Circulation

PERSPECTIVE

Summarizing Primary Results in Clinical Trials With a Time-to-Event End Point: Complementing Different Measures for a Comprehensive Assessment of Treatment Effect

Andrea Bellavia, PhD; Sabina A. Murphy, MPH

Long-term, phase 3 randomized clinical trials are often designed to prospectively compare the occurrence of a clinical end point between 2 or more treatment arms. The optimal statistical tool to visualize this comparison is the cumulative incidence function curve (or its complement to one, the survival curve), which combines information on the number of patients at risk, the risk of the event (or probability), and the time by which that risk is achieved. Together with presenting these curves by treatment arm, it is common to include an estimate of the hazard ratio (HR) as a summary measure of the treatment effect. In recent years, this approach has been heavily criticized, and various alternatives to the HR have been presented.[1,2] The goal of this commentary is to review the definition and interpretation of some core summary measures in time-to-event analysis, discussing their advantages and limitations and highlighting the different information each of them can provide for an in-depth assessment of treatment effects.

ESTIMATING TREATMENT EFFECTS UNDER DIFFERENT PERSPECTIVES

The Figure presents the cumulative incidence functions for a hypothetical randomized clinical trial comparing time to major adverse cardiovascular events between a novel treatment versus placebo over 5 years of follow-up. A formal treatment comparison can be explored under 3 different perspectives, each of them providing relevant and complementary information on the treatment effect: how treatment is affecting the risk of the event at a given time point (time-specific comparison); how treatment is affecting the difference in time by which a given risk is achieved (risk-specific comparison); and the overall treatment effect over the entire follow-up (overall, or average comparison).

An intuitive and effective way to communicate the clinical benefits of the new treatment is by focusing on the cumulative risk of the event at a specific time point. In the Figure, for example, we see that among participants assigned to placebo, there is a 19% risk of major adverse cardiovascular event at 2 years, and that the same cumulative risk is reduced to 14% in the treatment group. We can quantify the treatment effect by calculating their difference (absolute risk difference, sometimes referred to as absolute risk reduction, or simply risk difference), or their ratio (risk ratio, or relative risk). These measures can be calculated for several time points to fully appreciate the extent of the treatment's clinical benefits over time.

In addition to focusing on the risk scale, it can be useful to quantify the treatment effect in terms of time differences. This can be achieved by selecting a specific event probability and then comparing the time points at which that risk is achieved. For example, in the Figure, it takes 4 years for the first 25% of the treated individuals to experience the event, but only 2.8 years for individuals in the placebo group. The difference of 1.2 years provides

Key Words: biostatistics ■ data interpretation ■ randomized clinical trials ■ survival analysis

Correspondence to: Andrea Bellavia, PhD, TIMI Study Group, Brigham and Women's Hospital, Department of Medicine, Harvard Medical School, 60 Fenwood Rd, Ste 7022, Boston, MA 02115. Email abellavia@bwh.harvard.edu

For Sources of Funding and Disclosures, see page 1156.

Circulation is available at www.ahajournals.org/journal/circ

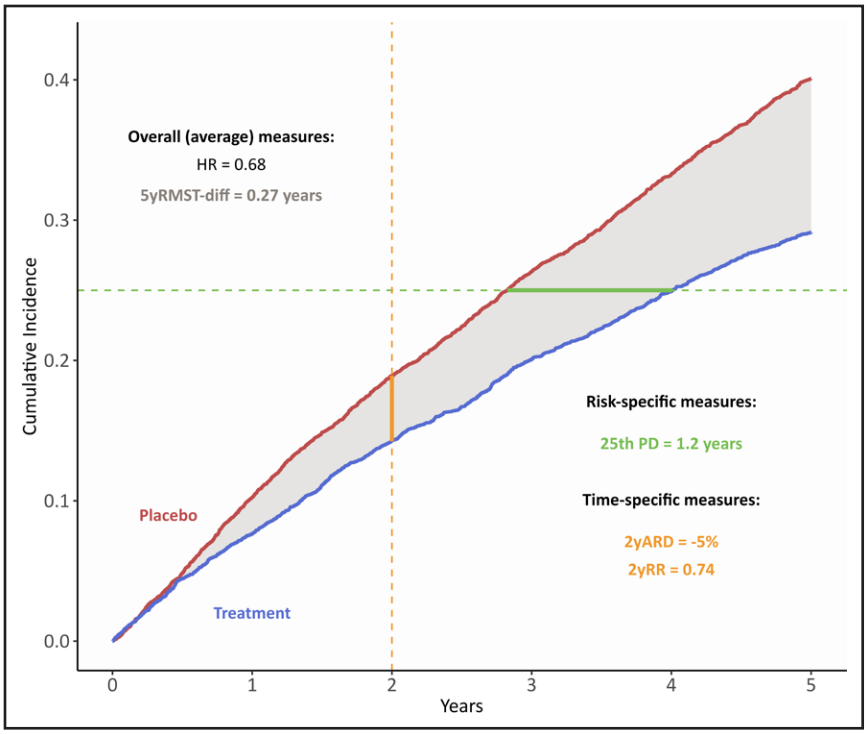

Figure. Comprehensive presentation of primary results in a hypothetical randomized clinical trial comparing time to major adverse cardiovascular events between a novel drug and placebo over 5 years of follow-up. Results are summarized by presenting the hazard ratio (HR) between the 2 groups, the difference in restricted mean survival time (RMST-diff) at 5 years (in gray, corresponding to the opposite of the shaded area between the 2 curves), the absolute risk difference (ARD) and risk ratio (RR) at a given time point (eg, 2 years, in yellow), and the percentile difference (PD) at a specific risk proportion (eg, 25%, in green).

