jaa Just accepted article

EuroIntervention

**Title: The "A, B, C" of multiple statistical methods for composite endpoints.**

**Authors**: Hironori Hara, M.D; David van Klaveren, PhD; Norihiro Kogame, M.D; Ply Chichareon, M.D; Rodrigo Modolo, M.D; Mariusz Tomaniak, M.D; Masafumi Ono, M.D; Hideyuki Kawashima, M.D; Kuniaki Takahashi, M.D; Davide Capodanno, M.D, PhD; Yoshinobu Onuma, M.D, PhD; Patrick W. Serruys, M.D, PhD

**DOI:** 10.4244/EIJ-D-19-00953

**Citation:** Hara H, van Klaveren D, Kogame N, Chichareon P, Modolo R, Tomaniak M, Ono M, Kawashima H, Takahashi K, Capodanno D, M.D, Onuma Y, M.D, Serruys PW**.** The "A, B, C" of multiple statistical methods for composite endpoints. *EuroIntervention* 2020; Jaa-770, 2020, doi: 10.4244/EIJ-D-19-00953



**Guest Editor:** Adnan Kastrati, M.D, PhD

**Manuscript submission date:** 21 October 2019

**Revisions received:** 04 February 2020, 23 March 2020

**Accepted date:** 23 April 2020

**Online publication date:** 28 April 2020

# The "A, B, C" of multiple statistical methods for composite endpoints

Hironori Hara[1, 2], MD; David van Klaveren[3, 4], PhD; Norihiro Kogame[1], MD;

Ply Chichareon[1], MD; Rodrigo Modolo[1], MD; Mariusz Tomaniak[5, 6], MD;

Masafumi Ono[1, 2], MD; Hideyuki Kawashima[1, 2], MD; Kuniaki Takahashi[1], MD;

Davide Capodanno[7], MD, PhD; Yoshinobu Onuma[2], MD, PhD;

Patrick W. Serruys[2, 8], MD, PhD

Affiliations

1. Department of Cardiology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
2. Department of cardiology, National University of Ireland, Galway (NUIG), Galway, Ireland
3. Department of Public Health, Center for Medical Decision Making, Erasmus MC, Rotterdam, The Netherlands
4. Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, USA
5. Department of Cardiology, Erasmus Medical Center, Erasmus University, Rotterdam, The Netherlands
6. First Department of Cardiology, Medical University of Warsaw, Warsaw, Poland
7. Division of Cardiology, Cardio-Thoraco-Vascular and Transplant Department, CAST, Rodolico Hospital, AOU "Policlinico-Vittorio Emanuele", University of Catania, Catania, Italy
8. NHLI, Imperial College London, London, United Kingdom



**Short running title:** Statistical methods for composite endpoints

**Address for correspondence:**
Patrick W. Serruys, MD. PhD.
Established Professor of Interventional Medicine and Innovation, National University of Ireland, Galway (NUIG), Galway, Ireland
University Road, Galway, H91 TK33, Ireland
E-mail: patrick.w.j.c.serruys@gmail.com

**Conflict of interest**

H.H. is supported by a grant for studying overseas from Japanese Circulation Society and a grant from Fukuda Foundation for Medical Technology. The other authors have no conflicts of interest to declare.

**Narrative Abstract**

Composite endpoints are commonly used in clinical trials, and time-to-first-event analysis has been the usual standard. Time-to-first-event analysis treats all components of the composite endpoint as having equal severity and is heavily influenced by short-term components. Over the last decade, novel statistical approaches have been introduced to overcome these limitations. We reviewed win ratio analysis, competing risk regression, negative binomial regression, Andersen-Gill regression, and weighted composite endpoint (WCE) analysis. Each method has both advantages and limitations.

The advantage of win ratio and WCE analyses is that they take event severity into account by assigning weights to each component of the composite endpoint. These weights should be pre-specified, because they strongly influence treatment effect estimates. Negative binomial regression and Andersen-Gill analyses consider all events for each patient – rather than only the first event – and tend to have more statistical power than time-to-first-event analysis.

Pre-specified novel statistical methods may enhance our understanding of novel therapy when components vary substantially in severity and timing. These methods consider the specific type of patients, drugs, devices, events, and follow-up duration.



**Classifications**

Clinical research; Clinical trials; Training and education

**Abbreviations**

weighted composite endpoint (WCE)

New York Heart Association (NYHA)

Wei-Lin-Weissfeld (WLW)

## Introduction

Composite endpoints are commonly used in clinical trials. Recently, the Academic Research Consortium-2 consensus stated that patient-oriented composite endpoints – the overall cardiovascular outcomes from the patient perspective, including all-cause death, any type of stroke, any myocardial infarction (MI), and any repeat revascularization – should constitute the foundation of novel coronary device or pharmacotherapeutic agent assessment.

Time-to-first-event method has been commonly used for the analysis of composite endpoints, but has the inherent limitation of treating all contributory endpoints as having equal severity and only gives weight to the first endpoint encountered in time. Thus, nonfatal events that occurred earlier have more impact than more serious events such as stroke or death that occur later. Furthermore, death may preclude or render impossible the observation of nonfatal events. Over the last decade, several novel statistical methods have been proposed to overcome these limitations. These methods consider all events occurring until follow-up, incorporate the severity of clinical events, and account for the competing risk nature of different events[1-10].

We aim to review the different statistical methods other than the traditional time-to-first-event analysis, including win ratio analysis, competing risk regression, negative binomial regression, Andersen-Gill regression, and weighted composite endpoint (WCE) analysis (Figure 1).



## Statistical approaches

### 1) Win ratio analysis

Win ratio analysis was introduced by Pocock et al. in 2012 and is a rank-based method, which puts more emphasis on the most clinically important component of the composite endpoints by ranking the constituent components[1]. This analysis requires 4 steps: 1) ranking events by their severity, 2) making patient pairs, 3) decision of a winner in each patient pair, and 4) calculation of win ratio.

First, the components of the composite endpoint are ranked on the basis of their perceived severity. Second, the concept is to match patients with different treatment assignment based on their individual risk estimates. Pocock et al. proposed to estimate a composite risk score for each patient based on pre-selected baseline prognostic factors[2]. Patients in the experimental treatment arm are matched to patients with similar risk score in the control arm on the condition that the follow-up durations do not differ greatly (Figure 2A-1). When the number of patients in the 2 groups differs, some patients are randomly excluded to equalize the number of patients in both groups.

The third step is to decide a winner in each matched patient pair (Figure 2A-2). The comparison of each pair is performed using every type of categorized event, either death, or stroke, or MI, or other event. The events of each patient pair are evaluated to decide whether one had the most severe event (usually death is applied). If this is not the case (both patients were alive at the end of follow-up), the remaining pairs are then evaluated for the occurrence of an event ranked second in severity, and so on for each ranking (third, fourth, or fifth rank). If there were no events until the time of last follow-up, the pair is treated as "tied"[1]. The win ratio emphasizes the more severe components when comparing composite endpoints between 2 groups of patients (Figure 2B).

Fourth, the win ratio is calculated as the number of winners divided by the number of losers, and a 95% confidence interval for the win ratio is easily obtainable[1] (Figure 2A-3). Since matched pairings are influenced by patients who are randomly excluded, it may be necessary to perform analyses repeatedly with different randomly excluded patients. Pocock et al. have described the formulas for these calculations[1] and these calculations do not require special software. In addition, Luo et al. presented a code for R software (R Foundation for Statistical Computing, Vienna, Austria) and this code could be helpful[11].

