When and how to make use of recurrent events in cardiovascular trials

Brief title: Use of recurrent events in cardiovascular trials

John Gregson PhD[a], Gregg W. Stone M.D.[b] , Deepak L. Bhatt M.D. M.P.H.[b], Milton Packer M.D.[c,d], Stefan D. Anker M.D., Ph.D.[e], Cordula Zeller Dipl. Math.[f], Bjorn Redfors M.D. PhD[g,h], Stuart J. Pocock PhD[a]

[a]Department of Medical Statistics, London School of Hygiene and Tropical Medicine
[b] Mount Sinai Heart, Icahn School of Medicine at Mount Sinai, New York
[c]Baylor Heart and Vascular Institute, Baylor University Medical Center, Dallas
[d]Imperial College, London, UK
[e]Department of Cardiology and Berlin Institute of Health Center for Regenerative Therapies, German Center for Cardiovascular Research Partner Site Berlin, Charité Universitätsmedizin, Berlin
[f] Boehringer Ingelheim Pharma, Biberach, Germany
[g]Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[h]Cardiovascular Research Foundation, New York

This work was funded by the British Heart Foundation.

Disclosures:

John Gregson has received: Consultancy fees from Boehringer Ingelheim, Amarin, Vifor Pharma. Research grants from Astra Zeneca.

Gregg Stone has received speaker honoraria from Medtronic, Pulnovo, Infraredx, Abiomed, Abbott, Amgen, Boehringer Ingelheim; has served as a consultant to Daiichi Sankyo, Ablative Solutions, CorFlow, Apollo Therapeutics, Cardiomech, Gore, Robocath, Miracor, Vectorious, Abiomed, Valfix, TherOx, HeartFlow, Neovasc, Ancora, Elucid Bio, Occlutech, Impulse Dynamics, Adona Medical, Millennia Biopharma, Oxitope; and has equity/options from Ancora, Cagent, Applied Therapeutics, Biostar family of funds, SpectraWave, Orchestra Biomed, Aria, Cardiac Success, Valfix, Xenter. Dr. Stone's employer, Mount Sinai Hospital, receives research support from Abbott, Abiomed, Bioventrix, Cardiovascular Systems Inc, Phillips, Biosense-Webster, Shockwave, Vascular Dynamics, Pulnovo and V-wave. Family disclosure: Dr. Stone's daughter is an employee at IQVIA.

Dr. Bhatt discloses the following relationships - Advisory Board: Angiowave, Bayer, Boehringer Ingelheim, Cardax, CellProthera, Cereno Scientific, Elsevier Practice Update Cardiology, High Enroll, Janssen, Level Ex, McKinsey, Medscape Cardiology, Merck, MyoKardia, NirvaMed, Novo Nordisk, PhaseBio, PLx Pharma, Regado Biosciences, Stasys; Board of Directors: Angiowave (stock options), Boston VA Research Institute, Bristol Myers Squibb (stock), DRS.LINQ (stock options), High Enroll (stock), Society of Cardiovascular Patient Care, TobeSoft; Chair: Inaugural Chair, American Heart Association Quality Oversight Committee; Consultant: Broadview Ventures, Hims; Data Monitoring Committees: Acesion Pharma, Assistance Publique-Hôpitaux de Paris, Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute, for the PORTICO trial, funded by St. Jude Medical, now Abbott), Boston Scientific (Chair, PEITHO trial), Cleveland Clinic (including for the ExCEED trial, funded by Edwards), Contego Medical (Chair, PERFORMANCE 2), Duke Clinical Research Institute, Mayo Clinic, Mount Sinai School of Medicine (for the ENVISAGE trial, funded by Daiichi Sankyo; for the ABILITY-DM trial, funded by Concept Medical), Novartis, Population Health Research Institute; Rutgers University (for the NIH-

funded MINT Trial); Honoraria: American College of Cardiology (Senior Associate Editor, Clinical Trials and News, ACC.org; Chair, ACC Accreditation Oversight Committee), Arnold and Porter law firm (work related to Sanofi/Bristol-Myers Squibb clopidogrel litigation), Baim Institute for Clinical Research (formerly Harvard Clinical Research Institute; RE-DUAL PCI clinical trial steering committee funded by Boehringer Ingelheim; AEGIS-II executive committee funded by CSL Behring), Belvoir Publications (Editor in Chief, Harvard Heart Letter), Canadian Medical and Surgical Knowledge Translation Research Group (clinical trial steering committees), Cowen and Company, Duke Clinical Research Institute (clinical trial steering committees, including for the PRONOUNCE trial, funded by Ferring Pharmaceuticals), HMP Global (Editor in Chief, Journal of Invasive Cardiology), Journal of the American College of Cardiology (Guest Editor; Associate Editor), K2P (Co-Chair, interdisciplinary curriculum), Level Ex, Medtelligence/ReachMD (CME steering committees), MJH Life Sciences, Oakstone CME (Course Director, Comprehensive Review of Interventional Cardiology), Piper Sandler, Population Health Research Institute (for the COMPASS operations committee, publications committee, steering committee, and USA national co-leader, funded by Bayer), Slack Publications (Chief Medical Editor, Cardiology Today's Intervention), Society of Cardiovascular Patient Care (Secretary/Treasurer), WebMD (CME steering committees), Wiley (steering committee); Other: Clinical Cardiology (Deputy Editor), NCDR-ACTION Registry Steering Committee (Chair), VA CART Research and Publications Committee (Chair); Patent: Sotagliflozin (named on a patent for sotagliflozin assigned to Brigham and Women's Hospital who assigned to Lexicon; neither I nor Brigham and Women's Hospital receive any income from this patent); Research Funding: Abbott, Acesion Pharma, Afimmune, Aker Biomarine, Amarin, Amgen, AstraZeneca, Bayer, Beren, Boehringer Ingelheim, Boston Scientific, Bristol-Myers Squibb, Cardax, CellProthera, Cereno Scientific, Chiesi, CinCor, Cleerly, CSL Behring, Eisai, Ethicon, Faraday Pharmaceuticals, Ferring Pharmaceuticals, Forest Laboratories, Fractyl, Garmin, HLS Therapeutics, Idorsia, Ironwood, Ischemix, Janssen, Javelin, Lexicon, Lilly, Medtronic, Merck, Moderna, MyoKardia, NirvaMed, Novartis, Novo Nordisk, Owkin, Pfizer, PhaseBio, PLx Pharma, Recardio, Regeneron, Reid Hoffman Foundation, Roche, Sanofi, Stasys, Synaptic, The Medicines Company, Youngene, 89Bio; Royalties: Elsevier (Editor, Braunwald's Heart Disease); Site Co-Investigator: Abbott, Biotronik, Boston Scientific, CSI, Endotronix, St. Jude Medical (now Abbott), Philips, SpectraWAVE, Svelte, Vascular Solutions; Trustee: American College of Cardiology; Unfunded Research: FlowCo, Takeda.Milton Packer has received personal fees from Abbvie, Altimmune, Amgen, Ardelyx, AstraZeneca, Boehringer Ingelheim, Caladrius, Casana, CSL Behring, Cytokinetics, Lilly, Moderna, Novartis, Reata, Regeneron, Relypsa, Salamandra

Cordula Zeller is an employee of Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach, Germany

Dr. Anker declares grants and personal fees from Vifor and Abbott Vascular, and personal fees for consultancies, trial committee work and/or lectures from Actimed, Amgen, Astra Zeneca, Bayer, Boehringer Ingelheim, Bioventrix, Brahms, Cardiac Dimensions, Cardior, Cordio, CVRx, Edwards, Farraday, Impulse Dynamics, Janssen, Novartis, Occlutech, Pfizer, Respicardia, Servier, Vectorious, and V-Wave. Dr. Anker declares that he is named co-inventor of two patent applications regarding MR-proANP (DE 102007010834 & DE 102007022367), but he does not benefit personally from the related issued patents.

Bjorn Redfors has no disclosures to report.

Stuart Pocock declares grants and personal fees from Astra Zeneca, and personal fees from Boehringer Ingelheim, Vifor, Abbot Vascular, Amarin and Novartis.

Word count : 8812

Address for correspondence:

John Gregson
London School of Hygiene and Tropical Medicine
Keppel Street
London
WC1E 7HT
+44 (0)20 7927 2162
john.gregson@lshtm.ac.uk

ABSTRACT

Many randomized trials in cardiovascular disease have repeat non-fatal events (such as hospitalizations) occurring during patient follow-up, yet it remains common practice to have time-to-first event as the primary outcome.

We explore the value of analyses that include repeat events. Do they help us understand the effect of treatment and total disease burden? Do they enhance statistical power? Should they become a trial's primary analysis?

It may also be difficult to choose which of the various statistical methods for analyzing repeat events to use, and we provide a non-technical guide to what each method is doing.

We compare several methods for repeat events: Lin Wei Yang Ying, negative binomial, joint frailty, win ratio and area under the curve. We illustrate their performance in five large cardiovascular trials and compare them to time-to-first-event analyses. We review their use in recently published heart failure trials and make recommendations for their use in future trials.

CONDENSED ABSTRACT

Many randomized trials in cardiovascular disease have repeat non-fatal events occurring during patient follow-up, yet it remains common practice to have time-to-first event as the primary outcome. We explore the value of analyses that include repeat events.

There are several methods for analyzing repeat events; we provide a non-technical guide to what each method is doing. We illustrate the performance of these methods, in five large cardiovascular trials and compare their use to time-to-first event analyses. We review their use in a larger set of recently published heart failure trials and make recommendations for their use in future trials.

KEY WORDS: trial design, recurrent events, statistics, heart failure

ABBREVIATIONS: AUC=area under the curve; CVD=cardiovascular death; DAOH=days alive and out of hospital; HFH=heart failure hospitalization; HR=hazard ratio; KCCQ=Kansas City Cardiomyopathy Questionnaire; MI=myocardial infarction; QoL=Quality of Life; RR=rate ratio

INTRODUCTION

Most large cardiovascular trials use time to the first event as the primary outcome. In patients who have a first non-fatal event, information about subsequent cardiovascular events is therefore ignored. But such 'repeat' events are clinically important. For example, fatal events may be amongst the events that are ignored. Therefore, analyses which include repeat events may better capture the effect of treatment on total disease burden. In addition, previous research has also suggested that using repeat events as the primary outcome may more efficiently determine whether a treatment is effective, with a potential gain in statistical power or reduction in required trial size.[1]

For these reasons analyses including repeat events is becoming more common[2], particularly in trials of heart failure.[3–8] However, how to best make use of repeat events in future trials is unclear. A decision must be made as to whether repeat event analyses are best used as a primary or as a secondary analysis. This may depend on whether a time to first event or repeat event outcome is more clinically relevant. For example, if the aim of treatment is to lengthen time to disease onset, then time-to-first events may make most sense, whereas if the aim is to reduce the total burden of a chronic condition, then a repeat events analysis may be preferred. Which approach has greater statistical power is also often a consideration. If repeat events are to be used, one must also decide which of the various statistical methods to use. The literature describing such methods is often highly technical. [9–11]

The article is structured as follows: 1) we describe in a relatively non-technical manner some of the most useful methods for repeat events analysis; 2) we then illustrate their use in five large cardiovascular trials and make comparisons with one another and also with use of only the first event; 3) we also review the use of repeat events in a larger set of recently published major heart failure trials; 4) we then discuss the pros and cons of each method and make recommendations as to how they can best be used in future cardiovascular trials.

