NEJM Evidence

Published June 30, 2022
NEJM Evid 2022; 1 (10)
DOI: 10.1056/EVIDoa2200047

ORIGINAL ARTICLE | INTERFACE SERIES

Quantifying Treatment Effects in Trials with Multiple Event-Time Outcomes

Brian Lee Claggett, Ph.D.,[1] Zachary R. McCaw, Ph.D,[2] Lu Tian, Sc.D.,[3] John J. V. McMurray, M.D.,[4] Pardeep S. Jhund, M.B. Ch.B., Ph.D.,[4] Hajime Uno, Ph.D.,[5] Marc A. Pfeffer, M.D., Ph.D.,[1] Scott D. Solomon, M.D.,[1] and Lee-Jen Wei, Ph.D.[6]

Abstract

BACKGROUND Data on the occurrence times of multiple outcomes, reflecting the temporal profile of disease burden/progression, have been used to estimate treatment effects in various recent randomized trials. Most procedures for analyzing these data require specific model assumptions. When the assumptions are not met, the results may be misleading. Robust, model-free procedures for study design and analysis that enable clinically meaningful interpretations are warranted.

METHODS For each treatment group, we constructed and summarized the estimated mean cumulative count of events over time by the area under the curve (AUC), which can be interpreted as the mean total event-free time lost from multiple undesirable outcomes. A higher curve, and resulting larger AUC, implies a worse treatment. The treatment effect is quantified by the ratio and/or difference of AUCs. The timing and occurrence of recurrent heart failure hospitalizations (HFHs) and cardiovascular (CV) death from Prospective Comparison of ARNI with ARB Global Outcomes in HF with Preserved Ejection Fraction (PARAGON-HF), comparing sacubitril/valsartan with valsartan, are presented for illustration. We also discuss the design of future studies on the basis of the proposed method.

RESULTS With 48 months of follow-up, estimated AUCs, representing the total event-free time lost to HFHs and CV death, were 11.3 and 13.1 event-months for sacubitril/valsartan and valsartan, respectively. The ratio of these AUCs was 0.86 (95% confidence interval, 0.75 to 1.00; P=0.049), a 14% reduction of disease burden favoring combination therapy. A future study, similar to PARAGON-HF, designed using the new proposal would require fewer patients would than a conventional time-to-first-event analysis.

CONCLUSIONS The proposed method is robust and model-free and provides a clinically interpretable, time-scale summary of the treatment effect. (Funded by National Institutes of Health.)

C. Corey Hardin, M.D., Ph.D.,
Interface Series Editor

Drs. Claggett and McCaw contributed equally to this article.

The author affiliations are listed at the end of the article.

Dr. Wei can be contacted at wei@hsph.harvard.edu or at Department of Biostatistics, Harvard University, 655 Huntington Ave., Boston, MA 02115.

The Problem

As the standard of care for a variety of conditions has improved, many formerly deadly conditions with short life expectancies have been transformed into chronic conditions characterized by multiple nonfatal clinical events. This has raised doubts about the value of conventional time-to-first-event analyses as an appropriate way to examine the disease course and to assess treatment efficacy.

The statistical properties of multiple event-time analysis, including the potential for increased statistical power, have been discussed extensively.[1-5] However, the validity of most existing methods depends on strong modeling assumptions (i.e., an assumption that has a reasonable likelihood of not being correct). If these assumptions are not met, the resulting estimates of treatment efficacy are difficult to interpret.[6] Various recent clinical studies have used multiple event-time outcomes as their end points (Table 1). The methods used in these trials all require some degree of modeling, and their modeling assumptions can be difficult to justify empirically. Moreover, patients in such trials may experience informative terminal events, such as premature study discontinuation or death, which prevent the occurrence of subsequent events of interest.[14] Most existing methods do not handle this challenging situation appropriately.

A Potential Solution Illustrated by Example

In this article, we recommend a simple, model-free, and clinically interpretable procedure that is a straightforward extension of the mean survival time, a summary measure used to quantify the effect of treatment in the time-to-first-event setting.[15-17] Our proposal appropriately accounts for terminal events, including the case in which a terminal event (e.g., cardiovascular [CV] death) is one of the events of interest. We use recurrent event-time data from the recent PARAGON-HF[7] trial to illustrate the limitations of existing analytic procedures and then present an alternative procedure that overcomes these limitations.

