. The clinical argument is that a win requires some clinically meaningful difference, a 5-point margin for KCCQ. We suspect the choice of margin, yes or no, will make little difference if the EMPULSE experience is generalizable.



Figure 4 Examples of three trials (A) EMPULSE,^{26,27} (B) TRILUMINATE,²⁸ and (C) ATTRIBUTE-CM²⁹ that extended the win ratio to include quantitative outcomes. Each panel shows the number of patients in each treatment group and the number of patient pair comparisons. The green and red boxes show the percentage of all patient pair comparisons that were wins or losses, respectively, for the active arm. The blue boxes show the percentage of ties. The right-hand column gives win differences both overall and for each component. HF, heart failure; KCCQ-TSS, Kansas City Cardiomyopathy Questionnaire-Total Symptom Score; CI, confidence interval; TEER, transcatheter edge-to-edge repair; TV, tricuspid valve

The second key issue is how to account for missing data. If the quantitative outcome is measured multiple times, the latest time point with common shared data may be used. If data are entirely missing for a patient, then all their paired comparisons count as ties. Compared to the primary EMPULSE result in [Figure 4A](#), frequency of ties increases from 6.1% to 18.7%. The consequent win ratio becomes 1.43 but with a wider 95% CI of 1.14–1.78. Instead, the pre-defined primary analysis used a multiple imputation algorithm.²⁷ Analysis becomes more complex but is generally thought more appropriate.

Our second example, the TRILUMINATE trial,²⁸ randomized 350 patients with tricuspid regurgitation to transcatheter edge-to-edge repair (TEER) or medical therapy (control). Over 1 year, the primary endpoint was the hierarchical composite of (i) time to death or tricuspid valve surgery, (ii) time to hospitalization, and (iii) improvement of ≥ 15 points in KCCQ score at 1 year ([Figure 4B](#)).

Note that the change in KCCQ score became binary, ≥ 15 points yes or no. Hence, for any patient pair, if both were yes (or both were no), then it is a tie. Only for those pairs with one yes and one no do we have a win or loss for TEER. Any dichotomizing of a quantitative outcome loses statistical power and insight into details of any treatment effect^{30,31} and hence is not recommended. Having said that, in TRILUMINATE, the treatment difference in KCCQ change was so marked that any analysis is highly significant. However, TRILUMINATE is open label so such a subjective outcome carries a risk of bias.

Another concern with TRILUMINATE is that the overall win ratio of 1.48 (95% CI 1.06–2.13; $P = .0204$) was driven by KCCQ with no evidence of a benefit on CV events, i.e. the win differences for the three levels of the hierarchy +.8%, -3.0% , and $+14.4\%$, respectively, yield an overall win difference of 12.1%. Like a conventional composite outcome with disparate findings across components, the consequent overall summary, the win ratio, does not fully describe what happened. For TRILUMINATE, the TEER group clearly had a better self-assessed health status at 12 months, but to what extent did bias due to lack of blinding play a part?

Our third example, the ATTRIBUTE-CM trial,²⁹ randomized 632 patients with transthyretin amyloid cardiomyopathy in a 2:1 ratio to acoramidis hydrochloride or placebo. Over 30 months, the primary hierarchical composite outcome comprised death, number of CV hospitalizations, change in NT-proBNP, and change in 6-min walk distance. For these two quantitative components, any paired comparison used the last visit where both patients had assessments. The win/loss criterion had a margin of ≥ 500 pg/mL difference for NT-proBNP, whereas 6-min walk distance had no margin.

The win ratio for ATTRIBUTE-CM was 1.8 (95% CI 1.4–2.2; $P < .001$) ([Figure 4C](#)). The overall win difference of 27.8% had its greatest contribution from change in NT-proBNP (win difference 16.3%) with number of CV hospitalizations also relevant (win difference 9.1%). With concern about relying on a biomarker change so much, sensitivity analyses were done with it removed. For death and CV hospitalization alone, the win ratio was 1.5 (95% CI 1.1–2.0), whereas adding in 6-min walk distance made little difference: win ratio 1.4 (95% CI 1.1–1.8).

ATTRIBUTE-CM illustrates the uncertainties in defining what clinical hierarchy to choose as primary: during the trial's course, the investigators twice changed their mind.

For these three examples, it is also informative to present conventional results for each separate component of the hierarchical composite (see [Table 4](#)) to complement the win ratio findings in [Figure 3](#).

For EMPULSE, there are fewer deaths and HF events on empagliflozin vs. placebo such that the conventional composite of time-to-first event was borderline significant: HR .65 (95% CI .43, .99). For

KCCQ-TSS, the baseline-adjusted 90-day mean change had a modest treatment effect, though this simple analysis is somewhat flawed by ignoring the deaths. Overall, one can see how the consistent signal for deaths, HF events, and KCCQ combined to give clear evidence of treatment benefit.

For TRILUMINATE, the conventional analyses of each component of the hierarchical composite reinforce what the win difference showed in [Figure 4B](#): the evidence of treatment benefit was confined to KCCQ improvement.

For ATTRIBUTE-CM, the conventional analyses of each component (see [Table 4](#)) were compatible with the win differences in [Figure 4C](#): treatment benefit being driven by fewer CV hospitalizations and a marked effect of NT-pro BNP. Acoramidis also improved 6-min walking distance, but this was less evident in the win ratio since it was last in the hierarchy with few ties remaining by then.

Other uses of the win ratio

We now explore some innovative uses of the win ratio.

