Generalized Pairwise Comparisons to Assess Treatment Effects: JACC
Review Topic of the Week
Peer-reviewed author version

Generalized Pairwise Comparisons to Assess Treatment Effects: JACC Review Topic of the Week

Johan Verbeeck, PhD, a,* Mickaël De Backer, PhD, b Jan Verwerft, MD, c   Samuel Salvaggio, PhD, b Marco Valgimigli, MD Prof PhD, d Pascal Vranckx, MD Prof PhD, c Marc Buyse, Prof ScD a,b, Edgar Brunner, Prof PhD, e

Total word count: 4993 words

Affiliations:

a DSI, I-Biostat, University Hasselt, Hasselt, Belgium
b International Drug Development Institute, Louvain-la-Neuve, Belgium.
c Department of Cardiology and Critical Care Medicine, Hartcentrum Hasselt, Jessa Ziekenhuis Hasselt, Faculty of Medicine and Life Sciences, University of Hasselt, Hasselt, Belgium
d Cardiocentro Institute, Ente Ospedaliero Cantonale, Università della Svizzera Italiana (USI), CH-6900 Lugano, Switzerland
e Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

Disclosures:

MV reports personal fees from Astra Zeneca, grants and personal fees from Terumo, personal fees from Alvimedica/CID, personal fees from Abbott Vascular, personal fees from Daiichi Sankyo, personal fees from Bayer, personal fees from CoreFLOW , personal fees from IDORSIA PHARMACEUTICALS LTD, personal fees from Universität Basel | Dept. Klinische Forschung, personal fees from Vifor, personal fees from Bristol Myers Squib SA, personal fees from Biotronik, personal fees from Boston scientific, personal fees from Medtronic, personal fees from Vesalio, personal fees from Novartis, personal fees from Chiesi, personal fees from PhaseBio, outside the submitted work.
PV reports personal fees from Astra Zeneca, CSL Behring, Daichii Sankyo, Janssen, Novartis, Pfizer & Bristol-Meyers Squibb alliance none related to this manuscript.
MB is stockholder of IDDI.
All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Address for correspondence:

* Johan Verbeeck, University Hasselt, Agoralaan Gebouw D - B-3590 Diepenbeek, Hasselt, Belgium. Email: johan.verbeeck@uhasselt. be tel: +32 11 26 82 68

Twitter handle: @DSI_Uhasselt

Tweet: The generalized pairwise comparisons method permits combining prioritized clinical events and multiple #QOL or mechanistic parameters in a single analysis; this #JACC focus seminar shows you the nuts and bolts of this method #Statistic #ClinicalTrials #Cardiotwitter

Abstract:

A time-to-first event composite endpoint analysis has well-known shortcomings in evaluating a treatment effect in cardiovascular clinical trials. It does not fully describe the clinical benefit of therapy as the severity of the events, events repeated over time, and clinically relevant non-survival outcomes cannot be considered. The generalized pairwise comparisons (GPC) method adds flexibility in defining the primary endpoint by including any number and type of outcomes that best capture the clinical benefit of a therapy as compared to standard of care. Clinically important outcomes, including, bleeding severity, number of interventions, and quality of life, can easily be integrated in one single analysis. The treatment effect in GPC can be expressed by the net treatment benefit, the success odds or the win ratio. This review provides guidance on the use of GPC and the choice of treatment-effect measures for the analysis and reporting of cardiovascular trials.

Condensed Abstract:

Generalized pairwise comparisons is a statistical method for outcome analysis and reporting of prioritized composite endpoints in cardiovascular trials. All clinically meaningful outcomes can be included in a single analysis. The three treatment measures, net treatment benefit, success odds and win ratio, lead to equal p-values but differ in their estimation of the treatment effect. For survival times, the three measures depend on follow-up time, except the win ratio under a constant hazard ratio. The net treatment benefit and success odds are recommended under a time-varying hazard ratio and when categorical (non-survival) outcomes are included in the analysis.

Abbreviations and acronyms:

CI: Confidence Interval

GPC: Generalized Pairwise Comparisons

HR: Hazard Ratio

KCCQ: Kansas City Cardiomyopathy Questionnaire

NNT: Number Needed to Treat

NTB: Net Treatment Benefit

PH: Proportional Hazards

QoL: Quality of Life

SO: Success Odds

WR: Win ratio

Consensus statements[1] recommend selecting a single, clinically meaningful endpoint to assess the efficacy and/or safety of a randomized treatment in a clinical trial, but in cardiovascular (CV) clinical trials the treatment effect is often evaluated with a primary composite endpoint, combining two or more related clinical events (e.g., all-cause death, disabling non-fatal stroke, non-fatal myocardial infarction (MI), or urgent coronary revascularization)[2-5]. Composite endpoints are often evaluated in a time-to-first event analysis, using conventional statistical methods such as the Kaplan–Meier estimator, log-rank test, and Cox proportional hazards regression[6]. However, this conventional reporting of composite endpoints has inherent limitations. It considers all contributory events of the primary endpoint as equally important and considers only each patient's first event; all further events are ignored.

Non-fatal events, irrespective of how serious they are, are considered just as important as fatal events and since they often occur earlier in the trial, the conventional analysis is often dominated by lesser important events[3-4, 7-9]. This may support the need to report the treatment effect on the composite endpoint as well as on all individual components[2,9-12], but it may lead to misleading conclusions if the composite endpoint is considered in the primary analysis.

In addition, the time-to-first event analysis ignores recurrent events[4] and disregards the association between events in the composite endpoint. Finally, the time-to-first event analysis is restricted to survival outcomes, even though categorical or continuous outcomes, such as quality of life (QoL), left ventricular ejection fraction, or 6-minute walk test, may also be of interest. Hence, the time-to-first event analysis is limited in assessing the clinical benefit of an experimental therapy.

The generalized pairwise comparisons (GPC) method was introduced to overcome these shortcomings[13-16]. GPC is a flexible statistical tool that allows prioritizing any number and type of outcomes by clinical severity to reflect the overall clinical benefit of a therapeutic intervention[14].

While initially proposed as a testing procedure in 1999 by Finkelstein and Schoenfeld[13], the GPC method was formally introduced in 2010 by Buyse[14] and the win ratio in 2012 by Pocock et al.[15]. Although it has been applied in other clinical areas, the CV disease area has witnessed a fast uptake of GPC, including post-hoc analyses of CV trials [12,17-27], clinical trial design[24,28,29], and primary endpoint analyses[30-34]. Health authorities have approved Tafamidis for cardiac amyloidosis based on a GPC analysis in the ATTR-ACT trial[31]. The treatment effect in a GPC analysis can be expressed in several measures, including the net treatment benefit, win ratio and success odds[35,36], but the win ratio has gained popularity in CV trials. However, these measures of treatment effect do not always provide the same insights into the clinical treatment effect.