In PARAGON-HF, a trial comparing the treatment of patients with heart failure with sacubitril/valsartan versus valsartan alone, the primary end point was the timing and occurrence of all HFHs, first and recurrent, as well as CV death. The primary analysis compared 2407 patients in the sacubitril/valsartan arm with 2389 patients in the valsartan arm.[5] The patients were followed for a median of 35 months. During the study period, 690 and 797 HFHs were observed for the sacubitril/valsartan and valsartan arm, respectively. There were 405 and 433 patients who had at least one HFH. Moreover, there were 204 and 212 CV deaths and 138 and 137 non-CV deaths in the sacubitril/valsartan and valsartan groups, respectively.

TYPICAL MULTIPLE-OUTCOME PATIENT PROFILES

As in many clinical settings, the occurrence of a terminal event in PARAGON-HF, namely death, precludes the observation of subsequent HFHs.[7] Figure 1 presents various event patterns for patients from this trial with a maximum of 48 months of follow-up. For case 1, the patient experienced two recurrent HFHs before administrative censoring at month 48. For case 2, observation was censored at month 12; for example, because of late entry into the study, without having experienced any fatal or nonfatal events. For case 3, the patient experienced HFH at month 12 and died at

Table 1. A Sample of Clinical Trials Using Multiple or Recurrent Event-Time End Points.*

Trial	Treatments	Multiple Event-Time Outcome	First Author and Reference
PARAGON-HF	Sacubitril/valsartan vs. valsartan	Total HFH and CV death	Solomon et al.[7]
L-Glutamine	L-Glutamine vs. placebo	Sickle cell pain crises	Niihara et al.[8]
Voxelotor	Voxelotor vs. placebo	Sickle cell pain crises	Vichinsky et al.[9]
PARADISE-MI	Sacubitril/valsartan vs. ramipril	Total no. of HFHs, outpatient HF events, and CV death	Pfeffer et al.[10]
SCORED, SOLOIST	Sotagliflozin vs. placebo	Recurrent HFH, nonfatal MI, nonfatal stroke, and CV death	Bhatt et al.[11,12]
ACT DMD	Ataluren vs. placebo	North Star Ambulatory Assessment	McDonald et al.[13]

* ACT DMD denotes Phase 3 Study of Ataluren in Participants With Nonsense Mutation Duchenne Muscular; CV, cardiovascular; HF, heart failure; HFH, heart failure hospitalization; MI, myocardial infarction; PARADISE-MI, Prospective ARNI vs ACE Inhibitor Trial to DetermIne Superiority in Reducing Heart Failure Events After MI; PARAGON-HF, Prospective Comparison of ARNI with ARB Global Outcomes in HF with Preserved Ejection Fraction; SCORED, Effect of Sotagliflozin on Cardiovascular and Renal Events in Patients With Type 2 Diabetes and Moderate Renal Impairment Who Are at Cardiovascular Risk; and SOLOIST-WHF Trial, Effect of Sotagliflozin on Cardiovascular Events in Patients With Type 2 Diabetes Post Worsening Heart Failure.

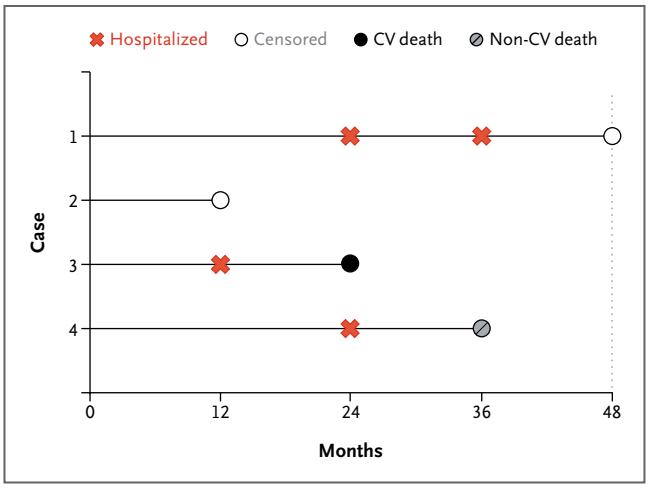

Figure 1. Typical Patterns for the Time-to-Heart-Failure Hospitalization or CV Death from PARAGON-HF.

CV denotes cardiovascular.

month 24 because of CV-related complications. For case 4, the patient experienced HFH at month 24 before dying at month 36 from a non-CV cause. In PARAGON-HF, non-CV death was treated as independent censoring for time to the event of interest. An alternative is to treat non-CV death as a competing risk for the events of interest,[14] HFH and CV death, which is the approach we take in this article.

