

Rethinking survival analysis: advancing beyond the hazard ratio?

Peer-reviewed author version

VERBEECK, Johan & Saad, Everardo D. (2024) Rethinking survival analysis: advancing beyond the hazard ratio?. In: European Heart Journal-Acute Cardiovascular Care, 13 (3) , p. 313 -315.

DOI: 10.1093/ehjacc/zuae017

Handle: <http://hdl.handle.net/1942/42715>

Rethinking survival analysis: advancing beyond the hazard ratio?

Johan Verbeeck ^{1*} and Everardo D. Saad²

¹Data Science Institute, UHasselt, Agoralaan Building D, 3590 Diepenbeek, Belgium; and ²International Drug Development Institute, Louvain-la-Neuve, Belgium

Randomized clinical trials are the foundation of evidence-based medicine, offering vital insights into the efficacy of interventions. Although these trials guide clinical decision-making through hypothesis testing, relying solely on a significant *P*-value is insufficient; it is essential to convey in addition the magnitude of the treatment effect. In cardiovascular trials, the primary endpoint frequently comprises (a composite of) time-to-event outcomes, for which the magnitude of the treatment effect is typically quantified using the hazard ratio (HR). Despite the omnipresence of HRs, their clinical interpretation is not straightforward and, unfortunately, often incorrect.^{1,2}

The HR is a ratio of the hazard rates in each group, the latter being the instantaneous risk of the occurrence of events under each treatment. Consequently, it is a rate ratio and should be interpreted as a relative rate reduction, rather than a relative risk reduction. Although the HR shares the same direction of the treatment effect with the risk ratio, and they converge if the event rate is rare, in general, the HR tends to overestimate the relative risk reduction.¹ For example, a HR of 0.7 indicates that the risk decreases, but not necessarily by 30%. Instead, the instantaneous event rate decreases by 30%, and the relative risk reduction may well vary over time. Unfortunately, this is not the only source of confusion, as the HR has also been interpreted as a reduction in absolute risk, or as events occurring later.² The distinction between these measures is presented in the table.

The appeal of the HR to express the treatment effect is that if the hazard rates are assumed proportional over time, the HR is independent of time, allowing it to be expressed as a single value. However, violations of the proportional hazards assumption complicate its interpretation, rendering the HR time-dependent and in many circumstances unsuitable to represent the magnitude of the treatment effect.^{1,3} This violation occurs when the population includes patients with varied treatment effects or when the treatment effect changes over time, something visually evident in survival curves displaying time-dependent effects like late separation, convergence, or crossing.

The HR may thus not be the most optimal measure for effectively communicating treatment effects in (composite) survival endpoints, prompting the consideration of alternative measures (see [Table 1](#)).

While relative measures are well-established in time-to-event outcomes, absolute measures are generally more appropriate for expressing treatment effects over time and are recommended for a correct interpretation of relative effects.³ Although straightforward in

interpretation and use, the absolute difference in median survival time is often impractical in cardiovascular trials, as the survival proportions commonly do not drop below 50%. In contrast, the difference in mean survival time is generally applicable, does not require proportional hazards, and quantifies the mean gain in event-free time. Two main methods can be distinguished to estimate this difference. The restricted mean survival time assesses the difference in the area under the survival curve over a limited period of time,^{4,5} while parametric survival models predict the mean survival in each group through modelling of entire survival curves.⁶ The latter are often more powerful in detecting a treatment difference, but depend on correct distributional assumptions, which can be readily assessed.⁶

