






# The win ratio approach for composite endpoints: practical guidance based on previous experience

Björn Redfors <sup>1,2,3</sup>, John Gregson <sup>4</sup>, Aaron Crowley <sup>1</sup>, Thomas McAndrew <sup>1</sup>, Ori Ben-Yehuda<sup>1,2</sup>, Gregg W. Stone <sup>1,5</sup>, and Stuart J. Pocock<sup>4\*</sup>

<sup>1</sup>Clinical Trials Center, Cardiovascular Research Foundation, New York, NY, USA; <sup>2</sup>Division of Cardiology, NewYork-Presbyterian Hospital/Columbia University Irving Medical Center, New York, NY, USA; <sup>3</sup>Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden; <sup>4</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E7HT, UK; and <sup>5</sup>The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Received 14 May 2020; revised 1 July 2020; editorial decision 25 July 2020; accepted 29 July 2020; online publish-ahead-of-print 9 September 2020

The win ratio was introduced in 2012 as a new method for examining composite endpoints and has since been widely adopted in cardiovascular (CV) trials. Improving upon conventional methods for analysing composite endpoints, the win ratio accounts for relative priorities of the components and allows the components to be different types of outcomes. For example, the win ratio can combine the time to death with the number of occurrences of a non-fatal outcome such as CV-related hospitalizations (CVHs) in a single hierarchical composite endpoint. The win ratio can provide greater statistical power to detect and quantify a treatment difference by using all available information contained in the component outcomes. The win ratio can also incorporate quantitative outcomes such as exercise tests or quality-of-life scores. There is a need for more practical guidance on how best to design trials using the win ratio approach. This manuscript provides an overview of the principles behind the win ratio and provides insights into how to implement the win ratio in CV trial design and reporting, including how to determine trial size.

## Keywords

Win ratio • Composite endpoints • Methods • Clinical trials

The efficacy of a randomized treatment in cardiovascular (CV) clinical trials is often evaluated using a primary composite endpoint consisting of several individual components (e.g. death, myocardial infarction (MI), or target vessel revascularization (TVR)). Conventional statistical methods such as the Kaplan–Meier estimator, log-rank test, and Cox proportional hazards regression focus on the time-to-first occurrence of any event in the composite. These analyses ignore the differences in clinical severity of the individual components.<sup>1–5</sup> For the composite endpoint of death, MI, or TVR, a patient that requires a target vessel re-intervention at 3 months is considered to have had the composite endpoint at 3 months, and whether that patient experiences a more serious event (e.g. death) thereafter is ignored. A re-intervention in one patient that occurs at 3 months is given a higher priority than a death occurring in another patient at 6 months. Thus, non-fatal events, irrespective of how serious they are, are considered just as important as fatal events and often occur earlier. Despite the

shortcomings of time-to-first event analyses of composite endpoints, they are still widely used in CV trials.<sup>1–5</sup> Finally, it is difficult to incorporate changes in quantitative measures such as exercise tests or quality-of-life scores in conventional clinical composite outcomes.

The win ratio was introduced in this journal as a new approach for analysing composite endpoints with varying severity, and to account for the relative priority of components.<sup>6–11</sup> The win ratio, unlike more conventional methods restricted to composites of a single variable type, can analyse composites composed of time-to-event, recurrent event, continuous, and/or categorical outcomes. For example, time to the first occurrence of one event can be combined with the number of occurrences of another event, or with a categorical or quantitative metric.<sup>6,7</sup>

The hierarchical structure, statistical power, and flexibility of the win ratio approach make it an attractive alternative for comparing the efficacy of randomized treatments. However, given its novelty,

\* Corresponding author. Tel: 0044 (0)20 7927 2413. Email: [stuart.pocock@shhtm.ac.uk](mailto:stuart.pocock@shhtm.ac.uk)

Published on behalf of the European Society of Cardiology. All rights reserved. © The Author(s) 2020. For permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

there is as yet little published advice available on how best to design trials using the win ratio approach. This manuscript provides such guidance.

## The win ratio

The win ratio was motivated by the Finkelstein–Schoenfeld (FS) test, with the aim of providing an estimate of the treatment effect (the win ratio) and confidence interval, in addition to a *P*-value.<sup>6,7</sup> Briefly, the general principle behind both the FS and the win ratio is as follows. One first forms every possible patient-to-patient pair: that is, every patient on the new treatment is compared with every patient on the control treatment. Then, within each pair one evaluates the component outcomes in descending order of importance until one of the pair shows a better outcome compared with the other. If the patient on the new treatment has the better outcome it is called a 'win', whereas if the control patient does better it is a 'loss'. Otherwise, it is a 'tie'. For instance, consider the composite endpoint of death, stroke, and number of hospitalizations due to heart failure. Death is the most severe event, followed by stroke and lastly heart failure hospitalizations (HFH). For each patient pair we assess who died first. If neither of the patients in that pair died, we assess who had a stroke first, and if neither patient had a stroke, we evaluate who was hospitalized due to heart failure (HF) the most times (Table 1). All three pairwise comparisons are over the pair's shared duration of follow-up. A simplified example with the composite endpoint death or number of HF hospitalizations is illustrated in Figure 1.

The win ratio approach may consider unmatched or matched patient pairs. In the unmatched win ratio approach (as described in the preceding example), every patient in the Treatment group is compared with every patient in the Control group. Specifically, if we let  $N_T$  and  $N_C$  be the number of patients in the Treatment and Control groups, respectively, then we make  $N_T \times N_C$  paired comparisons. The win ratio ( $R_W$ ) is then calculated as  $N_W/N_L$ , where  $N_W$  and  $N_L$  are the total number of pairwise wins and losses, respectively, for the Treatment group. In this regard, the unmatched win ratio is consistent with the FS test, providing a magnitude of the effect estimate with the same *P*-value. A more detailed description of the unmatched win ratio method, including formulae for calculating the effect size, 95% confidence interval and *P*-value, are provided elsewhere.<sup>6</sup>

Figure 2 is an illustrative example of the win ratio approach in the COAPT trial, which compared the MitraClip device plus guideline-recommended medical therapy vs. guideline-recommended medical therapy alone in 614 patients with heart failure and functional mitral regurgitation over 2 years of follow-up.<sup>12</sup> This was a pre-planned secondary analysis based on a hierarchical composite outcome of time to all-cause death then time to first HFH. Amongst the  $302 \times 312 = 94\,224$  pairs the consequent numbers of wins (in green) and losses (in red) are shown first for death (i.e. who died first) and then HFH (who experience an HFH first in pairs for whom neither patient died). The total wins 42 330 are divided by the total losses 26 277 to give a win ratio of 1.61 with 95% CI 1.29–2.04 and  $P < 0.0001$ .