Win ratio analysis is a rank-based method and could reflect the event severity in the analysis of composite endpoints. Therefore, it is valuable when the components of composite endpoint vary in their clinical severity and importance (e.g. composite endpoint of death, stroke, MI, and revascularization in an ischemic heart disease trial; composite endpoint of cardiovascular death and heart failure hospitalization in a heart failure trial). On the other hand, there are several limitations. Severity ranking of each adverse event affects the result of the composite endpoint and the ranking in itself is debatable without universal consensus (e.g. severity ranking of MI and major bleeding). In addition, it only can be applied to the comparison between 2 groups. An example used in EMPHASIS-HF study, which compared eplerenone (n=1364) and placebo (n=1373) in patients with New York Heart Association (NYHA) class II heart failure and ejection fraction ≦35%, is shown in Figure 2C[1].

Several options for making pairs have been proposed for comparing patients with similar anatomic and physio-pathological background. For example, prognostic scores, such as anatomic SYNTAX score and SYNTAX score II, have been applied, instead of composite relative risk scores[3, 4].

In long-term event-driven trials, patient follow-up durations vary greatly, and many pairs are often categorized as "tied" (Figure 2D). To reduce this problem, patients can be stratified into several follow-up duration categories and patients are matched in strata of similar follow-up duration[1].

When baseline risk factors are not well established, it is more difficult to match patients on the basis of risk. In this case, one can compare every patient in one group with every patient in the other group (unmatched pairs approach)[1, 2].

**2) Competing risk regression**

An event (e.g. non-cardiovascular death) which precludes less severe events or an event (e.g. heart transplant) which changes the possibility to observe events of interest (e.g. congestive

heart failure) are called competing risks (Figure 3A). Competing risk regression method takes these issues into account for composite endpoints and allows disentangling the contribution of an intervention on each type of event. The Fine-Gray model is the most popular model[12]. In this model, patients experiencing competing risk events remain in the risk set for the event of interest until experiencing events of interest or their censoring (Figure 3B, C). This analysis can be easily performed using free statistical software, EZR, and Kanda has described the detailed method[13].

This competing risk within clinical research was first introduced in the field of Oncology. In patients who underwent chemotherapy for cancer, failure events commonly studied are relapse of the cancer and treatment-related death. The interest is to estimate the probability of relapse. In this case, treatment-related death is a competing risk event (which would obviously not allow the investigators to observe any relapse of cancer in dead patients) and competing risk regression analysis is useful[14]. When the age of study population is high, death could be used as a competing risk since the rate of non-treatment related death is relatively high. In the substudy of prosthetic valve endocarditis from the PARTNER trial[15], age of patients was 83 years and death was used as the competing risk event. Incidences of prosthetic valve endocarditis after transcatheter and surgical aortic valve replacement were assessed using this competing risk regression model (Figure 3D). In the field of cardiology, all-cause death often may be less device or procedure specific than deaths adjudicated as cardiovascular death. Non-cardiovascular death could be used as a competing risk, although all-cause death is the most unbiased method to report deaths.

**3) Negative binomial regression**

The traditional time-to-first-event analysis only evaluates the first adverse event and does not capture the subsequent events. However, in the field of cardiology, some adverse events, such as revascularization, bleeding, hospitalization for heart failure, occur repeatedly. Incorporation

of all events are meaningful in terms of the evaluation of patients' quality of life and medical cost. In addition, increase of the number of events could yield additional statistical power. A simple method for the assessment of all adverse events between two groups is to compare the numbers of events.

In a book entitled "The Law of Small Numbers", Bortkiewicz investigated the annual deaths by horse kicks in the Prussian Army from 1875 to 1984, and noted that events with low frequency in a large population follow a Poisson distribution even when the probability varies (Supplementary figure 1A). The Poisson distribution has commonly been used to model the number of events in an interval of time (Supplementary figure 1A). The variance of clinical events in a trial is usually greater than the mean (Supplementary figure 1B). In other words, the distribution of the number of clinical events is better represented by an over-dispersed Poisson distribution. The negative binomial distribution is often used for modeling over-dispersed Poisson data. Negative binomial regression analysis has been used to estimate treatment effect in terms of the rate ratio of a composite endpoint[5-8](Figure 4A) and is valuable especially in high risk population since patients tend to experience repeated adverse events. For this analysis, "glm.nb" function from the "MASS" package in R software could be helpful[16]. In the PARADIGM-HF trial[7], the primary endpoint (a composite of cardiovascular death or hospitalization for congestive heart failure) was analyzed by a negative binomial regression analysis (Figure 4B). On the other hand, this analysis considers only the total account of events per patient. Therefore, the same follow-up duration should be applied per patient, which sometimes restricts the application of this method.

**4) Cox-based models for recurrent events**

Negative binomial regression analysis is not applicable if the follow-up duration differs from patient to patient. To overcome this limitation, several time-to-event methods have been proposed for the analysis of repeated events. The Andersen-Gill model is a simple extension

of the traditional Cox model and is based on a gap-time approach, in which the clock is reset after an event and the patient is at risk for the next event. This analysis assumes that the risk of an event is not affected by whether another event has already occurred[3, 4, 8]. The Wei-Lin-Weissfeld (WLW) model is different from the Andersen-Gill model in that it uses the time from study entry to the first, second and subsequent events (Figure 5A)[7, 8]. In the WLW model, each time-ordered event is analyzed on its own time-to-event basis, that is, for the 1st events in each patient, the 2nd events in each patient, the 3rd events in each patient, and so on. For these analyses, "coxph" function from the "survival" package in R software could be helpful[17]. These analyses consider all adverse events and time to events. Therefore, these analyses are valuable in high risk population like the negative binomial regression analysis. In addition, these analyses are applicable regardless of the follow-up duration of each patient. On the other hand, this methodological approach treats all adverse events as having equal severity, and severe adverse events, such as death, could be underestimated as well as time-to-first-event analysis. In the REDUCE-IT trial, the primary endpoint (a composite of cardiovascular death, MI, stroke, revascularization, or hospitalization for unstable angina) – including recurrent events – was analyzed using the Andersen-Gill and the WLW approaches (Figure 5B)[8].

**5) Weighted composite endpoint (WCE)**

The WCE methodology extends the standard time-to-event methodology by determining a weight for each of nonfatal events (event severity) and incorporating all adverse events into the analysis (recurrent events)[3, 4, 9, 10]. The WCE analysis requires 4 steps: 1) Decision of event weights, 2) Calculation of residual weight at the end of each day in each patient, 3) Creation of a modified life table with a weighted number of patients at risk, and 4) Comparison of groups (Figure 6A).

In the field of cardiovascular disease, two sets of event weights have been used[9, 10]. The first set gives a weight of 1.0 to death, 0.47 to stroke, 0.38 to MI, and 0.25 to target vessel

revascularization[3, 4], and in the second set, death has a weight of 1.0, shock has a weight of 0.5, congestive heart failure has a weight of 0.3, re-MI has a weight of 0.2, and re-ischemia has a weight of 0.1. These weights were decided based on Delphi panels to achieve consensus between clinician-investigators. Delphi panel is a panel of experts to achieve consensus in solving a problem or deciding the most appropriate strategy based on the results of multiple rounds of questionnaires.

For calculation of residual weights at each time point, each patient starts with a weight of 1.0, which remains unaltered if no event occurs until end of follow-up (Figure 6A-2-a). Nonfatal events reduce the residual weight of a patient by the weight of the event (Figure 6A-2-b, c, d). From the individual patient data, a modified life table with a weighted number of patients at risk is created, providing estimates of weighted event rates in each group and of a weighted hazard ratio for the reference group (Figure 6A-3). The WCE method allows the incorporation of repeated events in a single patient and distinguishes between the severity of components of the composite endpoint. Indication for this method is the same as that for time-to-first-event analysis and a representative analysis of this WCE in the DELTA registry[3] is shown in Figure 6B. This approach may better reflect all event information, but evidently depends on assigned event weights. Furthermore, weighing events reduces the number of effective events. Therefore, WCE could limit power and requires a larger sample size, although statistical power largely depends on severe outcomes, such as death[18]. To date, commercial statistical software do not support this analysis and there is no R package for this analysis in the Comprehensive R Archive Network or Bioconductor. Therefore, this analysis needs dedicated program.

