



HAL
open science

Use of the Win Ratio in Cardiovascular Trials

Joao Pedro Ferreira, Pardeep Jhund, Kevin Duarte, Brian Claggett, Scott Solomon, Stuart Pocock, Mark Petrie, Faiez Zannad, John J.V. McMurray

► **To cite this version:**

Joao Pedro Ferreira, Pardeep Jhund, Kevin Duarte, Brian Claggett, Scott Solomon, et al.. Use of the Win Ratio in Cardiovascular Trials. *JACC: Heart Failure*, 2020, 8 (6), pp.441-450. 10.1016/j.jchf.2020.02.010 . hal-02863669

HAL Id: hal-02863669

<https://hal.univ-lorraine.fr/hal-02863669>

Submitted on 10 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title: Use of the Win Ratio in Cardiovascular Trials

Short title: Win Ratio vs Hazard Ratio

Authors: João Pedro Ferreira MD PhD^{a,b}; Pardeep Jhund MB BCh PhD^a; Kévin Duarte PhD^b; Brian Claggett PhD^c; Scott Solomon MD^c; Stuart Pocock PhD^d; Mark Petrie MB ChB^a; Faiez Zannad MD PhD^{bc}; John J.V. McMurray MD^a.

Affiliations: ^aBHF Cardiovascular Research Centre, University of Glasgow, Glasgow, UK; ^bNational Institute of Health and Medical Research (INSERM), Center for Clinical Multidisciplinary Research 1433, INSERM U1116, University of Lorraine, Regional University Hospital of Nancy, French Clinical Research Infrastructure Network (F-CRIN) Investigation Network Initiative-Cardiovascular and Renal Clinical Trialists (INI-CRCT), Nancy, France; ^cCardiovascular Division, Brigham and Women's Hospital Boston, MA, USA; ^dLondon School of Hygiene and Tropical Medicine, London, UK.

Corresponding author: Professor John J.V. McMurray,
British Heart Foundation Cardiovascular Research Centre,
University of Glasgow,
126 University Place,
Glasgow, G12 8TA,
United Kingdom.
Tel: +44 141 330 3479 **Fax:** +44 141 330 6955
Email: john.mcmurray@glasgow.ac.uk

Word count (excluding references): 3469

ABSTRACT

Background: The primary outcome in many cardiovascular trials is a composite including non-fatal and fatal events. The time-to-first event analysis gives equal statistical weighting to each component event. The win ratio (WR), which takes into account the clinical importance and timing of the outcomes, has been suggested as an alternative approach.

Objectives: Compare the WR with the corresponding hazard ratios (HR) and 1/HR.

Methods: Cox PH models and WR.

Results: In the studied trials (n=16) the WR and HR differed only slightly. For example, in PARADIGM-HF (sacubitril/valsartan vs. enalapril), the primary-outcome of time-to-first HFH or CVD using Cox model gave a 1/HR=1.25 (1.12-1.41; z-score=4.8). Using WR for testing this composite in the hierarchical order of CVD and HFH gave a WR=1.27 (1.15-1.39; z-score=4.7), reflecting a similar effect of sacubitril/valsartan on CVD and HFH. In DIG (digoxin vs. placebo), the outcome of time-to-first HFH or CVD using Cox gave a 1/HR=1.18 (1.10-1.27; z-score=4.5). Using WR for testing this composite in the hierarchical order of CVD and HFH gave a WR=1.14 (1.05-1.20; z-score=3.1), reflecting a larger effect of digoxin on HFH than CVD. Several other trials and end-points including patient-reported measures were studied.

Conclusion: In 16 large cardiovascular outcome trials, HR and WR provided similar estimates of treatment effects. The WR allows the prioritization of fatal outcomes and the hierarchical testing of broader composite endpoints including patient-reported outcomes. In this way, the WR allows the incorporation of patient-centred and other outcomes while prioritizing the competing risk of death and hospital admission.

Key-words: hazard ratio; win ratio; comparison; heart failure; trials.

CONDENSED ABSTRACT

In large cardiovascular outcome trials Cox proportional and win ratio analysis provided similar estimates of treatment effects. The win ratio also allows hierarchical testing of broader composite endpoints including patient-reported outcomes and other less clinically important components such as biomarkers. In this way, the win ratio approach allows incorporation of patient-centred and other outcomes while prioritizing the competing risk of death and hospital admission.

List of abbreviations

HF, heart failure

HFH, heart failure hospitalization

CVD, cardiovascular death

MI, myocardial infarction

WR, win ratio

HR, hazard ratio

KCCQ-OSS, Kansas City Cardiomyopathy Questionnaire Overall Summary Score

NT-pro BNP, N-terminal pro brain natriuretic peptide

INTRODUCTION

The primary outcome in major cardiovascular outcome trials is frequently a composite that includes one or more non-fatal cardiovascular events and a fatal outcome, often cardiovascular death (CVD). By convention, evaluation of the treatment under investigation usually involves analysis of the components of this end-point by time-to-first event. Cox proportional-hazards models are commonly used to produce a hazard ratio (HR), the respective confidence interval and p-value. This approach has several advantages, including simplicity and reproducibility, and there is a large experience, over many years, of its use. However, this conventional approach also has some drawbacks.¹ An important problem is that the components of the end-point are given equal statistical weighting, even though they may not have equal clinical importance, e.g. non-fatal myocardial infarction (MI) is not usually considered as important as death. This is especially important where the effect of therapy differs for different component events and the less clinically important events occur earlier. For example, a cholesterol-lowering therapy might reduce the risk of a composite atherothrombotic outcome, overall, but the benefit may be driven completely by reduction in non-fatal events, with no reduction in death. Conversely, in a heart failure (HF) trial with the primary end-point of heart failure hospitalization (HFH) or death, an implantable cardioverter-defibrillator would have little or no effect on the primary end-point because it exclusively reduces death. An even more extreme hypothetical scenario is where a treatment has opposite effects on the components of a composite, so, for example, by preventing non-fatal events, a treatment might reduce the risk of a primary composite outcome yet increase mortality.

