The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities

Stuart J. Pocock*, Cono A. Ariti, Timothy J. Collier, and Duolao Wang

Department of Medical Statistics, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

The conventional reporting of composite endpoints in clinical trials has an inherent limitation in that it emphasizes each patient's first event, which is often the outcome of lesser clinical importance. To overcome this problem, we introduce the concept of the win ratio for reporting composite endpoints. Patients in the new treatment and control groups are formed into matched pairs based on their risk profiles. Consider a primary composite endpoint, e.g. cardiovascular (CV) death and heart failure hospitalization (HF hosp) in heart failure trials. For each matched pair, the new treatment patient is labelled a 'winner' or a 'loser' depending on who had a CV death first. If that is not known, only then they are labelled a 'winner' or 'loser' depending on who had a HF hosp first. Otherwise they are considered tied. The win ratio is the total number of winners divided by the total numbers of losers. A 95% confidence interval and *P*-value for the win ratio are readily obtained. If formation of matched pairs is impractical then an alternative win ratio can be obtained by comparing all possible unmatched pairs. This method is illustrated by re-analyses of the EMPHASIS-HF, PARTNER B, and CHARM trials. The win ratio is a new method for reporting composite endpoints, which is easy to use and gives appropriate priority to the more clinically important event, e.g. mortality. We encourage its use in future trial reports.

Keywords   Heart failure • Clinical trials • Composite endpoints • Statistical analysis • Trial reporting

Introduction

Major cardiovascular (CV) trials often use a primary composite endpoint for comparing the randomized treatment's efficacy; that is, it includes two or more types of related clinical events (e.g. CV death, myocardial infarction, stroke) and analysis focuses on the time to the first event, whichever that is. A Cox model, log-rank test, and Kaplan–Meier plot are used to produce a hazard ratio, confidence interval (CI), *P*-value, and graphical comparison of the time to first event.

This commonly used approach to analysis of composite endpoints has marked limitations. It handles all contributory endpoints as if of equal importance, and only notices the first endpoint. For instance, after a patient's non-fatal myocardial infarction, whether they subsequently died is ignored. Thus, non-fatal events occurring early get higher priority than later more serious events and deaths. Furthermore, non-fatal events can occur more than once. Nevertheless, composite endpoints and this conventional approach to their analysis are used in CV trials, despite critical commentaries.[1–6]

Here, we introduce a new approach to analysing composite endpoints that accounts for clinical priorities, i.e. CV deaths are considered more important than non-fatal events and get first priority. The method also recognizes that patients have differing risk profiles, by using risk-matched pairs. This is easy to use, and provides an informative estimate of treatment difference with CI and *P*-value. We apply the approach to several recent trials and compare with existing methods.

Methods

Consider a clinical trial comparing new vs. standard treatment with a composite primary endpoint, e.g. CV death and hospitalization for chronic heart failure (HF hosp), as commonly used in heart failure trials.

One recognizes that CV death is more important than HF hosp. Hence in comparing any two patients on new and standard treatment, one first determines whether either one had a CV death before the other. If that is not known, then one determines which patient had a

----

HF hosp first. This is the essence of the new approach to analysing composite endpoints.

In comparing two such patients, it seems appropriate to consider their underlying risks of the composite endpoint. It would be an unfair comparison if one patient is high risk (e.g. age 75, ejection fraction 25%, diabetic) and the other is lower risk (e.g. age 55, ejection fraction 40%, non-diabetic). Hence, we use a risk score or risk stratification to select matched pairs of patients on new and standard treatment. This preamble leads to the following method.

The matched pairs approach

There are three steps to analysis.

(1) One forms matched pairs of patients on new and standard treatment. The method of matching (see Results below) takes into account individual patient risk. There will usually be slightly unequal patient numbers in the two groups, leaving a small number of patients unmatched in the larger group.

(2) For each matched pair, one studies the more major event (CV death): is it known which patient had a shorter time from randomization to CV death? If both or neither had a CV death, this is straightforward. If one patient had a CV death, the other must be followed for longer in order to know definitely who had CV death first.

If it is not known who had CV death first, one then checks if it is known which patient had HF hosp first, using the same principles. So each matched pair fits into one of five categories:

(a) new patient had CV death first;
(b) standard patient had CV death first;
(c) new patient had HF hosp first;
(d) standard patient had HF hosp first;
(e) none of the above.