**Comparison of methods; How do we treat a composite endpoint?**

The differences in dealing with composite endpoints are shown in Figure 7. These statistical methods have recently been applied to several clinical trials in the field of cardiology (Figure

8, 9). The estimated treatment effect, using multiple statistical methods, showed similar tendencies, but, as expected, the significance of the treatment effect estimates was dependent on the statistical method used in trials. The negative binomial regression and the Andersen-Gill analyses tended to have more statistical power than time-to-first-event analysis, while the statistical power of the WCE method tended to be low. In particular, WCE did not demonstrate a significant difference between treatments (Figure 8), in contrast with time-to-first-event analyses.

The method of counting a "series of events" has to be defined in detail for analyses using all adverse events[19]. Whenever a revascularization is performed on the same day as MI, the number of serial events would depend on the methodological definition. Two events (MI and revascularization) occurring on the same day could even be counted as one event[3, 8]. Therefore, the method of event counting could affect the result.

The win ratio and WCE analyses depend on the severity ranking and weighting of events severity, which may induce arbitrariness of the comparison. On the other hand, a universal ranking is not appropriate because the event severity may depend on patient characteristics. For example, the impact of revascularization is different in the patients with and without a history of percutaneous coronary intervention. The way to determine event severity should be discussed in future trials. Pre-specification of weights is necessary to avoid any arbitrariness.



## Conclusion

All methods for the analysis of composite end points have strengths and weaknesses (95 10). Pre-specified novel statistical methods may enhance our understanding when components vary substantially in severity and timing. These methods should consider the specific type of patients, drugs, devices, events, and follow-up duration.

**Funding**

None.

**References**

1. 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):176-82.
2. Milojevic M, Head SJ, Andrinopoulou ER, Serruys PW, Mohr FW, Tijssen JG, Kappetein AP. Hierarchical testing of composite endpoints: applying the win ratio to percutaneous coronary intervention versus coronary artery bypass grafting in the SYNTAX trial. EuroIntervention 2017;**13**(1):106-114.
3. Capodanno D, Gargiulo G, Buccheri S, Chieffo A, Meliga E, Latib A, Park SJ, Onuma Y, Capranzano P, Valgimigli M, Narbute I, Makkar RR, Palacios IF, Kim YH, Buszman PE, Chakravarty T, Sheiban I, Mehran R, Naber C, Margey R, Agnihotri A, Marra S, Leon MB, Moses JW, Fajadet J, Lefevre T, Morice MC, Erglis A, Alfieri O, Serruys PW, Colombo A, Tamburino C, Investigators D. Computing Methods for Composite Clinical Endpoints in Unprotected Left Main Coronary Artery Revascularization: A Post Hoc Analysis of the DELTA Registry. JACC Cardiovasc Interv 2016;**9**(22):2280-2288.
4. Bakal JA, Roe MT, Ohman EM, Goodman SG, Fox KA, Zheng Y, Westerhout CM, Hochman JS, Lokhnygina Y, Brown EB, Armstrong PW. Applying novel methods to assess clinical outcomes: insights from the TRILOGY ACS trial. Eur Heart J 2015;**36**(6):385-92a.
5. Rogers JK, McMurray JJ, Pocock SJ, Zannad F, Krum H, van Veldhuisen DJ, Swedberg K, Shi H, Vincent J, Pitt B. Eplerenone in patients with systolic heart failure and mild symptoms: analysis of repeat hospitalizations. Circulation 2012;**126**(19):2317-23.
6. Rogers JK, Pocock SJ, McMurray JJ, Granger CB, Michelson EL, Ostergren J, Pfeffer MA, Solomon SD, Swedberg K, Yusuf S. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. Eur J Heart Fail 2014;**16**(1):33-40.
7. Mogensen UM, Gong J, Jhund PS, Shen L, Kober L, Desai AS, Lefkowitz MP, Packer M, Rouleau JL, Solomon SD, Claggett BL, Swedberg K, Zile MR, Mueller-Velten G, McMurray JJV. Effect of sacubitril/valsartan on recurrent events in the Prospective comparison of ARNI with ACEI to Determine Impact on Global Mortality and morbidity in Heart Failure trial (PARADIGM-HF). Eur J Heart Fail 2018;**20**(4):760-768.
8. Bhatt DL, Steg PG, Miller M, Brinton EA, Jacobson TA, Ketchum SB, Doyle RT, Jr., Juliano RA, Jiao L, Granowitz C, Tardif JC, Gregson J, Pocock SJ, Ballantyne CM, Investigators R-I. Effects of Icosapent Ethyl on Total Ischemic Events: From REDUCE-IT. J Am Coll Cardiol 2019;**73**(22):2791-2802.
9. Armstrong PW, Westerhout CM, Van de Werf F, Califf RM, Welsh RC, Wilcox RG, Bakal JA. Refining clinical trial composite outcomes: an application to the Assessment of the Safety and Efficacy of a New Thrombolytic-3 (ASSENT-3) trial. Am Heart J 2011;**161**(5):848-54.
10. Bakal JA, Westerhout CM, Cantor WJ, Fernandez-Aviles F, Welsh RC, Fitchett D, Goodman SG, Armstrong PW. Evaluation of early percutaneous coronary intervention vs. standard therapy after fibrinolysis for ST-segment elevation myocardial infarction: contribution of weighting the composite endpoint. Eur Heart J 2013;**34**(12):903-8.
11. Luo X, Tian H, Mohanty S, Tsai WY. An alternative approach to confidence interval estimation for the win ratio statistic. Biometrics 2015;**71**(1):139-145.

12. Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondre K, Heinze G. Competing risks analyses: objectives and approaches. Eur Heart J 2014;**35**(42):2936-41.
13. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant 2013;**48**(3):452-8.
14. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant 2007;**40**(4):381-7.
15. Summers MR, Leon MB, Smith CR, Kodali SK, Thourani VH, Herrmann HC, Makkar RR, Pibarot P, Webb JG, Leipsic J, Alu MC, Crowley A, Hahn RT, Kapadia SR, Tuzcu EM, Svensson L, Cremer PC, Jaber WA. Prosthetic Valve Endocarditis After TAVR and SAVR: Insights From the PARTNER Trials. Circulation 2019;**140**(24):1984-1994.
16. Venables WN, Ripley BD, Venables WN. *Modern applied statistics with S*. 4th ed. New York: Springer; 2002.
17. Therneau TM, Grambsch PM. *Modeling survival data : extending the Cox model*. New York: Springer; 2000.
18. Bakal JA, Westerhout CM, Armstrong PW. Impact of weighted composite compared to traditional composite endpoints for the design of randomized controlled trials. Stat Methods Med Res 2015;**24**(6):980-8.
19. Granger CB, Nelson AJ, Pagidipati NJ. Risk of Total Events With Icosapent Ethyl: Can We Reduce It? J Am Coll Cardiol 2019;**73**(22):2803-2805.

**Figure legends**

**Figure 1. Decision tree for statistical models**

WLW, Wei-Lin-Weissfeld; WCE, weighted composite endpoint.

**Figure 2. Win ratio**

(A) Flow chart for analysis. (A-1) Adjustment of each group. When there are slightly unequal sample sizes in Groups A (n=417) and B (n=419), respectively, 2 patients (*) are randomly excluded from Group B to equalize the number of patients. The patients are arranged and tabulated based on the decreasing ranking of their relative risk scores. (A-2) Patients' level assessment. Winners and losers are decided based on event severity within the censoring period. In the condition that decreasing ranking of event severity is death, stroke and myocardial infarction (MI), decisions in each pair are as follows.

(Pair 1) Death is the most severe event, so the patient figuring in the upper line is a loser.

(Pair 2) A death does not occur in neither patient. The event of stroke should be evaluated because stroke is more severe than MI but less than death, and the patient figuring in the upper line is a loser.