Analysis of trials with repeat events

Before deciding on a statistical method for the analysis of repeat events, a key first step is to decide how repeat events will be counted, i.e. what constitutes a distinct repeat event, rather than being simply a consequence of a previous event. For example, consider a trial with a composite primary outcome of cardiovascular death (CVD) or heart failure Hospitalization (HFH). If a patient dies whilst still hospitalised, it is unclear whether this should count as one event or two. A similar issue arises when considering how to handle closely spaced HFHs. Inclusion of events that are strongly related to one another has implications both from a clinical perspective (whether the resulting analysis remains meaningful), and from a statistical perspective (some models and sample size calculations assume events within a patient are unrelated).

A next step is to decide on a statistical method to use. We provide a summary of such methods below and in Tables 1 and 2 and a graphical display in the Central Illustration. Statistical programming code (including for power calculations for repeat events analyses) are provided in the Appendix.

Analysis using only the first event: A widely used method for analysing only the time to the first event is the Cox proportional hazards model. A key assumption is proportional hazards: that the ratio of the hazard of an event in the treatment and control groups is the same at all time points during follow-up. This hazard ratio for time-to-first event is the key summary measure. In practice, many trials still use the Cox proportional hazards model even when there is some doubt as to whether this assumption holds. Alternatives to the Cox model exist: for instance, the restricted mean (event-free) survival time, whereby one analyses the average event-free time up until a fixed milestone time, adjusting for loss to follow-up. Two limitations of this method are the need for an arbitrary fixed milestone time, and that it tends to (sometimes inappropriately) place far greater emphasis on early events than late events.[12]

Lin Wei Yang Ying (LWYY) models: An extension of the Cox proportional hazards model. Both models assume proportional hazards, but whereas the Cox model ignores repeat events, the LWYY model includes them. The resulting LWYY model yields a hazard ratio for the all events (rather than just the first) and has been used to analyze repeat events in several major cardiovascular trials. [3–5] The LWYY model is closely related to the Andersen-Gill model[13], with the only difference being in the estimation of the standard error. In the Andersen-Gill model one assumes that all events are independent of one another, but this is unrealistic in cardiovascular trials because events tend to cluster in high-risk patients. In the LWYY model a robust standard error is calculated for the hazard ratio that takes into account clustering of events in high risk patients. This results in wider 95% confidence intervals than in an Andersen-Gill model and allows for valid inference even when events are not independent.[14] Initial applications of LWYY and Andersen-Gill models were in modelling recurrent non-fatal outcomes. However, they have also been used to model composite repeat events which include a fatal component (e.g. HFH and CVD), whereby the fatal event is counted as the final event. [5]

Negative binomial model: An extension to a Poisson regression model which estimates a treatment effect as a rate ratio.[15] A key assumption is that events occur at a fixed rate throughout follow-up. This may well not be true in acute conditions where patients tend to be at higher risk of events in earlier follow-up, and may also be false in some chronic diseases. A Poisson regression model assumes that all patients have the same underlying rate of events, which is unrealistic because events often tend to cluster in high-risk patients. A Negative Binomial model relaxes this assumption and takes into account an underlying variation in outcome rates between patients. This allows each patient to have their own individual true rate of events. In most applications, the distribution of individual rates is highly skewed: a small number of patients have a very high rate of events and most patients have much lower rates (most having 0 or 1 event). To account for this the Negative Binomial models assumes that the

individual rates follow a highly skew (gamma) distribution. The comparison between treatment groups then focusses on whether the average rate differs between groups. As with the LWYY model initial applications focused on non-fatal events, but subsequent applications have used composite outcomes which include a fatal component.[7]

Joint frailty models: Joint frailty models simultaneously estimate the rate of a fatal outcome and a recurrent non-fatal outcome.[16, 17] 'Frailty' here refers to underlying statistical variation in disease risk. Joint frailty models account for this frailty but also recognize that the risk of fatal and non-fatal outcomes can be related to one another. This is helpful because patients who are high-risk for non-fatal outcomes tend to also be at high risk for fatal outcomes. Whereas other methodologies tend to (incorrectly) handle fatal outcomes as a non-informative censoring (i.e. assuming non-fatal outcomes are unrelated to fatal outcomes), joint frailty models allow for a relationship between the events. They are therefore useful models when the primary outcome is a non-fatal recurrent event and one wishes to adjust for a related fatal event as a competing risk. The summary measure is a hazard ratio comparing the risk of non-fatal events adjusted for a patient's (model-estimated) frailty.

The joint frailty approach can be implemented using a variety of statistical models.[16, 17]. Results can be sensitive to the choice of model, and even its specific implementation. It is therefore important when using joint frailty models to give a detailed description of the exact implementation in the pre-specified statistical analysis plan. Our implementation of the joint frailty model uses a piece-wise constant hazards model with 10 intervals, and is implemented using the SAS macro developed by Toenges et al.[17]

Win ratio: The win ratio is a method to analyse composite outcomes in which the components have a hierarchy of clinical importance (e.g. deaths are a higher clinical priority than hospitalizations).[18, 19] The win ratio works by comparing all pairs of patients: every patient from the treated group versus every patient from the control group (i.e. $N_1 \times N_2$) comparisons.

Within each pair one evaluates if the treated patient has a better outcome (a "win" for the treatment group), the control patient has a better outcome (a "loss" for the treatment group), or if the outcomes are the same (a "tie"). When classifying each patient pair, the component outcomes are considered in a descending hierarchy of importance (e.g. CVD before HFH) until one of the pair shows a better outcome compared to the other, a process best illustrated with a practical example (see Results section). The win ratio can be considered a method for analysing repeat events, because for non-fatal events the number of events can be used to break a tie – e.g. comparing a treated patient with two HFHs to a control patient with only one HFH results in a "loss". Using all pairs the win ratio is calculated as the total number of wins divided by the total number of losses. The win ratio is accompanied by its 95% confidence interval and a p-value. These can be calculated using the method of Finkelstein-Schoenfeld[20], although they are often calculated using other analytical methods.[21] The win ratio can be interpreted as the odds that for a randomly chosen pair of patients that are not tied, the patient in the treated arm has a better outcome. A win ratio of >1 therefore means that treated patients experience better outcomes more often than they experience worse outcomes (the treatment is beneficial). We consider here the unmatched win ratio, where one first forms every possible patient-to-patient pair: that is, every patient on the new treatment is compared with every patient on the control treatment. ATTR-ACT[22], EMPULSE[23] and TRILUMINATE[24] used a win ratio approach for their primary outcome assessment. EMPULSE and TRILUMINATE included quality of life (assessed using the KCCQ) in their primary outcome as an additional level in their hierarchical outcome, taking lower priority than mortality and HFH. Implementation of the win ratio can be carried out in R (e.g. using packages WinRatio or WINS both available at https://cran.r-project.org/) or using winratiotest in Stata available from the SSC archive.

Area under the curve (AUC): This method is an extension of restricted mean event-free survival time to repeat events. One chooses a fixed milestone follow-up time (e.g. 3 years) and works as follows. For every event, the amount of time spent between the event occurring and the fixed milestone time is calculated. This event-time is summed across all events occurring for each patient.[25] The average event-time in each treatment groups is then calculated and compared between groups. The AUC method estimates the difference or ratio of total time after events occur until a fixed milestone time. The 'area under the curve' name is used because the average event-time is equal to the area under a cumulative event rate curve showing the average number of events per patient after adjusting for loss to follow up. This AUC method is an extension of the restricted mean survival time method [26] to allow for repeat events. We present further details regarding how the AUC is calculated and interpreted using practical examples in the Results section. In most trials patients have a range of follow-up times reflecting the interval from recruitment to a fixed calendar date, so the method's choice of a fixed milestone time is a limitation. Our implementation of the AUC method used the R package described in *Claggett et al*[25] where possible, otherwise we calculated the area under the curve by integrating the mean cumulative functions and used 1000 bootstrap samples to estimate 95% confidence intervals.

Other methods for repeat events: There are several other methods used to analyse repeat events including the Wei-Lin-Weissfeld and Prentice-Williams-Peterson models .[27, 28] However, they are not suitable for the primary analysis of a randomized trial as they do not result in a single estimate of a treatment's effect. The Wei-Lin-Weissfeld model analyses each of the ordered repeat events by calculating a separate hazard ratio for each event (i.e. time to the first event, time to the second event and so on). Each individual is considered to be at risk for all recurrent events, starting at the time of randomisation. The Prentice, Williams and

Peterson (PWP) model also calculates a hazard ratio for each ordered event. But using this approach patients are considered at risk for the next event only if they have already experienced the previous event. For example, risk of a second event is only considered after a first event has occurred. The problem is that by restricting the focus to a subset of patients who have had a preceding event, the comparison of repeat events (i.e. those after the first event) is no longer randomized.

Comparison of repeat events methods in five cardiovascular trials

EMPEROR-Preserved: The Empagliflozin Outcome Trial in Patients with Chronic Heart Failure and a Preserved Ejection Fraction (EMPEROR-Preserved) trial[29] randomized heart failure patients with preserved ejection fraction to empagliflozin (n=2997) or placebo (n=2991). The primary outcome was a composite of HFH or CVD. There was a total of 1411 such events (463 CVD and 948 HFH), of which 926 were first events and 485 were repeat events. Figure 1A shows a Kaplan-Meier plot of cumulative incidence of the first primary event in EMPEROR-Preserved by treatment group (dashed lines). Alongside this is the mean cumulative function (solid lines), which looks similar to a Kaplan-Meier plot but includes all events per patient (i.e. also recurrent events). It is also helpful to understand the distribution of the number of events that occur within each patient, as shown in Figure 2A. We know from the Kaplan Meier curves that most patients have no event, and Figure 2A shows that among those who do have an event most patients experience only one event, while a few patients experience a substantial number of events. It can also be seen that there are fewer patients with events in the treatment group than the control group.