QUANTIFYING INDIVIDUAL PATIENTS' TEMPORAL RECURRENT-EVENT PROFILES

The cumulative event count curve for each patient from Figure 1 is displayed in Figure 2. For case 1, the curve has two one-unit jumps because HFHs occur at months 24 and 36. After the first HFH, this patient's disease burden was increased for the remaining 48 − 24 = 24 months of the study. In other words, the patient "lost" 24 months of event-free time because of the first HFH, meaning that without the HFH occurring at month 24, the patient would have spent the last 24 months of the study at a lower cumulative disease burden. The patient lost an additional 12 months of potential event-free time because of a second HFH occurring at 36 months, for a total of 24 + 12 = 36 event-months lost across the 48 months of follow-up. The total event-free time lost is the area under the curve (AUC) up to 48 months. For case 2, the patient was censored at 12 months without having experienced an event. Thus, it is only known that the patient's 48-month AUC is greater than or equal to 0. For case 3, the patient had a one-unit jump at month 12, because of an HFH, and another at month 24, because of CV death. The patient lost 36 months of potential event-free time because of the HFH plus an additional 24 months of potential event-free time because of CV death. The total event-free time lost was therefore 36 + 24 = 60 event-months. For case 4, the patient lost 24 months of potential event-free time because of the HFH occurring at month 24. Because the non-CV death occurring at month 36 was not part of the trial's outcomes of interest, its occurrence did not increment the cumulative event curve with respect to the CV disease burden/progression. Note that for cases 3 and 4, the patients died during the study and their curves remained flat thereafter, indicating the occurrence of no further events of interest.

Visually, a higher curve indicates a worse cumulative experience across the study for any given patient. Correspondingly, the larger the AUC, the worse the patient's experience. Note that, for an individual patient, the AUC is the total event-free time lost from all recurrent events or CV death. If the patient experiences numerous HFHs during follow-up, their total event-free time lost can exceed the duration of follow-up.

QUANTIFYING TREATMENT EFFECT WITH RECURRENT EVENT-TIME DATA

The PARAGON-HF trial considered the timings of recurrent HFHs and CV death as the end point. In the absence

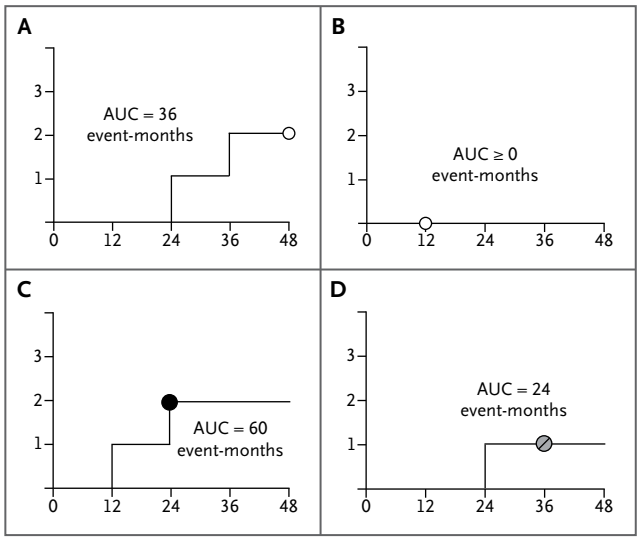

Figure 2. Cumulative Event Curves and AUCs, Corresponding to the Typical Patient Profiles from Figure 1.

Case 1 (Panel A), Case 2 (Panel B), Case 3 (Panel C), and Case 4 (Panel D) are shown. AUC denotes area under the curve. See text for details.

of censoring, the individual patient counting curves in Figure 2 could be directly aggregated by averaging across all patients in each treatment arm. However, in the presence of censoring, directly averaging the counting curves in Figure 2 is not possible because the trajectories of patients like case 2 are unobserved after censoring. Although these patients experienced neither HFH nor CV death, their time spent at risk before censoring contributes to estimating the mean cumulative count curve.