A relevant question is how do we interpret the value of the win ratio? That is, what does a win ratio of 1.61 actually mean? If any two patients are compared, one on device and one on control, and they are not a tie, then the odds that the device patient is the winner is

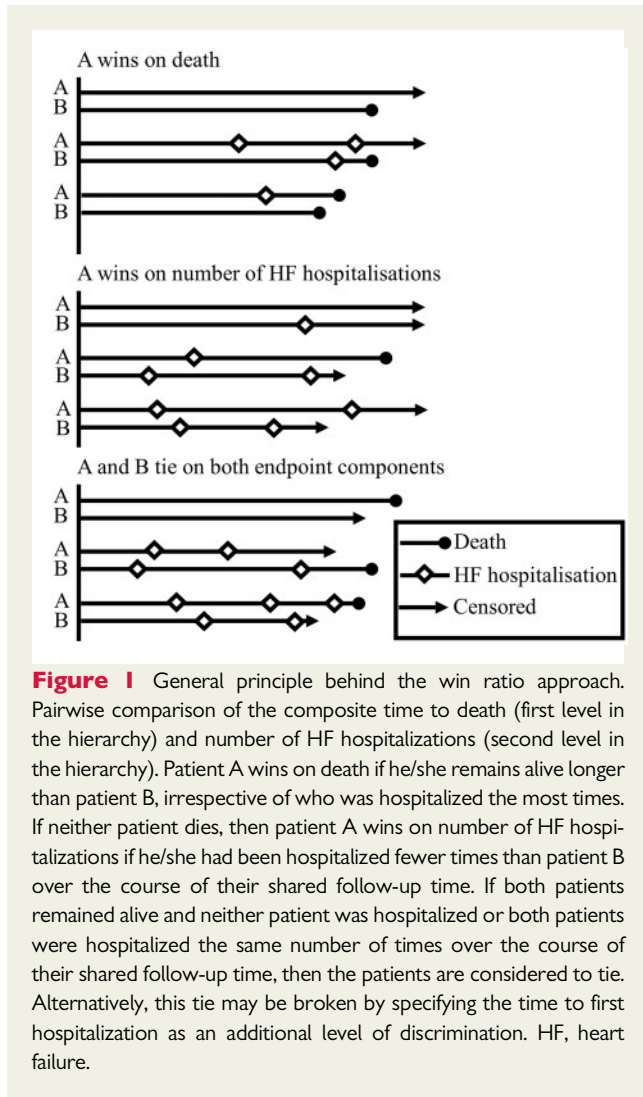
**Table 1** General principle behind the win ratio approach

Comparison of each potential patient pair (e.g. one from treatment group and one from control group)	How event is assessed
<b>Step 1</b>	
If 1 or both patients die	
Patient in treatment group dies first	Control wins
Patient in control group dies first	Treatment wins
Both patients die on the same day	Go to Step 2
If neither of the patients die	Go to Step 2
<b>Step 2</b>	
If no ranking yet available	
Patient in treatment group has a stroke first	Control wins
Patient in control group has a stroke first	Treatment wins
Both patients have a stroke on the same day	Go to Step 3
If neither of the patients have a stroke	Go to Step 3
<b>Step 3</b>	
If no ranking yet available	
Patient in treatment group had more hospitalizations for heart failure	Control wins
Patient in control group had more hospitalizations for heart failure	Treatment wins
Both patients were hospitalized equal number of times	Tie
Neither of the patients was hospitalized for heart failure	Tie

1.61. For those who prefer probabilities, the probability that the device patient wins is  $1.61/(1.61 + 1) = 0.62$ .

An alternative is the matched win ratio, which attempts to account for each patient's underlying risk of the composite endpoint by forming matched pairs of patients in the Treatment and Control groups who have similar risks. The underlying rationale for the matched win ratio is that avoiding pairwise comparisons between patients with different baseline risk would increase  $R_W$  and enhance statistical power. Unfortunately, experience has shown that it is difficult to objectively define the matching process in advance, and it is often not possible to match all patients. Hence, we favour the unmatched win ratio to which we devote the remainder of this article.

A better way to avoid comparisons of patient pairs with different baseline risk is to use the stratified win ratio, which is another variant of the unmatched win ratio that attempts to control for known prognostic variables by dividing patients into strata based on prognostically meaningful variables, and then perform pairwise comparisons and count the wins and losses within each stratum.<sup>13</sup> The stratum-specific numbers of wins and losses are then combined across strata to estimate the stratified win ratio. By comparing only patient pairs within the same strata, the influence of the stratification variables on prognosis can be reduced and statistical power can be enhanced. The stratified win ratio was used in the recent ATTR-ACT trial, which compared tafamidis to placebo for patients with transthyretin amyloid cardiomyopathy.<sup>14</sup> In ATTR-ACT, stratification was done