To overcome some of these limitations, Pocock and colleagues proposed a win ratio (WR) approach to the analysis of trials taking account of the clinical importance of the components of composite outcomes, as well as the relative timing of the component events by considering the most important component first.² The WR is an adaptation of a non-parametric test proposed by Schoenfeld and

Finkelstein that allows the combination of time-to-event and longitudinal measures.³ The time-to-first event Cox often considers potentially less serious events first, e.g. hospitalizations or other non-fatal events, ignoring the fatal events that occurred after; the non-fatal events often dominate trials' primary findings. The WR approach removes this dilemma by computing the more serious fatal events first, e.g. CVD gets greater priority than HFH. This is achieved by comparing pairs of patients on new and control treatment to determine which of the pair had death first, and if that is not known, only then which of the pair had hospitalization first.² For any random pair of patients who are not a tie, the odds that the active patient did better than the control patient is the WR. For example, in the HF trial mentioned above, patients in the treatment and the control groups are paired, and the first step is to determine whether one patient died before the other. Where CVD did not occur, the next step is to determine which patient had a HFH first. If a "winner" cannot be determined at either of these steps, then the two patients are considered "tied".

This alternative approach to analysis of composite outcomes has not yet been widely reported in cardiovascular clinical trials. To provide clinicians with examples of how the WR compares with conventional analyses, we have reanalysed several influential trials using both conventional Cox proportional-hazards models and the WR, considering the potential advantages and disadvantages to each approach.² In addition, we examined the use of the WR to look at broader composite outcomes, including additional measures such as quality of life and a biomarker in combination with hospitalization in death.

METHODS

Studies and outcomes

The WR (see below) and Cox proportional-hazards regression models were compared in 16 published cardiovascular trials available to us. A brief description of each of these trials and their outcomes is provided in the *Supplementary Appendix*. Five trials which are particularly illustrative of some of the important issues that might influence the choice between these analytical methods are reported in more detail in the *Supplementary Appendix* and the *Results* below.

Institutional ethics approval was obtained in each centre of the studied trials.

Statistical approaches

The statistical methods are described in detail in the *Supplementary Appendix*. The WR compares the “wins” vs. “losses” in the intervention group; therefore, the benefit of a given treatment (i.e. a treatment with significantly more wins than losses) is provided by a number that is greater than 1. The time-to-first event Cox model compares the survival times of patients in the intervention vs. the placebo (control) groups producing a HR; the benefit of a given treatment (i.e. a treatment with significantly lower event rate) is provided by a number that is smaller than 1. Hence, to facilitate the comparison of the estimates provided by WR with those of the HR, we also report the HR as 1/HR and respective 95% confidence interval. This represents the rate of events in the control group vs. the rate in the intervention group, whereby a $1/HR > 1$ indicates a higher event rate in the control group compared with the intervention group i.e. a benefit with the intervention. For example, a 1/HR of 1.25 indicates a higher event rate in the control group, and is the same as a HR of 0.80 indicating a lower event rate in the intervention group. For both the WR and Cox we provided the respective p-values and z-scores, the latter to ease the interpretation of small p-values. To facilitate the comprehension of these methods an illustration of the time-to-first and WR is provided in the *Figure 1*.

RESULTS

Analysis of two-component end-points

PARADIGM-HF: The 1/HR for the effect of sacubitril/valsartan, compared with enalapril, on the primary composite outcome of HFH or CVD, examined as time-to-first event, was 1.25 (1.12, 1.41), $z=4.8$, $p=0.00000048$ (HR =0.80 [0.71, 0.89]); where 66% and 67% of CVD contributed to the primary outcome in the enalapril and sacubitril/valsartan, respectively. The WR, ordered as CVD followed by HFH, was 1.27 (1.15, 1.39), $z=4.7$, $p=0.00000062$ (*Table 1*).

DIG: A *post hoc* analysis of the effect of digoxin on the same composite outcome of HFH or CVD, in a time-to-first event analysis, gave a 1/HR of 1.18 (1.10, 1.27), $z=4.5$, $p=0.0000034$ (HR =0.85 [0.79, 0.91]); driven by an effect on the non-fatal component, with no effect on mortality; where 47% and 58% of CVD contributed to the outcome in the placebo and digoxin groups, respectively. The corresponding WR, tested as CVD followed by HFH, was 1.14 (1.05, 1.20), $z=3.1$, $p=0.00096$ (*Table 1*); the WR had a smaller z -score/greater p -value reflecting the “neutral” effect of digoxin on CVD.

Analyses of multi-component end-points

SPRINT: In SPRINT, the effect of more-intensive blood pressure, compared with less-intensive blood pressure lowering, on the primary composite outcome of MI, stroke, HFH or CVD, analysed as time-to-first event, gave a 1/HR of 1.33 (1.12, 1.56), $z=3.1$, $p=0.00095$ (HR =0.75 [0.64, 0.89]). The 1/HR for the individual components varied considerably, ranging from 1.75 (1.18, 2.63) (HR =0.57 [0.38, 0.85]) for the CVD to 1.12 (0.80, 1.59) (HR =0.89 [0.63, 1.25]) for stroke (*Table 2*). The WR for the primary composite outcome tested in the order CVD, stroke, MI and HFH was 1.39 (1.16, 1.67), $z=3.3$, $p=0.00042$ (*Table 2*). Changing the hierarchy of the outcomes did not change the WR substantially; for example, testing in the order CVD, HFH, MI and stroke, the WR was 1.39 (1.16, 1.67), $z=3.4$, $p=0.00035$.