In the presence of censoring, the mean cumulative count curve could be estimated using the method proposed by Nelson,[18] provided there are no terminal events, such as death. However, in PARAGON-HF, there were 342 deaths in the sacubitril/valsartan arm (204 CV and 138 non-CV) and 349 deaths in the control arm (212 CV and 137 non-CV). Following death, a patient is no longer contributing to the recurrent HFH data. For this case, the mean cumulative count curve can instead be estimated using a modified version of Nelson's estimator introduced by Ghosh and Lin.[19] The Ghosh–Lin estimator provides the mean cumulative count curve for HFHs over time while treating death from any cause (i.e., CV and non-CV) as a competing risk, but without including CV death as an event of interest. To accommodate the current case, we modify the Ghosh–Lin estimator by counting CV death as an event of interest and also as a competing risk for nonfatal events. A detailed description of this estimate is provided in the Supplementary Appendix available with the full text of this article, and software for performing the analysis is freely and publicly available at https://github.com/zrmacc/MCC.

Figure 3A presents the mean cumulative count curves for PARAGON-HF. At 48 months, the estimated mean counts are 0.55 and 0.49 for valsartan and sacubitril/valsartan, respectively. Intuitively, the higher the curve, the worse total morbidity and mortality are for that therapy. The curve for sacubitril/valsartan was lower than that for valsartan over the 48 months of follow-up, suggesting visually and numerically that the new treatment was better than control. Figure 3B and 3C provide the areas under the individual mean cumulative count curves, which were 13.1 and 11.3 event-months for valsartan and sacubitril/valsartan, respectively. Comparing the treatment arms with respect to their mean cumulative counts at 48 months is analogous to cumulative incidence rates at the end of follow-up in the time-to-first-event setting, and comparing the curves with respect to their AUCs is analogous to comparing the 48-month mean event-free times lost.[15-17]

For PARAGON-HF, one of the analyses was based on a procedure proposed by Lin, Wei, Yang, and Ying (LWYY).[20] This analysis required an assumption that the ratio of the mean cumulative count curves between the two arms was constant across the 48 months of follow-up and was stratified by geographic region. In this study, for ease of exposition, we present an unstratified LWYY analysis. Detailed discussion of stratified analysis is deferred to the Supplementary Appendix. Recently, LWYY has also been applied in several other clinical studies.[2,8-13] For the end point of recurrent HFH or CV death from PARAGON-HF, the (unstratified) LWYY rate ratio is 0.88 (95% confidence interval [CI], 0.76 to 1.01; P=0.075) in favor of sacubitril/valsartan. That is, sacubitril/valsartan is estimated to reduce the cumulative rate of recurrent HFH or CV death by 12% relative to valsartan.

The LWYY rate ratio requires that the ratio of the mean cumulative count curves comparing the two treatment arms is constant across time. This requirement is similar to the proportional hazards assumption in the case of a time-to-first-event outcome.[21] If this assumption is not met, the resulting ratio estimate cannot be interpreted as a simple average of ratios over time and has no direct clinical interpretation.[22] This is undesirable and may provide misleading conclusions regarding treatment efficacy. Although the constant ratio assumption may be checked through conventional goodness-of-fit tests, such tests are typically underpowered for detecting violations of modeling assumptions. Moreover, such tests can only "fail to detect" a significant violation; there is no P value that would actually indicate that the model assumptions have been satisfied.[23]

As discussed above, the areas under the cumulative count curves in Figure 3B and 3C provide heuristic summaries of treatment efficacy, with a greater AUC reflecting a greater average event rate across the follow-up period. Moreover, the AUC has a clinically meaningful interpretation as the mean total event-free time lost during the course of follow-up. This is the expected total time that a patient is under an increased disease burden across all outcomes under consideration. The mean is obtained as the area under the mean cumulative count curve and is conceptually equivalent to the average of the AUCs for the individual patient-counting processes from Figure 2, accounting for censoring. Having obtained the AUC for each treatment arm, the difference and ratio of AUCs then quantifies the absolute and relative treatment effect in an assumption-free manner. The key idea of quantifying treatment efficacy