Generalized pairwise comparisons (GPC), an extension of the Wilcoxon–Mann–Whitney test, does not depend on the proportional hazards assumption for survival outcomes and offers both relative and absolute measures of treatment effects.⁷ The GPC method works by forming all possible pairs of patients, one from each arm of the trial, and comparing the two patients in a pair regarding the outcome(s) of interest. By doing this, the GPC method quantifies proportions of pairs of patients, one from each treatment arm, that are favourable or unfavourable to the treatment of interest. The net treatment benefit (NTB), an absolute measure, is the difference in proportions of favourable and unfavourable pairs and quantifies the difference in the probability that a patient on the treatment of interest will fare better. Unlike other methods, GPC can accommodate the analysis of outcome types beyond survival. Interestingly, in GPC, a threshold of clinical relevance can be assumed in the comparison of two patients, such that a difference between two patients is not confined to a single day. Importantly, the method allows the analysis of ‘time to worst outcome’ rather than ‘time to first outcome’, offering a solution to the criticism of conventional analyses of composite endpoints, which ignore serious events if they didn’t occur first in a patient. Moreover, a NTB restricted in time can be formulated.⁵

To enhance understanding of clinical trial results and to facilitate decision-making and effective communication to patients, it is important to acknowledge that no single metric can capture the entire profile of differences between treatments.^{3,4} Hence, it is crucial, in addition to the appropriate interpretation of the HR, to incorporate various measures of treatment effect, preferably absolute measures, for a more comprehensive assessment of trial results. Encouraging statisticians

* Corresponding author. Tel: +3211268624, Email: johan.verbeeck@uhasselt.be

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Table 1 Example of a two-arm clinical trial with the number of patients having the event of interest, to explain the different measures of risk and alternative measures for survival outcomes

Example			
	Event	No event	
Control	A	B	A total of n_1 patients followed for a cumulative time t_1
Active	C	D	A total of n_2 patients followed for a cumulative time t_2

Risk interpretation		Interpretation	Formula
Hazard ratio		Instantaneous risk reduction or relative rate reduction	$\frac{A}{n_1} - \frac{B}{n_2}$
Risk ratio		Relative risk reduction	$\frac{A}{n_1} - \frac{C}{n_2}$
Risk difference		Absolute risk difference	$\frac{A}{n_1} - \frac{C}{n_2}$

Alternative measures for survival outcomes			
Measure	Methodology	Interpretation	Advantages
Hazard ratio	Proportional hazards (PH) model	Relative rate reduction	Takes survival curve into account; experience
Difference in median survival	Median	Difference in median	Easy
Difference in restricted mean	Restricted mean survival time	Difference between areas under the curve	Absolute measure, does not depend on PH assumption
Difference in unrestricted mean	Parametric survival models	Several interpretations possible	Absolute measure, does not depend on PH assumption, takes full survival curve into account
Net treatment benefit	Generalized pairwise comparisons	Difference in probability of doing better	Absolute measure, does not depend on PH assumption, takes full survival curve into account, can prioritize endpoints by severity, can include a threshold of clinical relevance
			Disadvantages
			Relative measure, depends on PH assumption for correct interpretation
			Relative measure, requires the survival proportions to drop below 50%
			Restricted in time
			Requires correct assumption on the survival distribution
			Depends on the censoring distribution if not restricted

and clinical scientists to embrace diverse measures recognizes that the variability in clinical scenarios prompts a nuanced selection of treatment effect metrics based on specific clinical contexts.

Conflict of interest: Everardo D. Saad is an employee of IDDI.

Data availability

No new data were generated or analysed in support of this research.

References

1. Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Ann Epidemiol* 2018;**28**:54–57.
2. Weir IR, Marshall GD, Schneider JI, Sherer JA, Lord EM, Gyawali B, et al. Interpretation of time-to-event outcomes in randomized trials: an online randomized experiment. *Ann Oncol* 2019;**30**:96–102.
3. Saad ED, Zalcberg JR, Péron J, Coart E, Burzykowski T, Buyse M. Understanding and communicating measures of treatment effect on survival: can we do better? *J Natl Cancer Inst* 2018;**110**:232–240.
4. Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;**32**:2380–2385.
5. Mao L. On restricted mean time in favor of treatment. *Biometrics* 2023;**79**:61–72.
6. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data (second edition)*. Hoboken: Wiley; 2002.
7. Verbeeck J, De Backer M, Verwerft J, Salvaggio S, Valgimigli M, Vranckx P, et al. Generalized pairwise comparisons to assess treatment effects: JACC review topic of the week. *J Am Coll Cardiol* 2023;**82**:1360–1372.