a measure of effect in terms of time that is referred to as the 25th survival percentile difference. Like relative risks/absolute risk differences, percentile differences can be calculated for a range of values of event probabilities.

Another popular way to summarize treatment effects is by focusing on its average performance over the course of the study. The standard choice for a single average measure is the HR, which is especially useful when it is roughly constant over time (proportionality of the hazards). The interpretation of HR is not as straightforward as for relative risks/absolute risk differences and percentile differences. First, the hazard is a measure of the instantaneous rate of the event at a specific time *t*, conditional on being event-free up to *t*. As such, the HR, while informing on the relative difference in the instantaneous rate of the event between the randomization groups, does not provide any direct information on the cumulative risk of the event, and should not be interpreted as a relative risk.[3] Second, when the HR is estimated with a Cox regression, the hazard rate in the referent group is not specified, complicating its interpretation as a comparison measure.

An alternative summary measure that is becoming increasingly popular is the difference in restricted mean survival time (RMST Δ),[4,5] calculated as the difference in the areas under the survival curves between 0 and a prespecified time *t* (either the end of follow-up or any time of interest). When presenting cumulative incidence functions, RMST Δ corresponds to the opposite of the area between the curves. RMST Δ is interpreted as the difference in event-free survival between the treatment and the referent group during the specified time period, thus providing an appealing and intuitive measure of treatment effect. Nevertheless, RMST Δ is calculated as an average over a given timeframe. As such, interpretation must be cautious if the treatment effect is varying over time.

FINAL REMARKS

Assessing treatment effects under different perspectives (time-specific, risk-specific, overall) and metrics (risk, hazard, time) is necessary to comprehensively understand the clinical benefits of a treatment over time. We focused this review on common measurements of treatment effects for randomized clinical trials in cardiology. Except for the HR, which is traditionally estimated with a Cox regression, the discussed measures are commonly estimated with model-free estimators of the survival curve, such as Kaplan-Meier. In several contexts, however, regression models might be required to incorporate additional covariates or interactions. The popularity of the HR is partially attributable to the historical predominance of the Cox model as the standard regression approach for survival data. In recent years, however, several alternative approaches have been presented. Even though their description is beyond the scope of this commentary, flexible regression models are available for all the measures discussed. Additional information on regression approaches for the measures discussed can be found at https://timi.org/biostatistics.

Researchers should consider incorporating several perspectives into their analysis plan, accompanying HRs with other measures of associations that can improve their clinical interpretation in terms of actual risk benefits (time-specific relative risk or absolute risk difference) or

in terms of time differences (RMST Δ or risk-specific percentile difference).

ARTICLE INFORMATION

Affiliation

TIMI Study Group, Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA.

Acknowledgments

The authors thank Drs Andrea Discacciati and Siddharth M. Patel for their comments. The authors are members of the TIMI Study Group, which has received institutional research grant support through Brigham and Women's Hospital from Abbott, Abiomed, Inc, Amgen, Anthos Therapeutics, ARCA Biopharma, Inc, AstraZeneca, Boehringer Ingelheim, Daiichi-Sankyo, Ionis Pharmaceuticals, Inc, Janssen Research and Development, LLC, MedImmune, Merck, Novartis, Pfizer, Regeneron Pharmaceuticals, Inc, Roche, Saghmos Therapeutics, Inc, Siemens Healthcare Diagnostics, Inc, Softcell Medical Limited, The Medicines Company, Verve Therapeutics, Inc, and Zora Biosciences.

Sources of Funding

None.

Disclosures

None.

REFERENCES

1. Uno H, Wittes J, Fu H, Solomon SD, Claggett B, Tian L, Cai T, Pfeffer MA, Evans SR, Wei LJ. Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Ann Intern Med*. 2015;163:127–134. doi: 10.7326/M14-1741
2. Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. *Eur Heart J*. 2019;40:1378–1383. doi: 10.1093/eurheartj/ehy770
3. Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Ann Epidemiol*. 2018;28:54–57. doi: 10.1016/j.annepidem.2017.10.014
4. Han L. Breaking free from the hazard ratio: embracing the restricted mean survival time in clinical trials. *NEJM Evid*. 2023;2:EVIDe2300142. doi: 10.1056/EVIDe2300142
5. Perego C, Sbolli M, Specchia C, Fiuzat M, McCaw ZR, Metra M, Oriecuia C, Peveri G, Wei LJ, O'Connor CM, 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