Note that categories (a) and (b) take priority over (c) and (d): that is, any pair is only classified as HF hosp if it is not known who had CV death first. Category (e) mostly comprises pairs with neither patient having CV death nor HF hosp, but will include a few pairs where one had an event but the others' follow-up time was shorter. *Figure 1* illustrates the various scenarios classifying any pair into categories (a)–(e).

(3) The trial's composite endpoint results are summarized by N_a, N_b, N_c, N_d, and N_e, the numbers of matched pairs in categories (a), (b), (c), (d), and (e), respectively. $N_b + N_d = N_w$ is the number of 'winners' for the new treatment, i.e. those matched pairs where standard treatment fared worse. Similarly, $N_a + N_c = N_L$ is the number of 'losers' for the new treatment. We call $R_w = N_w/N_L$ the 'win ratio'.
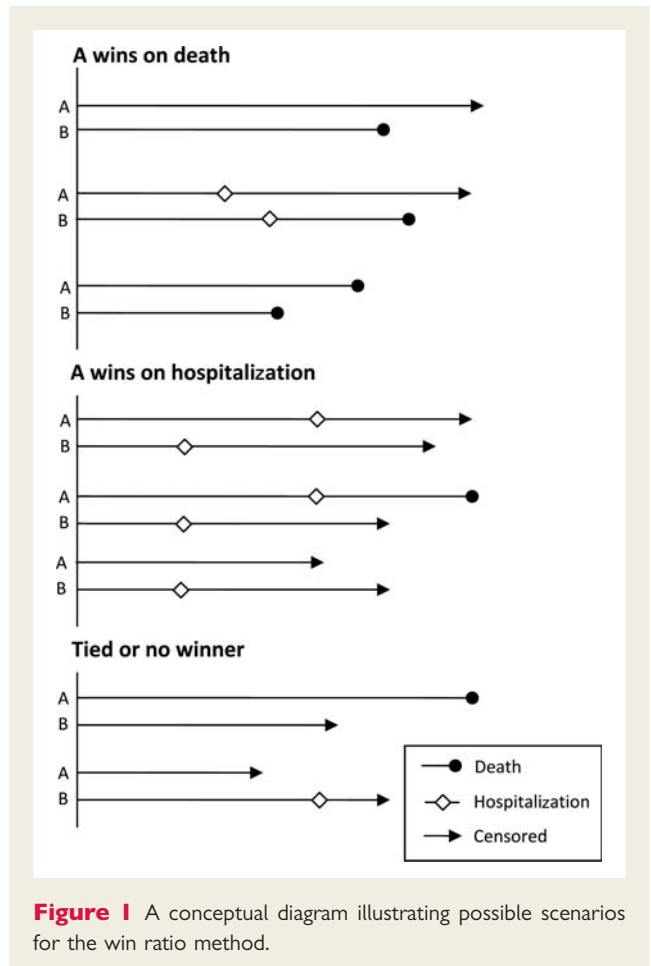
$$\frac{N_w}{N_w + N_L} = p_w$$

is a proportion, its 95% CI is:

$$p_w \pm 1.96 \left[\frac{p_w(1 - p_w)}{(N_w + N_L)} \right]^{1/2} = p_L, p_U$$

$R_w = p_w/(1 - p_w)$, so the 95% CI for the win ratio is $p_L/(1 - p_L)$, $p_U/(1 - p_U)$.

For a significance test, $z = (p_w - 0.5)/[p_w(1 - p_w)/(N_w + N_L)]^{\frac{1}{2}}$ is a standardized normal deviate under the null hypothesis, readily yielding the required P-value.

Figure 1 A conceptual diagram illustrating possible scenarios for the win ratio method.

One can also obtain a win ratio just for CV death, $R_D = N_b/N_a$ with corresponding 95% CI and P-value.

N_e the number of matched pairs in which neither patient 'won' mostly comprises pairs in which both patients were event free. Thence, a useful supplementary statistic is the proportion of pairs which were tied in this sense, $p_T = N_e/N$ where $N = N_a + N_b + N_c + N_d + N_e$. Its 95% CI uses the same formula as for p_w above replacing p_w by p_T and $N_w + N_L$ by N.

The above method can be readily extended to a composite with three or more components, provided they can be sensibly ranked in order of clinical importance.

For instance, in a hypertension trial with composite primary outcome CV death, stroke, and myocardial infarction, one could untie each pair first on CV death, then stroke, and then myocardial infarction, since strokes are typically more debilitating than myocardial infarctions.