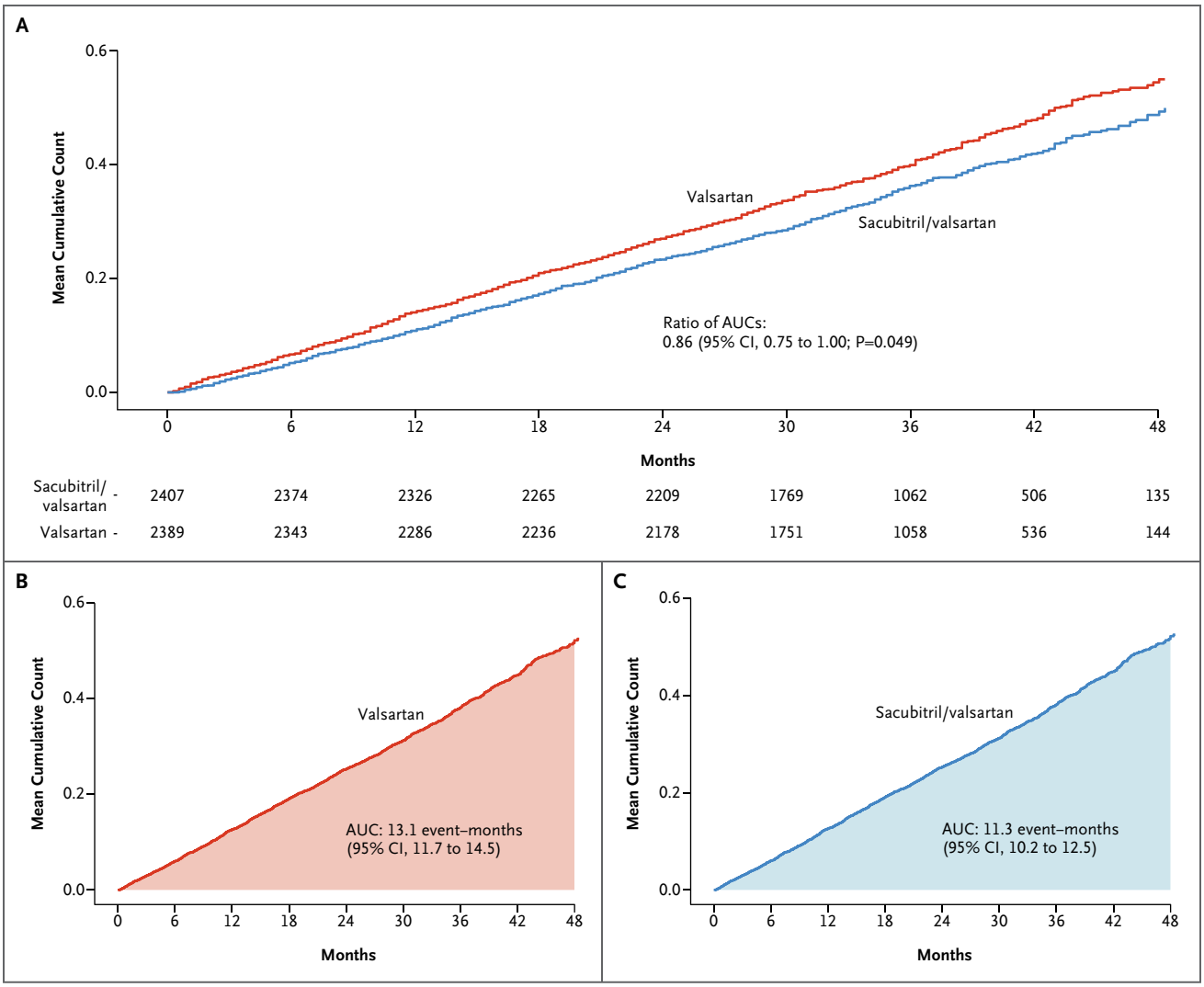

Figure 3. Mean Cumulative Count Curves for Heart Failure Hospitalizations and CV Death from PARAGON-HF.

Mean cumulative count curves for heart failure hospitalizations and cardiovascular (CV) death from PARAGON-HF are shown (Panel A). Non-CV death was regarded as a competing risk. Total event-free time lost for valsartan and sacubitril/valsartan through the area under the curve (AUC) is presented in Panels B and C. CI denotes confidence interval.

through an AUC is a natural extension of summarizing the survival curve by its AUC, which is the restricted mean survival time when dealing with a time-to-first-event outcome.[15-17]

The 48-month AUCs were 11.3 event-months for sacubitril/valsartan and 13.1 event-months for valsartan. That is, across 48 months, the total event-free time lost because of HFHs and CV death was, on average, 11.3 event-months for patients receiving sacubitril/valsartan versus 13.1 event-months for patients receiving valsartan alone. For comparing these therapies, the difference of AUCs was 1.8 event-months (95% CI, −0.01 to 3.6), while the ratio was 0.86 (95% CI, 0.75 to 1.00; P=0.049), both in favor of sacubitril/valsartan. Thus, on average, patients spent 1.8 fewer months under an increased disease burden with sacubitril/valsartan, and the corresponding total event-free time lost was reduced by 1 − 0.86 = 14% relative to valsartan alone. It is interesting to note that this ratio is quite similar to that obtained with LWYY. The difference and ratio of AUCs have clear clinical interpretations in terms of the absolute and relative reduction in total

morbidity and mortality. Moreover, the analysis based on AUCs provides a valid estimate without requiring any modeling assumptions.