EPHESUS: Examining the effect of eplerenone on the composite outcome of MI, stroke, HFH or CVD in EPHESUS, analysed as time-to-first event, gave a 1/HR of 1.15 (1.04, 1.27), $z=2.7$, $p=0.0033$ (HR =0.87 [0.79, 0.96]). Again, the 1/HR for the components varied considerably, from 1.20 (1.06, 1.39) (HR =0.83 [0.72, 0.94]) for CVD to 0.75 (0.52, 1.08) (HR =1.33 [0.92, 1.92]) for stroke (*Table 2*). The WR for the same composite outcome tested in the order CVD, stroke, MI and HFH was 1.15 (1.05, 1.27), $z=2.8$, $p=0.0026$ (*Table 2*). Changing the hierarchy to CVD, HFH, MI and stroke, gave a WR of 1.18 (1.06, 1.28), $z=3.0$, $p=0.0012$; the WR had a greater z-score/smaller p-value reflecting a higher magnitude of the eplerenone effect in CVD and HFH.

CORONA: In a *post hoc* analysis of CORONA, the effect of rosuvastatin, compared with placebo, on the composite outcome of MI, stroke, HFH or CVD, analysed as time-to-first event, gave a 1/HR of 1.11 (1.09, 1.20), $z=2.2$, $p=0.015$ (HR =0.90 [0.83, 0.92]). The 1/HR for the individual components again varied, ranging from 1.28 (1.00, 1.64) (HR =0.78 [0.61, 1.00]) for MI to 1.03 (0.92, 1.16) (HR =0.97 [0.86, 1.09]) for CVD (*Table 2*). The WR for the primary composite outcome tested in the hierarchy of CVD, stroke, MI, and HFH was 1.09 (1.00, 1.19), $z=1.7$, $p=0.044$ (*Table 2*). Changing the order of analysis to the hierarchy of CVD, HFH, stroke and MI gave a WR of 1.08 (0.99, 1.18), $z=1.4$, $p=0.067$; changing the “statistical significance” of the results at the p-value threshold of 0.05.

PARADIGM-HF: A *post hoc* analysis of the same broad composite outcome of MI, stroke, HFH or CVD in PARADIGM-HF, analysed as time-to-first event, gave a 1/HR of 1.20 (1.10, 1.30), $z=4.1$, $p=0.000024$ (HR =0.83 [0.77, 0.91]) (*Table 2*). The numbers of each of the components of the composite varied enormously (and the proportion of one type of the component event versus another was quite different in PARADIGM-HF than in SPRINT, EPHESUS or CORONA). The effect of sacubitril/valsartan on each of the components also differed substantially, with less prominent effects on stroke and MI, than on the other components (*Table 2*). The corresponding WR for the composite tested in the hierarchy of CVD, stroke, MI and HFH was 1.21 (1.11, 1.32),

$z=4.1$, $p=0.000019$ (*Table 2*). Changing the order to CVD, HFH, stroke and MI gave a WR of 1.22 (1.11, 1.33), $z=4.3$, $p=0.0000097$; reflecting a higher magnitude of the sacubitril/valsartan effect in CVD and HFH.

Additional analyses exploring expanded end-points

Non-hospitalised worsening HF: We examined the effect of sacubitril/valsartan on a *post hoc* composite outcome of outpatient therapy for worsening HF, an emergency department visit for worsening HF, HFH or CVD (*Table 3, upper panel*). Comparing *Table 3* to *Table 1* shows the additional number of “first events” (156 more patients in the enalapril group and 123 more in the sacubitril/valsartan group). In a time-to-first event analysis, the 1/HR was 1.27 (1.14, 1.37), $z=4.9$, $p=0.000000021$ (HR =0.79 [0.73, 0.88]). Again, the number of each of these components, and the effect of sacubitril/valsartan on each of them, varied much (*Table 3*). The WR for the hierarchy tested in the order CVD, HFH, emergency department visit for worsening HF or intravenous diuretic use for worsening HF was 1.27 (1.16, 1.39), $z=4.9$, $p=0.000000036$. The additional number of wins and losses with this expanded composite was 469882 and 347193, respectively, compared with the two-component primary composite outcome (*Table 3*).

Health-related quality-of-life: To illustrate the potential utility of the WR vs. Cox in phase II trials and/or trials with short follow-up and patient-centered outcomes, we have also examined the following outcomes in PARADIGM-HF at 8 months: 1) the primary composite outcome of HFH or CVD “capped” at 8 months, and 2) a *post hoc* composite outcome, with deterioration in health-related quality-of-life (≥ 5 points decrease in the Kansas City Cardiomyopathy Questionnaire Overall Summary Score [KCCQ-OSS]) at month 8 added as a third component. The third component added 1020 more patients to the enalapril group and 930 more patients to the sacubitril/valsartan group (*Table 3, lower panel*). The 1/HR for the effect of sacubitril/valsartan, compared with enalapril, on the primary, two-component, composite outcome at 8 months was 1.22

(1.06, 1.41), $z=2.5$, $p=0.0055$ (HR =0.82 [0.71, 0.94]). The WR for this composite at 8 months was 1.18 (1.08, 1.28), $z=2.5$, $p=0.0056$. For the expanded, three-component, composite, including KCCQ-OSS, the 1/HR was 1.14 (1.06, 1.23), $z=3.3$, $p=0.00056$ (HR =0.88 [0.81, 0.94]). Testing this composite in the hierarchical order of CVD, HFH and ≥ 5 points decrease in KCCQ-OSS, gave a WR of 1.23 (1.06, 1.43), $z=3.7$, $p=0.000114$ (*Table 3*).

Biomarkers: In the subset of PARADIGM-HF patients who had a measurement of NT-proBNP at baseline and at 8 months ($n=1820$), we examined the three-component composite outcome described above (CVD, HFH, ≥ 5 points decrease in KCCQ-OSS at month 8) and a four-component composite adding the failure to experience a more than 30% decrease in NT-proBNP with treatment at month 8 (*Table 4*). Here, because of the smaller sample size, few deaths and admissions, and modest effect of sacubitril/valsartan on KCCQ-OSS, the three-component composite outcome was not improved significantly by sacubitril/valsartan, compared with enalapril. Addition of NT-proBNP, which was reduced more by sacubitril/valsartan than by enalapril, resulted in a significant improvement in the four-component composite outcome: 1/HR =1.12 (1.02, 1.25), $z=2.0$, $p=0.024$ (HR =0.89 [0.80, 0.98]). Using the WR with an event hierarchy of CVD, HFH, ≥ 5 points decrease in KCCQ-OSS and, last, failure to decrease NT-proBNP by more than 30%, gave a greater z -score/smaller p -value than HR; WR =1.30 (1.14, 1.67), $z=3.4$, $p=0.0004$ (*Table 4*).