The aim of this manuscript is to review the available treatment effect estimates for GPC, pointing to differences between these estimates and suggesting recommendations for their implementation in CV trials.

Generalized pairwise comparisons (GPC) and treatment effect measures

What is GPC? The generalized pairwise comparisons method is a multivariate extension of the Mann-Whitney form of the Wilcoxon test[37]. It aims at comparing two groups of observations based on multiple prioritized outcomes. In our example all-cause death is the most severe event and disabling stroke may be ranked hierarchically higher than MI (Figure 1). Urgent coronary revascularization may be considered the less severe event and may be assessed by time (i.e., the time to first revascularization) or frequency (i.e., number of revascularizations). In a prioritized GPC, every possible patient-to-patient pair is formed. That is, every patient on the new treatment is compared with every patient on the control treatment. Within each pair, the outcome of the highest priority is compared to determine which of the

two patients had the more favorable outcome. If this cannot be assessed on the outcome of highest priority, the comparison is performed on the next prioritized outcome. The evaluation continues until a more favorable outcome can be determined, or until the last outcome results in no assignment of a more favorable outcome. In this case, there is a 'tie' (Figure 1). Ties occur when the two patients of the pair have the same observed value (e.g., the same event time, the same category of QoL improvement, etc.) or through the censoring of either value in a pair. In case of censoring, the outcomes are evaluated on the pair's common duration of observation (Figure 1).

The definition of a better outcome is determined per outcome and may depend on a threshold[14]. For example, a patient could have a more favorable outcome than another patient only if the difference in survival were longer than 1 month, or any other threshold considered clinically relevant. This feature makes the method highly flexible in defining outcomes that are considered clinically meaningful. Introducing a threshold will increase the number of ties due to equal observed values. If one patient dies at 6 month and the other at 6 months and 1 day, the pairwise comparison without threshold will result in a favorable outcome for the latter patient, while the pairwise comparison with a threshold of 1 month difference will result in a tie.

Time-to-worst event versus time-to-first event. In contrast to a time-to-first event analysis, when outcomes are prioritized from the most to the least clinically relevant, the GPC method results in a time-to-worst event analysis. The emphasis is on the severity of the event, rather than its time of occurrence (Figure 1).

Treatment effect measures in GPC. The pairwise comparisons result in a count of wins for the experimental treatment, a count of losses for the experimental treatment (or wins for the

control treatment) and a number of ties. With these three counts, several GPC treatment effect measures can be estimated (Table 1). Initially, Finkelstein and Schoenfeld proposed the difference in wins, which depends on the sample size [13]. Buyse later suggested dividing this difference by the number of pairwise comparisons, i.e., the difference in win proportions, to estimate the net treatment benefit (NTB).[14] This absolute measure of the treatment effect always lies between -1 and 1. Values above zero indicate a benefit for the experimental treatment and below zero harm. As the net treatment benefit is an absolute measure of treatment effect, its inverse provides the number needed to treat (NNT), frequently used in health technology assessment of devices and drugs. The win ratio (WR) is the ratio of the win proportions[15], resulting in a relative measure of treatment effect.

Finally, the success odds[38-42] (SO) is a ratio of the win proportions with half the number of ties added to each proportion (Table 1). Moreover, it is a transformation of the NTB to a ratio, . It is obvious that when there are no ties among the pairwise comparisons, the WR and SO are equal. In the presence of ties, the success odds is always smaller[38-42] (Table 2). The success odds is also called win odds[38,40], but from a patient's perspective, success/failure may sound better than win/loss.

A wide array of variance estimators has been suggested for the GPC treatment effect measures, based on either relations to rank procedures, asymptotic U-statistic theory, permutations or bootstrapping[36,43].

Odds interpretation. The interpretation of the win ratio has been illustrated by drawing the parallel to odds in betting on horses[27]. However, odds require the probabilities in the denominator and the numerator to be complementary, i.e., adding up to 1. By ignoring the ties, the probabilities in the win ratio will not add up to 1 and, therefore, cannot be interpreted as an odds, in contrast to the success odds.

Outcome contributions. A major difference between the GPC measures is that only the net treatment benefit has the convenient property that the contributions of all outcomes in the GPC analysis are additive (Table 3). In other words, the net treatment benefit can be decomposed in partial contributions of each outcome to the overall treatment effect. Note that the contribution of each outcome is not the treatment effect on this particular outcome, except for the highest prioritized outcome[16]: it is the treatment effect on this outcome conditional on no effect on higher priority outcomes.

Choice between treatment effect measures. When testing for a treatment effect in a randomized clinical trial, in general, trialists do not have to choose between the GPC measures upfront, since all GPC measures will always result in approximately identical p-values[35,36]. However, there are differences between the GPC measures for estimating and reporting the treatment effect. These differences are pointed out in two situations which lead to ties, one where only survival outcomes compose the multivariate endpoint (with ties due only to censoring), and one where non-survival outcomes complement the survival outcomes (with ties due to both censoring of the survival outcomes and equal observations of the categorical outcomes). While it is often claimed that the win ratio ignores the number of ties, an alternative view is that it redistributes the ties to wins and losses according to the observed win proportions (Supplementary material). Thus, the win ratio implicitly assumes that the still unobserved events of survival outcomes behave similarly to observed events. Similarly, and more problematic, ties due to equal observations are redistributed according to the observed win proportions. On the other hand, the success odds redistributes ties equally to wins and losses, which is more in line with the concept of a tied observation.

GPC measures for survival outcomes

Time-dependency of survival. The presence of censored observations hallmarks survival endpoints, since not all patients experience a clinical event during the prespecified follow-up period of a clinical trial. Therefore, only part of the survival curve is observed, and we are unaware of what would happen after the trial period (Figure 2).

The analysis of survival endpoints thus depends on the time of analysis. The hazard ratio (HR) of the events between the two treatment arms is often assumed constant to avoid this time dependence (often called "proportional hazards assumption"). In this situation, the hazard ratio and win ratio estimates remain constant regardless of the time of analysis (Figure 2 top row). The net treatment benefit and success odds do not assume a constant hazard ratio, but an alternative net treatment benefit can be defined that does (Supplementary material). When a constant hazard ratio is present, the net treatment benefit and success odds will differ from the win ratio and converge towards the win ratio with increasing time. However, when a constant hazard ratio is not present, e.g., when there is a late treatment effect (Figure 2 middle row) or when hazards are crossing (Figure 2 bottom row), all GPC treatment measures as well as the hazard ratio depend on time of analysis.