The primary analysis of PARAGON-HF was based on LWYY stratified by geographic region.[7] The validity of this analysis needs a further modeling assumption, that the ratios of the mean cumulative count curves (as in Fig. 3A) are constant across all regions. The Supplementary Appendix explains how the AUC procedure introduced above can be extended to account for stratification.

Points for Trial Design

We now illustrate how to design a randomized clinical trial using a multiple event-time end point to evaluate a hypothetical new therapy versus sacubitril/valsartan in a target patient population similar to that of PARAGON-HF. For each patient, the primary end point consists of the times to recurrent HFH and CV death. Suppose that patient accrual is uniform over the course of 2 years, with an additional 2 years of follow-up after randomly assigning the last patient. The data will be analyzed using the method discussed in this article, with events collected for up to 45 months. In PARAGON-HF, the observed 45-month AUC for sacubitril/valsartan was nearly 10 months. Assume that the anticipated effect of the new therapy is a 20% reduction of the total event-free time lost, resulting in a 45-month AUC of 8 months.

A simulation study can be used to estimate the study size needed to provide 80% power for detecting the anticipated treatment difference. Specifically, we propose a joint parametric model for the recurrent HFH, CV death, and non-CV death times. Parameters for the sacubitril/valsartan arm are estimated from the observed PARAGON-HF data. Appropriate parameters for the new therapy arm are selected such that the 45-month AUC in the new arm is approximately 8 months. Details of the proposed model are presented in the Supplementary Appendix. Assuming a 1:1 randomization ratio, we select an initial sample size (e.g., 2000 patients per arm) and simulate a large number of data sets (e.g., 1000) from the parametric model. Empirical power at the selected sample size is the proportion of data sets in which we detect a significant treatment difference (i.e., two-sided $P < 0.05$). If the power is less than 80%, we increase the sample size in each arm; if greater, we decrease the sample size.

Using this procedure, we find that a sample size of 2300 patients per arm (4600 patients in total) would provide 80% power for detecting a 20% improvement in the 45-month AUC over sacubitril/valsartan. At this sample size, a total of 1506 HFHs and CV deaths are expected. For context, a similar study based on time-to-first event analysis would observe only 993 first events and have 72% power. To achieve 80% power, the time-to-first events analysis would require a sample size of 2750 patients per arm (5500 patients in total), 20% more than the multiple event-time design. While these numbers are specific to the assumptions made regarding the data-generating process, the sample size required for other settings is readily determined by modifying factors such as the anticipated treatment difference, the accrual rate, and the administrative censoring or loss-to-follow-up patterns. Note that, as an alternative to targeting a particular power for detecting the treatment difference, we could estimate the sample size needed to make the CI for the treatment difference sufficiently narrow (e.g., the sample size needed for the CI for the ratio in AUCs to have width less than 25%).

Discussion

In the presence of terminal events, whether or not included as events of interest, the estimation of the mean cumulative count curve defined in this article must account for competing risks.[14] Specifically, when a terminal event such as CV death forms part of the primary end point, it is essential to distinguish such events from nonterminal events of interest (e.g., HFH). A patient can remain at risk for future events after an HFH but must be removed from the risk set after CV death. Moreover, removal from the risk set due to censoring, which leaves open the possibility that the patient may have experienced additional events in the future (even if those events are not observed), is different from removal from the risk set due to death, which precludes the possibility of any future events.

Another analysis of potential interest is to compare the treatment arms with respect to HFHs only, treating death from any cause (CV or non-CV) as a competing risk. For PARAGON-HF, the results of this analysis are very similar to those obtained while including CV death as an event of interest. This is likely because the rates of CV death are low and balanced across the treatment arms. Moreover, the proposed analysis is readily applicable in the absence of terminal events. The estimation procedure provided at https://github.com/zrmacc/MCC easily accommodates both the presence and the absence of terminal events.