For an ordinal primary outcome

The SOS-AMI trial³² will assess the clinical efficacy of self-administered selatogrel vs. placebo when symptoms suggestive of recurrent acute myocardial infarction (AMI) occur. The primary outcome is an ordinal scale, the six outcomes ranked from worst to best being:

- Death within 7 days
- AMI with compromised electro-haemodynamics within 2 days
- ST-elevation myocardial infarction (STEMI) within 2 days
- High-risk non-ST-elevation myocardial infarction (NSTEMI) within 2 days
- NSTEMI with peak cardiac troponin ≥ 10 times normal within 2 days
- None of the above

Based on Wang and Pocock,³³ the win ratio method will be applied to all patient pairs (selatogrel vs. placebo), the winner being the one with the lower rank. Many ties are anticipated since most patients will be event free. An alternative estimator for such an ordinal outcome is the common odds ratio,³⁴ but this assumes the same improvement (odds ratio) applies to all five cut-offs in the ordinal scale. Hence, regulators encouraged the win ratio approach. Note that a conventional Mann–Whitney test is applicable to such an ordinal outcome, but does not elicit an estimate of treatment effect.

Adjusting a quantitative outcome for the competing risk of death

Many HF trials have quantitative outcomes, e.g. KCCQ score, recorded at several visits, with death being an informative censoring. Simple quantitative analyses assuming deaths missing at random can mislead. To overcome this, the DELIVER trial²⁰ used a win ratio analysis for a hierarchical composite of death and KCCQ-TSS at 8 months, recognizing prior death as a worse outcome. The consequent win ratio of 1.11 (95% CI 1.03–1.21; $P = .009$) demonstrated that dapagliflozin improved HF symptoms, a valuable secondary finding.

The RECHARGE trials³⁵ extend the same principle, a hierarchical composite of death and quality of life, into a longer-term evaluation of percutaneous coronary intervention (PCI) vs. coronary artery bypass graft (CABG) in female and ethnic minority patients with multi-vessel or left main coronary artery disease. The hierarchy comprises (i) time to death over 3–5 years and (ii) time-averaged change from

Table 4 Results for conventional analysis of each component to relate to the hierarchical composite win ratio analyses of the three trials in Figure 4

EMPULSE^{26, 27}	Empagliflozin	Placebo	
No. of patients	265	265	
Deaths	11	22	
1+ HF event	28	39	
Composite	37	57	Hazard ratio .65 (95% CI .43, .99)
All HF events	36	52	
Mean baseline-adjusted 90-day KCCQ change	Difference +4.45 (95% CI .32, 8.59)		
TRILUMINATE²⁸	TEER	Control	
No. of patients	175	175	
Death or TV surgery	16	18	
Annual rate of HF hospitalizations	.21	.17	
KCCQ improvement ≥ 15 points	73	39	
Mean KCCQ change at 1 year	Difference +11.7 (95% CI 6.8, 16.6)		
ATTRIBUTE-CM²⁹	Acoramidis	Placebo	
No. of patients	421	211	
Deaths	39	22	
CV hospitalizations	Rate ratio .496 (95% CI .355, .695)		
Geometric mean change in NT-proBNP	Ratio .529 (95% CI .463, .604)		
Mean change in 6-min walk test, m	Difference +39.6 (95% CI 21.1, 58.2)		

HR, hazard ratio; CI, confidence interval; KCCQ, Kansas City Cardiovascular Questionnaire; TV, tricuspid valve; TEER, transcatheter edge-to-edge repair.

baseline in the 12-Item Short Form Survey (SF-12) version 2 questionnaire, which is measured at 1, 3, 6, and 12 months and annually thereafter (with 5-point margin). The consequent win ratio captures the essence of a hierarchy of relevance to patients: (i) do I know who lived longer? and (ii) if not, who experienced the better quality of life? This bypasses many of the complexities and controversies in previous trials of PCI vs. CABG using conventional composites of death, myocardial infarction (MI), stroke, and revascularization.^{36,37}

In general, for a quantitative outcome in the composite hierarchy, it is good practice to use an average value across several time points to better capture the overall effect.

Extending the clinical hierarchy to multiple outcomes

The DAPA-MI trial³⁸ randomized 4017 MI patients without diabetes or HF to dapagliflozin or placebo with mean follow-up of 11.6 months. The original primary endpoint was CV death and HF hospitalization, but blinded interim data revealed low event rates and a lack of statistical power. Hence, a hierarchical primary composite outcome was defined with the following seven components: death, HF hospitalization, myocardial infarction, atrial fibrillation/flutter, type 2 diabetes, New York Heart Association (NYHA) class, and weight decrease $\geq 5\%$.

The first five steps determined win or loss for any patient pair using time to event, while win or loss for NYHA class and weight decrease used data from the last visit done for both patients. The consequent win ratio of 1.34 (95% CI 1.20, 1.50; $P < .001$) led to the conclusion that 'there were significant improvements in cardiometabolic

outcomes'. The original primary had hazard ratio of .95 (95% CI .64, 1.40), showing no effect.

Figure 5 shows two alternative ways of displaying these win ratio findings. The first, from the trial publication, concentrates on cumulative results over the hierarchy. The second is in the same style as other win ratio plots. In particular, the win difference column helps clarify what is happening. The first four components make little contribution, with win differences $\pm .3\%$ or less. Modest treatment benefits for type 2 diabetes incidence and improved NYHA class contribute win differences around 1.5%, but much the biggest contribution is weight loss with win difference close to 5%.