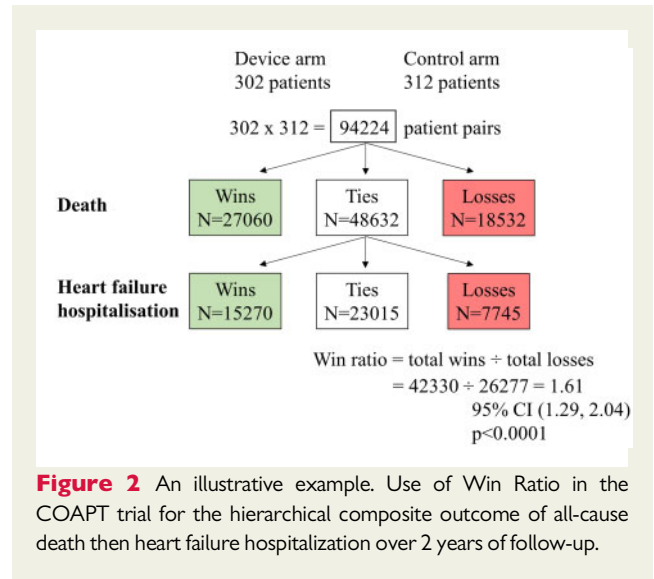
**Figure 1** General principle behind the win ratio approach. Pairwise comparison of the composite time to death (first level in the hierarchy) and number of HF hospitalizations (second level in the hierarchy). Patient A wins on death if he/she remains alive longer than patient B, irrespective of who was hospitalized the most times. If neither patient dies, then patient A wins on number of HF hospitalizations if he/she had been hospitalized fewer times than patient B over the course of their shared follow-up time. If both patients remained alive and neither patient was hospitalized or both patients were hospitalized the same number of times over the course of their shared follow-up time, then the patients are considered to tie. Alternatively, this tie may be broken by specifying the time to first hospitalization as an additional level of discrimination. HF, heart failure.

according to disease subtype (deposition of mutated vs. wild-type transthyretin protein) and New York Heart Association (NYHA) class (I or II vs. III), both of which are strongly associated with prognosis. The details of the consequent win ratio analysis and its interpretation have been described.<sup>15</sup>

## Examples of trials that have used the win ratio

Since the win ratio method was introduced in 2012, there has been a growth in its use, including several medical device trials aimed at FDA approval. The win ratio (or the closely related FS method) was used to evaluate the primary endpoint in three recently completed trials,<sup>1,14,16,17</sup> and is the pre-defined method for evaluating the primary endpoint in several ongoing trials (Table 2).<sup>18</sup>

The ATTR-ACT trial was a double-blind trial that randomized 441 patients with transthyretin amyloid cardiomyopathy to tafamidis ( $n = 264$ ) or placebo ( $n = 177$ ) in a 3:2 ratio.<sup>14</sup> As mentioned above, stratification was done according to disease subtype and NYHA class.



**Figure 2** An illustrative example. Use of Win Ratio in the COAPT trial for the hierarchical composite outcome of all-cause death then heart failure hospitalization over 2 years of follow-up.

The primary composite endpoint was the hierarchical occurrence of all-cause death followed by the frequency of CV-related hospitalizations (CVHs). In the tafamidis group, 72 patients (27.2%) died over the course of the 30-month study period and 267 CVHs occurred in 138 patients (52.3%). In the control group, 72 patients (40.7%) died and 231 CVHs occurred in 107 patients (60.5%).

Conventional methods for analysing composite endpoints could not fully account for these events and their clinical implications. A time-to-first event analysis of the occurrence of either death or CVHs would not account for the fact that death is more severe than a CVH, and it would disregard all CVHs or deaths that occurred after the first event. Hence, it did not show a clear signal of treatment efficacy. Since repeat CVHs were common in ATTR-ACT, a considerable amount of information would be lost with a conventional time-to-first-event analysis. An alternative repeated events analysis for comparing the occurrence of the composite endpoint death or CVH would count all CVHs, but would give greater weight to CVHs than death, since they are more frequent and tend to occur earlier in follow-up.<sup>19</sup> Hence, neither of these alternatives are ideal for analysing the ATTR-ACT trial's composite endpoint.

By incorporating two different variable types (time to mortality and the frequency of CVHs) in a hierarchical order, the win ratio accounts for the fact that death is more severe than CVH and uses information from all CVHs. The win ratio for treatment with tafamidis vs. placebo was highly significant at 1.70 [95% confidence interval (CI) 1.26–2.29,  $P = 0.00006$ ], a more impressive result than would be achieved by alternative analyses.

The win ratio has also been used to re-analyse the primary composite endpoints of several completed trials (Table 3).<sup>6,20–24</sup> For re-analysis of composite endpoints that were originally constrained to include only components of the same type (e.g. time-to-first event), the win ratio performs well compared with conventional approaches based on hazard ratios. Therefore, the win ratio appears to be a reasonable alternative to conventional methods for analysing composite endpoints in many clinical settings.

**Table 2** Trials that have applied the win ratio approach as the pre-defined method to analyse their primary composite endpoint

Trial	Population	Randomized treatment	Primary composite endpoint	Win ratio (95% CI)
ATTR-ACT <sup>14</sup>	Transthyretin amyloid cardiomyopathy	Tafamidis vs. placebo	All-cause mortality > number of heart failure hospitalizations	1.70 (1.26–2.29)
CHART-1 <sup>16</sup>	LVEF ≤35%	Cardiopoietic stem cells vs. placebo	Time to death > N of HF events > MLHFQ score ≥10-point improvement > 6MWT improvement ≥40 m > LVESV change ≥15 mL > LVEF change ≥4%.	1.17 (0.89–1.55)
TAVR-UNLOAD <sup>18</sup>	Moderate AS and reduced LVEF	TAVR vs. medical therapy	Time to death > disabling stroke > hospitalizations due to HF, aortic valve disease, or non-disabling stroke > change in KCCQ relative to baseline	Ongoing
RELIEVE-HF (NCT03499236)	NYHA class III and IV heart failure	Inter-atrial shunt vs. medical therapy	Time to death > time to heart transplant or LVAD > number and time of hospitalizations due to HF > improvement in 6MWT	Ongoing
CARILLION (NCT03142152)	Functional MR associated with HF	Carillion implant vs. medical therapy	Death > cardiac transplantation or LVAD > percutaneous or surgical mitral valve intervention > time to first HF hospitalization > improvement in 6MWT	Ongoing
ACTIVE (NCT03016975)	Functional MR associated with HF	Cardioband implant vs. medical therapy	Death > number of HF hospitalizations > improvement in 6MWT > improvement in KCCQ	Ongoing
PARACHUTE-HF (NCT04023227)	HF with reduced LVEF caused by chronic Chagas disease	Sacubitril/valsartan vs. enalapril	CV death > HF hospitalization > relative change in NT-proBNP from baseline to week 12	Ongoing

6MWT, 6-min walk test; AS, aortic stenosis; HF, heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; LVAD, left ventricular assist device; LVEF, left ventricular ejection fraction; LVESV, left ventricular end-systolic volume; MLHFQ, Minnesota Living with Heart Failure Questionnaire; MR, mitral regurgitation; TAVR, transcatheter aortic valve replacement. > Designates the order of the win ratio hierarchy, which decreases from left to right.