The proposed approach for comparing mean cumulative count curves through the difference or ratio of AUCs is a direct extension of summarizing a cumulative incidence curve, with or without competing risks, by its AUC when dealing with at most one event per patient. For this simple case, the AUC is the restricted mean time lost because of the undesirable event.[15-17] For AUC-based analyses, it is important to select a time window for analyzing the data. Ideally, this would be prespecified in the study protocol, but it may be chosen empirically at analysis time with clinical justification.[24]

As for all existing methods of recurrent event-time analysis, a limitation of the proposed analysis is that conclusions regarding treatment efficacy may be difficult to interpret if the terminal event profiles are imbalanced across treatment arms. For example, if sacubitril/valsartan had significantly extended patients' survival versus valsartan in PARAGON-HF, then patients in the sacubitril/valsartan arm would have had longer exposure periods during which to experience HFHs. This can introduce a "survival bias" that penalizes a treatment effective for extending survival.[25] For this case, the method proposed by Claggett et al.[26] may be a useful alternative, although it would limit the number of recurrent events a patient could contribute to the analysis. Further research along this line is needed.

Whether considering time-to-first or recurrent events analysis, many CV trials employ a composite end point, consisting of different types of undesirable outcomes with potentially varying degrees of severity. A general approach to accounting for differing severities is to assign these events different weights. For example, in the individual cumulative event rate plots in Figure 2, we could assign a larger "jump" for CV death than for HFH. Additionally, the size of a jump could be adjusted to depend on the length of hospital stay. For example, an HFH lasting 2 days would result in a smaller jump than an HFH that lasted 1 month. This approach is similar to the construction of a quality-of-life adjusted survival curve.[27,28] However, the assignment of "jump sizes" (that is, a weight for each event) requires careful clinical consideration, and identifying weights that are satisfactory to patients, clinicians, and payers may be difficult. A novel approach for comparing treatment arms with respect to multiple event-time end points while considering severity was recently proposed by Mao.[29] This approach quantifies the net time across follow-up during which patients in the treatment arm occupied a more favorable state than did patients in the control arm on average. Similar to win ratio, an analysis of composite endpoints in clinical trials based on clinical priorities,[30] this procedure does not automatically provide a reference value with which to compare the treatment difference. By contrast, when analyzing AUCs, we obtain not only the treatment difference, but also an estimate of the mean total event-free time lost in each arm.

Another conventional approach to analyzing recurrent events data is to examine incidence rates at the individual patient level. That is, for each patient, consider the number of recurrent events divided by the follow-up time as the study end point. Incidence rate data are commonly analyzed using a Poisson or negative binomial model.[4] These models assume a constant patient-specific incidence rate across time, which may be unrealistic for a condition that progresses. Moreover, when focusing on incidence rates, the occurrence times of the events are ignored. Therefore, this analysis may not effectively or appropriately quantify the extent to which a therapy delays the occurrences of undesirable events across the study period.

Conclusion

We have proposed using the area under the mean cumulative count curve for the design and analysis of comparative clinical trials with a multiple or recurrent event-time end point. This approach is model-free, potentially more efficient than time-to-first-event analysis, and provides a clinically interpretable, time-scale summary of the treatment effect.

Disclosures

Supported by National Institutes of Health Grant R01HL089778 (to Dr. Tian).

Author disclosures and other supplementary materials are available at evidence.nejm.org.

Author Affiliations

[1] Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston

[2] Insitro, South San Francisco

[3] Department of Biomedical Data Science, Stanford University, Stanford, CA

[4] British Heart Foundation Glasgow Cardiovascular Research Center, Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, Scotland

[5] Department of Data Science, Dana-Farber Cancer Institute, Boston

[6] Harvard T.H. Chan School of Public Health, Boston

References

1. Rogers JK, McMurray JJV, Pocock SJ, et al. Eplerenone in patients with systolic heart failure and mild symptoms: analysis of repeat

hospitalizations. Circulation 2012;126:2317–2323. DOI: 10.1161/CIRCULATIONAHA.112.110536.

2. Rogers JK, Pocock SJ, McMurray JJV, 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. DOI: 10.1002/ejhf.29.

3. Mogensen UM, Gong J, Jhund PS, et al. 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;20:760–768. DOI: 10.1002/ejhf.1139.

4. 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. DOI: 10.1161/CIRCULATIONAHA.117.033065.

5. Pfeffer MA, Claggett B, Lewis EF, et al. Impact of sacubitril/valsartan versus ramipril on total heart failure events in the PARADISE — MI trial. Circulation 2022;145:87–89. DOI: 10.1161/CIRCULATIONAHA.121.057429.