Table 3 shows the results for the primary outcome from the EMPEROR-Preserved trial using each of the statistical methods. A time-to-first event was the pre-specified primary analysis, and was highly statistically significant (HR=0.79, 95% CI 0.69-0.90, p=0.0003; z=3.62). A repeat events analysis was not pre-specified for the primary outcome, and so we first need to

define which cardiovascular deaths count as distinct events. For simplicity, we included all CV deaths as an additional event except those occurring on the same day as admission for HFH. We estimated a HR of 0.79 (95% CI 0.68-0.92; p=0.0029, z=2.98) using the LWYY model, and a rate ratio (RR) of 0.78 (95% CI 0.66-0.93; p=0.0044, z=2.85) using a negative binomial model. Neither result was as strongly significant as the time-to-first event analysis. The similarity of results from the LWYY model and negative binomial model was expected, as the two methods share a similar interpretation: they are both comparing the combined rate of CVD and HFH. A potential limitation of either approach is that they treat CVD and HFH as equal.

An alternative win ratio approach is to prioritize more clinically important events (CVD) over less important events (HFH), as illustrated in Figure 3A. One takes all 2997 patients in the empagliflozin arm and compares them to all 2991 patients in the control arm to form a total of 2997 x 2991 comparisons. Patients are first compared based on who remained free from CVD for the longest. Within each comparison, if the patient in the treated arm survived longest it is a win, whereas if the patient in the control arm survived longest it is a loss. The remaining comparisons are considered a tie based on time to CVD, which occurs if (unusually) both patients died on the same day, or if neither patient died during the period when both were in follow-up. For patients who tied based on CVD, the process is repeated based upon the total number of HFHs that occurred, i.e. when fewer HFHs occur in the treated patient it is considered a win. At the end of the process all wins and losses are added together, and the ratio is taken to estimate a win ratio of 1.25 (1.08-1.42, p=0.0013, z=3.22) which again in this case is not as statistically significant as the time-to-first event analysis.

Another alternative is to focus on the amount of time spent free from events, rather than focussing on the number of events by using the AUC method. This method works by calculating the total number of ‘event-months’ for each patient up until a fixed milestone time (we chose 30 months in EMPEROR-Preserved). Consider a patient with a HFH at 12 months

and a CV death at 29 months. This patient's total event-months would be (30-12) + (30-29) = 19 event-months. The AUC method calculates for each treatment group the mean number of event-months. In EMPEROR-Preserved this is 3.6 event-months per patient in the empagliflozin arm and 4.7 event-months in the control arm. This can be reported as an absolute improvement of 1.1 (95% CI 0.5-1.7) event-months or as a ratio of 0.76 (95% CI 0.66-0.88, p=0.002, z=3.12). The average difference in event-months can also be visualised as the area between the mean cumulative functions, as shown in Figure 4.

We note two limitations of this method. First is the need to choose a fixed milestone time. In trials with rolling recruitment, events that occur after the fixed milestone time do not contribute. In EMPEROR-Preserved 60 events occurred after the milestone time of 30 months, and so are ignored. Choosing a later milestone time lessens this issue but results in many patients being censored before the milestone time and hence yields a less precise estimate of the treatment effect. A second limitation is that the timing of events takes on far more importance than for the other methods. Events occurring late during follow-up contribute little. But the prognosis for patients in EMPEROR-Preserved is quite good, and so what matters most to patients is whether they have an event, rather than its timing. So in this example placing such emphasis on the timing of events seems inappropriate.

The joint frailty model can be used here to analyse HFHs whilst accounting for the competing risk of CVD (see Appendix Table 1). This analysis yielded a hazard ratio of 0.73 (95% CI 0.61-0.83) , p=0.0009; z=3.33. This result provided slightly greater evidence of treatment benefit than an analysis of HFH using either the LWYY model (HR=0.75, 95% CI 0.62-0.90; p=0.0023, z=3.05) or negative binomial models (RR=0.73, 95% CI 0.60-0.89, p=0.0017,z=3.13), a finding which is expected when numerically fewer CVDs occur in the treated arm. This occurs because although the three methods estimate a rate (or hazard) ratio for HFH, they differ in how they handle the competing risk of CVD. The LWYY and negative

binomial models assume CVD to be a non-informative censoring regarding a patient's rate of HFH, which is unrealistic. When fewer competing events occur in the treated arm this leads to an increase in statistical power for the joint frailty model relative to the other methods.[9]

We have illustrated the key options available for repeat events analysis. In this example, all analyses showed a clear treatment benefit, but using repeat events did not strengthen the evidence as compared with a conventional time-to-first event analysis. We next explore an example where repeat events did appear to offer such a benefit.

CHARM-Preserved: The Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity (CHARM)-Preserved trial[30] randomized patients with chronic heart failure and preserved left ventricular ejection fraction to either candesartan (n=1514) or placebo (n=1509). The primary outcome was a composite of CVD or HFH. Only first events were adjudicated in CHARM-Preserved, so in order to perform analyses using repeat events we focus here on investigator-reported primary outcome events. For our analysis of repeat events we counted CVD occurring during a HFH as a single event. This resulted in a total of 704 first events and 512 repeat events.

The time to first primary event analysis gives borderline evidence of treatment benefit: HR=0.86 (95% CI 0.74-1.00, p=0.050). Analyses of repeat events, however, tend to provide stronger evidence for a treatment benefit. Using an LWYY model resulted in a HR of 0.78 (95% CI, 0.65-0.93, p=0.006) and using a negative binomial model gave an RR of 0.76 (95% CI 0.62-0.92, p=0.007). Comparing results of the Cox and LWYY model, the smaller p-value results from a larger treatment effect (smaller hazard ratio), indicating that the estimated relative effect of candesartan on recurrent events is greater for recurrent events than it is for first events. This could be because a slightly greater proportion of repeat events (than first events) are HFH rather than CVD, and candesartan seems to reduce HFH but not CVD (see

Appendix Table 3). It does not result from improved precision, since the 95% confidence interval is wider in the LWYY model than using the Cox model.

The win ratio approach gave weaker evidence of treatment benefit (win ratio=1.15, 95% CI 0.99-1.34, p=0.062). This can be explained by the win ratio placing greater priority on CVD (top of the hierarchy) and less on HFH (next in the hierarchy).

The joint frailty model can be used to analyse the secondary outcome of total HFHs, whilst adjusting for CVD as a competing risk, and yields a hazard ratio of 0.69 (HR=0.56-0.85, p=0.0006). This is very similar to when total HFHs are analysed using either LWYY (HR=0.71, 95% CI 0.58-0.88, p=0.0018) or negative binomial models (RR=0.68, 95% CI 0.54-0.86, p=0.0012). This is to be expected, since the risk of the competing event (CVD) is well balanced between treatment and control groups.

Overall, CHARM-Preserved provides an example where a repeat event analysis enhanced the ability to identify a treatment benefit.

AFFIRM-AHF: AFFIRM-AHF[7] randomized patients with acute heart failure with iron deficiency to ferric carboxymaltose (n=558) or placebo (n=550). The primary outcome was a composite of total HFHs (including repeats) and CVD up to 52 weeks. When a heart failure hospitalization led to cardiovascular death this counted as two events, except where a patient died on the date of admission. Mean cumulative functions and the number of events per patient by treatment group are displayed in Figures 1B and 2B respectively. There was a total of 293 primary outcome events (181 first, 102 repeats) in the ferric carboxymaltose group compared to 372 (209 first, 163 repeats) in the placebo group.

The primary analysis used a negative binomial model and gave a rate ratio of 0.79 (95% CI 0.62-1.01, p=0.059). Some other statistical methods lent greater support to treatment benefit.

Using an LWYY model instead yielded a hazard ratio of 0.77 (0.62-0.95, p=0.010) and using only the first event in a Cox model yielded a hazard ratio of 0.80 (0.66-0.98, p=0.030).

The least statistically powerful analysis for AFFIRM-AHF was the win ratio (win ratio=1.18, 0.96-1.48; p=0.122. Figure 4B). Ferric carboxymaltose appeared not to impact CVD (see Appendix Table 2), which gets prioritized with the win ratio method. One advantage of the win ratio is that it enables combining repeat events with other quantitative outcomes such as KCCQ, allowing assessment of the impact of treatment not just on clinical events but also on patient quality-of-life (QoL). To do this, we add the KCCQ overall summary score at follow-up as a third level to the hierarchy of outcomes. It takes lower priority than CVD and HFH, and so only patients tied on these outcomes are evaluated using their KCCQ score. The results of this approach depends upon the timing of the KCCQ assessment. Using KCCQ at the end of the study (12 months), yields a non-significant win ratio 1.11 (95% CI 0.96, 1.29, P=0.162). Whereas, using KCCQ measured earlier in the trial, at 12 weeks as shown in Figure 3B, provides some evidence of treatment benefit: win ratio 1.17 (1.01 to 1.35; P=0.038). This post-hoc exploratory analysis should not change how we interpret the pre-specified findings from AFFIRM-AHF, but illustrates how information on time-to-event, repeat events and patient-reported outcomes can be combined using the win ratio.

COAPT: The Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation Trial[6] randomized patients with heart failure with moderate or severe mitral regurgitation to either a MitraClip device (n=302) or a control group (n=312), both with best medical therapy. In their main publication[6], patients were followed for up to 2 years and the primary outcome was the total number of HFHs (including repeat events). There were a total of 437 HFHs of which 243 were first events. The annualized rate of HFHs was 35.8% in the MitraClip group and 67.9% in the

control group. The mean cumulative function and the numbers of events per patient by treatment group are shown in Figures 1D and 2D respectively.

In COAPT there was a high mortality rate. The mean cumulative function is an estimate of the average number of HFHs that would occur in a patient who survives until the end of follow-up. One can instead use the method of Ghosh and Lind to estimate the number of events per patient that will actually occur accounting for the fact that some patients will die (see Appendix Figure 1).[31] This has the drawback that patients who die obviously cannot experience further HFHs, so if there are more deaths on placebo this can diminish the estimated treatment benefit.

In COAPT, the primary outcome (repeat HFH events) was a non-fatal outcome and there were frequent competing events (deaths). The trial pre-specified the primary outcome be analysed using a joint frailty model, which appropriately adjusts for the competing risk of death. Using a negative binomial or LWYY model here would instead mean handling death as non-informative censoring, thereby assuming (unrealistically) that it is unrelated to the underlying risk of HFH.