(Pair 3) A death occurs after the others' follow-up time, so the times to MI in absence of death or stroke occurrence should be compared. The upper line patient is a loser.

(Pair 4) An MI occurs after the others' follow-up time, and there are no events until censoring. Therefore, a winner and a loser are not established and we have a tie.

(A-3) Group assessment. The win ratio is provided by (total number of winners) / (total number of losers). See example: 1.30 (= (12+34+62) / (7+21+55)).

(B) The events used are different between the win ratio and traditional time-to-first-event analyses.

(C) The application of win ratio analysis in EMPHASIS-HF study.

(D) Time stratified approach. Whenever patient follow-up durations vary greatly, patients can be stratified into some follow-up duration categories (e.g. long follow-up group and short follow-up group) and pairs are matched in each category based on the decreasing ranking of each patient relative risk score.

NYHA, New York Heart Association; EF, ejection fraction; HF, heart failure; CV, cerebrovascular.



**Figure 3. Competing risks**

(A) Non-cardiovascular (CV) death precludes the possible subsequent events of congestive heart failure (CHF), and heart transplant changes the possibility of CHF occurrence. Therefore, these events are called competing risks.

(B) Flow chart for analysis. (B-1) Each group. (B-2) Patients' level assessment. In Fine-Gray model, patients experiencing competing risk events (e.g. heart transplant) remain in the risk set

for the event of interest (e.g. CHF) until either experiencing events of interest or their censoring. (B-3) Group assessment. From patient at risk and number of events, cumulative event rate is calculated. When we compare groups, the result is presented as hazard ratio and p-value.

(C) Flow chart for analysis. (C-1) Each group. (C-2) Patients' level assessment. In the case that competing risk event is a non-CV death, a competing risk event (non-CV death) is treated as a censoring because death isn't an event of interest (CHF) and also means the end of follow-up. (C-3) Group assessment.

(D) Application of competing risk regression to the subanalysis of PARTNER trial.

TAVR, transcatheter aortic valve replacement; SAVR, surgical aortic valve replacement.

**Figure 4. Negative binomial regression analysis**

(A) Flow chart for analysis. (A-1) Each group. (A-2) Patients' level assessment. Number of events is counted in each patient. (A-3) Group assessment. Negative binomial regression is a statistical method for the analysis of over-dispersed date. The comparison between groups is shown as rate ratio and p-value.

(B) Application of negative binomial regression to the PARADIGM-HF trial.

NYHA, New York Heart Association; EF, ejection fraction; CV, cerebrovascular; CHF, congestive heart failure; RR, rate ratio.

**Figure 5. Comparison of time-to-first-event, Andersen-Gill, and Wei-Lin-Weissfeld (WLW) methods**

(A) Flow chart for analysis. (A-1) Each group. (A-2) Patients' level assessment. (A-3) Group assessment. The time-to-first-event analysis uses only the first event and time to the first event. In this example, 2 step downs according to the first events in "patient 1" and "patient 2" are

shown in the Kaplan-Meier curve. In Andersen-Gill analysis, all events and the times between consecutive events (gap-time approach) are used. Five step downs according to 2 events in "patient 1" and 3 events in "patient 2" are demonstrated in this modified Kaplan-Meier curve. In WLW method, the analyses for the 1st events in each patient (e.g. 2 events in "patient 1 and 2"), the 2nd events in each patient (e.g. 2 events in "patient 1 and 2"), the 3rd events in each patient (e.g. 1 event in "patient 2"), and so on (e.g. the 4th event did not occur), are performed. When we compare groups, results are presented as hazard ratios and p-values.

(B) Application of time-to-first-event, Andersen-Gill, and WLW methods to the REDUCE-IT trial.

CV, cerebrovascular; MI, myocardial infarction; HR, hazard ratio.



**Figure 6. Weighted composite endpoint (WCE)**

(A) Flow chart for analysis. In this explanation, event weights of death, stroke, and myocardial infarction (MI) are assigned as 1.0, 0.47, and 0.38, respectively. (A-1) Each group. (A-2) Patients' level assessment. The residual weights and event weights in each patient are calculated as follows.

(a) No events occur at follow-up, a weight of 1.0 remains unaltered.

(b) A patient with a myocardial infarction on Day 1 and a non-disabling stroke on Day 11 has a cumulative weighting of 0.3286 = 1 – [(1 – 0.38) x (1 – 0.47)]. When the patient suffers the second stroke on Day 30, the patient has a residual weighting of 0.174158 = 0.3286 x (1 – 0.47).

(c) If a death is the only event, a weight of 1.0 is lost for a death event.

(d) If there is an event before a death, the residual weight is lost for a death.

(A-3) Group assessment.

(a) Calculation of weighted number of patients at risk (residual weight) and cumulative weighted event free rate.

(b) The table is an example when the number of patients is 20. A modified life table including weighted number of patients at risk and cumulative weighted event free rate is created by each patient data.

(B) Application of WCE method to the DELTA registry. This figure is reproduced with permission from Ref 3.

LMCA, left main coronary artery disease; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; MACCE, major adverse cardiac or cerebrovascular events; CI, confidence interval.

**Figure 7. The differences in dealing with composite endpoints**

MI, myocardial infarction; WLW, Wei-Lin-Weissfeld.

**Figure 8. Application of multiple statistical methods to cardiovascular disease trials**

Weight 1: death, 1.0; CVA or stroke, 0.47; MI, 0.38; TVR, 0.25. Weight 2: death, 1.0; severe stroke, 0.82; moderate stroke, 0.47; mild stroke, 0.23; severe MI, 0.59; moderate MI, 0.38; mild MI, 0.17. Weight 3: death, 1.0; shock, 0.5; CHF, 0.3; Re-MI, 0.2; RI, 0.1.

RCT, randomized control trial; LMCA, left main coronary artery disease; UA, unstable angina; NSTEMI, non-ST-segment elevation myocardial infarction; 3VD, 3 vessel disease; STEMI, ST-segment elevation myocardial infarction; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting; UFH, unfractionated heparin; CVA, cerebrovascular accident; MI, myocardial infarction; TVR, target vessel revascularization; Re-MI, recurrent myocardial infarction; RI, recurrent ischemia; CV, cardiovascular; TTE, time to first event; WR, win ratio; AG, Andersen-Gill; WCE, weighted composite endpoint; CR, competing risk; NB, negative

binomial; GRACE, Global Registry of Acute Cardiac Events; HR, hazard ratio; RR, rate ratio; CI, confidence interval; NA; not available.

**Figure 9. Application of multiple statistical methods to congestive heart failure trials**

RCT, randomized control trial; NYHA, New York Heart Association; EF, ejection fraction; CHF, congestive heart failure; ACE-I, Angiotensin-converting enzyme-inhibitor; CVA, cerebrovascular accident; IR, investigator reported; TTE, time to first event; WR, win ratio; NB, negative binomial; AG, Andersen-Gill; WLW, Wei-Lin-Weissfeld; HR, hazard ratio; RR, rate ratio; CI, confidence interval.

**Figure 10. Characteristics of statistical models and statistical power compared to time-to first event analysis**

WLW, Wei-Lin-Weissfeld; WCE, weighted composite endpoint.