The unmatched approach

There will be trials with no recognized strategy to choose matched pairs. Then the following unmatched approach, first described by Finkelstein and Schoenfeld[7] with an emphasis on significance testing, can be undertaken instead. One compares every patient on new treatment with every patient on standard treatment, each time noting who 'won'.

Let N_n and N_s be the number of patients on new and standard treatments. Then one makes all $N_n \times N_s$ paired comparisons. Again, each pair is classified into one of categories (a), (b), (c), (d), or (e). Based

on CV death and HF hospitalization, respectively, categories (b) and (d) 'are winners' for the new treatment while (a) and (b) are 'losers'.

N_a, N_b, N_c, N_d, and N_e are as before except now $N_a + N_b + N_c + N_d + N_e = N_n \times N_s$.

Again, $N_b + N_d = N_w$ and $N_a + N_c = N_L$ are the numbers of 'winners' and 'losers' for the new treatment and $R_w = N_w/N_L$ is the 'win ratio'.

In this unmatched approach, calculating the 95% CI and *P*-value for R_w is quite complex (see the Supplementary material online, Statistical Appendix). Finkelstein and Schoenfeld[7] describe a general framework for the significance test. The problem is that the $N_n \times N_s$ unmatched pairs are not independent comparisons, since each patient on new and standard treatments is used N_s and N_w times, respectively. Statistical software to perform these calculations is available from duolao.wang@lshtm.ac.uk.

Overall, we recommend the matched pairs approach, provided a pre-defined basis for matching exists. Calculations are easier to perform, and the matching means that each patient comparison is more fair and informative since patients are of comparable risk. Although the unmatched approach is still unbiased, the existence of 'unfair' comparisons, both for and against the new treatment, will dilute the win ratio R_w nearer to 1. The analogy is unadjusted vs. covariate-adjusted analyses for hazard ratios: the latter typically gives hazard ratio slightly further from 1 and smaller *P*-value.[8,9]

Results

We now apply the above methods to three recently published trials.

(1) **EMPHASIS-HF** compared eplerenone vs. placebo in 2737 patients with NYHA class II heart failure and ejection fraction ≤35% followed for a median of 21 months.[10] The composite primary outcome CV death or hospitalization for heart failure (HF hosp), occurred in 18.3% and 25.9% of eplerenone and placebo patients, respectively; hazard ratio 0.63 (95% CI 0.54, 0.74) *P* < 0.0001. But the hospitalizations tend to happen first and hence any impact of eplerenone on CV mortality gets lost in the composite.

We apply the new approach in three ways.

Matched pairs

The above hazard ratio was adjusted for nine baseline variables: (age, GFR, ejection fraction, body mass index, haemoglobin, heart rate, systolic blood pressure, diabetes, atrial fibrillation, and left bundle branch block or QRS duration > 130 ms), pre-selected because of their known associations with prognosis. From the Cox model's coefficients, one obtains a risk score for each patient in the trial. Although randomized treatment is in the model, its coefficient is not used in the risk score.

There were 1364 patients on eplerenone and 1373 on placebo. The first step is to get equal-sized groups by randomly removing nine patients from the placebo group. Then one risk-matches each eplerenone patient with each placebo patient using their risk scores. Specifically, patients on each treatment are ranked by their risk scores. Going from top-rank to bottom-rank, each eplerenone patients is paired with the same-ranked placebo patient.

Applying the matched pairs method, each pair fits into one of the five categories (a), (b), (c), (d), or (e); see left column of

Table I The EMPHASIS-HF trial: three alternative new approaches comparing eplerenone vs. placebo for the composite CV death and HF hospitalization.

	Matched pairs	Matched pairs time stratified	All unmatched pairs
(a) CV death on eplerenone first	90	105	124 825
(b) CV death on placebo first	118	148	163 129
(c) HF hosp on eplerenone first	61	61	86 127
(d) HF hosp on placebo first	131	137	175 606
(e) None of the above	964	913	1 323 085
Total no. of pairs	1364	1364	1 872 772
Win ratio for composite	1.65	1.72	1.61
95% CI	1.35, 2.03	1.42, 2.09	1.37, 1.89
z-score	5.05	5.81	5.45
Win ratio for CV death only	1.31	1.41	1.31
95% CI	1.00, 1.74	1.10, 1.82	1.04, 1.66
z-score	1.96	2.74	2.25

Table 1. For 208 pairs, we know which patient had CV death first: 90 eplerenone, 118 placebo. So the win ratio for CV death is 118/90 = 1.31, 95% CI 1.00, 1.74 (*P* = 0.05).