Table 3 shows results using each of the statistical methods. Regardless of the method chosen, MitraClip was highly protective for the primary outcome ($p<0.001$ for all). Hazard ratios comparing MitraClip vs. control were 0.51 ($p<0.0001$, $z=4.8$) using a joint frailty model, 0.54 ($p<0.001$, $z=4.4$) using an LWYY model, and the RR for the negative binomial model was 0.49 ($p<0.001$, $z=4.5$). The improved statistical power (as indicated by a larger Z-statistic) using a joint frailty model is expected when a treatment is protective for both the primary outcome and the competing risk of death.[9]

A win ratio based on HFH alone is 2.01 (95% CI 1.53, 2.63, $p<0.001$, $z=5.0$). But ignoring mortality in the evaluation of the win ratio goes against the underlying philosophy of the

method. Analyses including either CVD or all-cause death as the most important event in the hierarchy yield win ratios of 1.72 (1.35-2.20, p<0.001, z=4.38) and 1.61 (1.28 to 2.03, p<0.001, z=4.0) respectively.

Given the high rates of hospitalization and death, the timing of events is as important as whether or not one occurs. The AUC method is able to take this into account. There was an estimated reduction of 6.8 (95% CI 4.5-9.3, p<0.001, z=4.7) event-months in COAPT comparing MitraClip to control, or equivalently a ratio of 0.53 (95% 0.42-0.65).

Regardless of the choice of method, a repeat events analysis did not appear more statistically powerful (smaller Z-statistic) than using a time-to-first event analysis (hazard ratio=0.52, 95% CI 0.40 to 0.67, p<0.0001, z=5.0). Notably, there was no increase in precision of estimates for repeat events analyses.

Overall, COAPT offers shows how one may handle repeats events analysis in the presence of a common competing risk (high mortality), Given that there is little treatment crossover and a high degree of heterogeneity in underlying patient risk, one would also expect that repeat events analysis would offer a gains in statistical power,[1] but this did not occur. We consider why such a benefit may not have materialized in the discussion section, as it could have implications for future trial design.

REDUCE-IT: The REDUCE-IT trial[32, 33] randomized statin-stabilized patients with elevated fasting triglyceride and either established cardiovascular disease or diabetes and other risk factors to either icosapent ethyl (n=4089) or placebo (n=4090). The primary outcome was a composite of CVD, non-fatal stroke, non-fatal myocardial infarction (MI), hospitalization for unstable angina or coronary revascularisation. Mean cumulative functions and the numbers of events per patients by treatment group are shown in Figures 1E and 2E. In our analysis we

count all events with different dates of onset as distinct events, although we later discuss limitations of this approach.

Results in REDUCE-IT were highly statistically significant ($p<0.001$) regardless of the choice of statistical analysis (Table 3). Hazard ratios were 0.75 (95% CI 0.68-0.83) and 0.69 (95% CI 0.62-0.77) using a Cox proportional hazards model and LWYY model respectively, indicating that the relative effect of treatment was enhanced using repeat events. The AUC method gave similar conclusions (ratio=0.71, 95% CI=0.62-0.82, $p<0.001$).

The purpose of repeated events analysis here is to characterise the effect of treatment on total disease burden. Therefore, using a methodology that can estimate both relative and absolute effects of treatment is helpful. The negative binomial model is a natural choice because it allows calculation of rate differences as well as rate ratios. Using a negative binomial model we find a RR of 0.70 (95% CI, 0.62-0.78) and a difference in rates of 123 distinct events (95% CI: 84-162) per 1000 patients treated for 5 years. This difference is far larger than the number of first cardiovascular events prevented: 56 (95% CI 36-75) per 1000 patients treated for 5 years. The impact on total disease burden is therefore larger than the impact on just the first event.

In REDUCE-IT repeat events analysis gave greater evidence of treatment benefit than using only the first event. One may therefore wonder why such analyses are rarely used as the primary analysis in trials with major adverse coronary events as the primary outcome; we offer two possible reasons.

First is that the expected additional number of events included is often quite low because the ratio of recurrent to first events is often small. Although this was not so for the primary outcome in REDUCE-IT, where 38.7% of events were recurrent events, it is the case for the key secondary outcome of CVD, MI or stroke where only 19.6% of events were recurrent

events. If related events or those occurring close together in time were discounted, there would be fewer still.

A second reason is that events are often linked, making it hard to determine what constitutes a 'recurrent' event, rather than a clinical consequence of a previous event (even more so than in heart failure trials). For instance, a patient in REDUCE-IT had an MI 253 days after randomisation, coronary revascularization on day 259, stroke on day 261, and CV death on day 269. It is likely that all 4 events are related, a consequence of the first, but it is hard to tell without a detailed review of the patient's records. An independent adjudication committee could consider such clinical factors in a blinded manner to determine which are distinct events, but this adds further logistical burden to the trial.

The win ratio does not require that events are independent, but trialists, clinicians and patients may disagree on the proper hierarchy of importance when there are many components in the primary composite outcome. For simplicity we define a hierarchy by comparing patients first on time to CVD and then on the number of non-fatal CV events. This yields a win ratio of 1.34 (1.21-1.48, $p<0.001$).

This example illustrates how repeat events can help to better quantify the total benefit of a treatment, but also highlights how complexities arise when analysing repeat events in trials with major coronary adverse events as the primary outcome.

Review of the use of methods for repeat events in recent heart failure trials

To further investigate use of repeat events analyses in recent trials beyond those included here, we surveyed randomized trials in heart failure published in the *New England Journal of Medicine*, *The Lancet* or *Nature Medicine* between 1st July 2019 and 1st January 2023. Trials were eligible for inclusion if they were conducted in patients with heart failure and had HFH

as a component of the primary outcome. Search terms and reasons for exclusion of ineligible studies are given in Appendix B.

We identified a total of 18 eligible trials. Of these, 14 (78%) studies included at least one analysis that used repeat events in the primary publication (Table 4[3–5, 7, 23, 29, 34–41]) and in 8 (44%) of these trials it was the primary analysis. The LWYY model was used in eight trials, the win ratio was used in three trials, joint frailty models in two trials, and one trial used a negative binomial model.

To compare the repeat events analyses to time-to-first event analyses, we compared results in the ten studies where results were reported for the same outcome analysed using both methods. In two trials of these trials, the results were statistically non-significant (at p=0.05) irrespective of the approach taken. Of the remaining eight trials, four had stronger evidence of treatment benefit (smaller p-value and larger Z-statistic) using time-to-first event, two had stronger evidence of treatment benefit when using repeat events analysis, and in two trials (SCORED, SOLOIST-WH) the approaches provided near identical evidence of treatment benefit.

In eight studies that used an LWYY model, we calculated the standard error of the log hazard ratio under this approach and compared it to the standard error when using a Cox model for time-to-first event. We found that the more precise estimate (lower standard error) was using the Cox model in seven of these eight trials. One might have anticipated that adding in more (repeat) events would enhance precision. But because they have a skew distribution (see Figure 2) such heterogeneity (and imprecision in its estimation) counteracts this, and counter-intuitively leads typically to a slight loss of precision in the hazard ratio estimate.

DISCUSSION

In clinical trials of cardiovascular disease, and in particular trials of heart failure, there is an increasing use of repeat events in the presentation of results, either as a primary or secondary outcome. Our article illustrates the most commonly used statistical methods for such analyses along with the advantages and disadvantages of each approach.

When considering whether and how to use repeat events as a trial's primary endpoint it is important to consider how meaningful it is to patients. When studying patients without established disease or with an acute disease (e.g. post MI), the main purpose of treatment is usually to prevent or delay disease onset, and so time-to-first event analysis seems an intuitive choice. In contrast, in patients with more established chronic disease the main purpose of treatment is often to reduce the total number of subsequent events, making analyses including repeat events more meaningful. In patients with very advanced disease, QoL may be as, or more important than the occurrence of clinical events. The win ratio is able to combine measures of QoL with clinical events and/or mortality and may therefore be an attractive option in this context.

Many previous heart failure trials using repeat events have chosen a composite outcome of CVD and total HFH . In these cases an LWYY model, or alternatively a negative binomial model (if event rates are approximately constant over time), are logical choices. One consideration to be pre-specified is how to handle HFH events that result in CVD (i.e. does this count as one or two events, depending on how close in time the two events occurred?).

An alternative to counting CVD and HFH equally is to place greater emphasis on CVD using the win ratio approach. There is potential for wider application of the win ratio in cardiovascular fields outside of the heart failure trials where it most common (e.g. ATTR-ACT, EMPULSE, TRILUMINATE[22–24]). For example, in surgical trials, it could be used to

prioritize spontaneous MI over procedural MI (the latter typically occur first and so are prioritized in Cox models). The win ratio can also be used to incorporate quantitative patient-reported outcomes or functional measures (e.g. exercise time), with these outcomes taking less priority than death or non-fatal clinical events. As with other methodologies, one needs to consider which of the individual components of a composite outcome contribute evidence of treatment benefit. For example, whether the result is primarily driven by reduction in less clinically-impactful events (e.g. revascularisation) or QoL. In unblinded studies, a cautious interpretation may be required if the effect is driven by patient-assessed outcomes like QoL.

In some instances, one may wish to emphasize the timing rather than the number of events. In such instances, the AUC method may be appropriate because it emphasizes early events far more then later events. It also has the benefit of not requiring statistical assumptions. But in most trials only a minority of patients have an event, and in these cases it is usually the frequency rather than timing of events that is of prime importance. This makes the AUC method less attractive. The need for a fixed milestone time for follow-up and the reduced statistical power (in the examples explored here) are further drawbacks.

In some trials the aim is to estimate the effect of treatment on total HFHs. A key issue then becomes how best to handle the competing risk of CV death, particularly if it is common. We note that trials which use a composite outcome of CVD and HFH do not completely avoid the issue of competing events, since patients may die from non-cardiovascular causes. But this has traditionally been viewed as an acceptable limitation provided that non-cardiovascular events are unlikely to be influenced by treatment. A joint frailty approach is helpful for handling competing events, because it estimates a hazard ratio for non-fatal events that takes into account the relationship between the risk of fatal and non-fatal events. In contrast LWYY and negative binomial models inappropriately handle CVD as unrelated to the rate of HFH. In the trials included in this review, analyses focussing on HFH alone tended to give more

pronounced treatment effects than those focussing on a composite of CVD and HFH. Presumably many treatments influence non-fatal events more than cardiovascular death.

In addition to the methods described here, several other methods can be used to analyse repeat events, but some are inappropriate for a primary analysis of typical cardiovascular trials. The Ghosh and Lin model[31] is an approach similar to an LWYY model, except that it treats death as a terminating event, meaning that patients who have died remain in the risk set for future repeat events of course without any further events occurring. This approach may be useful for calculating the average number of events per patient over time, but it is inappropriate for calculating hazard ratios between groups because a treatment may appear protective for two very different reasons: either because it reduces the hazard of the primary outcome, or because by increasing the risk of death patients inevitably experience fewer non-fatal primary outcome events. The same drawback applies to the Fine and Gray model for handling competing risks in time-to-first event analyses.[42] Another statistical method is to analyse the number of days a patient is alive (discharged) and out of hospital (DAOH) before a fixed time from randomization.[43, 44] This approach tends to put *far* more emphasis on mortality than on non-fatal events, particularly in trials with long-term follow-up. The tendency of many treatments to have a greater impact on non-fatal outcomes than fatal outcomes, means that trials using DAOH may require large numbers of patients in order to achieve adequate statistical power. An additional consideration is that deaths occurring early during follow-up have a far greater impact on DAOH than those occurring later. This needs to be taken into account when considering whether DAOH is an appropriate outcome, noting that a treatment that slightly reduces early mortality but greatly increases late mortality may appear protective.