WR and its components over time

For clarity this is illustrated only for the two two-component HF trials PARADIGM-HF and DIG (*Table 1 & Figure 2*). As can be seen, the WR pattern over time was quite different between the two trials. In PARADIGM-HF we can observe a steady and favourable effect of sacubitril/valsartan vs. enalapril over time (*Figure 2a*). This effect is observed in both of the components of the primary outcome (*Figure 2b*). In DIG, we can see an initial favourable effect of digoxin, reflecting an early effect on HFH, that is reduced over time by an absence of effect on CVD (*Figure 2c*). This can be

visualized in the *Figure 2d*, where the effect of digoxin is only observed on hospitalizations, with no effect on mortality.

The main features of the time-to-first event Cox and the WR are depicted in the *Central Illustration*.

Supplemental analyses

Comparisons of WR and conventional Cox statistics for a range of other trials are provided in the Supplementary Appendix (*Supplemental Tables 1 and 2*), with consistently similar results.

A head-to-head comparison of the p-values obtained with the WR and HR taking 50% and 30% random subsamples of the studied trials is presented in the *Supplemental Table 2*. These comparisons show p-value Spearman correlations >0.8 for all trials suggesting that both approaches may provide similar statistical power when considering similar scenarios.

DISCUSSION

We explored the use of WR analysis in large cardiovascular outcome trials. We examined how WR analysis compared to time-to-first event Cox proportional-hazards analysis in exemplar trials. We also investigated the use of the WR to analyse expanded composite outcomes using non-typical components.

For conventional composite outcomes, incorporating hospitalization and death, we found that the WR analysis and time-to-first event Cox analysis gave very similar findings with only minor differences in treatment effect size and p-values. Because CVD was considered the most clinically important event in the composites examined, the effect of treatment on CVD was a critical determinant of the WR, especially if CVD was a common occurrence, relative to other events. In all cases where treatment reduced CVD, the most favourable WR testing hierarchy (i.e. testing in the order of the largest to the smallest treatment effect on the non-fatal components, after CVD) gave a WR that had a greater z-score/smaller p-value than the HR from the time-to-first event analysis. Conversely, in the DIG trial, where digoxin did not reduce CVD (but did reduce HFH), the WR for the composite of these two events gave a smaller z-score/greater p-value than the corresponding HR. These case-illustrations show that, by considering the fatal events as the most clinically relevant, the WR may be more clinically meaningful than the time-to-first event Cox that ignores the primary outcome fatal events that occur after a non-fatal one⁴. For example, in the PARADIGM-HF and DIG trials more than 30% and 40% of the CVD were not considered for the reported HR, respectively. In PARADIGM-HF the treatment effect favoured sacubitril/valsartan similarly in the fatal and non-fatal events; however, in DIG, digoxin reduced hospitalization but not CVD, which led to an “overestimation” of the effect using Cox (that ignored the CVD that occurred after the HF hospitalizations) compared with the WR that counted all the CVD. The representation of the WR over time allows identification of the treatment effect on each individual component of the outcome (i.e. all fatal and non-fatal) even when these are not in the same direction or when they

occur at different times. As an example, and to simplify the message to the individual patient, a WR of 1.2 with a 95% confidence interval that does not cross the unit may be stated to a patient as “your treatment was 20% better in reducing death and re-hospitalizations than the placebo, considering death as the priority”. Changing the hierarchy of WR testing resulted in a small difference between the two contrasting WRs examined, again reflecting the proportion of different types of component events in the composite outcomes and the effect of treatment on each of these events. In another example, CORONA, for the outcome of CVD, stroke, MI or HFH the HR gave a greater z-score/smaller p-value than the WR. By capturing the effect on non-fatal outcomes that directionally favoured rosuvastatin, the time-to-first Cox model increases the statistical significance of the results, whereas the WR gives a more modest “borderline significant” result by considering the CVD first (i.e. reflecting more the lack of effect on CVD). By switching the hierarchical order of HFH (small effect) to appear first in the hierarchy right after death with MI (large effect) to appear last in the hierarchy, the WR analysis led to a nominally statistically significant treatment effect becoming non-significant; although in all other trials the treatment effect size was not altered in an important way due to larger effect sizes and directionally similar outcomes. While it might still seem attractive to trial sponsors to obtain the most favourable treatment effect-size, the prospect of achieving this goal will depend on an accurate, *a priori*, prediction of the differential impact of a study treatment on the different components of the composite endpoint used, something that may not always be possible. There may also be disagreement about the relative clinical importance of different non-fatal events (e.g. whether MI should be ranked above HFH, although most would agree that stroke should follow death in terms of clinical relevance) and this may influence not only choice of hierarchy but also the value placed on the hierarchy tested.