Among the remaining 1156 matched pairs, for 192 we know which patient had HF hosp first: 61 eplerenone and 131 placebo. Hence, the win ratio for the composite of CV death and HF hosp is (118 + 131)/(90 + 61) = 1.65 with 95% CI 1.35, 2.05 (*P* < 0.0001).

Note 964 (70.7%) of matched pairs were tied for the composite outcome. In 814 of these tied pairs neither patient had a primary event. This reflects the relatively good prognosis of patients in EMPHASIS-HF.

Matched pairs, time stratified

One problem with the above is that recruitment lasted over 4 years, so that patient follow-up times vary enormously. So in 150 pairs, one patient had CV death and/or HF hosp but beyond the follow-up time of the other patient, meaning that the pair remained tied, leaving 68 'unused' CV deaths and 82 'unused' HF hosps. To reduce this problem, patients can be stratified into several intervals of randomization dates, each interval containing similar patient numbers; risk-matching is then done within each interval. Choosing five such intervals halved the problem, down to 33 'unused' CV deaths and 44 'unused' HF hosps (see middle column of *Table 1*).

The win ratios for the composite and for CV death alone have now increased to 1.72 and 1.41, respectively, both with smaller *P*-values as indicated by the greater z-scores.

Unmatched pairs

A third approach compares every eplerenone patient with every placebo patient: 1364 × 1373 = 1872772 unmatched pairs classified into outcome categories (a)−(e); see right column of

Table 2 The PARTNER trial: unmatched analysis for the composite, all-cause death, or hospitalization due to valve or procedure-related deterioration (Hosp)

(a) Death on TAVI first	8498
(b) Death on standard first	14 466
(c) Hosp on TAVI first	1345
(d) Hosp on standard first	3979
(e) None of the above	3753
Total no. of pairs	$179 \times 179 = 32041$
Win ratio for composite	1.87
95% CI	1.35, 2.54
z-score	4.58
Win ratio for death only	1.70
95% CI	1.23, 2.38
z-score	3.37

Table 3 Conventional analysis of CHARM trials' primary composite endpoint using hazard ratios

	CHARM Added		CHARM Alternative		CHARM Preserved	
Unadjusted HR	0.85		0.77		0.89	
95% CI	0.75−0.96		0.67−0.89		0.77−1.03	
P-value	0.011		0.0004		0.118	
Adjusted HR	0.85		0.70		0.86	
95% CI	0.75−0.96		0.60−0.81		0.77−1.00	
P-value	0.010		<0.0001		0.051	
	C	Pl	C	Pl	C	Pl
No. of patients	1276	1272	1013	1015	1514	1509
No. with primary composite event	483	538	334	406	333	366
No. of these which were CV death[a]	174	182	127	120	92	90
Total no. with CV death[a]	302	347	219	252	170	170

C, candesartan; Pl, placebo.
[a]For each column, the difference in these two numbers is the number of CV deaths occurring after HF hosp.

Table 1. The consequent composite win ratio is 1.61 smaller than for either matched pair analysis. This is expected since many unmatched pairs will be 'unfair' (in both directions), comparing patients with different underlying risks. However, statistical significance (z-score) of the unmatched analysis is not dissimilar.

(2) **PARTNER B** trial compared transcatheter aortic-valve implant (TAVI) with standard treatment for aortic stenosis patients who cannot undergo surgery[11]: 179 patients in each group, median follow-up 1.6 years. A composite co-primary endpoint was all-cause death or hospitalization due to valve- or procedure-related deterioration (Hosp). The conventional Cox model gave hazard ratio 0.46 (95% CI 0.35, 0.59) $P < 0.0001$. However, the pre-defined primary analysis was our unmatched pairs approach as in Finkelstein and Schoenfeld[7]: for detailed results (not previously published) see *Table 2*.

For this composite, the win ratio is 1.87 (95% CI 1.35, 2.54), $P < 0.0001$, and for death only the win ratio is 1.70 (95% CI 1.23, 2.38) $P < 0.0001$, very strong evidence for TAVI patients living longer, and in addition avoiding hospitalizations. Note that only 3753 (11.7%) pairs were tied for the composite, of which 2752 pairs had neither patient experiencing Death nor Hosp.