There is enthusiasm regarding the use of repeat events in cardiovascular trials, and a perception that their use might provide greater statistical power, potentially allowing trial size

to be reduced. [1] However, evidence from the trials reviewed here suggests that benefits for statistical power may be rare in practice. We found greater evidence of treatment benefit in CHARM-Preserved when using repeat events analyses, but in the majority of studies evidence of treatment benefit was similar or weaker than when using time-to-first event analyses. The reasons why this is the case merits consideration.

Claggett et al.[1] previously studied repeat events using simulated trials to identify where their use may be most useful. In studies with frequent discontinuation or crossover following the first event, there may be little benefit to using repeat events. This occurs in trials of anticoagulants for atrial fibrillation, where patients tend to go onto open-label therapy after a first event. In these trials repeat events analyses may be best avoided, since there is unlikely to be any effect of randomized treatment on recurrent events, unless there is a 'legacy' carryover effect. Claggett et al.[1] found the largest benefit to be in trials with a large numbers of repeat events (relative to first events) and infrequent discontinuation or crossover. Several trials included in this review (e.g. COAPT) met these criteria but did not gain in terms of statistical power by using repeats.

We suspect a major reason for a lack of gain in statistical power is that a large number of repeat events occur in a relatively small number of patients. This provides less new information than if repeat events were more evenly distributed across a larger number of patients. It is also likely that repeat events for the same patients are not really independent from one another. If later events are partly caused by earlier events, they contain less new information. When repeat events analysis is primary it is therefore important to ensure the sample size is adequate given that some dependency between events may exist. Future research to inform the degree of likely dependency could be useful to inform such considerations.

Where repeat events analyses were more powerful, the key reason was a larger relative effect of treatment on the outcome (i.e. a smaller hazard ratio), rather than the treatment effect being more precisely estimated. This suggests that the pattern of treatment effect over time may influence the relative efficiency of using first events or repeat events. An example of this is the CORONA trial in which analyses for repeat events were more powerful than analyses using only the first event.[45] In CORONA, the cumulative incidence of heart failure was similar with rosuvastatin and placebo during the first year of the trial, but appeared to diverge thereafter. Such a 'delayed' effect is more likely to favour the use of repeat events, since a larger proportion of recurrent events (than first events) will occur later in follow-up. In contrast, when the benefit of treatment emerges early (as was the case in COAPT), one would expect any benefit from using repeat events to be more limited. A difficulty is that the time pattern of treatment effect may often be difficult to predict in advance.

Any trial that includes repeat event analyses needs to pre-specify in the protocol and statistical analysis plan (SAP) what the precise intentions are. This requires clarification as to whether each such analysis is the primary outcome, a key secondary outcome or an exploratory (sensitivity) analysis. Furthermore, it is insufficient to just name the method, e.g. joint frailty or win ratio. Full details of the exact approach need to be explained in the SAP. This is important since methods that superficially look similar can give different results and post-hoc selections across a range of alternatives need to be avoided. Of course, exploratory post-hoc analyses are still permitted provided that they are perceived as such.

A key question for trialists will be how analyses using repeat events are viewed by regulators. FDA draft guidance permits the use of the number of HFHs or the time to recurrent hospitalizations as an acceptable endpoint in trials of heart failure[46], and such analyses are now commonplace in heart failure trials, including among industry-sponsored trials aiming for FDA regulatory approval (e.g. PARAGON-HF[5]). EMA is less clear-cut, stating that

"endpoints accounting for recurrent HFH events may under certain conditions better characterise the prognosis of patients with chronic HF", but that when used as a primary outcome they require "further justification, adjudication of the events and a clear methodological strategy".[47]The win ratio incorporating deaths, recurrent events and QoL or functional measures is being widely used in ongoing trials of new devices in valve disease. For trials using major adverse cardiovascular events in patients with ischemic heart disease, trials seeking regulatory approval rarely use repeat events analyses for the primary outcome.

Our article has limitations. First, with only five datasets explored in depth, caution is needed in drawing generalizable conclusions. Our examples were designed to illustrate some of the practical issues that may arise when applying repeat events analysis. Second, some previous articles have already explored the use of repeat events.[1, 11, 48] Our article adds to this work by exploring relatively new statistical methods such as the win ratio and AUC, as well as by providing a non-technical overview of the available methods and their pros and cons.

In summary, repeat event analyses can be helpful for quantifying the total benefit of a new treatment. In trials of heart failure, they rarely offer a substantial improvement in statistical power. Trialists should take the considerations described in this article into account when planning future studies.

HIGHLIGHTS

- Repeat non-fatal events occur in many trials, but often only the first event is considered
- The pros and cons of various methods for repeat event analysis are reviewed
- Topical examples are presented, especially in heart failure
- Including repeat events sometimes better captures the total benefit of a treatment, but rarely improve statistical power

Acknowledgements: We thank James Bell for his helpful comments on this paper.

REFERENCES

1. Claggett B, Pocock S, Wei LJ, Pfeffer MA, McMurray JJV, Solomon SD. Comparison of Time-to-First Event and Recurrent-Event Methods in Randomized Clinical Trials. *Circulation*. 2018;138:570–577.

2. Anker SD, McMurray JJV. Time to move on from "time-to-first": Should all events be included in the analysis of clinical trials? *Eur Heart J*. 2012;33:2764–2765.

3. Bhatt DL, Szarek M, Pitt B, et al. Sotagliflozin in Patients with Diabetes and Chronic Kidney Disease. *N Engl J Med*. 2021;384:129–139.

4. Bhatt DL, Szarek M, Steg PG, et al. Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure. *N Engl J Med*. 2021;384:117–128.

5. Solomon SD, McMurray JJV, Anand IS, et al. Angiotensin–Neprilysin Inhibition in Heart Failure with Preserved Ejection Fraction. *N Engl J Med*. 2019;381:1609–1620.

6. Stone GW, Lindenfeld J, Abraham WT, et al. Transcatheter Mitral-Valve Repair in Patients with Heart Failure. *N Engl J Med*. 2018;379:2307–2318.

7. Ponikowski P, Kirwan B-A, Anker SD, et al. Ferric carboxymaltose for iron deficiency at discharge after acute heart failure: a multicentre, double-blind, randomised, controlled trial. *Lancet*. 2020;396:1895–1904.

8. Abraham WT, Adamson PB, Bourge RC, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet* 2011;377:658–666.

9. Fritsch A, Schlömer P, Mendolia F, Mütze T, Jahn-Eimermacher A. Efficiency Comparison of Analysis Methods for Recurrent Event and Time-to-First Event Endpoints in the Presence of Terminal Events—Application to Clinical Trials in Chronic Heart Failure. *Stat Biopharm Res*. 2021;0:1–29.

10. Schmidli H, Roger JH, Akacha M. Estimands for Recurrent Event Endpoints in the Presence of a Terminal Event. *Stat Biopharm Res*. 2021;0:1–29.

11. Amorim LDAF, Cai J. Modelling recurrent events: A tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44:324–333.

12. Gregson J, Sharples L, Stone GW, Burman C-F, Öhrn F, Pocock S. Nonproportional Hazards for Time-to-Event Outcomes in Clinical Trials: JACC Review Topic of the Week. *J Am Coll Cardiol*. 2019;74:2102–2112.

13. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *Ann Stat*. 2007;10:1100–1120.

14. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J R Stat Soc Ser B Stat Methodol*. 2000;62:711–730.

15. Hilbe JM. *Negative Binomial Regression*. Cambridge University Press, 2011.

16. Rondeau V, Pignon J-P, Michiels S. A joint model for the dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer. *Stat Methods Med Res*. 2015;24:711–729.

17. Toenges G, Jahn-Eimermacher A. Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. *Comput Methods Programs Biomed*. 2020;188:105259.

18. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33:176–182.

19. 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;41:4391–4399.

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

21. Dong G, Li D, Ballerstedt S, Vandemeulebroecke M. A generalized analytic solution to the win ratio to analyze a composite endpoint considering the clinical importance order among components. *Pharm Stat*. 2016;15:430–437.

22. Maurer MS, Schwartz JH, Gundapaneni B, et al. Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *N Engl J Med*. 2018;379:1007–1016.

23. Voors A, Angermann C, Teerlink J, et al. Efficacy and Safety of Empagliflozin in Hospitalized Heart Failure Patients: Main Results from The Empulse Trial. *Nat Med*. 2022;55:172.

24. Sorajja P, Whisenant B, Hamid N, et al. Transcatheter Repair for Patients with Tricuspid Regurgitation. *N Engl J Med*. 2023. Published onlineMarch 4, 2023.

25. Claggett BL, McCaw ZR, Tian L, et al. Quantifying Treatment Effects in Trials with Multiple Event-Time Outcomes. *NEJM Evid*. 2022;1:EVIDoa2200047.

26. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13:152.

27. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc*. 1989;84:1065–1073.

28. Prentice RL, Williams BJ, Peterson A V. On the regression analysis of multivariate failure time data. *Biometrika*. 1981;68:373–379.

29. Anker SD, Butler J, Filippatos G, et al. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. *N Engl J Med*. 2021;385:1451–1461.

30. Yusuf S, Pfeffer MA, Swedberg K, et al. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet*. 2003;362:777–781.

31. Ghosh D, Lin DY. Marginal Regression Models for Recurrent and Terminal Events. *Stat Sin*. 2002;12:663–688.

32. Bhatt DL, Steg PG, Miller M, et al. Cardiovascular Risk Reduction with Icosapent Ethyl for Hypertriglyceridemia. *N Engl J Med*. 2019;380:11–22.

33. Bhatt DL, Steg PG, Miller M, et al. Effects of Icosapent Ethyl on Total Ischemic Events: From REDUCE-IT. *J Am Coll Cardiol*. 2019;73:2791–2802.

34. Solomon SD, McMurray JJV, Claggett B, et al. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. *N Engl J Med*. 2022;387:1089–1098.

35. McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2019;381:1995–2008.

36. Packer M, Anker SD, Butler J, et al. Cardiovascular and Renal Outcomes with Empagliflozin in Heart Failure. *N Engl J Med*. 2020;383:1413–1424.