Hazard ratio versus GPC. Conveniently, when the constant hazard ratio assumption applies, the win ratio equals the inverse of the hazard ratio. However, for multiple endpoints, the win ratio will only be equal to the inverse of the hazard ratio when the hazard ratio is constant in both the joint survival outcome and each survival outcome individually[44]. It is unlikely that these assumptions are met over the entire survival curve. Nevertheless, they may be present during the limited observation time in a clinical trial period[27]. The advantage of using the time-to-worst rather than the time-to-first event analysis is that the interpretation may be more

clinically intuitive. Moreover, the number of clinically more important events is increasing over time in GPC, while they're not in a time-to-first event analysis. This may be of interest when authorities request additional follow-up of a trial to investigate the effect of the treatment on severe clinical events, such as death. In a time-to-first event analysis, many deaths will remain hidden over time by minor events, while ultimately, GPC will evaluate survival. The net treatment benefit and success odds are also related to the hazard ratio, but via the amount of censoring.

MATRIX trial. The MATRIX trial was a randomized, multicenter, superiority trial comparing trans-radial (n= 4197) against transfemoral access (n= 4207) in acute coronary syndrome patients who underwent coronary angiography and percutaneous coronary intervention[45], evaluating if radial artery access decreases severe bleeding events and clinical outcomes under combined antithrombotic therapy, compared with femoral artery access. The 30-day coprimary endpoints were major adverse cardiovascular events (MACE), including all-cause death, MI, stroke, and net adverse clinical events (NACE), including major bleeding events according to Bleeding Academic Research Consortium (BARC) grading (BARC 3 or 5). The time-to-first event analysis showed superiority on the NACE endpoint (p=0.0079), but not on the MACE endpoint (p=0.0278) at the prespecified two-sided confidence level of 0.025. In the NACE endpoint, BARC 3 bleeding was the most frequent first event, masking later occurring MACE events. Clinically, however, it seems more appropriate to account for the event severity, rather than the first event. GPC allows to define an endpoint prioritizing the MACE outcomes and the severity of bleeding events (Figure 3).

We re-analyzed the MATRIX trial at 30 days and 365 days with a prioritized MACE of all-cause death (including BARC 5), hemorrhagic stroke (including BARC 3c), ischemic stroke and MI, and a prioritized NACE endpoint, adding BARC3a/b and BARC 2 bleeding (Figure

3).

Both at 30 and 365 days, there is a decrease in NACE events in the radial arm (Table 4), which is mainly established within the first ten days (Figure 4). However, at a confidence level of 0.025, there is no evidence for a decrease in MACE events (Table 4). Although visually there is no major violation against a constant hazard ratio assumption (Figure 4), the win ratio is the most variable effect size measure over time (Table 4). The treatment effect is smaller in the success odds and net treatment benefit since the event rate is low (~10-15%), and thus, the number of censored observations is large. The main advantage of the net treatment benefit is that it shows immediately that the trans-radial access mainly decreases BARC2 bleeding events, followed by MI and all-cause death (Table 4). Notice that the partial contributions of the win ratio are not intuitive and attribute a large negative effect to hemorrhagic and ischemic stroke, despite a small absolute difference. While the conclusion is not different from the original time-to-first event analysis, the prioritization of the outcomes is clinically more sound.

GPC measures for survival and non-survival outcomes

Categorical outcomes. In many recent designs of clinical trials with a prioritized GPC analysis as a primary endpoint, categorical outcomes with two or more categories (quality of life, left ventricular ejection fraction, improvement of 6-minute walk test, or of NT-proBNP, etc.) are considered alongside survival outcomes[30]. However, for categorical data, (many) ties due to equal observations are expected in the subjects pairwise comparisons. Because the win ratio redistributes these ties according to the observed win proportions (Supplementary material), it should not be overinterpreted in the presence of many equal observations[35,38-41] (Tables 2,3) In this case, the net treatment benefit and success odds are more appropriate. [35,38-41]

Continuous outcomes. In case a continuous outcome is added to survival outcome(s), ties due to equal observations are not likely to occur and the win ratio may be useful, given the constant hazard ratio constraint. The ongoing PARACHUTE-HF trial (NCT04023227), for example, evaluates in addition to CV death and time to first heart failure hospitalization, a continuous outcome defined as the relative change in NT-proBNP. Notice that ties are induced when thresholds of clinical importance are introduced in the pairwise comparisons of continuous outcomes (including survival outcomes), which favours the use of the net treatment benefit or success odds as a treatment effect measure.

TAVR UNLOAD trial. Since most of the cardiovascular clinical trials that combine survival and non-survival outcomes are still ongoing, a simulation of the TAVR UNLOAD trial[28] is presented as an illustration. In severe aortic stenosis patients with heart failure with reduced ejection fraction (HFrEF), transcatheter aortic valve replacement (TAVR) improves clinical outcomes, shows a faster recovery and a more improved QoL compared to surgical valve replacement. In patients with moderate aortic stenosis and HFrEF, current heart failure guidelines recommend drug therapy[46,47], which reduces the vascular resistance and prolongs survival, but does not affect the aortic valve impairment. Hence, hemodynamic burden persists in these patients, negatively affecting QoL. The TAVR UNLOAD trial is an ongoing, international, multicentric, open-label two-arm trial investigating whether transfemoral TAVR in addition to optimal heart failure therapy, improves clinical outcomes, including QoL in moderate aortic stenosis patients with HFrEF[28]. The primary endpoint is defined as the prioritized occurrence at one year of all-cause death, severity and time to disabling stroke, frequency and duration of hospitalizations related to HF, symptomatic aortic valve disease or

non-disabling stroke, and the change in QoL measured by the Kansas City Cardiomyopathy Questionnaire (KCCQ) compared to baseline (Figure 3). Hence, the primary endpoint comprises survival and non-survival outcomes, where ties due to equal observations are expected.

A simulation of the TAVR UNLOAD trial (Table 5 and scenario 2 in Verbeeck et al. [16]) shows that while the test for a treatment effect is approximately the same between the three GPC treatment effect measures, the size of the treatment effect measures is different. As the win ratio implicitly redistributes the ties due to equal QoL change (24% of the pairs) according to the observed win proportions, it should not be overinterpreted and the net treatment benefit and/or the success odds should be preferred for a correct interpretation (Table 6). The simulation shows that a patient treated with TAVR has a 24% higher probability of a better outcome than a control patient (95% CI [9%, 39%]), p = 0.0018 and the NNT=4. This translates to a relative risk reduction or the complement of the inverse of the SO of 39% (95% CI [17%, 55%]), p = 0.0018. The importance of combining QoL with survival outcomes is underlined by the required sample size to detect the treatment effect: without QoL, NTB = 0.12, while with QoL, NTB = 0.24 (twice as large). Hence, the sample size required to achieve the same power is approximately quadrupled since the sample size is inversely proportional to the square of the treatment effect [16].