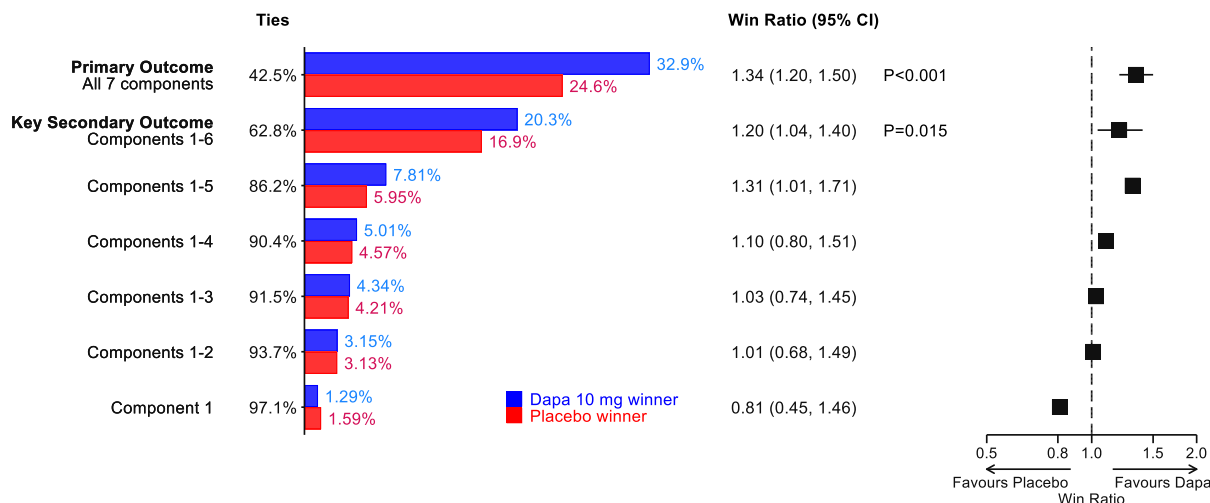
One feels this elaborate win ratio analysis did not capture a meaningful benefit of dapagliflozin. Its effects on weight loss³⁹ and avoiding diabetes⁴⁰ are already known, change in NYHA class is a 'soft' outcome, and evidence on reducing CV events or deaths is lacking.

Post hoc exploratory win ratio analysis

When a trial fails to show benefit for its primary endpoint, exploratory analyses are often pursued, seeking more positive findings.⁴¹ Post hoc win ratio analyses are an option.

One example is the SPYRAL HTN-ON MED trial⁴² of renal denervation vs. sham control in 337 patients with resistant hypertension on anti-hypertensive medication. The primary endpoint, mean change in ambulatory systolic blood pressure (SBP) over 6 months, had a treatment difference of -1.9 mmHg (95% CI $-4.4, .5$; $P = .12$). Increases in

A Bar Chart and Forest Plot of Cumulative Results



B Conventional Plot of Wins, Losses and Win Difference Dapagliflozin versus Placebo

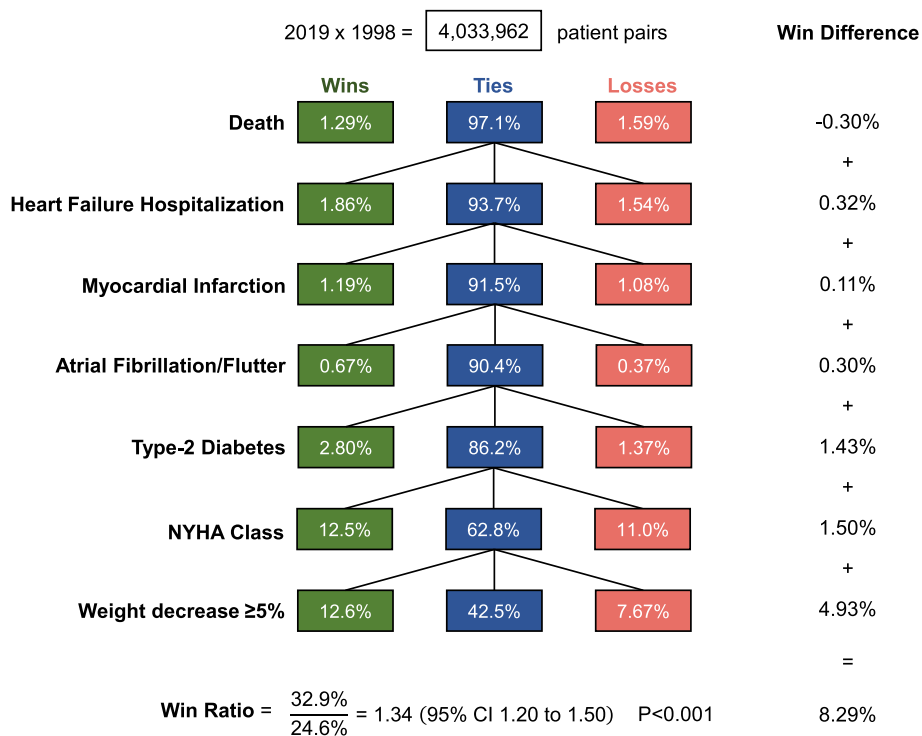


Figure 5 Two alternative ways of plotting win ratio findings from the DAPA-MI trial.³⁸ (A), from the trial publication, shows the cumulative results over the hierarchy, but does not explicitly show the percentage of ties. (B), adopting the style most commonly used, shows the percentage of wins, losses, and ties at each level of the hierarchy. The right column helps clarify what is happening by showing the win difference at each level. DAPA, dapagliflozin; CI, confidence interval; NYHA, New York Heart Association

medication intensity among sham controls were substantial, and hence may have diluted the true effect of renal denervation. Hence, a *post hoc* win ratio analysis was undertaken with hierarchy (i) change in ambulatory SBP at 6 months, with a margin of 5 mmHg, and (ii) change in

medication burden at 6 months, with no margin. The win ratio is 1.50 (95% CI 1.13–1.99; *P* = .005), with both differences in SBP and medication burden contributing win differences 9.1% and 7.3%, respectively (Figure 6).