Figure 1

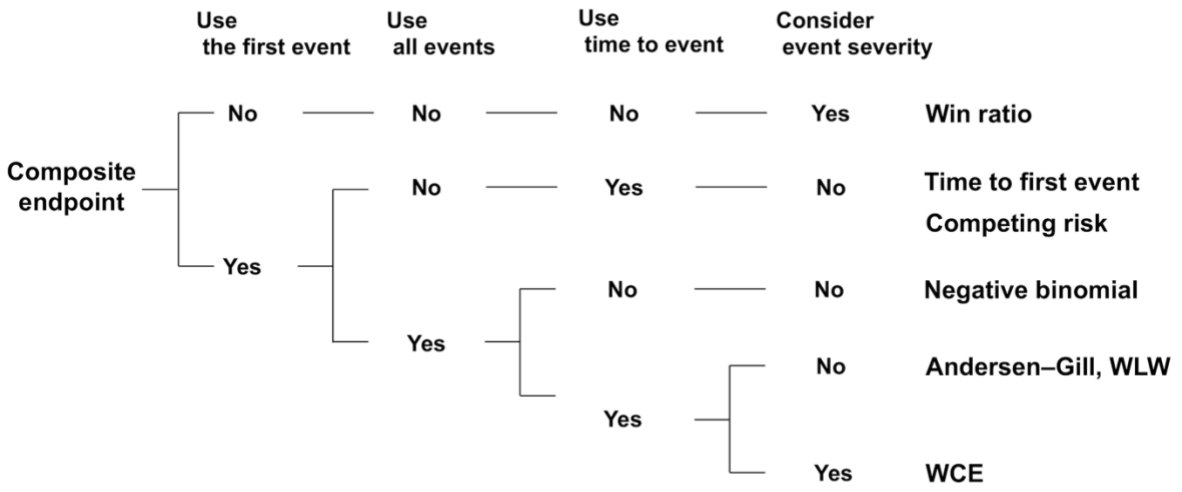


Figure 2AB

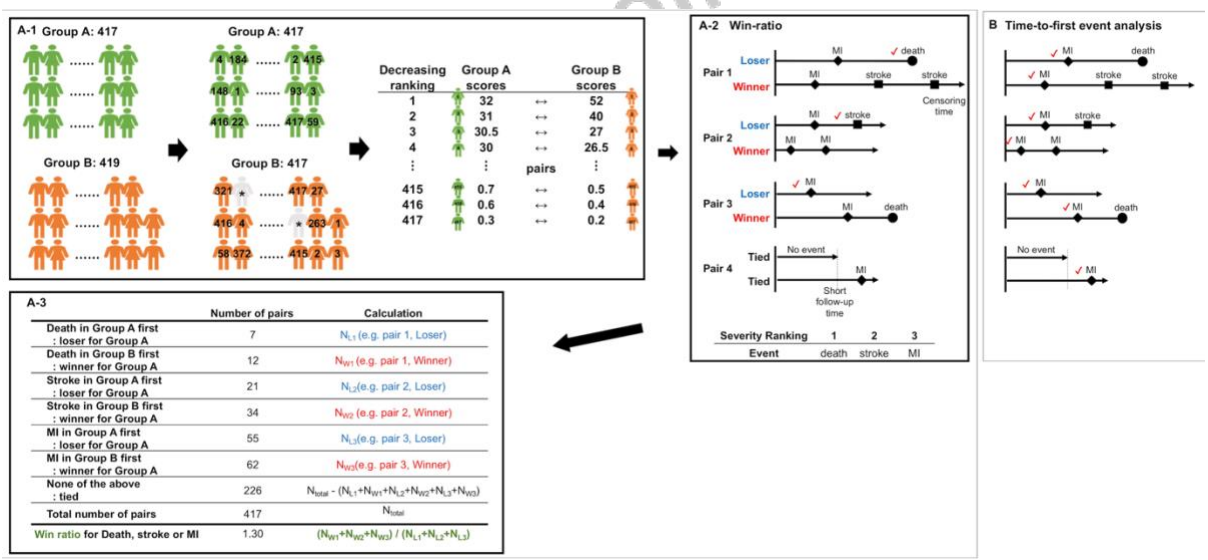

Figure 2CD

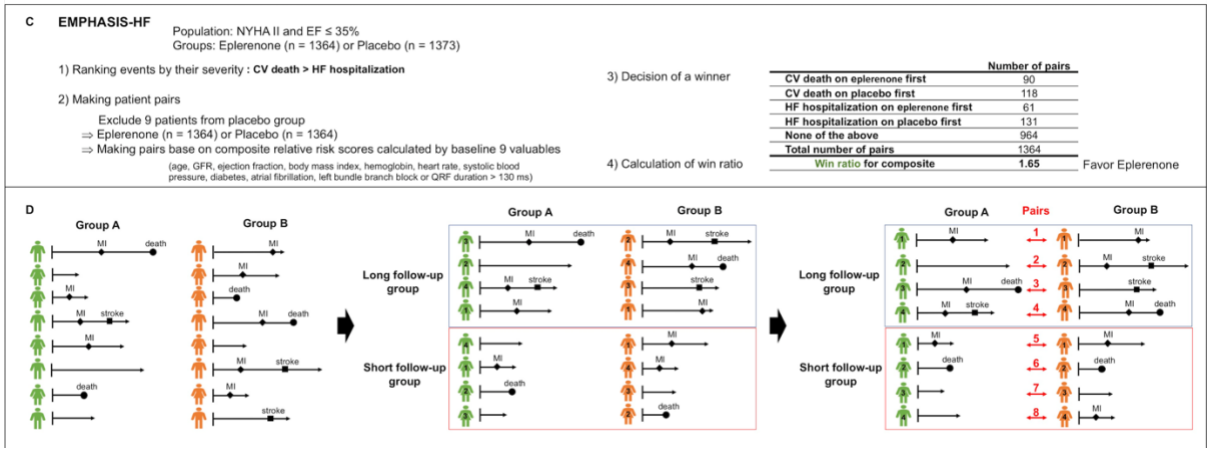


Figure 3

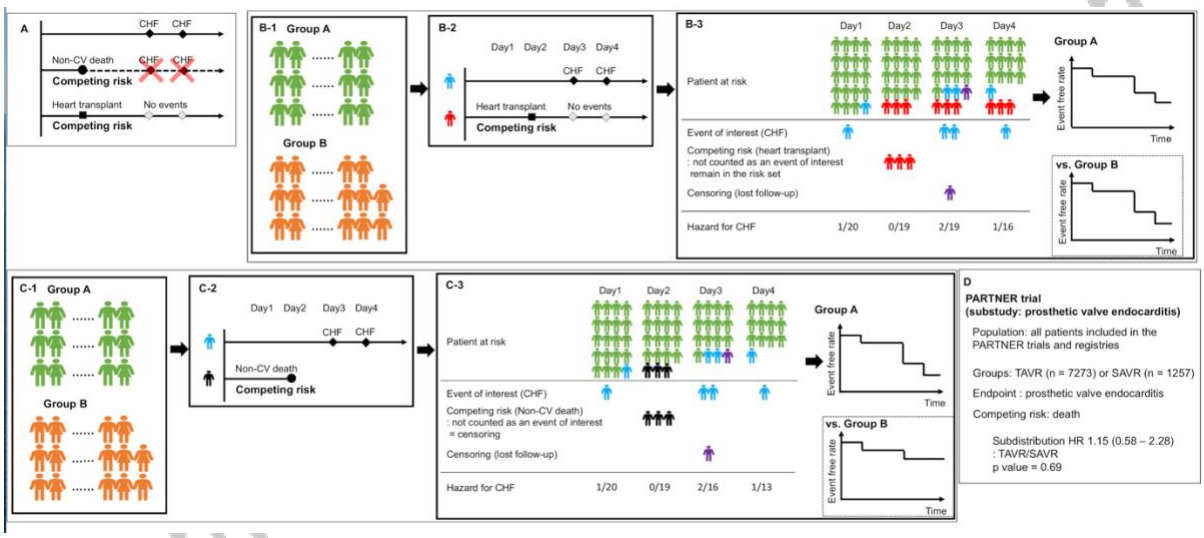


Figure 4

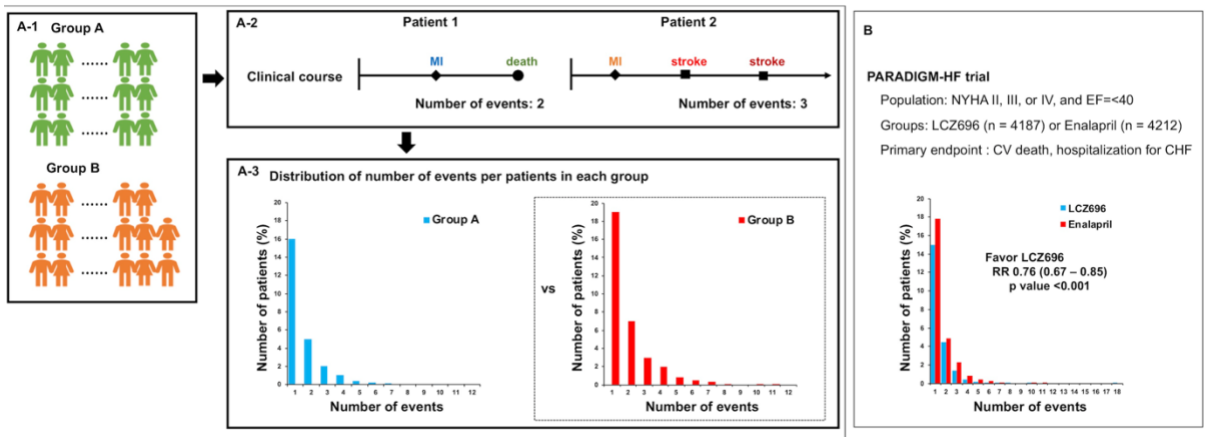

Figure 5

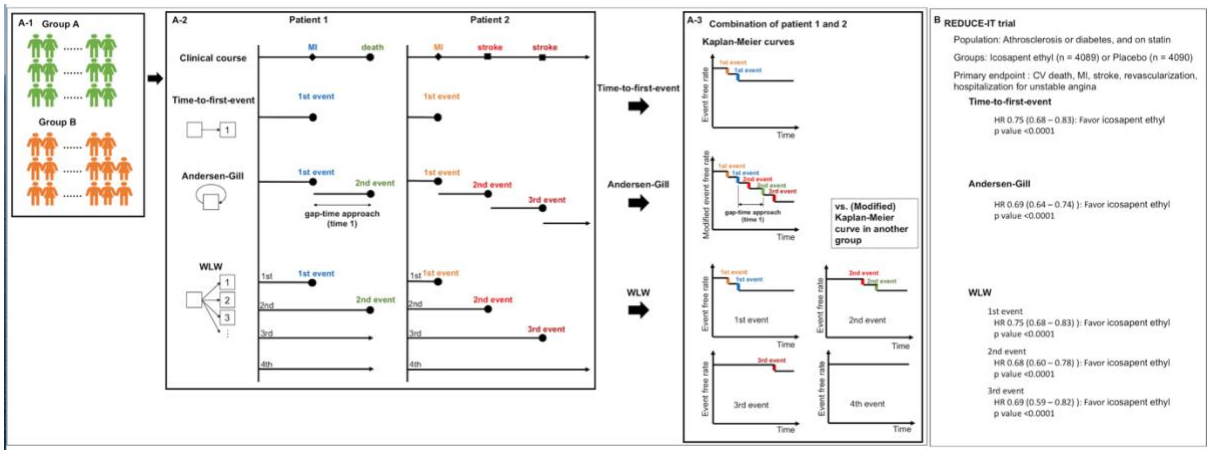


Figure 6A

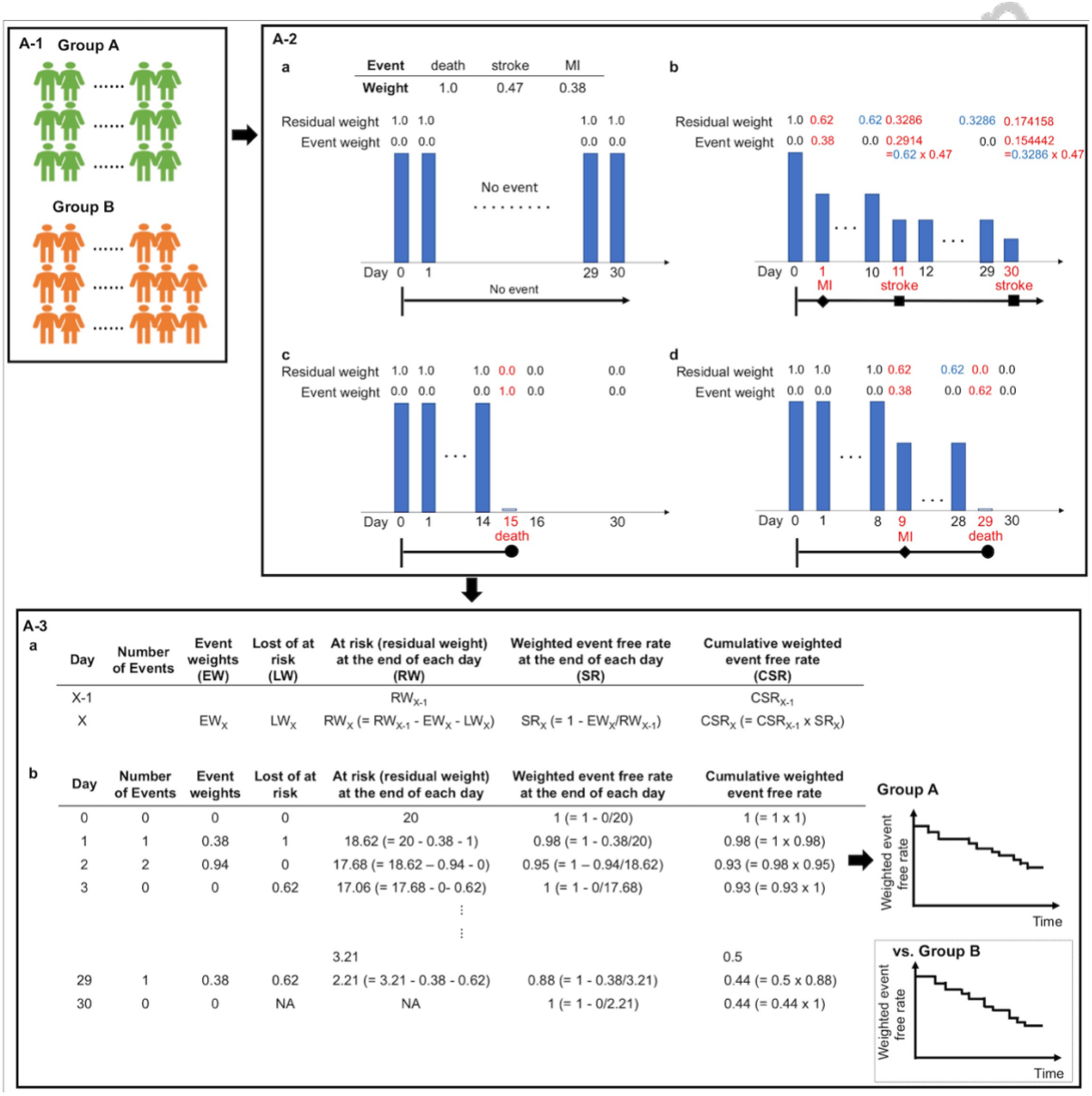

Figure 6B

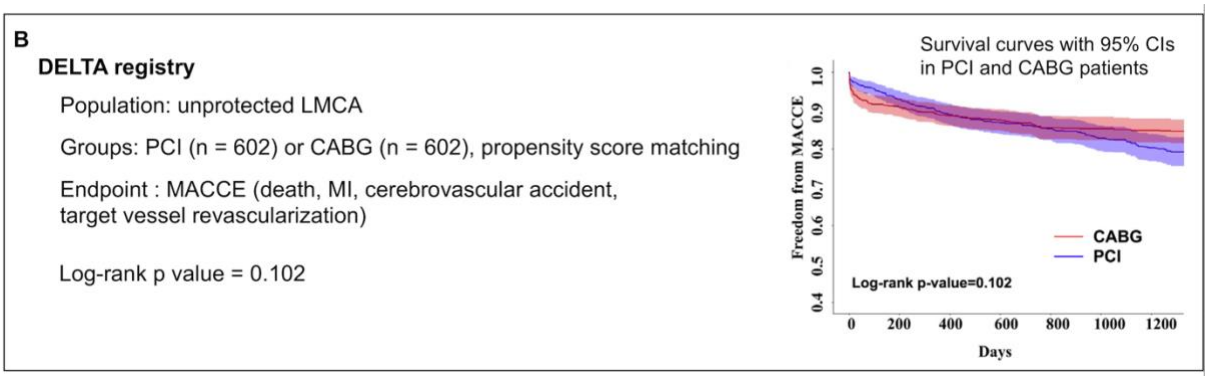


Figure 7

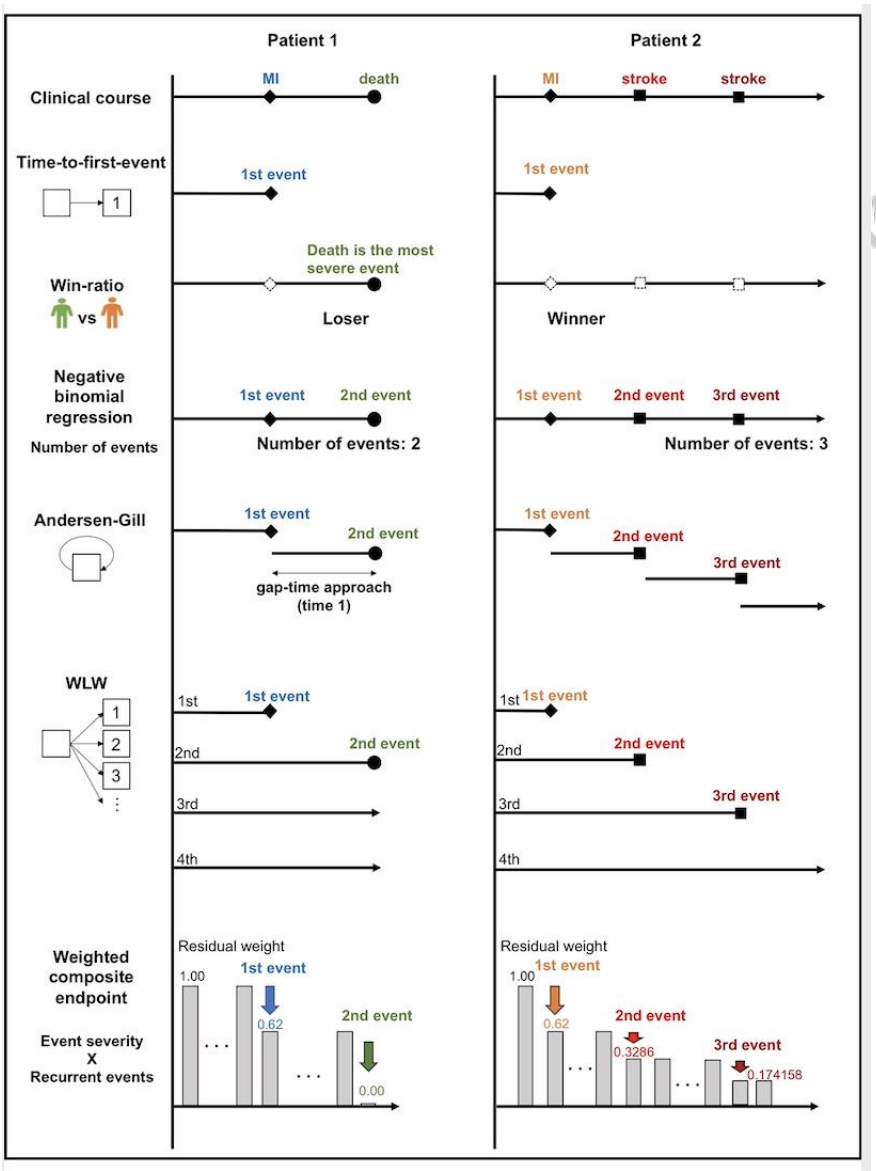

Figure 8

| Trial | Population | Treatment | follow up | Composite endpoints | Number of patients | Statistical method | HR, RR, Invese of WR (95% CI) | p-value |
|---|---|---|---|---|---|---|---|---|
| **DELTA (registry)**[3] | unprotected LMCA | PCI or CABG | 3 years (median) | Death, CVA, MI | 602 vs 602 | TTE propensity score matching | 0.91 (0.66 - 1.26) | 0.57 |
| | | | | | 602 vs 602 | WR propensity score matching | 0.96 (0.72 - 1.30) | 0.82 |
| | | | | | 602 vs 602 | AG propensity score matching | 1.01 (0.76 - 1.36) | 0.93 |
| | | | | | 602 vs 602 | WCE propensity score matching, weight 1 | NA | 0.80 |
| | | | | Nonfatal CVA, MI | 602 vs 602 | CR propensity score matching | 0.85 (0.46 - 1.58) | 0.61 |
| | | | | Death, CVA, MI, TVR | 602 vs 602 | TTE propensity score matching | 1.35 (1.03 - 1.76) | 0.03 |