Conclusions

The GPC method is an appealing statistical method for designing future CV trials, as it permits the simultaneous analysis of any number of survival and non-survival outcomes of varying severity. Hence, it allows tailoring the most optimal endpoint to evaluate the clinical benefit of a therapy. GPC is applicable to randomized and non-randomized trials, although the results of the latter should be considered exploratory. An advantage of GPC compared to a

time-to-first event analysis is a higher power (and thus lower required sample size) to detect a treatment effect when combining non-survival outcomes with survival or in the presence of survival outcome(s) with a time-varying hazard ratio. There is no difference between the three GPC measures (net treatment benefit, success odds and win ratio) in assessing the presence of a treatment effect. However, there is a difference in estimating the treatment effect size when ties are frequent (due to censoring, equal observations and thresholds). The treatment effect measure does not need to be chosen at the outset of the trial design and allows a tailored treatment effect measure representation. One single measure may not be sufficient in communicating the treatment effect for each stakeholder[48].

Expressing the treatment effect as a win ratio may be valuable when only survival or continuous outcomes are considered, without thresholds and under a constant hazard ratio of the joint outcome and the individual outcomes[40,44]. The net treatment benefit and success odds are recommended when the constant hazard ratio assumption is violated, in the presence of a categorical outcome or discrete survival times and when thresholds are used (Central figure). Compared to the success odds, the net treatment benefit has the advantage that it can be decomposed in an additive fashion over its components, clearly reflecting how much each component contributes to the overall treatment effect. Additionally, the net treatment benefit is an absolute risk reduction; hence, the inverse of the net treatment benefit (total pairs/absolute difference in wins) is the number needed to treat.

Table 7 summarizes the characteristics and limitations of the time-to-first event analysis and the three GPC measures. For further considerations and limitations of the GPC we refer to the Supplementary material.

Data availability statement

The data supporting this study's findings on the simulation of the TAVR-UNLOAD trial are available upon request from the corresponding author at johan.verbeeck@uhasselt.be. The data that support the findings of this study on the MATRIX study are available upon request from Marco Valgimigli at marco.valgimigli@cardiocentro.org.

Highlight Bullet points

- Generalized pairwise comparisons (GPC) is a statistical method that overcomes shortcomings of time-to-first event analyses.

- GPC can assess treatment effects as net treatment benefit, success odds and win ratio.

- This review provides guidance on the use of GPC and choice of treatment effect measures

References

1. International Conference on Harmonisation E9 Expert Working Group. ICH harmonised tripartite guideline for statistical principles for clinical trials. 1998. Accessed November 16, 2022. https://database.ich.org/sites/default/files/E9_Guideline.pdf.

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, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: Systematic review of randomised controlled trials. *BMJ*. 2007; 334(7597):786–788

4. Lim E, Brown A, Helmy A, Mussa S, Altman DG. Composite outcomes in cardiovascular research: A survey of randomized trials. *Ann Intern Med.* 2008;149(9):612–617.

5. Armstrong PW, Westerhout CM. Composite end points in clinical research. A time for reappraisal. *Circulation.* 2017;135(23):2299–2307

6. Cox D. Regression models and life-tables. *J R Stat Soc Series B*. 1972;34(2):187–220.

7. Kleist P. Composite endpoints for clinical trials. *Int J Pharmaceut Med*. 2007;21:187–98.

8. Tan N, Ali S, Lebovic G, Mamdani M, Laupacis A, Yan A. Temporal trends in use of composite end points in major cardiovascular randomized clinical trials in prominent medical journal. *Circ Cardiovasc Qual Outcomes*. 2017 Oct;10(10):e003753. doi: 10.1161/CIRCOUTCOMES.117.003753.

9. Anker SD, Schroeder S, Atar D, et al. Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. *Eur J Heart Fail*. 2016;18(5):482–489.

10. European Medicines Agency Committee For Human Medicinal Products (CHMP). Guideline on multiplicity issues in clinical trials. 2017. Accessed November 16, 2022. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-

multiplicity-issues-clinical-trials_en.pdf

11. McCoy CE. Understanding the use of composite endpoints in clinical trials. *West J Emerg Med.* 2018;19(4):631-634.

12. Rauch G, Rauch B, Schüler S, Kieser M. Opportunities and challenges of clinical trials in cardiology using composite primary endpoints. *World J Cardiol.* 2015;26(7):1–5

13. Finkelstein D, Schoenfeld D. Combining mortality and longitudinal measures in clinical trials. *Stat Med.* 1999;18:1341–1354.

14. Buyse M. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med.* 2010;29(30):3245-3257.

15. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal.* 2012;33(2):176-182.

16. Verbeeck J, Spitzer E, de Vries T, et al. Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints. *Stat Med.* 2019;38(30):5641-5656

17. Rogers JK, Pocock SJ, McMurray JJ, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to charm-preserved. *Eur J Heart Fail.* 2014;16(1):33–40

18. Capodanno D, Gargiulo G, Buccheri S, et al. Computing methods for composite clinical endpoints in unprotected left main coronary artery revascularization. *JACC Cardiovasc Interv.* 2016;9(22):2280–2288.

19. Bhatt DL, Steg PG, Miller M, et al. Effects of icosapent ethyl on total ischemic events. *J Am Coll Cardiol.* 2019;73(22):2791– 2802.

20. 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

21. Milojevic M, Head SJ, Andrinopoulou ER, et al. Hierarchical testing of composite endpoints: applying the win ratio to percutaneous coronary intervention versus coronary artery bypass grafting in the SYNTAX trial. *EuroIntervention*. 2017;3(1):106–114.

22. Kotalik A, Eaton A, Lian Q, Serrano C, Connett J, Neaton JD. A win ratio approach to the re-analysis of multiple risk factor intervention trial. *Clin Trials*. 2019;16(6):626–634.

23. Hironori H, van Klaveren D, Takahashi K, et al. Comparative methodological assessment of the randomized GLOBAL LEADERS trial using total ischemic and bleeding events. *Circ Cardiovasc Qual Outcomes*. 2020;13(8):e006660.

24. Ferreira JP, Jhund PS, Duarte K, et al. Use of the win ratio in cardiovascular trials. *JACC Heart Fail*. 2020;8(6):441–450

25. Kandzari DE, Hickey GL, Pocock SJ, et al. Prioritised endpoints for device-based hypertension trials: the win ratio methodology. *EuroIntervention*. 2021;16(18):e1496–e1502.

26. Berwanger O, Pfeffer M, Claggett B, et al. Sacubitril/valsartan versus ramipril for patients with acute myocardial infarction: win-ratio analysis of the PARADISE-MI trial. *Eur J Heart Fail*. 2022; 24(10):1918-1927

27. Redfors B, Gregson J, Crowley A, et al. The win ratio approach for composite endpoints: practical guidance based on previous experience. *Eur Heart J*. 2020;41(46):4391-4399.

28. Spitzer E, Van Mieghem NM, Pibarot P, et al. 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.

29. Pocock SJ, Collier TJ. Statistical Appraisal of 6 Recent Clinical Trials in Cardiology: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2019 Jun 4;73(21):2740-2755.