6. McCaw ZR, Tian L, Wei J, et al. Choosing clinically interpretable summary measures and robust analytic procedures for quantifying the treatment difference in comparative clinical studies. Stat Med 2021;40:6235–6242. DOI: 10.1002/sim.8971.

7. 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. DOI: 10.1056/NEJMoa1908655.

8. Niihara Y, Miller ST, Kanter J, et al. A phase 3 trial of L-glutamine in sickle cell disease. N Engl J Med 2018;379:226–235. DOI: 10.1056/NEJMoa1715971.

9. Vichinsky E, Hoppe CC, Ataga KI, et al. A phase 3 randomized trial of voxelotor in sickle cell disease. N Engl J Med 2019;381:509–519. DOI: 10.1056/NEJMoa1903212.

10. Pfeffer MA, Claggett B, Lewis EF, et al. Angiotensin receptor–neprilysin inhibition in acute myocardial infarction. N Engl J Med 2021;385:1845–1855. DOI: 10.1056/NEJMoa2104508.

11. 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. DOI: 10.1056/NEJMoa2030186.

12. 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. DOI: 10.1056/NEJMoa2030183.

13. McDonald CM, Wei LJ, Flanigan KM, Elfring G, Trifillis P, Muntoni F. Evaluating longitudinal therapy effects via the North Star Ambulatory Assessment. Muscle Nerve 2021;64:614–619. DOI: 10.1002/mus.27396.

14. McCaw ZR, Claggett BL, Tian L, et al. Practical recommendations on quantifying and interpreting treatment effects in the presence of terminal competing risks: a review. JAMA Cardiol 2022;7:450–456. DOI:10.1001/jamacardio.2021.4932.

15. Kim DH, Uno H, Wei LJ. Restricted mean survival time as a measure to interpret clinical trial results. JAMA Cardiol 2017;2:1179–1180. DOI: 10.1001/jamacardio.2017.2922.

16. McCaw ZR, Yin G, Wei LJ. Using the restricted mean survival time difference as an alternative to the hazard ratio for analyzing clinical cardiovascular studies. Circulation 2019;140:1366–1368. DOI: 10.1161/CIRCULATIONAHA.119.040680.

17. Perego C, Sbolli M, Specchia C, et al. Utility of restricted mean survival time analysis for heart failure clinical trial evaluation and interpretation. JACC Heart Fail 2020;8:973–983. DOI: 10.1016/j.jchf.2020.07.005.

18. Nelson WB. Recurrent-events data analysis for product repairs; disease recurrences, and other applications. Philadelphia: Society for Industrial and Applied Mathematics, 2003.

19. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. Biometrics 2000;56:554–562. DOI: 10.1111/j.0006-341x.2000.00554.x.

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

21. Cox D. Regression models and life-tables. J R Stat Soc Series B Stat Methodol 1972;34:187–220.

22. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. Biometrika 1986;73:363–369.

23. Wasserstein RL, Lazar NA. The ASA's statement on *p*-values: context, process, and purpose. Am Stat 2016;70:129–133. DOI: 10.1080/00031305.2016.1154108.

24. Tian L, Jin H, Uno H, et al. On the empirical choice of the time window for restricted mean survival time. Biometrics 2020;76:1157–1166. DOI: 10.1111/biom.13237.

25. McCaw ZR, Tian L, Sheth KN, Hsu WT, Kimberly WT, Wei LJ. Selecting appropriate endpoints for assessing treatment effects in comparative clinical studies for COVID-19. Contemp Clin Trials 2020;97:106145. DOI: 10.1016/j.cct.2020.106145.

26. Claggett B, Tian L, Fu H, Solomon SD, Wei LJ. Quantifying the totality of treatment effect with multiple event-time observations in the presence of a terminal event from a comparative clinical study. Stat Med 2018;37:3589–3598. DOI: 10.1002/sim.7907.

27. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. Stat Med 1990;9:1259–1276. DOI: 10.1002/sim.4780091106.

28. Gelber RD, Cole BF, Gelber S, Goldhirsch A. Comparing treatments using quality-adjusted survival: the Q-TWiST method. Am Stat 1995;49:161–169. DOI: 10.2307/2684631.

29. Mao L. On restricted mean time in favor of treatment. Biometrics 2021 September 25 (Epub ahead of print). DOI: 10.1111/biom.13570.

30. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. Eur Heart J. 2012;33(2):176–182. DOI: 10.1093/eurheartj/ehr352.