Perhaps the more interesting use of the WR analysis is to examine the effect of treatment on expanded composite outcomes using non-typical components. These additional components might include events or measures of less certain clinical significance. In this situation, the WR approach

provides clinicians and regulators with the assurance that the “softer” or less important components (e.g. outpatient treatment, quality-of-life, biomarkers) of the expanded composite are only tested after more traditional events (e.g. death and hospitalization) have been accounted for. However, for sponsors and trialists, these broader composites may increase the study power (because more patients potentially have an “event”) and, of course, some of these additional measures are in themselves clinically important e.g. change in health-related quality-of-life. We investigated three examples of this approach in PARADIGM-HF. Expanding the primary composite outcome of CVD or HFH to include emergency department visits and outpatient treatment for worsening HF, with ranking of these two additional components second-last and last, respectively, in the hierarchy to testing gave a WR identical to the HR, which from a regulatory, clinical and payer perspective may be reassuring. Of course, the WR and HR may not always be similar and were in this example because of the similar effect of sacubitril/valsartan on each of the four components of the expanded composite outcome. We also examined a composite outcome including health-related quality-of-life at 8 months after randomization (this was the time point pre-specified for evaluation of quality-of-life). Here the size of the treatment effect on KCCQ-OSS was notably smaller than on both CVD and HFH, but many more patients experienced an “event” by incorporating decrease in KCCQ-OSS score (1427 patients in the enalapril group compared with 407 for the two-component composite). Because of the greater importance attributed to CVD (compared first) and HFH (compared second), in this example the WR gave a greater z-score/smaller p-value than the corresponding HR. Arguably, the WR approach was advantageous here compared with analysis of time-to-first event, in weighting the components of the composite outcome. Specifically, some patients experienced worsening of quality-of-life before hospitalization or death, and, thus, those subsequent events were “masked” in time-to-first analysis. In other words, here the WR approach may have given a more clinically relevant, estimate of the size of treatment effect. In a further analysis of the subset of PARADIGM-HF patients (n=1820) with a baseline and 8-month measurement of NT-proBNP,

addition of failure to experience a more than 30% decrease in NT-proBNP to the composite of CVD, HFH, and ≥ 5 points decrease in KCCQ-OSS allowed demonstration of a favourable effect of sacubitril/valsartan compared with enalapril. This superiority was not demonstrable with the narrower 3-component composite (CVD, HFH, and ≥ 5 points decrease in KCCQ-OSS) due to a loss of statistical power because of a smaller sample size. Both conventional statistics and the WR approach were significant for the 4-component composite but not for the 3-component composite. In this example, the addition of a “soft” (biomarker) component to the composite outcome was sufficiently important to allow demonstration of the benefit of sacubitril/valsartan in a modest sample-size, and the WR analysis ensured the effect of treatment on “hard” outcomes was prioritized and not masked.

In several other, individual, trials, a secondary WR analysis has been conducted. Generally, the WR findings reported were consistent with the primary time-to-first event analysis. However, the WR approach has rarely been used as the pre-specified primary analysis in trials. One example using a two-component composite was the recent Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy trial (ATTR-ACT).⁵ The primary analysis consisted of the hierarchical assessment of all-cause mortality, followed by the frequency of cardiovascular-related hospitalizations, using the Finkelstein–Schoenfeld method.³ The WR for the primary outcome was 1.70 (1.26-2.29) in favour of tafamidis. Interestingly, the investigators we asked to do additional, conventional, analyses to confirm this favourable effect. The primary findings were supported by a 1/HR of 1.43 (1.04-1.96) (HR= 0.70 [0.51-0.96]) for all-cause death and a 1/HR of 1.25 (0.97-1.61) (HR= 0.80 [0.62-1.03]) for cardiovascular hospitalizations (the relative risk-ratio for cardiovascular hospitalizations was 0.68 [0.56-0.81] in a Poisson regression analysis and 0.71 [0.54-0.93] in a negative binomial analysis). In keeping with our earlier discussion, the favourable WR reflects the prominent effect of tafamidis on mortality.

Another, ongoing trial, is using an expanded composite outcome with more components, including change in quality-of-life. The Transcatheter Aortic Valve Replacement to UNload the Left Ventricle in Patients With ADvanced HF trial (NCT02661451) will examine the hierarchical occurrence of all-cause death, disabling stroke, hospitalization (for HF, symptomatic aortic valve disease or nondisabling stroke) and change in the Kansas City Cardiomyopathy Questionnaire at 1 year.

As with any test, there are limitations inherent to both Cox and the WR. There are no accepted methods for the determination of sample size and power when using the WR approach. There is no accepted approach to adjusting for covariates, although this is of less concern in large randomized trials where baseline variables are likely to be balanced between the treatment groups. Because of this, we used the unmatched pairs approach, although a matched paired approach does exist. Clinicians are still not very familiar with the WR, but the familiarity with this approach may be improved by its increasing use. Lastly, the WR does not use a precise time-to-event, but only whether the event of interest occurred before or after the same event in the patient's pair. There are also limitations to our analysis of the WR as we did not look at other potential uses, including in the evaluation of recurrent events and continuous outcomes, among others.

Conclusions

In the large cardiovascular outcome trials, using composite outcomes including death and hospitalization, Cox and WR analysis provided similar estimates of treatment effects. The WR allows the prioritization of fatal events and also the hierarchical testing of broader composite endpoints including patient-reported outcomes and other less clinically significant components such as biomarkers. In this way, the WR approach allows incorporation of patient-centred and other outcomes, while prioritizing the competing risk of death and hospital admission.

Clinical Perspectives

Competence in Medical Knowledge: Cox models are the commonly used method to assess treatment effects in cardiovascular trials. The WR has been proposed as an alternative to Cox to assess treatment effects because it takes into account the clinical priority of the outcomes. The WR and Cox provided similar treatment effect estimates, however the WR offered advantages when comparing soft end-points such as biomarkers and/or quality of life, because it considers death and hospitalizations as priority.

Translational outlook: This study leads to a better understanding of the use of WR in clinical trials and can be applied in future studies where the clinical importance of the events is considered relevant.

Funding

JPF is funded by an ESC research grant for collaboration with the University of Glasgow. All other authors report no specific funding for this project.

Disclosures

None.

REFERENCES

1. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *Bmj*. 2007;334(7597):786.
2. 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(2):176-182.
3. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med*. 1999;18(11):1341-1354.
4. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation*. 2016;133(6):601-609.
5. Maurer MS, Schwartz JH, Gundapaneni B. Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *N Engl J Med*. 2018;379(11):1007-1016.