| | | | | | 602 vs 602 | WR propensity score matching | 1.33 (1.04 - 1.72) | 0.03 |
| | | | | | 602 vs 602 | AG propensity score matching | 1.41 (1.12 - 1.79) | 0.04 |
| | | | | | 602 vs 602 | WCE propensity score matching, weight 1 | NA | 0.10 |
| | | | | Nonfatal CVA, MI, TVR | 602 vs 602 | CR propensity score matching | 1.89 (1.33 - 2.68) | <0.0001 |
| **TRILOGY ACS (RCT)**[4] | UA/NSTEMI without planned | prasugrel or clopidogrel | 2.5 years | Death, stroke, MI | 4663 vs 4663 | TTE | 0.96 (0.86 - 1.06) | >0.05 |
| | | | | | 4663 vs 4663 | WR GRACE Risk Score, time stratified | 0.95 (0.85 - 1.06) | >0.05 |
| | | | | | 4663 vs 4663 | AG | 0.86 (0.72 - 0.97) | <0.05 |
| | | | | | 4663 vs 4663 | WCE weight 1 or 2 | NA | >0.05 |
| **SYNTAX (RCT)**[2] | de novo LMCA or 3VD | PCI or CABG | 5 years | Death, stroke, MI | 903 vs 897 | TTE | 1.29 (1.04 - 1.62) | 0.03 |
| | | | | | 880 vs 880 | WR HR matching | 1.39 (1.10 - 1.77) | 0.006 |
| | | | | | 903 vs 897 | WR unmatched | 1.28 (1.11 - 1.53) | <0.05 |
| | | | | Death, stroke, MI, revascularization | 903 vs 897 | TTE | 1.43 (1.21 - 1.70) | <0.0001 |
| | | | | | 880 vs 880 | WR HR matching | 1.61 (1.34 - 1.96) | <0.0001 |
| | | | | | 903 vs 897 | WR unmatched | 1.49 (1.23 - 1.79) | <0.05 |