## Guidance on how to choose the hierarchy of endpoints

### General principle (clinical priorities)

The essence of the win ratio approach is that it accounts for clinical priorities among the component endpoints. Once we have decided which variables should sensibly contribute to a primary win ratio analysis, we then need to rank their clinical priorities. Consider an HF trial with the composite primary endpoint death, number of HFHs, and improvement in the Kansas City Cardiomyopathy Questionnaire (KCCQ) score. Death is the most severe event, and given the prognostic implications of recurrent HFHs, we consider them to be more clinically important than any lack of improvement in KCCQ score.

Selecting appropriate components for the composite endpoint should be done with care: power would be adversely affected if a component were to truly have no treatment effect or an inverse effect. The estimate of the win ratio may be sensitive to the chosen order of outcomes, as reported by Ferreira *et al.*<sup>24</sup> in an analysis from several HF and hypertension trials. The relative contribution of each component of the primary outcome is also sensitive to the duration of follow-up. More patients will die in a longer trial, and so the relative contribution of death will be greater.

## Guidance on how to compare individual components

### Adverse clinical events

Adverse clinical events are the most important outcomes in most clinical trials. Most non-fatal adverse events may occur several times in the same patient and the same event-type may vary considerably in severity. When conducting the win ratio analysis, we must decide whether to compare patients in regard to (i) simply whether they experienced the event, (ii) how soon they experienced the event, (iii) how many events they experienced, or (iv) how severe the events were. In general, we discourage simply comparing patients in regard to whether they had the event (option i), **since ignoring information on timing or frequency of events omits important information.** The choice as to how to prioritize outcomes in regard to the time a patient remains free of the event, vs. the number of events experienced over the study period, vs. the clinical severity of the event should be guided by clinical reasoning. For example, for the endpoint of TVR in a trial that compared two coronary stent types, it may be reasonable to prioritize comparing pairs of patients in regard to who had a target vessel re-intervention first—since any subsequent re-intervention of the same vessel may be related to the second procedure rather than

**Table 3** Trials that have re-analysed their primary composite endpoint using the win ratio approach

Trial	Population	Randomized treatment	Primary composite endpoint	Outcome	
				Primary analysis	Win ratio
PARTNER B* <sup>6</sup>	Severe symptomatic aortic stenosis	TAVR vs. OMT	Death, or hospitalization due to valve- or procedure-related clinical deterioration	1/HR 2.17 (1.69, 2.86)	1.87 (1.35, 2.54)
EMPHASIS-HF <sup>6</sup>	NYHA class II and ejection fraction $\leq 35\%$	Eplerenone vs. placebo	CV death or HF hospitalization	1/HR 1.59 (1.35, 1.85)	1.61 (1.37, 1.89)
CHARM <sup>6</sup>					
Added	HF, LVEF $<40\%$ and on ACEI	Candesartan vs. placebo	CV death or HF hospitalization	1/HR 1.18 (1.04, 1.18)	1.30 (1.13, 1.50)
Alternative	HF, LVEF $<40\%$ and intolerant to ACEI	Candesartan vs. placebo	CV death or HF hospitalization	1/HR 1.30 (1.12, 1.49)	1.42 (1.20, 1.70)
Preserved	HF, LVEF $\geq 40\%$	Candesartan vs. placebo	CV death or HF hospitalization	1/HR 1.12 (0.97, 1.30)	1.17 (0.99, 1.39)
EVOLVE <sup>20</sup>	End-stage CKD and $\geq$ moderate hyperparathyroidism	Cinacalcet vs. placebo	Death, MI, hospitalization due to unstable angina, HF or peripheral vascular event	1/HR 1.08 (0.98, 1.18)	1.09 (0.97, 1.21)
ATLAS ACS <sup>22</sup>	ACS	Rivaroxaban vs. placebo	Death, stroke, MI	1/HR 1.28 (1.03, 1.61)	1.30 (1.06, 1.59)
SYNTAX <sup>21</sup>	De-novo 3-vessel or left main coronary artery disease	CABG vs. PCI for multi-vessel CAD	Death, stroke, MI, revascularization	1/HR 1.29 (1.04, 1.62)	1.28 (1.11, 1.53)
TRILOGY ACS <sup>23</sup>	NSTEMI, medically treated	Prasugrel vs. Clopidogrel	CV death, MI, or stroke	1/HR 1.04 (0.94, 1.16)	1.05 (0.94, 1.18)
DIG <sup>24</sup>	LVEF $\leq 45\%$	Digoxin vs. placebo	CV death or HF hospitalization	1/HR 1.18 (1.10, 1.27)	1.14 (1.05, 1.20)
SPRINT <sup>24</sup>	SBP $\geq 130$ mmHg and increased CV risk	Intense vs. less intense BP control	CV death	1/HR 1.33 (1.12, 1.56)	1.39 (1.16, 1.67)
EPHESUS <sup>24</sup>	Acute MI	Eplerenone vs. placebo		1/HR 1.15 (1.04, 1.27)	1.15 (1.05, 1.27)
CORONA <sup>24</sup>	Ischaemic heart failure, NYHA class $\geq 2$	Rosuvastatin vs. placebo	CV death, stroke, MI, HF hospitalization	1/HR 1.11 (1.09, 1.20)	1.09 (1.00, 1.19)
PARADIGM-HF <sup>24</sup>	HF, NYHA class $\geq 2$ , LVEF $\leq 40\%$ .	Entresto vs. Enalapril	CV death or HF hospitalization	1/HR 1.25 (1.12, 1.41)	1.27 (1.15, 1.39)
			CV death, stroke, MI, HF hospitalization	1/HR 1.20 (1.10, 1.30)	1.21 (1.11, 1.32)
			CV death, HF hospitalization, KCCQ $\geq 5$ -point decrease, absence of $>30\%$ decrease in NT-pro BNP	1/HR 1.12 (1.02, 1.25)	1.30 (1.14, 1.67)