PARTNER B had no matched pairs analysis, since baseline risk factors were not well established. *Post hoc* Cox analyses within PARTNER B did not reveal any especially strong predictors. Under such circumstances, the unmatched approach is preferred.

(3) The **CHARM** program compared candesartan with placebo in chronic heart failure,[12] the three patient types being, the following:

CHARM Added: ejection fraction <40% and on ACE-inhibitor;
CHARM Alternative: ejection fraction <40% and intolerant to ACE-inhibitor;
CHARM Preserved: ejection fraction ≥40%.

For the composite primary endpoint, CV death, or HF hosp, the published hazard ratios (HR) with and without covariate adjustment are in *Table 3*. The CHARM Added and CHARM Alternative treatment differences are highly significant, but evidence is weaker for CHARM Preserved. Covariate adjustment enhanced treatment effects and their significance, but CHARM Preserved still had a tantalizing $P = 0.051$.

Table 3 also lists the numbers of CV deaths and HF hosps in each trial by treatment groups, and whether they were the first occurring primary event. Overall, only 54% of CV deaths contributed to the primary composite endpoint: the rest were ignored because they occurred after a patient's HF hosp. Only 32% of first-occurring primary endpoints were CV deaths.

Now we re-analyse CHARM using the win ratio concept. There were 32 pre-defined baseline covariates, and a multivariable Cox model for the composite outcome (stratified by trial and treatment group) provides a risk score. Then within each trial, one obtains risk-matched pairs of candesartan and placebo patients (having first achieved equal-size treatment group by randomly removing a few patients from the larger group).

Table 4 shows the consequent win ratio results. All three trials' win ratios for both the composite and CV death alone are greater than 1. For CHARM Preserved, the composite's win ratio was smaller and less significant, $P = 0.065$. CHARM Preserved had 64% matched pairs tied, compared with 41% and 47% tied in CHARM Added and CHARM Alternative, respectively, indicating the better prognosis in CHARM Preserved.

In CHARM Added, the win ratio (*Table 4*) was more highly significant than the conventional hazard ratio (*Table 3*), because the CV deaths get a higher priority, and there is clear evidence of fewer CV deaths on candesartan.

Table 4 The CHARM Program: results using a matched pairs approach to the win ratio

	CHARM Added	CHARM Alternative	CHARM Preserved
CV death on candesartan first	220	148	136
CV death on placebo first	289	202	150
HF hosp on candesartan first	104	74	115
HF hosp on placebo first	132	114	144
None of the above	527	475	964
Total no. of pairs	1272	1013	1509
Win ratio for composite	1.30	1.42	1.17
95% CI	1.13, 1.50	1.20, 1.70	0.99, 1.39
P-value	<0.0001	<0.0001	0.065
Win ratio for CV death only	1.31	1.37	1.10
95% CI	1.10, 1.57	1.11, 1.70	0.88, 1.39
P-value	0.002	0.003	0.40

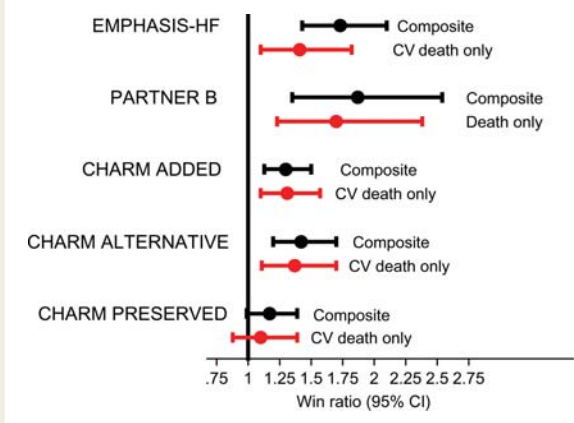**Figure 2** A graphical display of win ratio estimates (95% CI) for the three trials.

Overall, the win ratio approach made greater use of CV death data than the conventional approach. The three win ratios included 1145 CV deaths, 504 on candesartan 641 on placebo, yielding a CV death win ratio of 1.27 ($P < 0.001$) for the whole CHARM program.

We also calculated win ratios in each CHARM trial using the unmatched approach (results not shown). The unmatched win ratios were all smaller and less significant than the matched win ratios, since there are many strong risk factors in heart failure (e.g. age, ejection fraction, diabetes, NYHA class, etc.) which risk-matching utilizes.