| **ASSENT-3 (RCT)**[9] | STEMI | enoxaparin, abciximab, or UFH | 30 days | Death, shock, CHF, Re-MI | 2040 vs 2017 vs 2038 | TTE | NA | <0.05 |
| | | | | | 2040 vs 2017 vs 2038 | WCE weight 3 | NA | 0.2 |
| **TRANSFER-AMI (RCT)**[10] | STEMI and receiving fibrinolysis | early PCI or standard therapy | 30 days | Death, shock, CHF, Re-MI, RI | 537 vs522 | TTE | 0.64 (0.47 - 0.87) | 0.004 |
| | | | | | 537 vs522 | WCE weight 3 | NA | 0.54 |
| **TRANSFER-AMI, WEST, GRACIA-1 (RCT)**[10] | STEMI and receiving fibrinolysis | early PCI or standard therapy | 30 days | Death, shock, CHF, Re-MI | 883 vs897 | TTE | NA | 0.01 |
| | | | | | 883 vs897 | WCE weight 3 | NA | 0.44 |
| **REDUCE-IT (RCT)**[8] | Athrosclerosis or diabetes, and on statin | icosapent ethyl or placebo | 4.9 years (median) | CV death, MI, stroke, revascularization, hospitalization for UA | 4089 vs 4090 | TTE | 0.75 (0.68 - 0.83) | <0.0001 |
| | | | | | 4089 vs 4090 | NB | 0.70 (0.62 - 0.78) | <0.0001 |
| | | | | | 4089 vs 4090 | AG with cluster-robust standard errors | 0.69 (0.64 - 0.77) | <0.0001 |
| | | | | CV death, MI, stroke | 4089 vs 4090 | TTE | 0.74 (0.65 - 0.83) | <0.0001 |
| | | | | | 4089 vs 4090 | NB | 0.72 (0.63 - 0.82) | <0.0001 |
| | | | | | 4089 vs 4090 | AG with cluster-robust standard errors | 0.72 (0.63 - 0.82) | <0.0001 |