ACEI, angiotensin-converting enzyme inhibitor; ACS, acute coronary syndrome; CAD, coronary artery disease; CKD, chronic kidney disease; HF, heart failure; HR, hazard ratio; LVEF, left ventricular ejection fraction; MI, myocardial infarction; NSTEMI, non-ST-elevation acute coronary syndrome; NYHA, New York Heart Association; OMT, optimal medical therapy.\*Specified in the study protocol as a co-primary analysis to be analyzed using the Finkelstein-Schoenfeld test, and subsequently re-analyzed using the win ratio approach.

the study stent. In contrast, for the endpoint of recurrent HFH in a trial examining the effect of two therapies for HF, it may be more reasonable that the number of HFHs is the first level of comparison (rather than the time to first HFH), since the number of HFHs is strongly associated with prognosis in patients with HF. Patients could be compared first for the number of HFHs, and if that is a tie, then subsequently for the time to the first HFH as the 'tie-breaker'. Lastly, for the endpoint of ischaemic stroke in a trial examining the effect of cerebral protection devices for reduction of peri-procedural stroke it may be more meaningful to compare patients with regard to the severity of the stroke rather than the exact timing or number of strokes.

The method used to compare each outcome may affect study power, depending on the expected effect of a treatment on the endpoint itself compared with the endpoints lower in the hierarchy. It may therefore be attractive to choose a means of comparison that maximizes study power.

Irrespective of whether patients are compared in regard to the time of the first event, the number of events, or the severity of the event, it is important to take into consideration that each patient pair can only be compared for the shared follow-up duration they both achieved. For example, if one patient had an event only after the other patient was lost to follow-up, that event should not be considered, and a 'tie' would be declared for that pairwise comparison.

## Patient-reported outcomes

In recognition of the importance of symptom relief and overall well-being for patients with heart disease,<sup>25</sup> there has been an increasing interest in patient-reported outcomes in CV studies.<sup>26</sup> Quality-of-life questionnaires like the KCCQ are increasingly used to assess health status in clinical trials.<sup>27</sup> One advantage of the win ratio method is that it can incorporate patient-reported outcomes such as the KCCQ and adverse events in the same composite endpoint. Usually, patient-reported outcomes are placed lower in the hierarchy than clinical adverse events since the latter have greater prognostic implications.

Patient-reported outcomes can be assessed using either generic or disease-specific instruments. These instruments attempt to score patient well-being based on a series of self-reported items and such quantitative scores can be re-assessed at several follow-up times. Patients can then be compared for the extent to which they improved or worsened over the course of their shared follow-up time, or the comparison can be assessed at one specific time-point after randomization.

## Pathophysiological measures

Pathophysiological measures, which include indices of cardiac function (e.g. left ventricular ejection fraction), myocardial infarct size or tests of endurance and aerobic capacity (e.g. 6-min walk test) are generally considered less important than clinical adverse events. These variables should therefore be placed lower than adverse events in the win ratio hierarchy. The relative importance of pathophysiological measures compared with patient-reported outcomes such as the KCCQ is less obvious.

Like patient-oriented outcomes, pathophysiological measures are usually quantitative variables. Such variables can be useful to include in the win ratio hierarchy of outcomes because their diversity of values means that most pairwise comparisons identify a winner. This may substantially increase the power of the win ratio approach if there exists a treatment effect at that level of the hierarchy.

An alternative approach for such quantitative variables is to categorize them. Patients can be classified into those that responded or worsened to therapy (by improving or deteriorating by a clinically meaningful amount) and others who did not. More than two categories can be used. But a consequence of categorization is that considerably more patient pairs will tie with a potential loss of statistical power.

Another use of the win ratio approach is for a single continuous non-normal outcome.<sup>28</sup> While the Mann–Whitney test provides a *P*-value, use of the win ratio provides a useful estimate and CI. For instance, in trials of new treatments for COVID-19 patients, a 7-point ordinal scale 15 days after randomization has been proposed as the primary outcome. A proportional odds model is one analytical approach, yielding a common odds ratio. But this makes potentially unjustifiable assumptions whereas the win ratio method does not.

## The use of a margin

It has been suggested that one should declare a winner or loser only if the patient pair differs by a clinically meaningful quantity, or margin. The use of such a margin increases the number of ties at that level in the hierarchy rather than allowing any difference, no matter how

small, to discriminate. The clinical rationale for using a margin is that if one patient in a pair experiences a given adverse event e.g. 1 day earlier or later than the other patient or improves a patient-reported outcome by a single point more or less than the other patient, it is not a clinically meaningful win or loss for that pair. But it may be challenging to define what the margin should be, i.e. what constitutes a clinically meaningful difference.

There are also statistical arguments against the use of such margins. First, their use will likely reduce statistical power by reducing the number of decisions that are made at that level. Second, the win ratio (and FS test) can be seen as a type of non-parametric test, building on simpler methods such as the two-sample Wilcoxon test for a single quantitative outcome. There, the ranking of subjects from highest to lowest value takes into account the close proximity of some pairs of values but does not consider them equal (by using a margin of difference). Hence, imposing a clinically meaningful margin is contrary to the principle of rank-based statistical testing and is best avoided.