In CHARM repeat hospitalizations can be incorporated into the win ratio method, using investigator-assessed HF hosps, since the clinical events committee only adjudicated the first HF hosp. Hence, we now untie each pair on who had fewer HF hosps over their shared follow-up time. For instance, in CHARM Preserved the number of winners vs. losers on candesartan now becomes 150 vs. 106 (rather than the 144 vs. 115 shown in *Table 4* for first hospitalization time only). Combining this with CV death winners and losers (150 vs. 136) gives a win ratio 1.23, CI 1.05 to 1.48 $P = 0.012$. One could instead untie on the number of days in hospital due to HF, which gives a similarly significant result for CHARM Preserved.

Finally, *Figure 2* illustrates how the win ratio estimates in *Tables 1, 2,* and *4* can be presented graphically. In all trials, one can see that the CI is wider for the mortality win ratio compared with the composite' win ratio. This is because the non-fatal events increase the numbers of patients with a primary endpoint, and hence increase the numbers of winners and losers.

Whether adding in the non-fatal events increases or decreases the estimated win ratio depends on whether treatment has a greater effect on morbidity or mortality. In these examples, win ratios were somewhat greater for the composite, except for CHARM Added.

Discussion

The widespread use of composite primary endpoints in randomized trials provokes controversy regarding their suitability and difficulties in interpreting results.[1–6] The conventional practice of analysing the time to first event has limitations. Often first events are less clinically important, e.g. hospitalizations or other non-fatal events occur before death, and these lesser events dominate the trials' primary findings more than seems clinically appropriate.

Our new approach removes this dilemma by putting emphasis on the more clinically important component of the composite, e.g. CV deaths get greater priority than HF hospitalizations in heart failure trials. This is achieved by comparing pairs of patients on new and control treatment to determine which of the pair had CV death first, and if that is not known, only then which of the pair had HF hospitalization first. Our main approach uses matched pairs of patients, matching being on the risk of primary events (as in CHARM) and if necessary (as in EMPHASIS-HF, which stopped early) matching can also be stratified by follow-up time.

Each pair is 'untied' first on the basis of the most important event (e.g. CV death) and secondly (if necessary) on the lesser event (e.g. HF hospitalization). The numbers of pairs in which the patient on new treatment 'won' and 'lost' are compared to produce the 'win ratio'. The 95% CI and *P*-value for the win ratio are readily obtained.

This approach has several advantages over existing methods:

(1) It prioritizes the more major component of the composite (e.g. CV death).
(2) More of those major-component events are included in the analysis.
(3) The win ratio concept is easy to understand, easier than explaining what a hazard ratio means.
(4) Calculations are easily performed, not requiring special software.

Any slightly unequal treatment groups means a few patients are not included in matched pairs. If this is considered a problem, one

could perform many repeat analyses with different randomly removed patients and take the median win ratio etc. of these repeats as the declared estimate.

For the matched pairs win ratio to have credibility in any future trial, the method of matching (and development of any risk score, and time stratification if required) needs to be rigorously pre-defined in a Statistical Analysis Plan. Such statistical robustness is essential in ensuring that win ratio estimates in a trial report, especially in a regulatory context, are precisely defined and reproducible.

For some trials, risk matching can be based on a pre-existing risk score obtained from an earlier study, not necessarily on the same composite outcome. Otherwise, such a risk score can be obtained from the trial data themselves with appropriate modelling of pre-defined predictors of the composite outcome. The essential point is to try and ensure that each matched pair of patients have a similar prognosis, thereby making all paired comparisons intrinsically fair.

The win ratio estimate could be inverted so that values below 1 indicate evidence of treatment superiority, as is the case with hazard ratios, e.g. a win ratio of 1.5 becomes a 'lose ratio' of 0.667. However, we feel that 'win ratio' conveys better what the method is about.

One could give more weight to pairs untied by death rather than a non-fatal event. The consequent statistical calculations are not difficult, but we find such artificial weightings an unappealing complication, given that the win ratio method already gives priority to deaths.

The win ratio method can be adapted to utilize repeat events, e.g. hospitalizations, as illustrated in CHARM. The principle is unaltered: one simply unties any pair of patients on the number of hospitalizations (or number of days in hospital) rather than the first hospitalization. Indeed, other aspects of follow-up, e.g. changes in NYHA class, could be incorporated.