Figure 1. Illustration with a simplification of the Win Ratio method

Note: In the analysed trials most of the comparisons are “ties” i.e. most patients do not have any event during the follow-up; a “tie” may also happen in a situation where a patient in one group dies but its pair stopped follow-up before.

Figure 2. Representation of the Win Ratio over time

Note: These figures are representations of the treatment effect over time i.e. panels “a” and “c” represent the effect on the globality of the composite outcome over time and panels “b” and “d” represent the effect on each individual component of the outcome. In PARADIGM-HF (panels “a” & “b”) sacubitril/valsartan had a stable effect over time and similar effect on cardiovascular death and heart failure hospitalizations. In DIG (panels “c” & “d”) digoxin had an effect that decreases over time, reflecting an early effect on heart failure hospitalizations but a lack of effect on cardiovascular death that occurred later in the follow-up.

Central Illustration. Key features of the time-to-first Hazard Ratio versus the Win Ratio

Note: By prioritizing the fatal events, the WR is particularly useful for studying hierarchical composite outcomes that include fatal and non-fatal events.

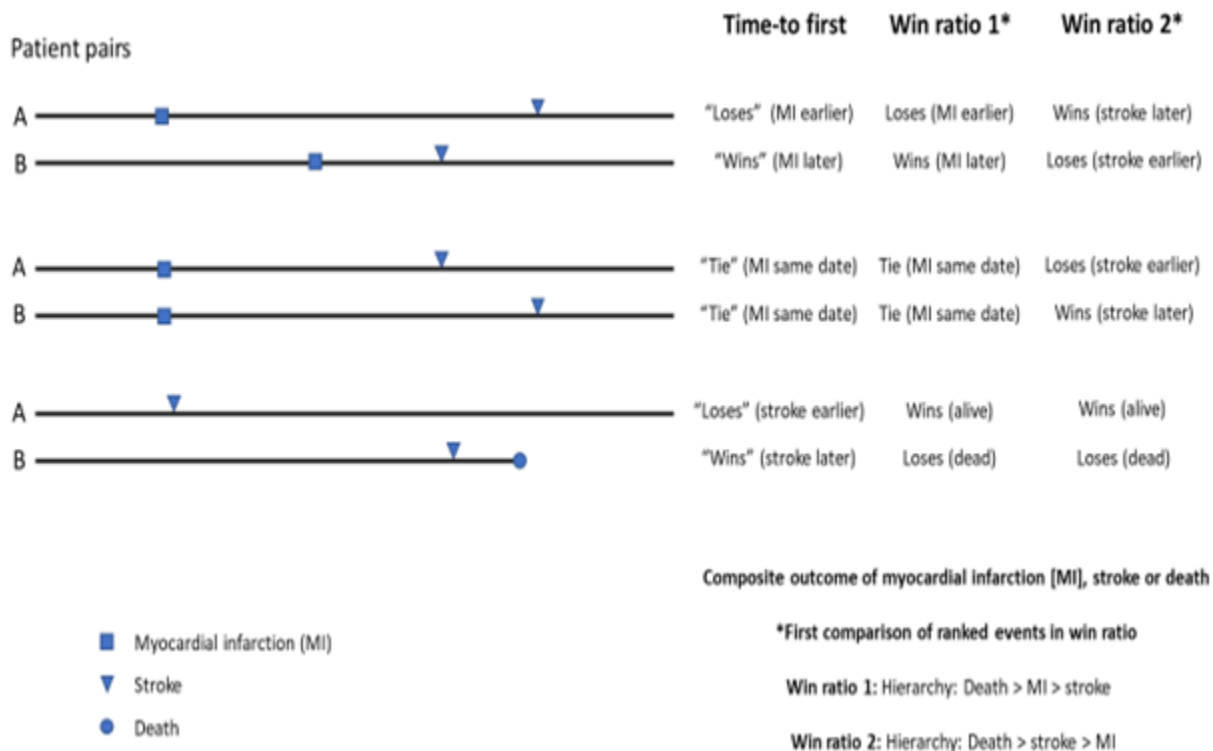
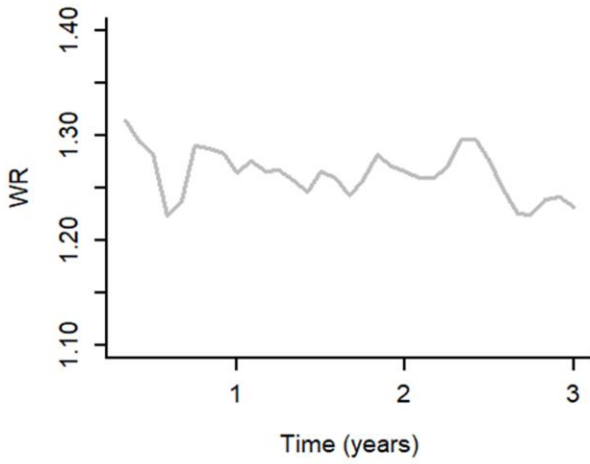
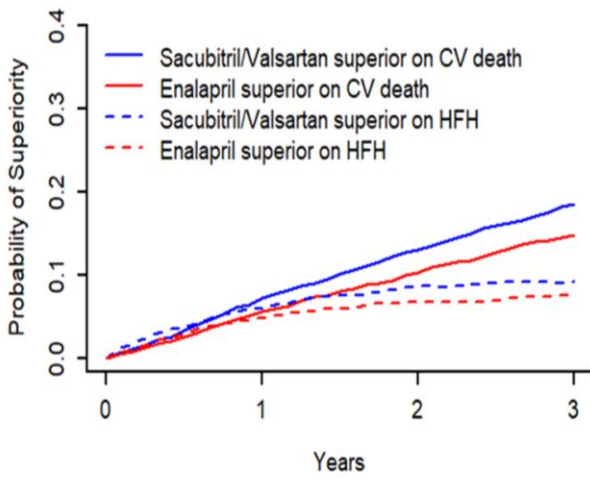