37. Lindenfeld J, Zile MR, Desai AS, et al. Haemodynamic-guided management of heart failure (GUIDE-HF): a randomised controlled trial. *Lancet*. 2021;398:991–1001.

38. Kalra PR, Cleland JGF, Petrie MC, et al. Intravenous ferric derisomaltose in patients with heart failure and iron deficiency in the UK (IRONMAN): an investigator-initiated, prospective, randomised, open-label, blinded-endpoint trial. *Lancet*. 2022;400:2199–2209.

39. Shah SJ, Borlaug BA, Chung ES, et al. Atrial shunt device for heart failure with preserved and mildly reduced ejection fraction (REDUCE LAP-HF II): a randomised, multicentre, blinded, sham-controlled trial. *Lancet*. 2022;399:1130–1140.

40. Mebazaa A, Davison B, Chioncel O, et al. Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure (STRONG-HF): a multinational, open-label, randomised, trial. *Lancet*. 2022;400:1938–1952.

41. Armstrong PW, Pieske B, Anstrom KJ, et al. Vericiguat in Patients with Heart Failure and Reduced Ejection Fraction. *N Engl J Med*. 2020;382:1883–1893.

42. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc*. 1999;94:496–509.

43. Chen Y, Lawrence J, Stockbridge N. Days alive out of hospital in heart failure: Insights from the PARADIGM-HF and CHARM trials. *Am Heart J*. 2021;241:108–119.

44. Szarek M, Bhatt DL, Steg PG, et al. Effect of Sotagliflozin on Total Hospitalizations in Patients With Type 2 Diabetes and Worsening Heart Failure : A Randomized Trial. *Ann Intern Med*. 2021;174:1065–1072.

45. Rogers JK, Jhund PS, Perez A-C, et al. Effect of rosuvastatin on repeat heart failure hospitalizations: the CORONA Trial (Controlled Rosuvastatin Multinational Trial in Heart Failure). *JACC Heart Fail*. 2014;2:289–297.

46. United States Food and Drug Administration. Treatment for Heart Failure: Endpoints for Drug Development Guidance for Industry (draft) Accessed February 22, 2022. https://www.fda.gov/media/128372/download.

47. European Medicines Agency. Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure CPMP/EWP/235/95, Rev.2

48. Rogers JK, Pocock SJ, McMurray JJ V, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail*. 2014;16:33–40.

Central illustration: A graphical representation of methods for repeat events

Caption: Each of the first five panel show graphical representations of a method for repeat events. The bottom right panel shows a comparison of these methods when applied to the EMPEROR-Preserved trial

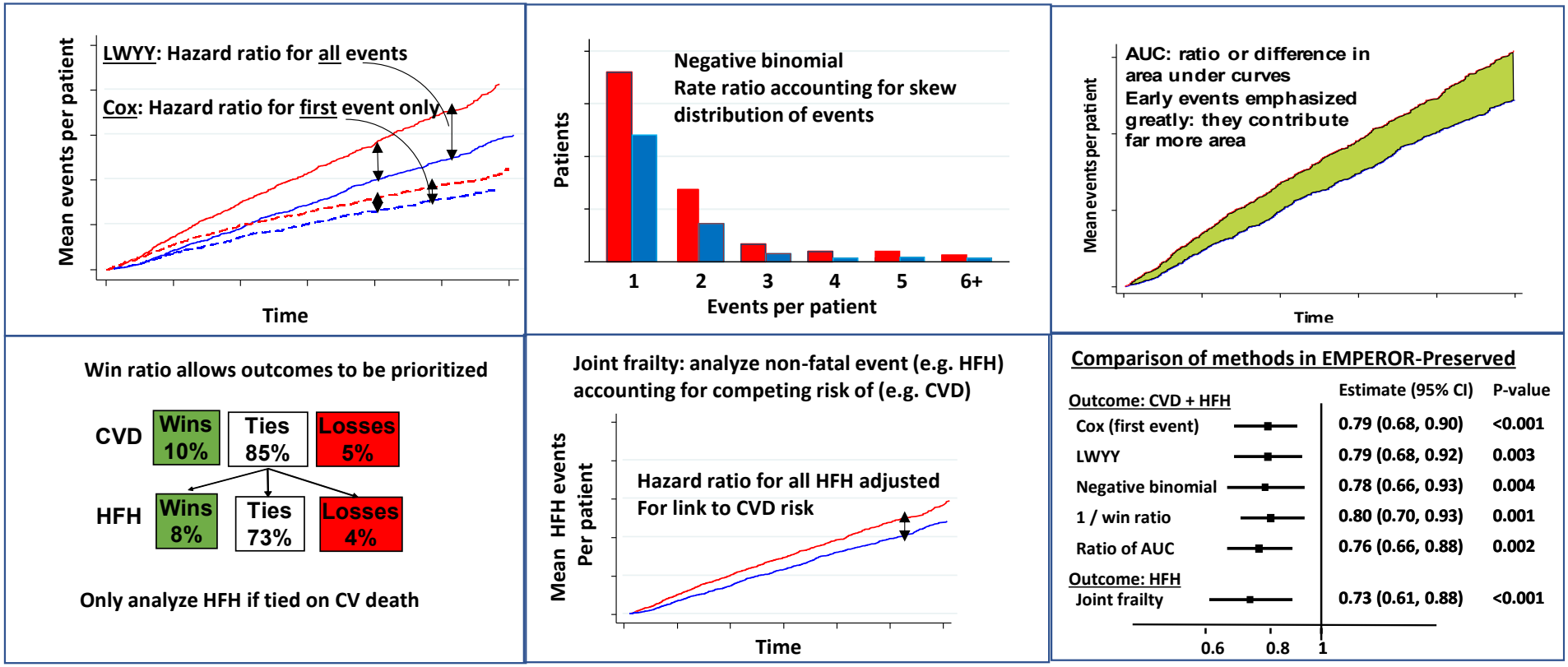


Figure 1: Cumulative first events and mean events per patient

Caption: The graphs show Kaplan Meier cumulative first events, and the mean cumulative number of events in (A) EMPEROR-Preserved; (B) CHARM-Preserved; (C) AFFIRM-AHF; (D) COAPT; (E) REDUCE-IT

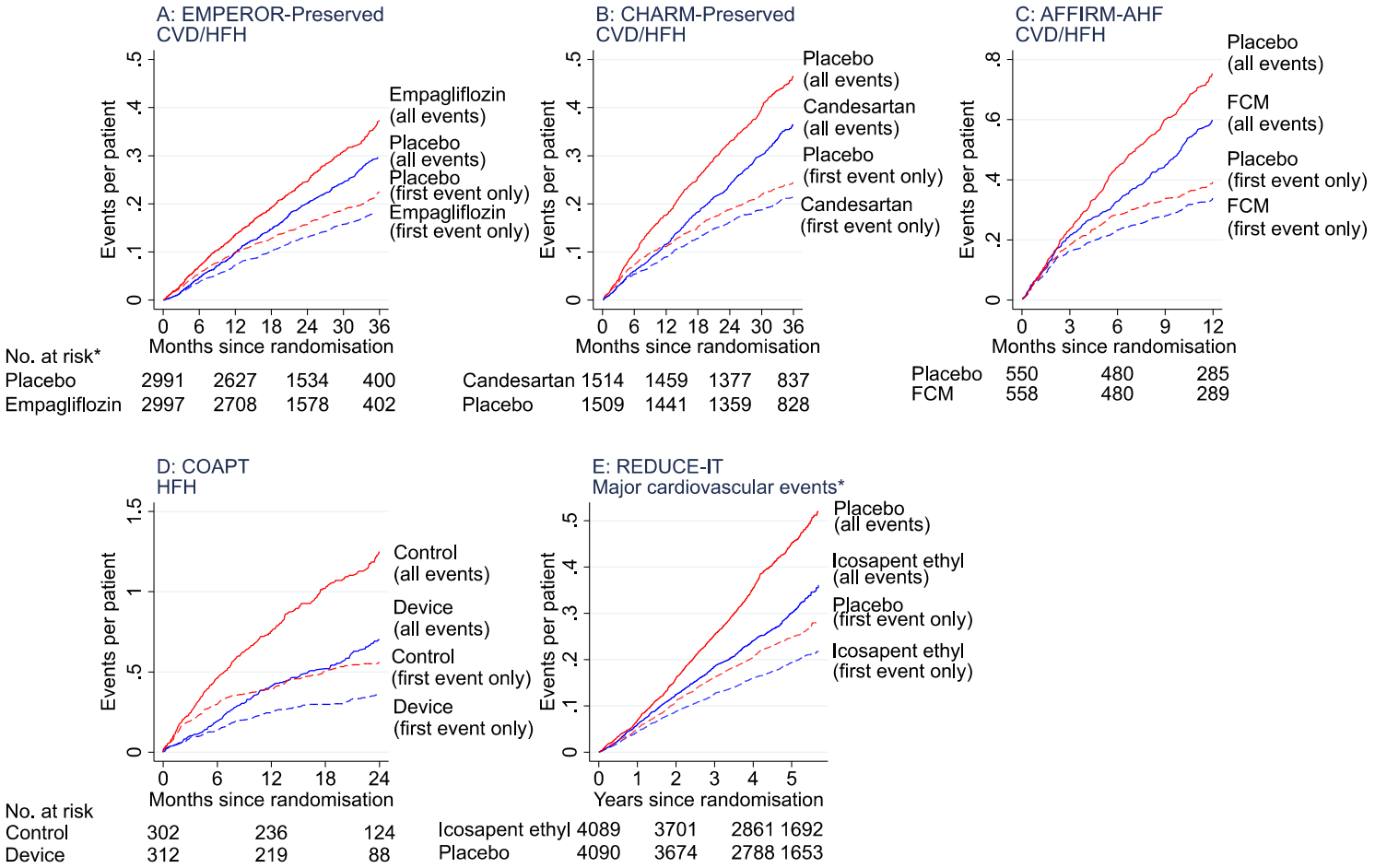


Figure 2: Events per patient in each of five cardiovascular trials

Caption: The graphs show the distribution of the number of events occurring per patient in five trials. The great majority only have 1 or 2. The distribution is highly skew with a small number of patients contributing lots events. The five trials are (A) EMPEROR-Preserved; (B) CHARM-Preserved; (C) AFFIRM-AHF; (D) COAPT; (E) REDUCE-IT