Renal Denervation versus Sham Control

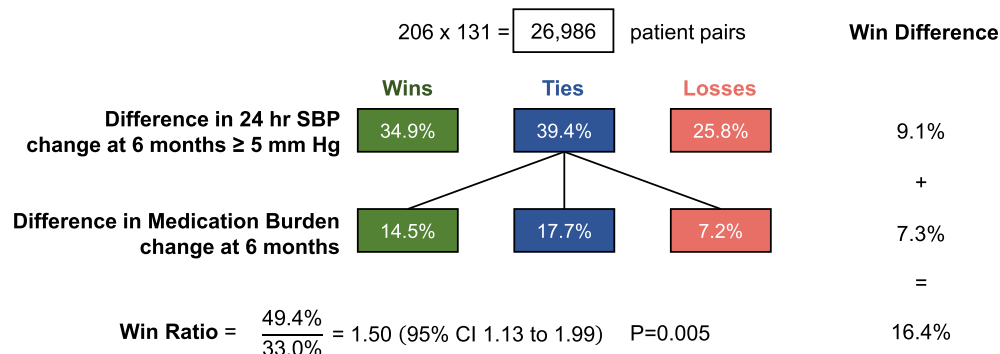


Figure 6 An example of a *post hoc* win ratio analysis of 24-h ambulatory systolic blood pressure and medication burden in the SPYRAL HTN-ON MED trial.⁴² The primary endpoint, mean change in ambulatory systolic blood pressure over 6 months, did not reach statistical significance. Increases in medication intensity among sham controls were substantial, diluting a true effect. Hence, this *post hoc* win ratio analysis has hierarchy (i) change in ambulatory systolic blood pressure at 6 months, with a margin of 5 mmHg, and (ii) change in medication burden at 6 months, with no margin. SBP, systolic blood pressure; CI, confidence interval; hr, hour

While such *post hoc* analyses are interesting, they do not directly influence a trial's conclusions. This example illustrates how a clinical hierarchy of endpoints and its win ratio can be useful in trials where deaths and clinical events are not relevant.

Many trials have published *post hoc* win ratio analyses when the original primary endpoint is a conventional composite.³ If every component of the composite shows treatment benefit, e.g. COMPASS,⁴³ then inevitably, analyses using win ratio and hazard ratio will be consistent. But if treatment benefit is confined to events lower in the hierarchy with no influence on mortality, then win ratio analysis will rightly show a lesser effect.

Stratified win ratio, matched win ratio, and subgroup analysis

Many trials have stratified randomization. It is then appropriate to stratify analyses, either based on the same factors or others of key interest. We explain how this works for win ratio analyses.

The EMPULSE trial's^{26,27} unstratified analysis is already presented in Figure 4A. The pre-defined primary analysis was stratified by *de novo* acute HF and decompensated chronic HF (Figure 7). The two subgroup win ratios 1.29 and 1.39, respectively (Figure 7A), are plotted with CIs in Figure 7B. An overall stratified win ratio CI and P-value were then obtained (see Supplementary data online, Appendix, for details).

Note that the stratified and unstratified win ratios are very similar. We expect this, unless the strata have markedly different risk profiles across the hierarchy. A quick check is that the two strata's % ties 7.6% and 5.7% are similar.

Lastly, an interaction test comparing the stratum specific win ratios yields interaction $P = .76$ for EMPULSE. Such subgroup analysis interaction tests can be done comparing the log (win ratios) and their standard errors (see Supplementary data online, Appendix).

In general, stratifying a win ratio analysis leads to a slight loss of statistical power if the strata are of equal risk. On the other hand, if one or two patient factors strongly relate to patient risk, then stratified analysis will slightly enhance statistical power. Further examples and insights concerning the stratified win ratio are in the Supplementary data online, Appendix.

The RELIEVE-HF trial²⁵ randomized 508 patients to interatrial shunt or sham control and has interesting subgroup analyses for the win ratio. The primary endpoint over 12–24 months was a hierarchical composite of death, cardiac transplant or left ventricular assist device (LVAD), all HF hospitalizations, all outpatient HF events, and change in KCCQ-OS at 12 months (with 5-point margin). The overall win ratio showed no effect, but subgroup analyses by left ventricular ejection fraction (LVEF) \leq 40% or $>$ 40% showed evidence of heterogeneity (Table 5). For LVEF $>$ 40% (302 patients), there was evidence of harm: win ratio .70 (95% CI .54, .92), whereas for LVEF \leq 40% (206 patients), there is a suggestion of benefit: win ratio 1.40 (95% CI .80, 2.46); interaction $P = .0275$.

Secondary analyses revealed no effect on KCCQ-OS, so removing this from the hierarchy enhanced the interaction to $P = .01$. However, the most striking evidence of interaction was a conventional (non-hierarchical) analysis of all the clinical events. For LVEF $>$ 40% and LVEF \leq 40%, the consequent rate ratios were 1.68 (95% CI 1.29, 2.19) and .55 (95% CI .42, .73), respectively (interaction $P < .0001$).