When risk-matching cannot be pre-defined (as in the pioneering PARTNER B trial), then an unmatched approach is adopted: every new treatment patient is compared with every standard treatment patient. The win ratio is readily obtained, but calculating the 95% CI and *P*-value is more complex (see Supplementary material online, Statistical Appendix).

In general, risk-matching is preferable, giving a more patient-specific interpretation: suppose you are on the new treatment and there is another patient with similar prognosis on standard treatment. A win ratio of 1.5 means that there are 50% more wins on the new treatment. Provided there exist strong predictors of prognosis, the risk-matched approach will have greater statistical power.

Any new statistical approach to clinical trials needs evaluating where it fits in with the standard repertoire. We propose that clinicians, statisticians, sponsors, regulators, and journals familiarize themselves with the win ratio concept and see how it works in other specific clinical trials. As in the examples presented here, it may begin as a supplementary analysis alongside existing methods, although in PARTNER B, the unmatched approach was the primary pre-defined hypothesis for the composite.[11]

As experience grows, we would hope win ratio methods become a common primary analysis for future trials with composite endpoints. This would require appropriate statistical power calculations that are not complex in the matched pairs case since the win ratio is based on a binary proportion: methods of sample size determination are currently being developed and will be published shortly.

The extension to composite endpoints with three or more components is straightforward: the main issue is to have a sensible ranking of the components. For instance with CV death, stroke, and myocardial infarction as a primary composite endpoint, would there be agreement to rank stroke as a more clinically important event than myocardial infarction, and less important than death? That may depend on the definitions used.

There are several limitations to the win ratio method. First, the lack of familiarity among trialists and, as for any new method, that inevitably takes time to overcome. Secondly, it does not actually use the precise times from randomization to event occurrence, only whether the event occurred sooner in one patient compared with their matched partner. This would matter if one lost statistical power compared with conventional methods, but from our examples that is not evident. The 'magnitude of win' in terms of event-free time gained is an interesting concept worth considering in future research. We have also considered alternative life table methods with hierarchical ordering of events and time to worst event as outcome, but feel they would lead to greater complexity and potential logical inconsistency. Further technical properties of win ratio methods need to be explored in future articles in statistical journals, although the methods and results presented here are enough for researchers to apply in their own trials.

In conclusion, we need to rethink the way that composite endpoints are utilized in clinical trials. The win ratio method introduced here is conceptually clear, statistically straightforward, easy to use, and clinically relevant, giving appropriate priority to more clinically important component(s) of any composite, e.g. mortality. The win ratio performed well in the trials presented here. We encourage others to analyse results of their own trials using this approach whether studies are in planning, in progress, or already completed.

Supplementary material

Supplementary material is available at *European Heart Journal* online.

Pfizer Pharmaceuticals. D.W. and C.A.A. report no conflict of interest.

References

1. Montori VN, Permanyer-Miralda G, Ferreira-Gonzalez I, Busse JW, Pacheco-Huergo V, Bryant D, Alonso J, Akl EA, Domingo-Salvany A, Mills E, Wu P, Schunemann HJ, Jaeschke R, Guyatt GH. Validity of composite end points in clinical trials. *BMJ* 2005;**330**:594–596.
2. 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.
3. Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coelho P, Walter SD, Guyatt GH. Methodologic discussions for using and interpreting endpoints are limited, but still identify major concerns. *J Clin Epidemiol* 2007;**60**:651–657.
4. Ferreira-González I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonson J, Worster A, Upadhye S, Jaeschke R, Schunemann HJ, Permanyer-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**:786–792.
5. 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.
6. Lubsen J, Kirwan BA. Combined endpoints: can we use them?. *Stat Med* 2002;**21**:2959–2970.
7. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med* 1999;**18**:1341–1354.
8. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparison in clinical trial reporting: current practice and problems. *Stat Med* 2002;**21**:2917–2930.
9. Ford I, Norrie J. The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Stat Med* 2002;**21**:2899–2908.
10. Zannad F, McMurray JJV, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pocock SJ, Pitt B for the EMPHASIS-HF Study Group. Eplerenone in patients with systolic heart failure and mild symptoms. *New Engl J Med* 2011;**364**:11–21.
11. Leon MB, Smith CR, Mack M, Miller DC, Mosses JW, Svensson LG, Tuzcu M, 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 SJ for the PARTNER Trial Investigators. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med* 2010;**363**:1597–1607.
12. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, Michelson EL, Olofsson B, Ostergren J, Yusuf S for the CHARM Investigators and Committees. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 2003;**362**:759–66.