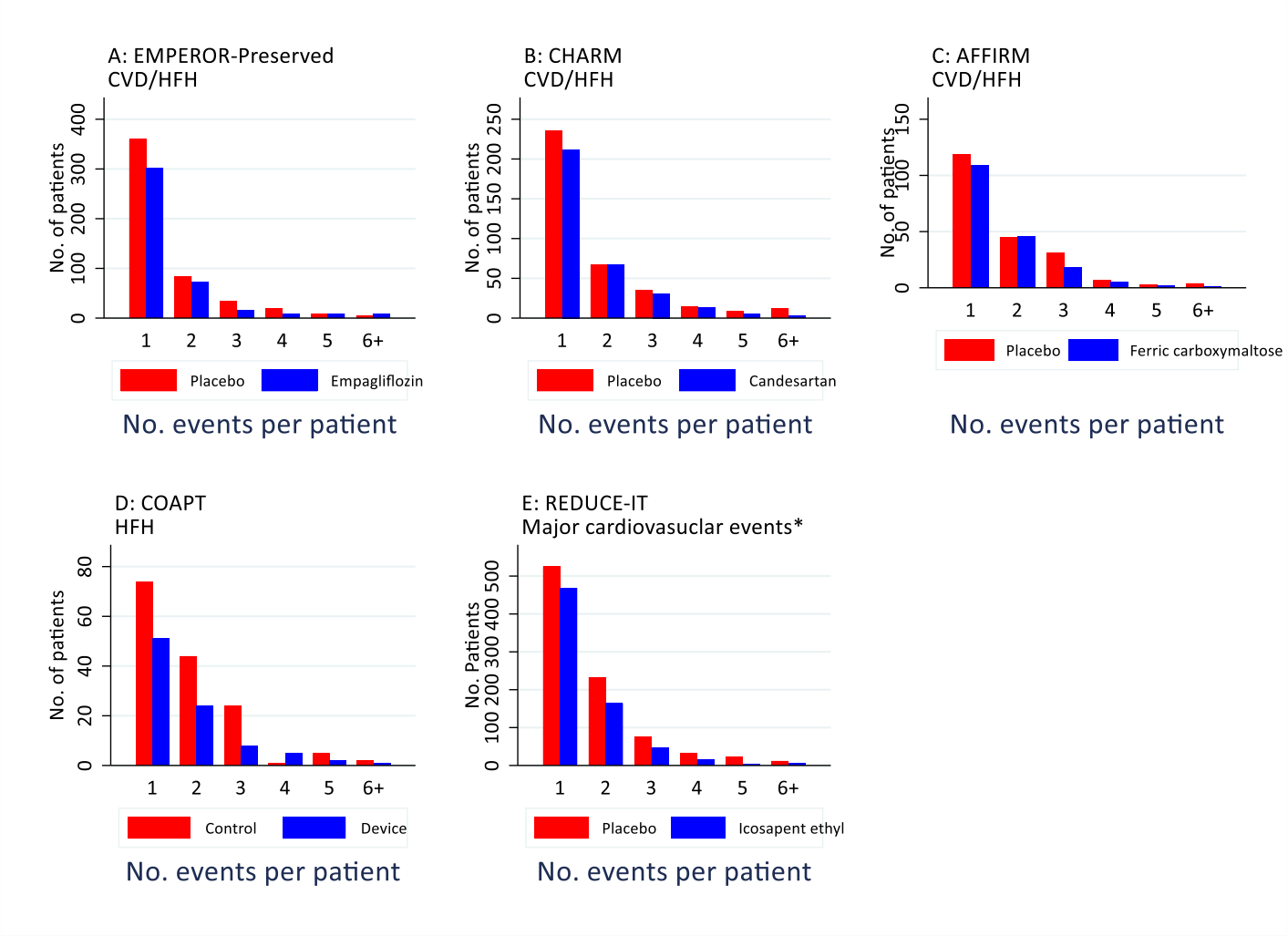
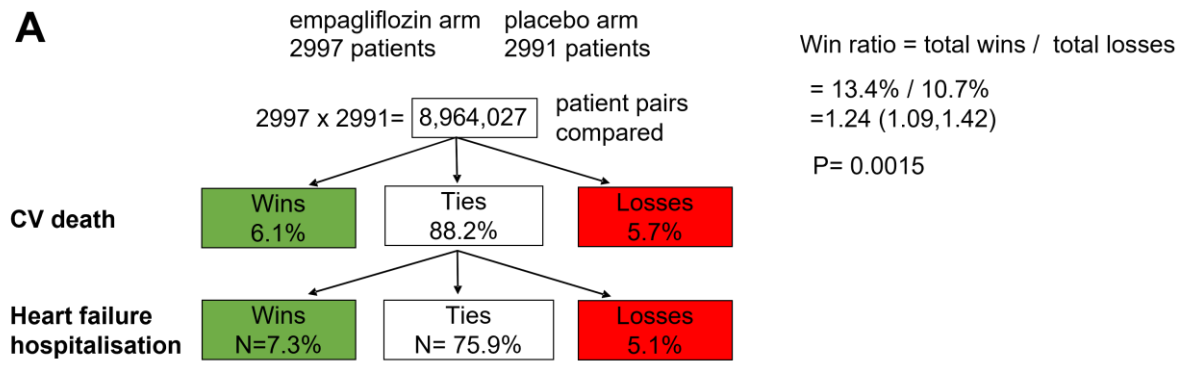


Figure 3: Results of analyses using the win ratio

Caption: Every patient in the active group is compared to every patient in the control group. For each pair, it is determined whether the patient in the active group 'wins' or 'loses' on cardiovascular death. If they are 'tied', then the patients are compared based on the number of heart failure Hospitalization. In AFFIRM-AHF, patients are compared based on their KCCQ at 12 weeks if they are tied based on CV deaths and number of heart failure hospitalization.

(A) EMPEROR-Preserved



(B) AFFIRM-AHF

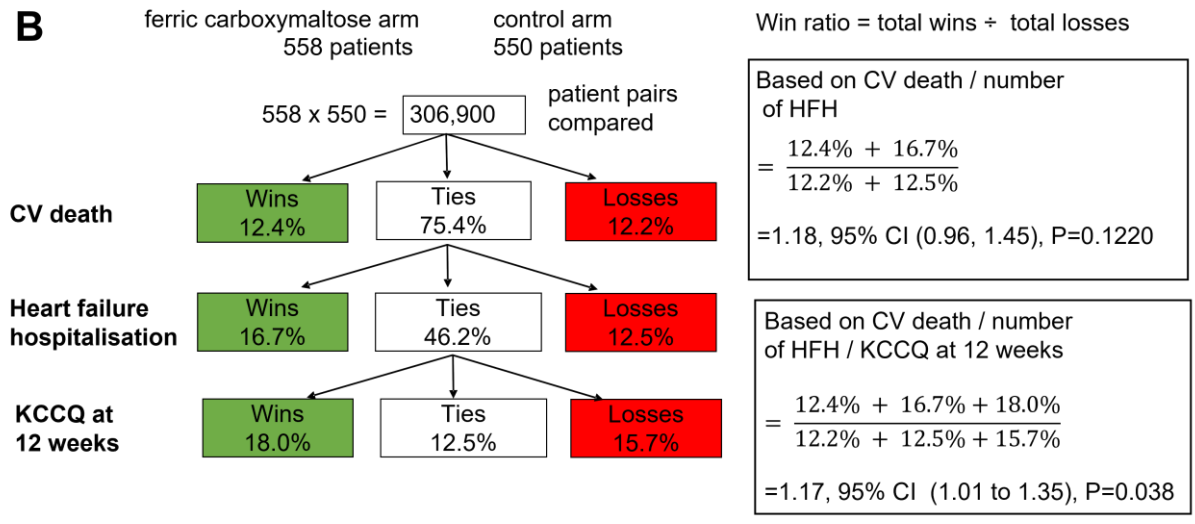


Figure 4: Area under the curve (AUC) in EMPEROR-Preserved.

Caption: A representation of the difference in AUC as the difference in areas under the mean cumulative function (i.e. the shaded area). The total size of this area is 1.1 event-months (95% CI 0.5-1.7). Alternatively, one can calculate the ratio of the AUCs for each arm (the AUC for the candesartan group divided by the AUC in the placebo group), which is equal to 0.76 (95% CI 0.66-0.88) .

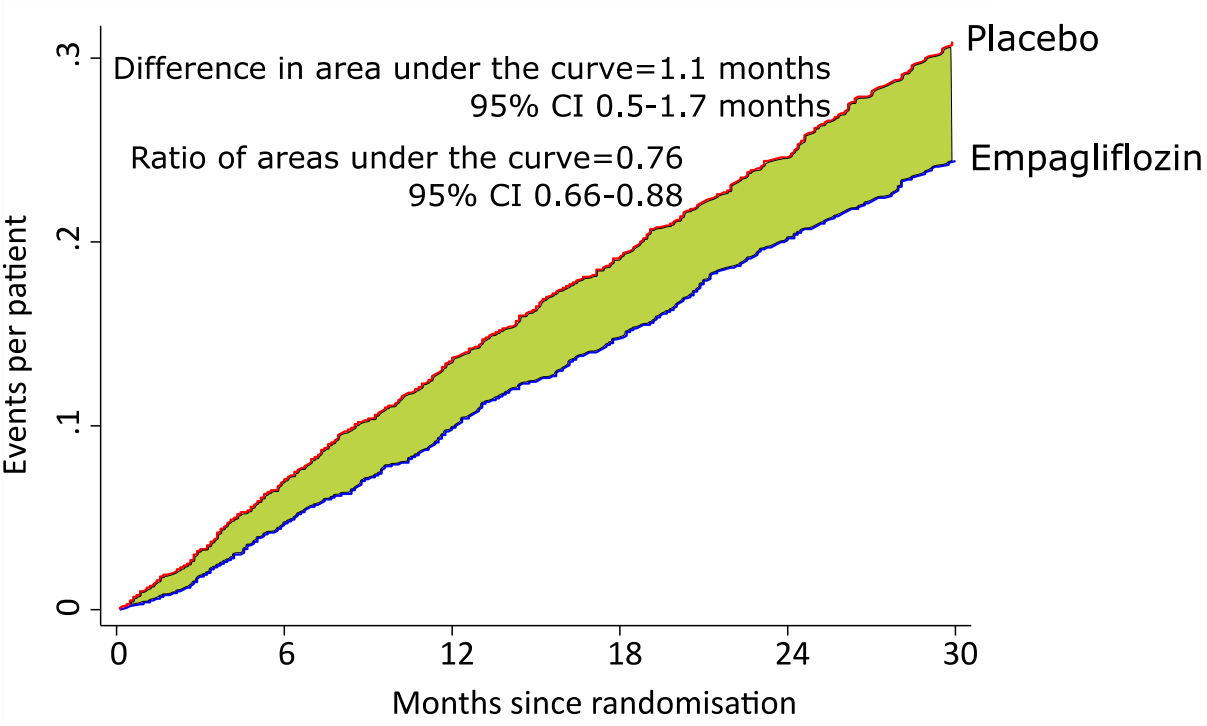


Table 1: Description of methods for repeat events analysis

*Note Lei Wei Yang Ying and Negative Binomial can either include or exclude fatal events

Method	Description
Lin Wei Yang Ying (LWYY)*	Extension to the commonly used Cox proportional hazards model for first events to allow inclusion of recurrent events. Assumes proportional hazards. Uses robust standard errors to allow for valid inference when the underlying risk of events varies between patients.
Negative binomial*	Extension to Poisson regression. Assumes constant rate of events over time. More appropriate than Poisson regression for modelling recurrent events because it allows the underlying risk of an event to vary between patients. Also provides estimates of rates and rate differences.
Win ratio	Compares outcomes between all pairs of patients: each in treatment and control groups. An algorithm to clearly define a 'better outcome' is pre-specified, with a hierarchy of components reflecting clinical priorities. For each pair, it is a "win" if the treated patient has the better outcome, and a "loss" if the control patient has the better outcome.
Joint frailty	Models non-fatal events (including repeats) allowing for variation in patient risk which is also linked to the risk of a fatal event, thereby helping to adjust for competing events. Encompasses a family of statistical models: the exact model to be employed should be pre-specified.
Area under the curve	For every event, the amount of time spent between the event occurring and a fixed milestone time is calculated. The event-time is summed across all events occurring for each patient. The method is called 'the area under the curve' because it is equal to the area under the cumulative event rate curve

Table 2: Pros and cons of methods for repeat events and time-to-first event analyses.