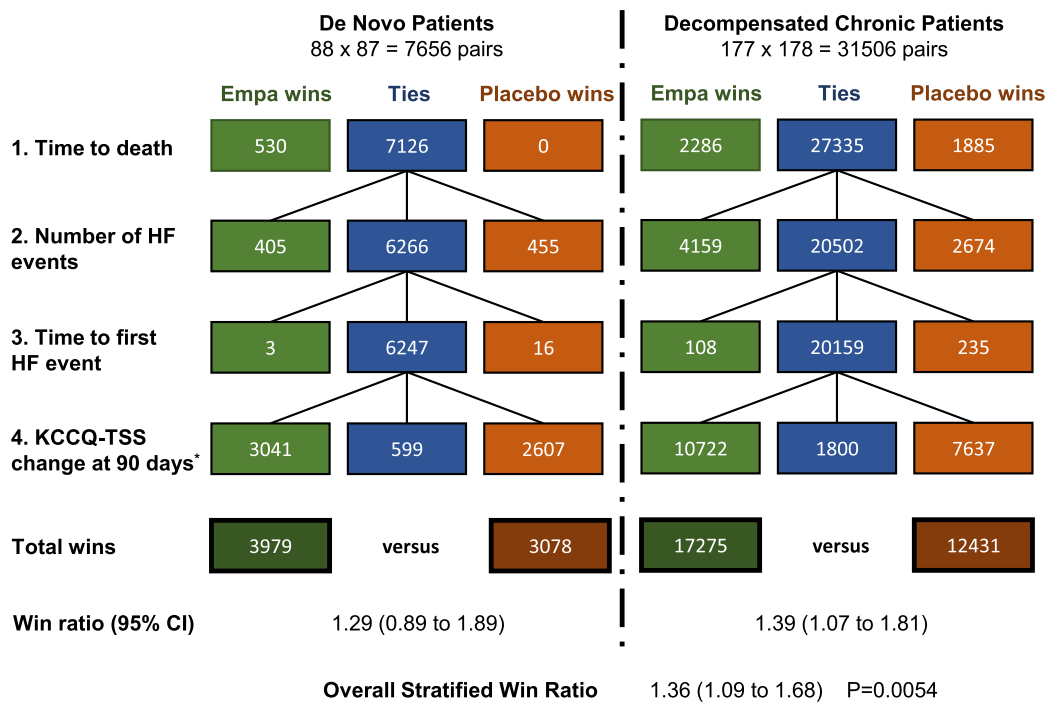
While such subgroup findings need cautious interpretation, it appears the overall win ratio was ill-suited to capture this diversity of effects across subgroups and components of the primary outcome.

Lastly, the matched win ratio and covariate adjustment for the win ratio are topics of methodological interest, but as yet rarely used in practice (see Supplementary data online, Appendix, for details).

Statistical software for win ratio analyses

Statistical software packages for analysing data using the win ratio are available in R and Stata. Several packages allow a flexible, user-defined

A Detail of Win Ratio Analysis for Each Stratum



* a win requires at least 5 points difference between patients

B Combining the Evidence into an Overall (Stratified) Win Ratio

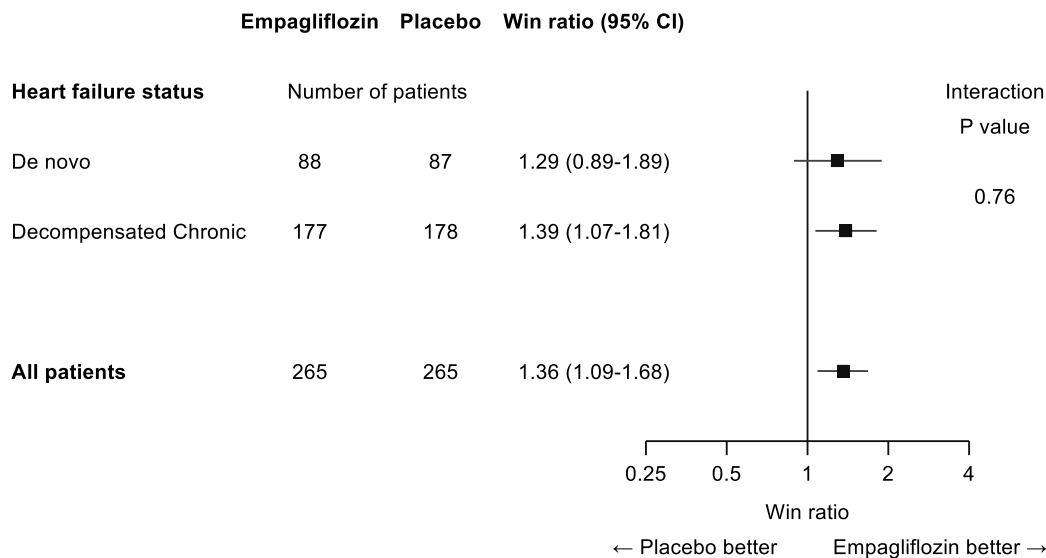


Figure 7 The stratified win ratio in the EMPULSE trial^{26,27}: combining win ratio estimates for the two strata into an overall estimate. (A) shows details of the win ratio analysis for each stratum (*de novo* and decompensated chronic patients) with stratum-specific win ratios and an overall stratified win ratio. (B) shows the stratum-specific win ratios with 95% confidence intervals and an interaction P-value, plus the overall stratified win ratio with 95% confidence interval. Empa, empagliflozin; HF, heart failure; KCCQ-TSS, Kansas City Cardiomyopathy Questionnaire-Total Symptom Score; CI, confidence interval