## The win ratio approach for a treatment that is not expected to affect mortality

When using the win ratio, death should generally be included as the top of the hierarchy, since a death should be considered the ultimate loss. The use of the win ratio to evaluate the effect of a treatment that is not expected to affect mortality poses a dilemma, since inclusion of death in the top of the hierarchy then may lead to loss of statistical power. This issue may be small if the death rates are low. If death rates are high, the loss of statistical power may be substantial, and it may be tempting to exclude death from the hierarchy. However, excluding death from the primary outcome may be deemed untenable. Even if acceptable, deaths would then become a competing risk for the remaining outcomes in the hierarchy. Thus, if mortality is substantial and the anticipated treatment effect is confined to non-fatal outcomes, one might be better off choosing other methods than the win ratio.

A related issue is that in CV trials the treatment may be unlikely to affect the risk of non-CV death, and one could consider replacing all-cause death by CV death. But one should bear in mind that non-CV deaths would then constitute a competing risk for all outcomes in the hierarchy. If the relative incidence of non-CV deaths is expected to be low and CV death is expected to benefit from treatment, we recommend the use of all-cause death at the top of the hierarchy.

## Study power and sample size determination

Power calculations to determine the required number of patients are an important part of any trial protocol, but at present there is little guidance available on how to calculate sample size for the win ratio approach. This section provides recommendations for how to do this, using simulation techniques along with specific examples. An R program for calculating the sample size for the win ratio is included in the [Supplementary material online, Statistical Appendix](#).

**Table 4** Examples of sample size estimation for the win ratio

Follow-up	First tier: death			Second tier: number of CVH			Third tier: reduction in KCCQ			Power (%)	Sample size
	Control rate (per year) (%)	Hazard ratio	Decisions <sup>a</sup> (%)	Control rate (per year)	Rate ratio	Decisions <sup>a</sup> (%)	Control, mean $\pm$ SD	Treatment, mean $\pm$ SD	Decisions <sup>a</sup> (%)		
3 years <sup>b</sup>	21.5	0.75	67.8	0.70	0.75	32.1				80	1050
1 year <sup>c</sup>	21.5	0.80	47.1	0.70	0.80	52.9				80	1948
1 year <sup>c</sup>	21.5	0.80	31.0	0.70	0.80	36.6	5 $\pm$ 12.0	0 $\pm$ 12.0	32.4%	80	590

<sup>a</sup>The decisions (%) columns refer to the percentage of all non-tied patient pairs that have a winner/loser at that level in the hierarchy.

<sup>b</sup>Median follow-up of 3 years; variable between 2.5 and 3.5 years.

<sup>c</sup>Fixed follow-up of 1 year.

CVH, cardiovascular hospitalization; KCCQ, Kansas City Cardiomyopathy Questionnaire; SD, standard deviation.

## Examples of sample size estimations

To illustrate how the sample size for the win ratio is derived, we present three hypothetical trial scenarios in *Table 4*. First, we consider a trial designed to evaluate the effect of a new therapy vs. control for patients with HF on the hierarchical composite of all-cause death and the number of CVHs with patient follow-up between 2.5 and 3.5 years. To derive the necessary sample size for the trial to achieve 80% power to demonstrate an effect of the therapy on the primary composite endpoint using the win ratio, we must first propose what the rates of death and HF would be for the patients randomized to the control arm. For illustration, we use similar expected death and CVH rates as those observed in the control arm from the ATTR-ACT trial: an annual death rate of 21.5% and an annual rate of CVH of 0.7 per patient. We plan on treatment conferring 25% relative reductions in the risks of death and CVH and calculate the statistical power to show superiority of our intended therapy under this alternative hypothesis. It is important to account for the non-independence of CVHs (i.e. their tendency to occur more frequently in high risk, more frail patients) as it can have a major impact on the required sample size. To achieve this, we simulated data which allow for such variation in patient frailty.<sup>29</sup> We assumed that around a third of HFHs would occur in the 10% of patients at highest risk. We also modelled a link between HFH and death, such that those at high risk of HFH were more likely to die early (further details are given in the [Supplementary material online, Statistical Appendix](#)). Whereas ATTR-ACT had an unbalanced randomization, we herein consider sample size for a conventional trial with 1:1 randomization.

Once the above parameters have been specified, an iterative simulation process can be used to determine sample size (see [Supplementary material online, Statistical Appendix](#)). For 80% power, we calculated a required sample size of 1050 patients. This is substantially smaller than the sample size required to achieve 80% power for a conventional time-to-first-event analysis using Cox proportional hazards regression, which would be 1284 patients.

Now we consider a trial that plans to examine the effect of a new HF therapy in the same population, but that is restricted to examine the treatment effect at 1 year, and for which the expected treatment effect is smaller. With only a fixed 1-year follow-up, and an expected relative risk reduction of 20% rather than 25% in the treatment arm, this trial would require a sample size of 1948 patients to achieve 80%

power for the hierarchical composite endpoint we used in the earlier example.

A means of reducing the sample size required for the win ratio is to include an additional endpoint in the hierarchical composite. Since quality-of-life questionnaires such as the KCCQ have gained considerable recognition as important endpoint in clinical trials,<sup>25–27</sup> and since the win ratio allows different types of variables to be included in the hierarchical composite endpoint, KCCQ would be a reasonable addition that would increase the statistical power of the win ratio. To achieve a smaller trial size, we therefore choose the hierarchical composite endpoint of death, number of CVHs and KCCQ at 1 year. We simulate the change in KCCQ scores for the patients in the control arm based on data from ATTR-ACT, as a mean reduction (worsening) of 5.0 points (standard deviation 12.0 points) and assume a reduction of 0 points in the treatment arm. With this expanded hierarchical composite endpoint, the sample size required to achieve 80% power would be 590 patients. This reduction in trial size comes at the expense of incorporating a subjective ‘softer’ outcome measure, the KCCQ, into the hierarchical composite endpoint. The reduction in trial size also means that comparisons of individual endpoints will be less precise.