Method	Can be used with	Summary measure	Pros	Cons
Cox proportional hazards	First event only	Hazard ratio for first event	• Simple • Sometimes more powerful than including repeat events • No need to adjudicate recurrent events	• Ignores repeat events
Lin Wei Yang Ying (LWYY)	All types of repeat events	Hazard for total events	• Flexible	• Counts fatal and non-fatal outcomes equally
Negative binomial	Repeat events occurring at a constant rate	Rate ratio for total events	• Simple • Useful for quantifying treatment benefit: can calculate rates as well as rate ratios	• Assumes a constant event rate • Counts fatal and non-fatal outcomes equally
Win ratio	Any type of outcome	Win ratio: odds of a better outcome in a treated patient compared to a control patient	• Flexible • Prioritizes the most important outcomes • Can combine clinical events with quantitative outcomes (e.g. quality-of-life measures)	• Less powerful if treatment less effective on death • Need to decide priority of outcomes • Interpretation less intuitive than some other methods
Joint frailty	Non-fatal repeat events with a fatal competing risk	Hazard ratio for total non-fatal events	• Naturally handles competing risk of death	• Technically difficult to implement • Death cannot be part of the outcome
Area under the curve	All types of repeat events	Ratio or difference in time after events occur until a fixed milestone time	• Assumption free • Emphasis on early events can be helpful in populations with a poor prognosis	• Discards events after a fixed milestone time • Too much emphasis on early events in populations with a good prognosis

Table 3: Time-to-first event analysis and repeat event analyses in five cardiovascular trials

Trial	Primary outcome	First events (active vs. control) / subsequent events (active vs. control)	Analysis considering only first event	Analyses using information on repeat events				
			Cox proportional model (first event only) Hazard ratio	LWYY model hazard ratio	Negative binomial model rate ratio	Joint frailty model hazard ratio	Win ratio	Ratio of AUC (areas under the curve)
EMPEROR-Preserved[29]	CVD or HFH	926 (415 vs. 511) / 485 (211 vs. 274)	**0.79 (0.69–0.90); p<0.001; z=3.62**	0.79 (0.68-0.92) p=0.0029; z=2.98	0.78 (0.66-0.93) P=0.0044; z=2.85	N/A-contains a fatal event	1.25 (1.08-1.42) P=0.001; z=3.22	0.76 (0.66-0.88) p=0.002; z=3.12
CHARM-Preserved[30]	CVD or HFH[1]	704 (331 vs. 373) / 512 (205 vs.307)	**0.86 (0.74-1.00) p=0.050; z=1.96**	0.78 (0.65-0.93) p=0.0064; z=2.73	0.76 (0.62-0.92) p=0.0070; z=2.70	N/A-contains a fatal event	1.15 (0.99-1.34) p=0.0617; z=1.81	0.73 (0.56-0.95), p=0.020; z=2.32
AFFIRM-AHF[7]	CVD or HFH	390 (181 vs. 209) / 355 (112 vs. 163)	HR 0·80 (0·66–0·98) p=0·030; z=2.17	0.77 (0.62-0.95) p=0.014, z=2.45	**0.79 (0·62–1·01) p=0·059; z=1.88**	N/A-contains a fatal event	1.18 (0.96-1.45) p=0.122; z=1.55	0.76 (0.57-1.01) p=0.058; z=1.89
COAPT[6]	HFH	243 (92 vs. 151) / 200 (68 vs. 132)	0.52 (0.40 to 0.67) p<0.001; z=5.0	0.54 (0.41 to 0.71); p<0.001; z=4.4	0.49 (0.36 to 0.67) p<0.001; z=4.5	**0.51 (0.39 to 0.67) p<0.001; z=4.8**	2.01 (1.53-2.63) p<0.001; z=5.0	0.53 (0.42-0.66); p<0.001; z=4.7
REDUCE-IT[32, 33]	CVD, stroke, MI, unstable angina or coronary revascularis-ation	1606 (705 vs. 901) / 1016 (371 vs.645)	**0.75 (0.68-0.83) p<0.001; z=5.65**	0.69 (0.62 –0.77) p<0.001; z=6.56	0.70 (0.62–0.78) p<0.001; z=6.30	N/A-contains a fatal event	1.34 (1.21 - 1.48) p<0.001; z=5.53[2]	0.71 (0.62-0.82) p<0.001; z=4.56

Results **in bold and underlined** font are the primary analysis for each trial**.**

[1]Investigator reported rather than adjudicated events were used for this analysis.

[2]Using a hierarchy of CV death then number of non-fatal CV events.

Abbreviations: AFFIRM-AHF= A Randomised, Double-blind Placebo Controlled Trial Comparing the Effect of Intravenous Ferric Carboxymaltose on hospitalizations and Mortality in Iron Deficient Subjects Admitted for Acute Heart Failure; AUC=Area Under the Curve; EMPEROR-Preserved :The Empagliflozin Outcome Trial in Patients with Chronic Heart Failure and a Preserved Ejection Fraction (EMPEROR-Preserved); CHARM = Candesartan in Heart failure Assessment of Reduction in Mortality and morbidity; COAPT= Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients with Functional Mitral Regurgitation Trial; CVD=Cardiovascular death; HFH=heart failure hospitalization; LWYY =Lin Wei Yang Ying; REDUCE-IT= Reduction of Cardiovascular Events with Icosapent Ethyl–Intervention Trial

Table 4: Summary of recent heart failure trials using repeat events analyses

	Investigational product	Repeat events method used	Repeat events analysis primary?	Endpoint used in repeat events analysis*	Results from repeat events analysis			Results using Cox regression for the time-to-first-event		
					Effect estimate (95% CI)	Z statistic	P value	Hazard ratio (95% CI)	Z statistic	P value
AFFIRM-AHF[7]	Ferric Carboxymaltose	Negative binomial	Yes	CVD and total HFH	0·79 (0·62–1·01)	1.88	0.059	0.80 (0.66 to 0.98)	2.17	0.03
DELIVER[34]	Dapagliflozin	LWYY	No	CVD and total heart failure events	0.77 (0.67 to 0.89)	3.60	<0.001	0.82 (0.73 to 0.92)	3.36	<0.001
DAPA-HF[35]	Dapagliflozin	LWYY	No	CVD and total HFH	0.75 (0.65 to 0.88)	3.72	<0.001	0.75 (0.65 to 0.85)	4.20	<0.001
EMPEROR-Reduced[36]	Empagliflozin	Joint frailty	No	Total HFH	0.70 (0.58 to 0.85)	3.61	<0.001	0.69 (0.59 to 0.81)	4.41	<0.001
EMPEROR-Preserved[29]	Empagliflozin	Joint frailty	No	Total HFH	0.73 (0.61-0.88)	3.32	<0.001	0.71 (0.60-0.83)	4.13	<0.001
EMPULSE[23]	Empagliflozin	Win ratio	Yes	Hierarchical: (1) Death (2) Total HFH (3) time-to-first HFH (4) KCCQ	1.36 (1.09 to 1.68)	2.78	0.0054	No comparable results		
GUIDE-HF[37]	Implantable PAP monitor	LWYY	Yes	Death and total heart failure events	0.88 (0.74 to 1.05)	1.41	0.16	No comparable results		
IRONMAN[38]	Ferric Carboxymaltose	LWYY	Yes	CVD or total HFH	0.82 (0.66-1.02)	1.81	0.070	0.84 (0.70-1.02)	1.74	0.081
PARAGON-HF[5]	Sacubritil-Valsartan	LWYY	Yes	CVD or total HFH	0.87 (0.75 to 1.01)	1.89	0.058	0.92 (0.81 to 1.03)	1.43	0.153
REDUCE LAP-HF II[39]	Interatrial shunt	Win ratio	Yes	Hierarchical: (1) CV death or ischemic stroke, (2) Total HFH (3) KCCQ	1.0 (0.8-1.2)	0.19	0.85	No comparable results		
SCORED[3]	Sotagliflozin	LWYY	Yes	CVD and total heart failure events	0.74 (0.63 to 0.88)	3.53	<0.001	0.77 (0.67 to 0.90)	3.47	<0.001
SOLOIST-WHF[4]	Sotagliflozin	LWYY	Yes	CVD or total HFH	0.68 (0.52 to 0.88)	2.98	0.0039	0.71 (0.56 to 0.89)	2.89	0.0029
STRONG-HF[40]	High intensity vs. standard care*	Win Ratio	No	Hierarchical: (1) CV death (2) Total HFH (3) EQ5D	1.37 (1.16 t o1.62)	3.72	<0.001	No comparable results		
VICTORIA[41]	Vericiguat	LWYY	No	Total HFH	0.91 (0.84 to 0.99)	2.25	0.025	0.90 (0.81 to 1.00)	1.96	0.05

Trials published in NEJM, Lancet or Nature Medicine between July 1st 2019-January 1st 2023. All time to first event analyses used Cox proportional hazards models.
A list of study abbreviations and references can be found in Appendix C. CVD=cardiovascular death; HFH=heart failure hospitalization: LWYY=Lin Wei Yang Ying

* We used the highest priority outcome (according to the trial's testing strategy) with both time-to first and repeat events analyses. Where there were no comparable analyses, we present the highest priority repeat events analysis. ** Z-statistics are calculated either from the p-value, or where an exact p-value was not reported, we divided the reported log hazard ratio by its estimated standard error (calculated from the 95% confidence on the log scale)