Table 5 Subgroup analyses in the RELIEVE-HF trial²⁵ by left ventricular ejection fraction for the primary win ratio and a composite of clinical events

Win ratio analysis								
	LVEF ≤40%				LVEF >40%			
	Shunt N = 101 Wins	Patient pairs 10 606 Ties	Placebo N = 106 Losses	Win difference	Shunt N = 149 Wins	Patient pairs 22 797 Ties	Placebo N = 153 Losses	Win difference
Death	14.6%	74.6%	10.9%	+3.7%	4.0%	83.4%	12.6%	-8.6%
CT or LVAD	4.4%	69.6%	.6%	+3.8%	0	83.4%	0	0
HF hospitalization	15.8%	40.6%	13.3%	+2.5%	13.5%	51.0%	18.9%	-5.5%
Outpatient HF events	4.8%	30.4%	5.4%	-.6%	7.0%	37.8%	6.2%	+8%
Change in KCCQ	12.6%	4.7%	13.1%	-.5%	14.2%	6.2%	17.4%	-3.2%
Overall	52.1%		43.2%	+8.9%	38.6%		55.1%	-16.5%
Win ratio (95% CI)								
Overall	.89 (.72, 1.09)							
LVEF ≤40%	1.21 (.87, 1.67)							
LVEF >40%	.70 (.54, .92) Interaction P = .0275							
Rate ratio analysis of clinical events								
	LVEF ≤40%		LVEF >40%					
	Shunt N = 101	Placebo N = 106	Shunt N = 149	Placebo N = 153				
Death	13	20	22	7				
CT or LVAD	1	6	0	0				
HF hospitalization	41	78	87	47				
Outpatients' HF events	21	30	34	34				
Total	76	134	143	88				
Rate ratio (95% CI)								
LVEF ≤40%	.55 (.42, .73)							
LVEF >40%	1.68 (1.29, 2.19) Interaction P < .0001							

in the primary publication, the win ratio was weighted before and after an adaptive design interim analysis and thus the results are slightly different. LVEF, left ventricular ejection fraction; CI, confidence interval; CT, cardiac transplant; LVAD, left ventricular assist device; HF, heart failure; KCCQ-TSS, Kansas City Cardiomyopathy Questionnaire.

hierarchy and include analysis of quantitative, binary, and time-to-event outcomes. Some packages offer analysis of repeat events, stratified analyses, or use of a margin for quantitative outcomes. Results obtained from software packages may differ due to how they handle missing data for quantitative outcomes. Therefore, the approach to missing data needs pre-specifying and an appropriate software package chosen (see [Supplementary data online, Appendix](#), for details).

Determining trial size for win ratio primary outcome

Trials based on the win ratio require a means of determining trial size. The principles behind such power calculations are unchanged, but often require simulations for the win ratio. The details of this methodology and its statistical software are in the [Supplementary data online, Appendix](#).

Adaptive sample size re-estimation is often advocated for trials. This can best be done for a win ratio outcome when the interim analysis contains complete primary outcome data for each patient, using the promising zone methodology⁴⁴ (see [Supplementary data online, Appendix](#), for details).

The win ratio approach can be adapted to non-inferiority trials, a key step being to define an appropriate non-inferiority margin for the win ratio⁴⁵ and a consequent justification of trial size. We intend to document this in a subsequent publication.

Limitations

Some of our attempts to provide guidance on various aspects of the win ratio approach are based on experience from specific examples, so some caution is warranted until they are backed up by further investigation, including simulation studies. Our claim about increased use of

the win ratio is based on anecdotal evidence but will be supported by a systematic review of published trials with the win ratio that is currently in progress.

Conclusions

The value of hierarchical composite outcomes and the win ratio in cardiology trials is being increasingly recognized. This article has elucidated the diversity of its application, including hierarchies of death and clinical events, repeat events, and quantitative outcomes. The consequent statistical developments are also documented with added detail in the [Supplementary data online](#), [Appendix](#). Lessons learnt regarding potential misuses of the win ratio are also tackled. In order to aid wise future use of the win ratio, the [Graphical Abstract](#) provides a set of recommendations.

Supplementary data

[Supplementary data](#) are available at *European Heart Journal* online.

Declarations

Disclosure of Interest

S.J.P. declares grants and personal fees from AstraZeneca and personal fees from Boehringer Ingelheim, Vifor, Abbott Vascular, Amarin, Novartis, Abiomed, Edwards, Janssen, and Medtronic. J.G. has received consultancy fees from Boehringer Ingelheim, Amarin, and Vifor Pharma and research grants from AstraZeneca. G.W.S. has received speaker honoraria from Medtronic, Pulnovo, Infraredx, Abiomed, Amgen, and Boehringer Ingelheim; has served as a consultant to Abbott, Daiichi Sankyo, Ablative Solutions, CorFlow, Cardiomech, Robocath, Miracor, Vectorious, Apollo Therapeutics, Elucid Bio, Cardiac Success, Valfix, TherOx, HeartFlow, Neovasc, Ancora, Occlutech, Impulse Dynamics, Adona Medical, Millennia Biopharma, Oxitope, HighLife, Elixir, Remote Cardiac Enablement, and Aria; and has equity/options from Cardiac Success, Ancora, Cagent, Applied Therapeutics, Biostar family of funds, SpectraWave, Orchestra Biomed, Aria, Valfix, and Xenter. G.W.S.'s employer, Mount Sinai Hospital, receives research grants from Shockwave, Abbott, Abiomed, Bioventrix, Cardiovascular Systems Inc., Phillips, Biosense Webster, Vascular Dynamics, Pulnovo, and V-wave. T.J.C. and J.P.F. have no disclosures related to this article.