Some patients may discontinue the study or may be missing data on one of the component endpoints. An advantage of the win ratio is that it handles missing data easily, under the usual assumption of missing at random. If one patient in a patient pair is missing data for an endpoint, then we simply proceed to compare the pair at the next level. Of course, the presence of missing data is likely to reduce study power and one should consider inflating the proposed sample size to account for this. Missing data or withdrawals alter the proportion of decisions that are decided at each level of the hierarchy, which can alter the estimated win ratio.

## Limitations of the win ratio approach

One limitation of the win ratio approach is its novelty, whereby some trialists may have difficulty in conceptualizing what it means in terms of clinical relevance. The interpretation we offer is to take any patient on the new treatment and compare them with any control patient. If not tied in their outcome, then the win ratio is the odds that the new

Advantages	Challenges
<p><b>1. All Key Elements Included</b> The win ratio recognizes all events, not just the first one, e.g. a death after a non-fatal event gets included in the analysis</p>	<p><b>1. Lack of Familiarity</b> The win ratio is a relatively new statistical method: <i>This article should facilitate a better understanding of the concept and its potential value.</i></p>
<p><b>2. Clinical Priorities Recognized</b> The win ratio forms the component outcomes into a hierarchy based on their relative clinical importance, e.g. death gets top priority.</p>	
<p><b>3. Repeat Events Easily Incorporated</b> The win ratio can be readily extended to account for recurrent events (e.g. hospitalizations) without statistical complexity.</p>	<p><b>2. Statistical Software</b> Calculation of the win ratio (and its CI and p-value) requires statistical programs being readily available: <i>We provide links to such software.</i></p>
<p><b>4. Non-Event Outcomes can be Included</b> The win ratio can be extended to include visit-related items, e.g. quality of life scores and physiological measures.</p>	
<p><b>5. Conceptually Straightforward</b> Counting up the “winners” and “losers” across all pairwise comparisons is a simple concept, compared to explaining what a hazard ratio means.</p>	<p><b>3. Determining Sample Size</b> Power calculations for the win ratio entail simulations: <i>We have created new software to facilitate this task</i></p>

**Take home figure** The win ratio method's key advantages and challenges compared with conventional methods for composite outcomes in randomized controlled trials. CI, confidence interval.

patient did better. For those who bet on horses, a win ratio of 2 means that a patient on the new treatment has odds '2 to 1' of doing better than control. The win ratio is therefore a relative (rather than absolute) measure of treatment effect in that patients with ties do not contribute. Alternative methods<sup>11,30</sup> for a hierarchy of clinical outcomes provide an absolute measure of treatment effect exist (see [Supplementary material online, Appendix](#)). Buyse's method may be included as a secondary analysis of absolute benefit, alongside the primary win ratio analysis.<sup>30</sup> We also support supplementing the win ratio with conventional analyses for each component of the composite outcome, such as time-to-first-event comparisons. A caveat, however, is that such standard analyses may be under-powered given the reduced sample size a primary endpoint based on the win ratio enables. A second limitation (exemplified in the prior section) is that determining sample size for trials using the win ratio is not straightforward: no simple formulae exist, and simulation techniques are required. We anticipate, however, that with increasing uptake of the win ratio as a primary analysis method, standard statistical packages will incorporate programming to simplify these calculations. Our provision of statistical software for sample size calculations in the [Supplementary material online, Appendix](#) is a step forward in this regard.

## Technical issues for the win ratio and alternatives

This article aims to provide practical guidance for users of the win ratio approach. For those wanting to know more technical details, we provide a [Supplementary material online, Statistical Appendix](#). This covers a brief summary of its theoretical basis and discusses some alternative rank-based approaches to pairwise comparisons. In addition, the [Supplementary material online, Statistical Appendix](#) describes

readily available statistical software for calculating the win ratio, its 95% confidence interval and corresponding *P*-value.

## Conclusions

Unlike conventional methods for comparing composite endpoints, the hierarchical win ratio approach accounts for the differing clinical importance of individual endpoint components. The win ratio is also more flexible than conventional methods in that it can incorporate different types of outcomes into its composite endpoint. When designing a trial using the win ratio approach, one must first rank the individual endpoint components by their clinical importance. The statistical power and required number of patients for a trial using the win ratio approach can be determined using simulations. The *Take home figure* documents the key advantages of the win ratio method along with challenges to be overcome.

From real experiences with the win ratio to date, and practical insights on how to select prioritized endpoints and determine trial size in any specific trial setting, we enhance the means whereby the win ratio approach can be more confidently and widely implemented in future randomized trials of CV (and other) interventions.

## Supplementary material

[Supplementary material](#) is available at *European Heart Journal* online.

**Conflict of interest:** J.G.: Consultant—BioSensors Europe SA, Amarin Corporation, Edwards Lifesciences, MvRX, and Boehringer-Ingelheim. G.W.S.: Speaker or other honoraria from Cook, Terumo, QOOL Therapeutics and Orchestra Biomed; Consultant to Valfix, TherOx, Vascular Dynamics, Robocath, HeartFlow, Gore, Ablative Solutions, Miracor, Neovasc, V-Wave, Abiomed, Ancora, MAIA Pharmaceuticals, Vectorious, Reva, Matrizyme, Cardiomech; Equity/



options from Ancora, Qool Therapeutics, Cagent, Applied Therapeutics, Biostar family of funds, SpectraWave, Orchestra Biomed, Aria, Cardiac Success, MedFocus family of funds, Valfix. Other authors: None.