30. Leon MB, Smith CR, Mack M, et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363(17):1597–1607

31. Maurer MS, Schwartz JH, Gundapaneni B, et al. Tafamidis treatment for patients with transthyretin amyloid cardiomyopathy. *N Engl J Med.* 2018;379(11):1007–1016.

32. Berry N, Mauri L, Feldman T, et al. Transcatheter interatrial shunt device for the treatment of heart failure: Rationale and design of the pivotal randomized trial to REDUCE elevated left atrial pressure in patients with heart failure II (REDUCE LAP-HF II). *Am Heart J.* 2020;226:222–231.

33. Lansky AJ, Makkar R, Nazif T, et al. A randomized evaluation of the TriGuard™ HDH cerebral embolic protection device to reduce the impact of cerebral embolic lesions after transcatheter aortic valve implantation: the REFLECT I trial. *Eur Heart J.* 2021;42(27):2670–2679.

34. Nazif TM, Moses J, Sharma R, et al. Randomized evaluation of triguard 3 cerebral embolic protection after transcatheter aortic valve replacement: REFLECT II. *JACC Cardiovasc Interv.* 2021;14(5):515–527.

35. Dong G, Huang B, Verbeeck J, et al. Win statistics (win ratio, win odds, and net benefit) can complement one another to show the strength of the treatment effect on time-to-event outcomes. *Pharm Stat.* 2023; 22(1):20-33.

36. Verbeeck J. (Non)-Prioritizing Generalized Pairwise Comparisons and COVID-19 (Doctoral dissertation). 2022. Accessed November 16 2022. http://hdl.handle.net/1942/38048.

37. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat.* 1947;18(1):50–60.

38. Dong G, Hoaglin DC, Qiu, J, et al. The win ratio: on interpretation and handling of ties. *Stat Biopharm Res.* 2020;12(1):99-106.

39. Brunner E. Success-odds: an improved Win-Ratio. 2020. Accessed November 16 2022.

https://arxiv.org/abs/2002.09273

40. Brunner E, Vandemeulebroecke M, Mütze T. Win odds: An adaptation of the win ratio to include ties. *Stat Med.* 2021;40(14):3367-3384.

41. Gasparyan SB, Folkvaljon F, Bengtsson O, Buenconsejo J, Koch GG. Adjusted win ratio with stratification: Calculation methods and interpretation. *Stat Methods Med Res.* 2021;30(2):580-611

42. Song J, Verbeeck J, Huang B, et al. The win odds: statistical inference and regression. *J Biopharm Stat.* 2023;33(2):140-150.

43. Verbeeck J, Ozenne B, Anderson WN. Evaluation of inferential methods for the net benefit and win ratio statistics. *J Biopharm Stat.* 2020;30(5):765-782.

44. Oakes D. On the win-ratio statistic in clinical trials with multiple types of event. *Biometrika.* 2016;103(3):742–745.

45. Valgimigli M, Gagnor A, Calabró P, et al. MATRIX Investigators. Radial versus femoral access in patients with acute coronary syndromes undergoing invasive management: a randomised multicentre trial. *Lancet.* 2015;385(9986):2465-76.

46. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation.* 2022;145(18): e895–e1032

47. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail.* 2022; 24(1):4-131.

48. Saad ED, Zalcberg JR, Péron J, Coart E, Burzykowski T, Buyse M. Understanding and

Communicating Measures of Treatment Effect on Survival: Can We Do Better? *J Natl Cancer Inst*. 2018 Mar 1;110(3):232-240

Figure 1: Two examples of clinical events at 12 months in a pair of patients.

Clinical events of a composite endpoint in one patient under experimental treatment (E) and one patient under the control treatment (C) in a clinical trial. A time-to-first event analysis considers only the first event per patient, irrespective of its clinical severity, and ignores all subsequent events. A patient with a later death is thus favored over a patient surviving with an early non-fatal MI (Example 1). Additionally, when events of varying severity occur at the same time in two patients within the common follow-up time, e.g., urgent revascularization and a disabling stroke, a time-to-first event analysis will treat these events as equal, while the time-to-worst event favors the patient with the revascularization (Example 2).

Figure 2: Survival analysis at 1 and 2 years under varying hazards.

Potential Kaplan-Meier survival curves and survival analysis at 1 year (first column) and 2 years (second column) under proportional hazards (top row), late treatment effect (middle row) and crossing of hazards at 18 months (bottom row). The analyses at 1 year are the same for each scenario, irrespective of what happens after the time of analysis. If data up to 2 years are analyzed the results may change drastically, except for the win ratio and hazard ratio under proportional hazards.

Figure 3: GPC flowchart for the MATRIX trial.

"E" = Experimental, "C" = Control, "NA" = Not Applicable.

Figure 4: Kaplan-Meier survival curves of the MATRIX trial.

MACE (left) and NACE endpoint (right).

Figure 5: GPC flowchart for the TAVR UNLOAD trial.

"E" = Experimental, "C" = Control, "NA" = Not Applicable. Stroke severity according to modified Rank Score (mRS). Hospitalizations are for heart failure, symptomatic aortic valve disease, or non-disabling stroke (mRS 0 or 1). KCCQ = Kansas City Cardiomyopathy Questionnaire.

Central illustration: Flowchart on the choice of GPC treatment effect for multivariate outcomes.

Table 1.

Summary of GPC treatment effect measures

GPC measure	Estimated by	Comments
Finkelstein-Schoenfeld	win difference	Depends on the sample size and is therefore not comparable between trials.
Net treatment benefit (NTB)		Or the Finkelstein-Schoenfeld measure divided by the number of pairs (=. An absolute measure of treatment effect with values ranging between -1 and 1. Equal to 0 if there is no treatment effect. Its inverse is the NNT.
Win ratio (WR)	ratio of wins	A relative measure of treatment effect with values ranging from 0 to infinity. Equal to 1 if there is no treatment effect. Is equal to the inverse of the hazard ratio under proportional hazards. Caution for overinterpretation in the presence of ties. Not interpretable as odds in the presence of ties.
Success odds (SO)	ratio of wins plus half the ties	A relative expression of the NTB with values ranging from 0 to infinity. Equal to 1 if there is no treatment effect. Includes the number of ties and interpretable as odds.

Number of wins for the Experimental treatment (), the number of losses for the experimental treatment or wins for the Control treatment (), and the number of Ties (). NNT = Number Needed to Treat.

Table 2.

Differences between the GPC treatment effect measures in two examples of a clinical trial

	Wins	Losses	Ties	NTB	SO	WR
Trial 1	3 (0.06%)	1 (0.02%)	4,996 (99.92%)	0.0004	1.0008	3.00
Trial 2	3,000 (60%)	1,000 (20%)	1,000 (20%)	0.40	2.33	3.00

Two examples of potential results of a clinical trial with 5,000 patients. The win ratios are identical in Trials 1 and 2, yet there is more robust evidence of a treatment effect in Trial 2, as indicated by the net treatment benefit and success odds. NTB= net treatment benefit, WR=win ratio, SO=success odds.