Data Availability

No data were generated or analysed for this manuscript.

Funding

All authors declare no funding for this contribution.

References

- Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012;**33**:176–82. <https://doi.org/10.1093/eurheartj/ehr352>
- Gregson J, Stone GW, Bhatt DL, Packer M, Anker SD, Zeller C, et al. Recurrent events in cardiovascular trials: JACC state-of-the-art review. *J Am Coll Cardiol* 2023;**82**:1445–63. <https://doi.org/10.1016/j.jacc.2023.07.024>
- Ferreira JP, Jhund PS, Duarte K, Claggett BL, Solomon SD, Pocock S, et al. Use of the win ratio in cardiovascular trials. *JACC Heart Fail* 2020;**8**:441–50. <https://doi.org/10.1016/j.jchf.2020.02.010>
- Redfors B, Gregson J, Crowley A, McAndrew T, Ben-Yehuda O, Stone GW, et al. The win ratio approach for composite endpoints: practical guidance based on previous experience. *Eur Heart J* 2020;**41**:4391–9. <https://doi.org/10.1093/eurheartj/ehaa665>
- van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basic concepts and methods of Cox regression. *Kidney Int* 2008;**74**:705–9. <https://doi.org/10.1038/ki.2008.294>
- Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, et al. Empagliflozin in heart failure with a preserved ejection fraction. *N Engl J Med* 2021;**385**:1451–61. <https://doi.org/10.1056/NEJMoa2107038>
- Nissen SE, Lincoff AM, Brennan D, Ray KK, Mason D, Kastelein JJP, et al. Bempedoic acid and cardiovascular outcomes in statin-intolerant patients. *N Engl J Med* 2023;**388**:1353–64. <https://doi.org/10.1056/NEJMoa2215024>
- McCoy CE. Understanding the use of composite endpoints in clinical trials. *West J Emerg Med* 2018;**19**:631–4. <https://doi.org/10.5811/westjem.2018.4.38383>
- Rauch G, Kunzmann K, Kieser M, Wegscheider K, König J, Eulenburger C. A weighted combined effect measure for the analysis of a composite time-to-first-event endpoint with components of different clinical relevance. *Stat Med* 2018;**37**:749–67. <https://doi.org/10.1002/sim.7531>
- Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med* 1999;**18**:1341–54. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990615\)18:11<1341::AID-SIM129>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0258(19990615)18:11<1341::AID-SIM129>3.0.CO;2-7)
- Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med* 2010;**363**:1597–607. <https://doi.org/10.1056/NEJMoa1008232>
- Stone GW, Lindenfeld J, Abraham WT, Kar S, Lim DS, Mishell JM, et al. Transcatheter mitral-valve repair in patients with heart failure. *N Engl J Med* 2018;**379**:2307–18. <https://doi.org/10.1056/NEJMoa1806640>
- Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med* 2010;**29**:3245–57. <https://doi.org/10.1002/sim.3923>
- Ajufo E, Nayak A, Mehra MR. Fallacies of using the win ratio in cardiovascular trials: challenges and solutions. *JACC Basic Transl Sci* 2023;**8**:720–7. <https://doi.org/10.1016/j.jacbt.2023.05.004>
- Butler J, Stockbridge N, Packer M. Win ratio: a seductive but potentially misleading method for evaluating evidence from clinical trials. *Circulation* 2024;**149**:1546–8. <https://doi.org/10.1161/CIRCULATIONAHA.123.067786>
- Pocock SJ, McMurray JJ, Collier TJ. Making sense of statistics in clinical trial reports: part 1 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol* 2015;**66**:2536–49. <https://doi.org/10.1016/j.jacc.2015.10.014>
- Brunner E, Vandemeulebroecke M, Mütze T. Win odds: an adaptation of the win ratio to include ties. *Stat Med* 2021;**40**:3367–84. <https://doi.org/10.1002/sim.8967>
- Solomon SD, McMurray JJV, Anand IS, Ge J, Lam CSP, Maggioni AP, et al. Angiotensin-neprilysin inhibition in heart failure with preserved ejection fraction. *N Engl J Med* 2019;**381**:1609–20. <https://doi.org/10.1056/NEJMoa1908655>
- Packer M, Anker SD, Butler J, Filippatos G, Pocock SJ, Carson P, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. *N Engl J Med* 2020;**383**:1413–24. <https://doi.org/10.1056/NEJMoa2022190>
- Solomon SD, McMurray JJV, Claggett B, de Boer RA, DeMets D, Hernandez AF, et al. Dapagliflozin in heart failure with mildly reduced or preserved ejection fraction. *N Engl J Med* 2022;**387**:1089–98. <https://doi.org/10.1056/NEJMoa2206286>
- Bhatt DL, Steg PG, Miller M, Brinton EA, Jacobson TA, Ketchum SB, et al. Cardiovascular risk reduction with icosapent ethyl for hypertriglyceridemia. *N Engl J Med* 2019;**380**:11–22. <https://doi.org/10.1056/NEJMoa1812792>
- Claggett B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. *Circulation* 2018;**138**:570–7. <https://doi.org/10.1161/CIRCULATIONAHA.117.033065>
- Maurer MS, Schwartz JH, Gundapaneni B, Elliott PM, Merlini G, Waddington-Cruz M, et al. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med* 2018;**379**:1007–16. <https://doi.org/10.1056/NEJMoa1805689>
- Pocock SJ, Collier TJ. Statistical appraisal of 6 recent clinical trials in cardiology: JACC state-of-the-art review. *J Am Coll Cardiol* 2019;**73**:2740–55. <https://doi.org/10.1016/j.jacc.2019.03.484>
- Stone GW, Lindenfeld J, Rodés-Cabau J, Anker SD, Zile MR, Kar S, et al. Interatrial shunt treatment for heart failure: The randomized RELIEVE-HF trial. *Circulation* 2024; **In Press**. <https://doi.org/10.1161/CIRCULATIONAHA.124.070870>
- 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. *Eur J Heart Fail* 2023;**25**:632–41. <https://doi.org/10.1002/ehf.2853>
- 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;**28**:568–74. <https://doi.org/10.1038/s41591-021-01659-1>
- Sorajja P, Whisenant B, Hamid N, Naik H, Makkar R, Tadros P, et al. Transcatheter repair for patients with tricuspid regurgitation. *N Engl J Med* 2023;**388**:1833–42. <https://doi.org/10.1056/NEJMoa2300525>
- Gillmore JD, Judge DP, Cappelli F, Fontana M, Garcia-Pavia P, Gibbs S, et al. Efficacy and safety of acoramidis in transthyretin amyloid cardiomyopathy. *N Engl J Med* 2024;**390**:132–42. <https://doi.org/10.1056/NEJMoa2305434>

30. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;**332**: 1080. <https://doi.org/10.1136/bmj.332.7549.1080>
31. Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. *Epidemiology* 1992;**3**: 434–40. <https://doi.org/10.1097/00001648-199209000-00009>
32. ClinicalTrials.gov. *Selatogrel Outcome Study in Suspected Acute Myocardial Infarction (SOS-AMI)*. NCT04957719. <https://clinicaltrials.gov/study/NCT04957719> (20 September 2024).
33. Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharm Stat* 2016;**15**:238–45. <https://doi.org/10.1002/pst.1743>
34. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, et al. NXY-059 for acute ischemic stroke. *N Engl J Med* 2006;**354**:588–600. <https://doi.org/10.1056/NEJMoa052980>
35. ClinicalTrials.gov. *REvascularization CHOices Among Under-Represented Groups Evaluation: The RECHARGE Trial*. NCT06399705. <https://clinicaltrials.gov/study/NCT06399705> (20 September 2024)
36. Holm NR, Mäkikallio T, Lindsay MM, Spence MS, Erglis A, Menown IBA, et al. Percutaneous coronary angioplasty versus coronary artery bypass grafting in the treatment of unprotected left main stenosis: updated 5-year outcomes from the randomised, non-inferiority NOBLE trial. *Lancet* 2020;**395**:191–9. [https://doi.org/10.1016/S0140-6736\(19\)32972-1](https://doi.org/10.1016/S0140-6736(19)32972-1)
37. Stone GW, Kappetein AP, Sabik JF, Pocock SJ, Morice MC, Puskas J, et al. Five-year outcomes after PCI or CABG for left main coronary disease. *N Engl J Med* 2019;**381**: 1820–30. <https://doi.org/10.1056/NEJMoa1909406>
38. James S, Erlinge D, Storey RF, McGuire DK, de Belder M, Eriksson N, et al. Dapagliflozin in myocardial infarction without diabetes or heart failure. *NEJM Evid* 2024;**3**: EVIDoA2300286. <https://doi.org/10.1056/EVIDoA2300286>
39. Grandy S, Hashemi M, Langkilde AM, Parikh S, Sjöström CD. Changes in weight loss-related quality of life among type 2 diabetes mellitus patients treated with dapagliflozin. *Diabetes Obes Metab* 2014;**16**:645–50. <https://doi.org/10.1111/dom.12263>
40. Inzucchi SE, Docherty KF, Køber L, Kosiborod MN, Martinez FA, Ponikowski P, et al. Dapagliflozin and the incidence of type 2 diabetes in patients with heart failure and reduced ejection fraction: an exploratory analysis from DAPA-HF. *Diabetes Care* 2021;**44**: 586–94. <https://doi.org/10.2337/dc20-1675>
41. Pocock SJ, Stone GW. The primary outcome fails—what next? *N Engl J Med* 2016;**375**: 861–70. <https://doi.org/10.1056/NEJMra1510064>
42. Kandzari DE, Townsend RR, Kario K, Mahfoud F, Weber MA, Schmieder RE, et al. Safety and efficacy of renal denervation in patients taking antihypertensive medications. *J Am Coll Cardiol* 2023;**82**:1809–23. <https://doi.org/10.1016/j.jacc.2023.08.045>
43. Eikelboom JW, Yi Q, McIntyre WF, Bosch J, Whitlock R, Connolly SJ, et al. Results of the COMPASS trial analyzed using win ratio compared with conventional analytic approaches. *Canadian Journal of Cardiology* 2024;**In Press**. <https://doi.org/10.1016/j.cjca.2024.07.002>
44. Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: a practical guide with examples. *Stat Med* 2011;**30**:3267–84. <https://doi.org/10.1002/sim.4102>
45. Althunian TA, de Boer A, Groenwold RHH, Klungel OH. Defining the noninferiority margin and analysing noninferiority: an overview. *Br J Clin Pharmacol* 2017;**83**: 1636–42. <https://doi.org/10.1111/bcp.13280>