Figure 1

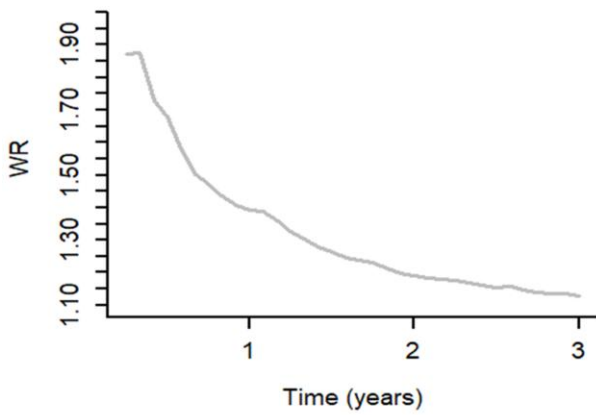
2a) PARADIGM-HF: global Win Ratio over time



2b) PARADIGM-HF: individual events Win Ratio in two component composite outcome



2c) DIG: global Win Ratio over time



2d) DIG: individual events Win Ratio in two component composite outcome

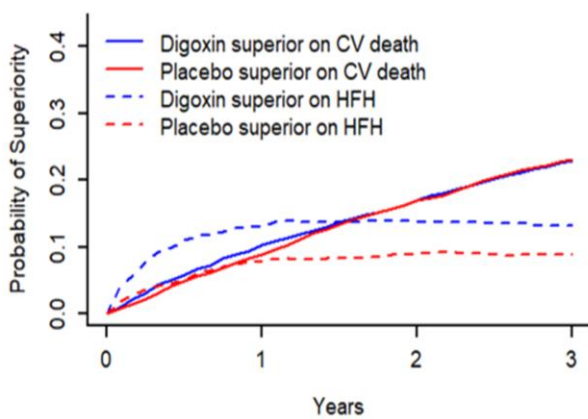


Figure 2

Hazard Ratio

- Simple, reproducible, large experience
- Requires proportionality of hazards
- Ignores fatal events occurring after non-fatal ones
- Does not incorporate event hierarchy or clinical relevance

Win Ratio

- Less experience, lack of familiarity
- Does not require proportionality of hazards (uses pairwise comparisons)
- Incorporates all fatal events of interest in the outcome
- Allows the study of hierarchical outcomes, including patient-centered measures

Central Illustration

	PARADIGM-HF				DIG			
	Enalapril (n=4212)	Sac./val. (n=4187)		z-score	Placebo (n=3403)	Digoxin (n=3397)		z-score
Event	n (%)	n (%)	1/HR (95% CI)		n (%)	n (%)	1/HR (95% CI)	
Cardiovascular death	693 (16.5)	558 (13.3)	-		1004 (29.5)	1016 (29.9)	-	
CV death contributing to the primary outcome	459 (66.2)	377 (67.6)	-		473 (47.1)	591 (58.2)	-	
Heart failure hospitalization	658 (15.6)	537 (12.8)	-		1180 (34.7)	910 (26.8)	-	
<i>Composite outcome</i>	<i>1117 (26.5)</i>	<i>914 (21.8)</i>	<i>1.25 (1.12, 1.41)¹</i>	4.8	<i>1653 (48.6)</i>	<i>1501 (44.2)</i>	<i>1.18 (1.10, 1.27)²</i>	4.5
Wins/Losses	Wins	Losses	WR (95% CI)		Wins	Losses	WR (95% CI)	
Cardiovascular death	2261469	1798909	-		2620721	2641543	-	
Heart failure hospitalization	1410767	1118983	-		1492666	1002474	-	
<i>Composite outcome</i>	<i>3672811</i>	<i>2918490</i>	<i>1.27 (1.15, 1.39)</i>	4.7	<i>4113387</i>	<i>3644017</i>	<i>1.14 (1.05, 1.20)</i>	3.1

Table 1 Comparison of time-to-first event analyses (1/HR) and the win-ratio (WR) in exemplar two-component heart failure trials

Legend: CV, cardiovascular; Sac./val., sacubitril valsartan.

The HRs were computed using Cox proportional hazards models.

¹HR =0.80 (0.71, 0.89); ²HR =0.85 (0.79, 0.91).

Table 2 Comparison of time-to-first event analyses (1/ HR) and the win-ratio (WR) in exemplar cardiovascular trials using multi-component composite outcomes

	SPRINT (n=9361)		EPHESUS (n=6632)		CORONA (n=5011)		PARADIGM-HF (n=8399)	
Disease	Hypertension		Acute myocardial infarction		Ischemic heart failure		Heart failure	
Treatment comparison	More- vs. less-intense blood pressure lowering		Eplerenone vs. placebo		Rosuvastatin vs. placebo		Sac./val. enalapril	
Event*	n (%)	1/HR (95% CI)	n (%)	1/HR (95% CI)	n (%)	1/HR (95% CI)	n (%)	1/HR (95% CI)
Cardiovascular death n, (%)	65 (1.4)	1.75 (1.18, 2.63)	483 (14.6)	1.20 (1.06, 1.39)	593 (23.7)	1.03 (0.92, 1.16)	693 (16.5)	1.25 (1.12, 1.41)
Stroke n, (%)	70 (1.5)	1.12 (0.80, 1.59)	51 (1.5)	0.75 (0.52, 1.08)	106 (4.2)	1.19 (0.89, 1.56)	107 (2.5)	1.02 (0.78, 1.33)
Myocardial infarction n, (%)	116 (2.5)	1.20 (0.92, 1.56)	229 (6.9)	1.04 (0.86, 1.25)	145 (5.8)	1.28 (1.00, 1.64)	105 (2.5)	0.99 (0.75, 1.30)
Heart failure hospitalization n, (%)	100 (2.1)	1.61 (1.19, 2.22)	391 (11.8)	1.18 (1.01, 1.35)	669 (26.7)	1.10 (0.98, 1.22)	658 (15.6)	1.25 (1.12, 1.41)
Treatment effect on composite outcome								
1/HR (Cox)	1.33 (1.12, 1.56) ¹		1.15 (1.04, 1.27) ²		1.11 (1.09, 1.20) ³		1.20 (1.10, 1.30) ⁴	
WR **	1.39 (1.16, 1.67)		1.15 (1.05, 1.27)		1.09 (1.00, 1.19)		1.21 (1.11, 1.32)	

* In the control group.