## References

- Montori VM, Permyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, Alonso J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schünemann HJ, Jaeschke R, Guyatt GH. Validity of composite end points in clinical trials. *BMJ* 2005;**330**:594–596.
- Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;**289**:2554–2559.
- Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Permyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007;**334**:786.
- Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. *J Card Fail* 2005;**11**:567–575.
- Lubsen J, Kirwan BA. Combined endpoints: can we use them? *Stat Med* 2002;**21**:2959–2970.
- 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**:176–182.
- Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med* 1999;**18**:1341–1354.
- Finkelstein DM, Schoenfeld DA. Graphing the Win Ratio and its components over time. *Stat Med* 2019;**38**:53–61.
- Wang H, Peng J, Zheng JZ, Wang B, Lu X, Chen C, Tu XM, Feng C. Win ratio—an intuitive and easy-to-interpret composite outcome in medical studies. *Shanghai Arch Psychiatry* 2017;**29**:55–60.
- Oakes D. On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika* 2016;**103**:742–745.
- Dong G, Hoaglin DC, Qiu J, Matsouka RA, Chang Y-W, Wang J, Vandemeulebroecke M. The win ratio: on interpretation and handling of ties. *Stat Biopharm Res* 2020;**12**:99–106.
- Stone GW, Lindenfeld J, Abraham WT, Kar S, Lim DS, Mishell JM, Whisenant B, Grayburn PA, Rinaldi M, Kapadia SR, Rajagopal V, Sarembock IJ, Briek A, Marx SO, Cohen DJ, Weissman NJ, Mack MJ, COAPT Investigators. Transcatheter mitral-valve repair in patients with heart failure. *N Engl J Med* 2018;**379**:2307–2318.
- Dong G, Qiu J, Wang D, Vandemeulebroecke M. The stratified win ratio. *J Biopharm Stat* 2018;**28**:778–796.
- Maurer MS, Schwartz JH, Gundapaneni B, Elliott PM, Merlini G, Waddington-Cruz M, Kristen AV, Grogan M, Witteles R, Damy T, Drachman BM, Shah SJ, Hanna M, Judge DP, Barsdorf AI, Huber P, Patterson TA, Riley S, Schumacher J, Stewart M, Sultan MB, Rapezzi C, ATTR-ACT Study Investigators. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med* 2018;**379**:1007–1016.
- 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**:2740–2755.
- Bartunek J, Terzic A, Davison BA, Filippatos GS, Radovanovic S, Beleslin B, Merkely B, Musialek P, Wojakowski W, Andreaka P, Horvath IG, Katz A, Dolatabadi D, El Nakadi B, Arandjelovic A, Edes I, Seferovic PM, Obradovic S, Vanderheyden M, Jagic N, Petrov I, Atar S, Halabi M, Gelev VL, Shochat MK, Kasprzak JD, Sanz-Ruiz R, Heyndrickx GR, Nyolczas N, Legrand V, Guedes A, Heyse A, Moccetti T, Fernandez-Aviles F, Jimenez-Quevedo P, Bayes-Genis A, Hernandez-Garcia JM, Ribichini F, Gruchala M, Waldman SA, Teerlink JR, Gersh BJ, Povsic TJ, Henry TD, Metra M, Hajjar RJ, Tendera M, Behfar A, Alexandre B, Seron A, Stough WG, Sherman W, Cotter G, Wijns W, CHART Program. Cardiopoietic cell therapy for advanced ischaemic heart failure: results at 39 weeks of the prospective, randomized, double blind, sham-controlled CHART-1 clinical trial. *Eur Heart J* 2017;**38**:648–660.
- Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, Brown DL, Block PC, Guyton RA, Pichard AD, Bavaria JE, Herrmann HC, Douglas PS, Petersen JL, Akin JJ, Anderson WN, Wang D, Pocock S; PARTNER Investigators. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med* 2010;**363**:1597–1607.
- Spitzer E, Van Mieghem NM, Pibarot P, Hahn RT, Kodali S, Maurer MS, Nazif TM, Rodes-Cabau J, Paradis JM, Kappetein AP, Ben-Yehuda O, van Es GA, Kallel F, Anderson WN, Tijssen J, Leon MB. Rationale and design of the Transcatheter Aortic Valve Replacement to UNload the Left ventricle in patients with Advanced heart failure (TAVR UNLOAD) trial. *Am Heart J* 2016;**182**:80–88.
- Rogers JK, Pocock SJ, McMurray JJ, Granger CB, Michelson EL, Ostergren J, Pfeffer MA, Solomon SD, Swedberg K, Yusuf S. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail* 2014;**16**:33–40.
- Abdalla S, Montez-Rath ME, Parfrey PS, Chertow GM. The win ratio approach to analyzing composite outcomes: an application to the EVOLVE trial. *Contemp Clin Trials* 2016;**48**:119–124.
- Milojevic M, Head SJ, Andrinopoulou ER, Serruys PW, Mohr FW, Tijssen JG, Kappetein AP. Hierarchical testing of composite endpoints: applying the win ratio to percutaneous coronary intervention versus coronary artery bypass grafting in the SYNTAX trial. *EuroIntervention* 2017;**13**:106–114.
- Luo X, Tian H, Mohanty S, Tsai WY. An alternative approach to confidence interval estimation for the win ratio statistic. *Biometrics* 2015;**71**:139–145.
- Bakal JA, Roe MT, Ohman EM, Goodman SG, Fox KA, Zheng Y, Westerhout CM, Hochman JS, Lokhnygina Y, Brown EB, Armstrong PW. Applying novel methods to assess clinical outcomes: insights from the TRILOGY ACS trial. *Eur Heart J* 2015;**36**:385–392a.
- Ferreira JP, Jhund PS, Duarte K, Claggett BL, Solomon SD, Pocock S, Petrie MC, Zannad F, McMurray JJV. Use of the win ratio in cardiovascular trials. *JACC Heart Fail* 2020;**8**:441–450.
- Institute of Medicine Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
- Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, Valderas JM, Guillemin F, Revicki D, Alonso J. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. *Heart Fail Rev* 2014;**19**:359–367.
- Spertus JV, Hatfield LA, Cohen DJ, Arnold SV, Ho M, Jones PG, Leon M, Zuckerman B, Spertus JA. Integrating quality of life and survival outcomes cardiovascular clinical trials. *Circ Cardiovasc Qual Outcomes* 2019;**12**:e005420.
- Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. *Pharm Stat* 2016;**15**:238–245.
- Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: application of the joint frailty model. *Stat Med* 2016;**35**:2195–2205.
- Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med* 2010;**29**:3245–3257.