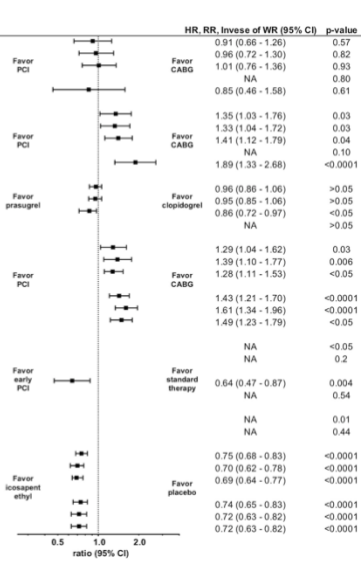


Figure 9

| Trial | Population | Treatment | follow up | Composite endpoints | Number of patients | Statistical method | HR, RR, Invese of WR (95% CI) | p-value |
|---|---|---|---|---|---|---|---|---|
| **EMPHASIS-HF (RCT)**[1,5] | NYHA II and EF ≤ 35% | eplerenone or placebo | 3.1 years (median) | CV death, hospitalization for CHF | 1364 vs 1373 | TTE | 0.63 (0.54 - 0.74) | <0.0001 |
| | | | | | 1364 vs 1364 | WR HR matching | 0.61 (0.49 - 0.74) | <0.0001 |
| | | | | | 1364 vs 1364 | WR HR matching, time stratified | 0.58 (0.48 - 0.70) | <0.0001 |
| | | | | | 1364 vs 1373 | WR unmatched | 0.62 (0.53 - 0.73) | <0.0001 |
| | | | | hospitalization for CHF | 1364 vs 1373 | TTE | 0.63 (0.53 - 0.76) | <0.05 |
| | | | | | 1364 vs 1373 | NB | 0.53 (0.42 - 0.66) | <0.0001 |
| **CHARM Added (RCT)**[1,6] | CHF, EF < 40%, and on ACE-I | candesartan or placebo | 3.1 years (median) | CV death, hospitalization for CHF | 1276 vs 1272 | TTE | 0.85 (0.75 - 0.96) | 0.011 |
| | | | | | 1272 vs 1272 | WR HR matching | 0.77 (0.67 - 0.88) | <0.0001 |
| | | | | CV death, IR hospitalization for CHF | 1514 vs 1509 | TTE | 0.83 (0.74 - 0.94) | 0.003 |
| | | | | | 1514 vs 1509 | NB | 0.75 (0.62 - 0.91) | 0.003 |
| **CHARM Alternative (RCT)**[1,6] | CHF, EF < 40%, and intolerant to ACE-I | candesartan or placebo | 3.1 years (median) | CV death, hospitalization for CHF | 1013 vs 1015 | TTE | 0.77 (0.67 - 0.89) | 0.0004 |
| | | | | | 1013 vs 1013 | WR HR matching | 0.70 (0.59 - 0.83) | <0.0001 |
| | | | | CV death, IR hospitalization for CHF | 1514 vs 1509 | TTE | 0.77 (0.67 - 0.89) | <0.001 |
| | | | | | 1514 vs 1509 | NB | 0.65 (0.51 - 0.82) | <0.001 |
| **CHARM Preserved (RCT)**[1,6] | CHF, EF ≥ 40% | candesartan or placebo | 3.1 years (median) | CV death, hospitalization for CHF | 1514 vs 1509 | TTE | 0.89 (0.77 - 1.03) | 0.118 |
| | | | | | 1509 vs 1509 | WR HR matching | 0.85 (0.72 - 1.01) | 0.065 |
| | | | | CV death, IR hospitalization for CHF | 1514 vs 1509 | TTE | 0.86 (0.74 - 1.00) | 0.050 |
| | | | | | 1509 vs 1509 | WR HR matching | 0.85 (0.71 - 1.00) | 0.049 |
| | | | | | 1514 vs 1509 | WR unmatched | 0.86 (0.74 - 1.00) | 0.062 |
| | | | | | 1514 vs 1509 | NB | 0.75 (0.62 - 0.91) | 0.003 |
| | | | | | 1514 vs 1509 | AG with cluster-robust standard errors | 0.78 (0.65 - 0.93) | 0.006 |
| **PARADIGM-HF (RCT)**[7] | NYHA II, III, or IV, and EF=<40 | LCZ696 or enalapril | 2.3 years (median) | CV death, hospitalization for CHF | 4187 vs 4212 | TTE | 0.80 (0.73 - 0.87) | <0.001 |
| | | | | | 4187 vs 4212 | NB | 0.76 (0.67 - 0.85) | <0.001 |
| | | | | | 4187 vs 4212 | AG with cluster-robust standard errors | 0.79 (0.71 - 0.87) | <0.001 |
| | | | | | 4187 vs 4212 | WLW average (1st-6th) | 0.80 (0.73 - 0.87) | <0.001 |

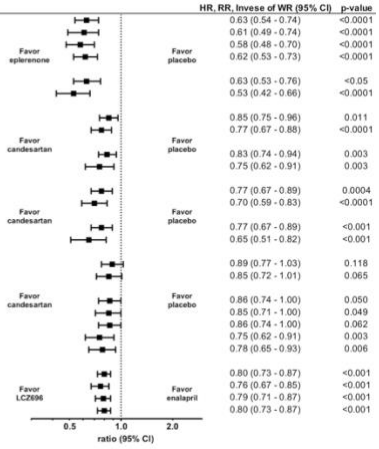

Figure 10

| | Time-to-first-event | Win ratio | Competing risk | Negative binomial | Andersen-Gill, WLW | WCE |
|---|---|---|---|---|---|---|
| Uses first event | Yes | No | Yes | Yes | Yes | Yes |
| Uses all events | No | No | No | Yes | Yes | Yes |
| Death as most important | No | Yes | No | No | No | Yes |
| Uses time to event | Yes | No | Yes | No | Yes | Yes |
| Distribute weight | No | No | No | No | No | Yes |
| Statistical efficacy | → | → or ↑ | → | ↑ | ↑ | → or ↓ |

**Supplementary data**

**Figure legends**

**Supplementary figure 1. Poisson distribution**

(A) The number of events in an interval of time may be represented by a Poisson distribution. The basic shapes of a Poisson distribution (gray and yellow) change according to the probability of events, unlike a Gaussian distribution, which is always symmetric. (B) When the frequency of events is very small (e.g. blue or orange), the variance may be greater than the mean. In this case, the data will approximately follow an over-dispersed Poisson distribution.