** WR tested in hierarchy: cardiovascular death, stroke, myocardial infarction, and heart failure hospitalization.

The HRs were computed using Cox proportional hazards models.

¹ HR =0.75 (0.64,0.89); ² HR =0.87 (0.79,0.96); ³ HR =0.90 (0.83,0.92); ⁴ HR =0.83 (0.77, 0.91).

z-scores: SPRINT: HR 3.1, WR 3.3; EPHESUS: HR 2.7, WR 2.8; CORONA: HR 2.2, WR 1.7; PARADIGM-HF: HR 4.1, WR 4.1.

Table 3 Analyses of PARADIGM-HF using expanded composite outcomes

	Enalapril (n=4212)	Sac./val. (n=4187)	
1. Event	n (%)	n (%)	1/HR (95% CI)
Cardiovascular death §	693 (16.5)	558 (13.3)	-
Heart failure hospitalization	658 (15.6)	537 (12.8)	-
ED visit	148 (3.5)	102 (2.4)	-
Outpatient therapy	604 (14.3)	520 (12.4)	-
<i>Composite</i>	<i>1273 (30.2)</i>	<i>1037 (24.8)</i>	<i>1.27 (1.14, 1.37)¹</i>
Wins/Losses	Wins	Losses	WR (95% CI)
Cardiovascular death	2261469	1798909	-
Heart failure hospitalization	1410767	1118983	-
ED visit	106604	56929	-
Outpatient therapy	363853	290682	-
<i>Composite</i>	<i>4142693</i>	<i>3265683</i>	<i>1.27 (1.16, 1.39)</i>
2. Event (at 8 months)	n (%)	n (%)	1/HR (95% CI)
Cardiovascular death	194 (4.6)	161 (3.9)	-
Heart failure hospitalization	24.8 (5.9)	212 (5.1)	-
KCCQ-OSS ≥5-point decrease	1283 (30.7)	1124 (27.0)	-
<i>Composite a)*</i>	<i>407 (9.7)</i>	<i>334 (8.0)</i>	<i>1.22 (1.06, 1.41)²</i>
<i>Composite b)*</i>	<i>1427 (34.2)</i>	<i>1264 (30.4)</i>	<i>1.14 (1.06, 1.23)³</i>
Wins/Losses	Wins	Losses	WR (95% CI)
Cardiovascular death	790036	654289	-
Heart failure hospitalization	832454	667974	-
KCCQ-OSS ≥5-point decrease	2957806	2559952	-
<i>Composite a)*</i>	<i>4580296</i>	<i>388215</i>	<i>1.18 (1.08, 1.28)</i>
<i>Composite b)*</i>	<i>1622779</i>	<i>1322550</i>	<i>1.23 (1.06, 1.43)</i>

--	--	--	--

*Composite a) is cardiovascular death or heart failure hospitalization; composite b) is composite a) with the addition of a ≥ 5 -point decrease in the KCCQ-OSS.

§ Enalapril: 420 (60.6%) vs. Sac./val.: 335 (60.0%) cardiovascular deaths contributing to the composite outcome.

ED = emergency department.

KCCQ-OSS = Kansas City Cardiomyopathy Questionnaire Overall Summary Score.

The HRs were computed using Cox proportional hazards models.

¹ HR =0.79 (0.73, 0.88); ² HR =0.82 (0.71, 0.94); ³ HR =0.88 (0.81, 0.94).

z-scores: Model 1 HR 4.9, WR 4.9; Model 2 HR a) 2.5 b) 3.3; WR a) 2.5 b) 3.7.

Table 4. Analysis of PARADIGM-HF including biomarkers

	Enalapril (n=905)	Sac./val. (n=915)	
Event (at 8 months)	n (%)	n (%)	1/HR (95% CI)
Cardiovascular death	2 (0.2)	1 (0.1)	-
Heart failure hospitalization	42 (4.6)	32 (3.5)	-
KCCQ-OSS \geq 5-point decrease	281 (31.1)	270 (29.5)	-
No decrease $>$ 30% in NT-pro BNP	724 (80.0)	617 (67.4)	-
<i>Composite a)*</i>	<i>308 (34.0)</i>	<i>291 (31.8)</i>	<i>1.08 (0.91, 1.25)¹</i>
<i>Composite b)*</i>	<i>770 (85.1)</i>	<i>701 (76.6)</i>	<i>1.12 (1.02, 1.25)²</i>
Wins/Losses	Wins	Losses	WR (95% CI)
Cardiovascular death	1828	905	-
Heart failure hospitalization	37003	28051	-
KCCQ-OSS \geq 5-point decrease	165361	153769	-
No decrease \geq 30% NT-pro BNP	116058	61902	-
<i>Composite a)*</i>	<i>203280</i>	<i>182698</i>	<i>1.11 (0.66, 1.33)</i>
<i>Composite b)*</i>	<i>319338</i>	<i>244591</i>	<i>1.30 (1.14, 1.67)</i>

*Composite a) is cardiovascular death, heart failure hospitalization or \geq 5-point decrease in the KCCQ-OSS; composite b) is composite a) with the addition of no decrease in NT-pro BNP of \geq 30%.

KCCQ-OSS = Kansas City Cardiomyopathy Questionnaire Overall Summary Score.

NT-pro BNP = N-terminal pro-B-type natriuretic peptide.

The HRs were computed using Cox proportional hazards models.

¹ HR =0.93 (0.80,1.10); ² HR =0.89 (0.80,0.98).

z-scores: HR a) 0.1 b) 2.0; WR a) 0.6 b) 3.4