Table 3.

Partial contributions of GPC measures

	Wins	Losses	Ties	NTB	SO	WR
Death	684 (6.84%)	295 (2.95%)	9021 (90.21%)	0.039	1.08	2.32
Stroke	0 (0%)	93 (0.93%)	8928 (89.28%)	-0.009	0.98	0.00
MI	1700 (17%)	787 (7.87%)	6441 (64.41%)	0.091	1.23	2.16
Revascularization	2625 (26.25%)	1413 (14.13%)	2403 (24.03%)	0.121	1.46	1.86
Overall	5009 (50.09%)	2588 (25.88%)	2403 (24.03%)	0.242	1.64	1.94

The partial contributions of each outcome to the overall measure in a CV trial with 4 outcomes are additive for the NTB, but not for the ratios. In the partial contributions of the NTB, it is easy to see that revascularization contributes most to the overall effect size, followed by MI, death and stroke. The partial contributions of the SO provides similar information, though the components of the SO are not additive. In contrast, partial contributions of the WR mistakenly give the impression that death contributes most to the overall treatment effect and attributes no effect by stroke. NTB=net treatment benefit, WR=win ratio, SO=success odds.

Table 4.

GPC analysis of prioritized MACE and NACE endpoint of the MATRIX trial at 30 and

365 days.

30 days	%wins	%losses	%ties	NTB (95%CI)	SO (95%CI)	WR (95%CI)	Logrank 1/HR
Death	2.10	1.55	96.35	0.0055	1.01	1.35	
Hemor. Stroke	0.05	0.07	96.23	-0.0002	1.00	0.67	
Isch. Stroke	0.21	0.26	95.77	-0.0005	1.00	0.82	
MI	7.23	6.31	82.22	0.0093	1.02	1.15	
Total MACE	9.59	8.19	82.22	0.014 (0.002-0.027)	1.03 (1.00-1.05)	1.17 (1.02-1.35)	1.17 (1.02-1.34)
p-value MACE				0.0281	0.0281	0.0282	0.0278
BARC3a/3b	1.22	0.87	80.13	0.0034	1.01	1.39	
BARC 2	3.95	2.24	73.95	0.0171	1.03	1.76	
Total NACE	14.76	11.30	73.95	0.035 (0.020-0.050)	1.07 (1.04-1.10)	1.31 (1.16-1.47)	1.32 (1.18-1.47)
p-value NACE				< 0.0001	< 0.0001	< 0.0001	< 0.0001

365 days	% wins	% losses	% ties	NTB (95%CI)	SO (95%CI)	WR (95%CI)	Logrank 1/HR
Death	4.31	3.62	92.07	0.0069	1.01	1.19	
Hemor. Stroke	0.05	0.07	91.95	-0.0002	1.00	0.67	
Isch. Stroke	0.41	0.29	91.25	0.0011	1.00	1.39	
MI	9.70	8.90	72.65	0.0080	1.02	1.09	
Total MACE	14.47	12.88	72.65	0.016 (0.000-0.031)	1.03 (1.00-1.06)	1.12 (1.00-1.26)	1.12 (1.00-1.25)
p-value MACE				0.0413	0.0413	0.0414	0.0501
BARC3a/3b	1.33	0.88	70.44	0.0045	1.01	1.51	
BARC 2	4.22	2.66	63.56	0.0155	1.03	1.58	
Total NACE	20.02	16.42	63.56	0.036 (0.019-0.053)	1.07 (1.04-1.11)	1.22 (1.11-1.47)	1.23 (1.12-1.35)
p-value NACE				< 0.0001	< 0.0001	< 0.0001	< 0.0001

NTB= net treatment benefit, WR=win ratio, SO=success odds, HR = hazard ratio.

Table 5.

Prioritized endpoint for the TAVR UNLOAD trial with the simulated treatment effects.

Outcome		TAVR	Control	Absolute Effect	Relative Effect
Death		5.7%	7%	-1.3%	-19%
Stroke*		1.5%	1%	+0.5%	+50%
Hospitalization**		14.7%	22%	-7.3%	-33%
KCCQ change	≥ 10	10%	5%	+5%	+100%
	≥ 5 and < 10	20%	5%	+15%	+300%
	< 5 and > -5	50%	50%	0	0
	≤ 5 and < -10	10%	20%	-10%	-50%
	≤ -10	10%	20%	-10%	-50%

The survival outcomes were simulated with a constant hazard ratio. * mRS 2-5 were equally likely simulated. ** If hospitalized the chance of 1,2, or 3 hospitalization was ½,1/3 and 1/6 and the days of hospitalization was simulated from a negative binomial distribution (NB (7,10)). KCCQ = Kansas City Cardiomyopathy Questionnaire.

Table 6.

GPC analysis of a simulation of the TAVR UNLOAD trial

Outcome	% wins	% losses	% ties	NTB (95%CI)	SO (95%CI)	WR (95%CI)	Logrank *
Death	6.84	2.95	90.21	0.04	1.08	2.32	
Stroke	0.00	0.93	89.28	-0.01	0.98	0.00	
Hospitalization	17.00	7.87	64.41	0.09	1.20	2.16	
KCCQ change	26.25	14.13	24.03	0.12	1.28	1.86	
Total	50.09	25.88	24.03	0.24 (0.09-0.39)	1.64 (1.21-2.23)	1.94 (1.29-2.89)	
p-value				0.0018	0.0018	0.0021	0.0158

Simulation of 100 patients in each treatment arm following the treatment effects in Figure 5.

NTB= net treatment benefit, WR=win ratio, SO=success odds, KCCQ = Kansas City

Cardiomyopathy Questionnaire. * The logrank test compares times to first event, it cannot

take the KCCQ change into account.

Table 7

Characteristics of the time-to-first event analysis and the three GPC measures

	Time-to-first	WR	SO*	NTB*
Prioritization of multiple outcomes	-	+	+	+
Is appropriate for categorical outcomes	-	-	+	+
Relative treatment effect	+	+	+	-
Absolute treatment effect	+	-	-	+
Evenly redistributes ties	-	-	+	+
Best under constant hazard ratio	+	+	-	-
Shows additive contributions of each outcome	-	-	-	+
Can be interpreted as odds	-	-	+	-

Win ratio (WR), success odds (SO) and net treatment benefit (NTB). * The net treatment benefit (NTB) and success odds (SO) are transformations of each other.
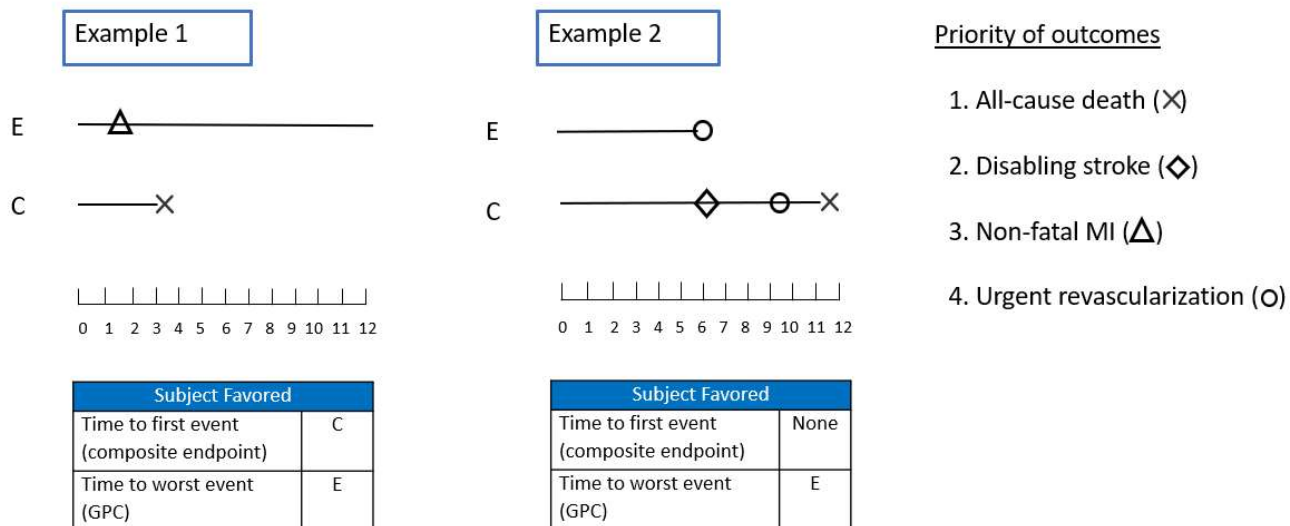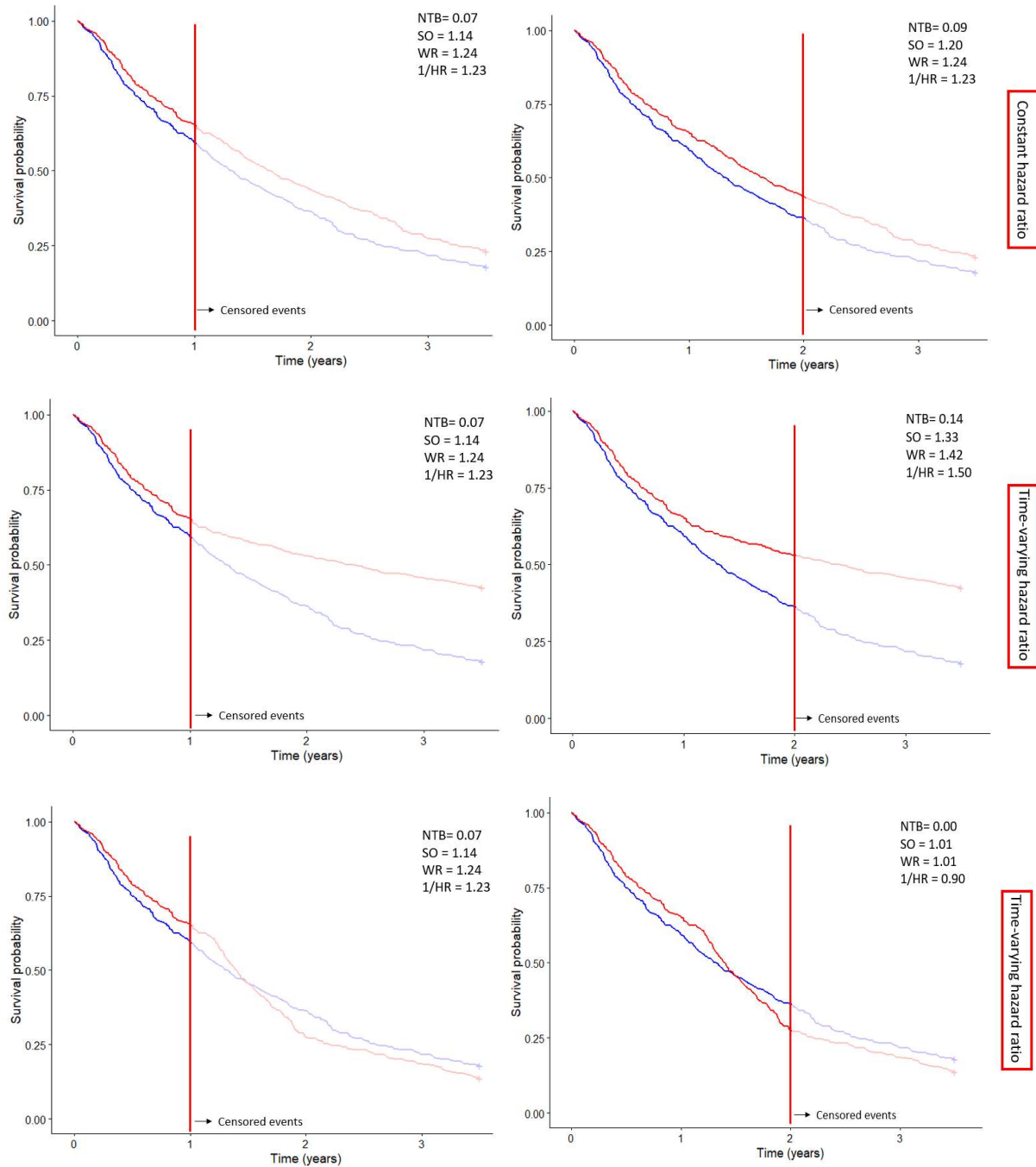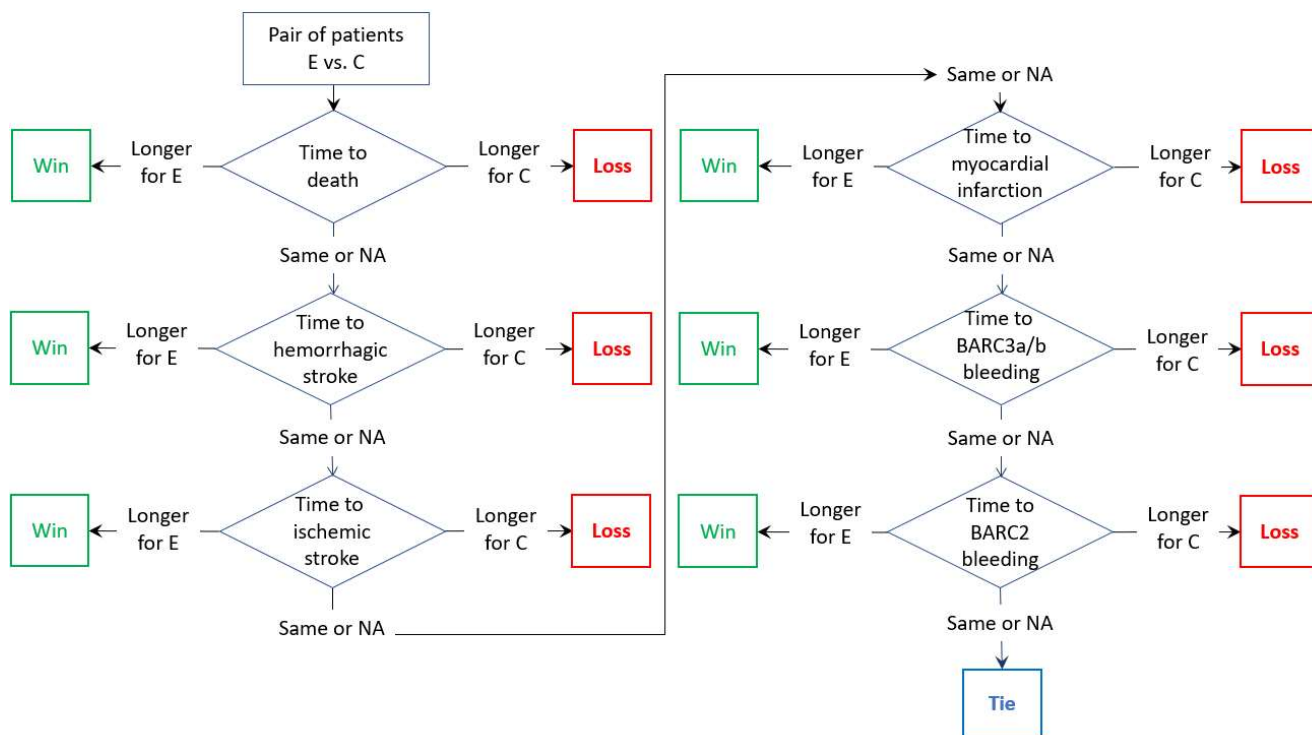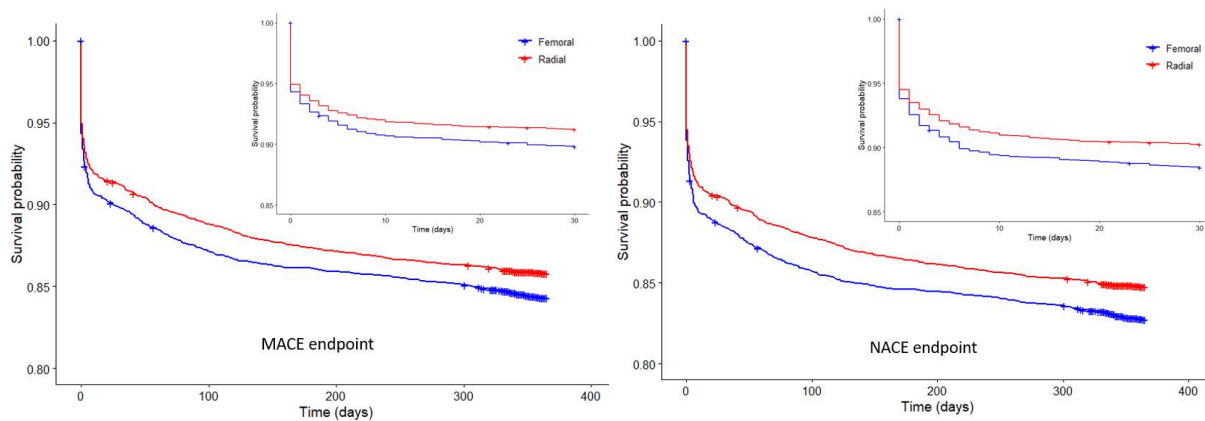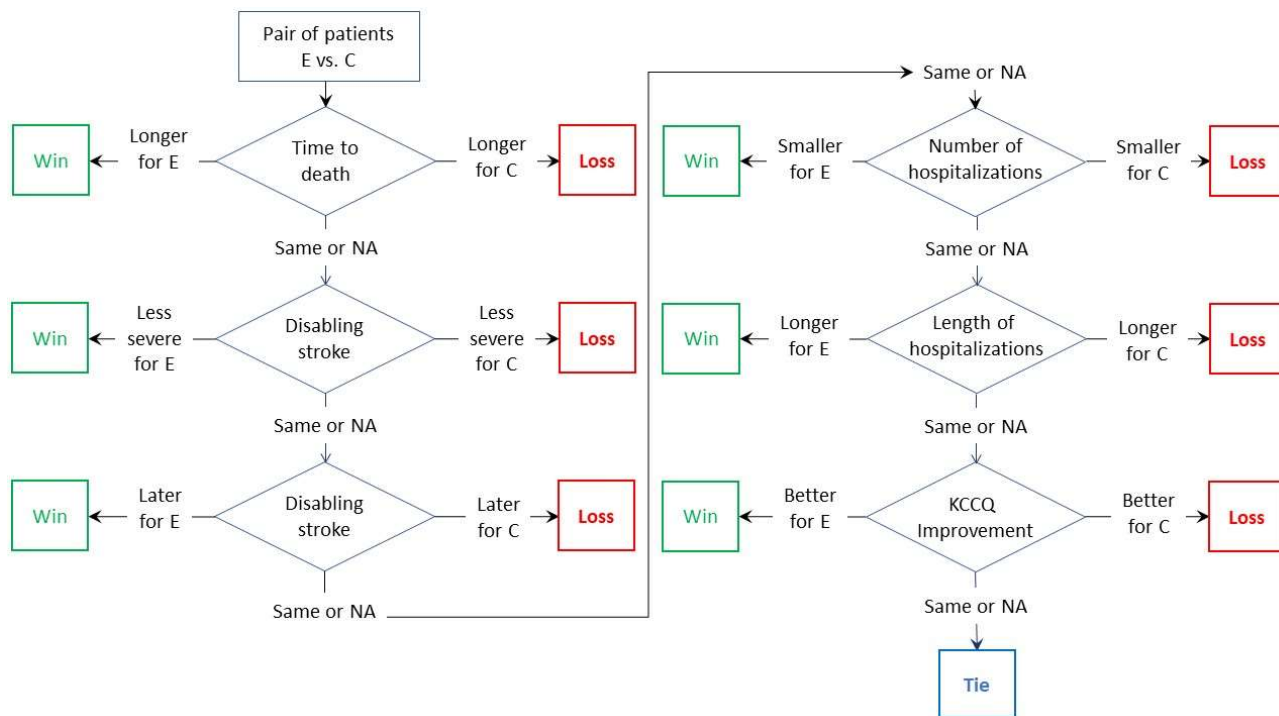
Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Central illustration