



Looking for a better measure of the benefit in clinical trials: a never-ending journey

Gennaro Daniele¹, Diana Giannarelli², Emilio Bria^{3,4}

¹Scientific Directorate, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy; ²Biostatistical Unit, Regina Elena National Cancer Institute IRCCS, Rome, Italy; ³Comprehensive Cancer Center, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy; ⁴Department of Translational Medicine and Surgery, Università Cattolica del Sacro Cuore, Roma, Italy

Correspondence to: Prof. Emilio Bria, MD. Comprehensive Cancer Center, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy; Department of Translational Medicine and Surgery, Università Cattolica del Sacro Cuore, L.go A. Gemelli 8, 00168, Roma Italy. Email: emilio.bria@unicatt.it.

Provenance and Peer Review: This article was commissioned and reviewed by the Section Editor Dr. Jianrong Zhang, MD, MPH (PhD Candidate, Cancer in Primary Care Research Group, Centre for Cancer Research; Department of General Practice, Melbourne Medical School, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne; Victorian Comprehensive Cancer Centre, Melbourne, Australia).

Comment on: Ben-Aharon O, Magnezi R, Leshno M, *et al.* Median Survival or Mean Survival: Which Measure Is the Most Appropriate for Patients, Physicians, and Policymakers? *Oncologist* 2019;24:1469-78.

Submitted Feb 24, 2020. Accepted for publication Mar 25, 2020.

doi: 10.21037/atm.2020.03.118

View this article at: <http://dx.doi.org/10.21037/atm.2020.03.118>

The hazard ratio (HR) is calculated as the ratio of the hazard rates (the chance for a patient to experience the event) in the treatment over the control arm, when the outcome is a time to event [overall survival (OS), progression-free-survival (PFS) and others]. An HR =1 means that the tested treatments (experimental and control arms) do not differ for the measured effect, while an HR <1 or an HR >1 means that the experimental treatment is better than the control and vice-versa.

The HR is the most commonly used summary measure of benefit in randomised clinical trial. Its widespread use is mainly linked to the ease of implementation of HR as summary measure of benefit during the design of RCT. However, HR is a relative measure that could be misleading in communicating the net benefit for both doctors and patients. Indeed, an HR =0.5 means that the risk of the event is halved but it is not immediate to calculate the real advantage for patient because it depends on survival of control arm (i.e., similar HRs can be obtained from very different absolute improvements in survival times).

Easier to be understood by both doctors and patients is the increase in the median survival times. This measure can be directly communicated to patients in order to give them reference to an expected benefit with a particular treatment.

Indeed, it could be easier to understand an advantage of 3 months over 8 months of survival rather than an HR of 0.72.

However, the median survival time represents only the time the half of the patients are surviving and is not dependent on the survival of all the patients. However, modern oncology teased this concept with the evidence that for certain drugs, only a small fraction of patients derives a greater benefit with a long-term gain in survival.

Moreover, since HR is a model-based measure (i.e., is calculated through a Cox model), its robustness relies on the assumption that the hazards in the treatment and control arm remain proportional over the time.

Unfortunately, this is increasingly less the case in the modern clinical trials in oncology. The clinical trials with immune checkpoint inhibitors can be regarded as the main examples of this situation. Several authors, proposed an alternative approach in designing clinical trials where the proportional hazard assumption is not prospectively respected. Among these, the use of milestone survival (i.e., the proportion of patients surviving at a certain time point) is one of the possible alternatives to HR and median values, thus it might represent an easy measure to quantify the benefit of treatments (1).

In this scenario, methodologists and biostatisticians explored new ways to describe time-to-event benefit, and

difference in restricted mean survival time is one of these.

Mainly, Restricted Mean Survival Time (RMST) represents the area under the curve of survival for a certain arm in a trial. It can be regarded as the life expectancy of patients in that arm while they are in the study. Comparison of RMST between arms of a trial has an unprecedented advantage over the past measures of being dependent on all the patients included in the trial and observed within the time of study and easily understood by both doctors and patients. It has also the additive property and this peculiarity allows the design of strategy trials being possible to sum up the RSMT of different treatments to obtain the RSMT of a sequence of therapy lines (2). This is not true for median values.

In their manuscript Ben-Aharon (3) *et al.*, compared the median survival time with RMST and mean survival time calculated using Weibull distribution, for 44 drugs which granted approvals by FDA between 2013 and 2017. Interestingly, the authors found that RMST invariantly underscores the improvements as compared with median and mean survival times, with the last ones being the greatest (3.6 vs. 4.6 vs. 6.1 months). These results are not particularly surprising since RMST is limited at a time t (i.e. t is the time when results are reported for regulatory approval) and include all the early failure times excluding long survivors experience. More interestingly, this scenario is confirmed for all the subgroup analyses, performed by mechanism of action, apart from antiangiogenics where the improvements were very similar (2.1 vs. 2.0 vs. 2.2 months).

This manuscript raises, at least, 2 important points. The first one is that we are actively looking for a better way of summarizing results of clinical trials, since the limitations of the most used ones. Which of these is the optimal one we do not know yet, but, perhaps, the best one does not exist. The evidence that the mean improvements, these too based on a model assumption, are invariantly better the median and RMST is unsurprising being more dependent on the outliers. Whether mean survival is more meaningful for clinical or regulatory decision is not clear. Perhaps, the answer to this is that one size does not fit all, as witnessed by the case of antiangiogenics. In fact, for these drugs it is neither expected that only a small fraction of patients derives a long-term benefit nor that the clinical effect starts after and 'induction' time. In cases like this, in other words, a proportion of hazards between the arm could be assumed and then HR could be also used as a useful summary measure.

Anyway, the use of absolute measures in the time domain such as medians, means and rates at a prespecified time t are more tangible than relative measures and more direct to

describe treatment efficacy to patients. From the physician standpoint, a good solution could be imported from the meta-analytic approach where the NTT (Number of patients to be Treated for one to benefit) summarize in a single figure the real impact of treatments making possible a ranking of alternatives.

Acknowledgments

Funding: EB is currently supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) under Investigator Grant (IG) No. IG20583 and by institutional funds of Università Cattolica del Sacro Cuore (UCSC-project D1-2018/2019).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm.2020.03.118>). EB has received honoraria or speaker's fees from Merck-Sharp and Dohme, AstraZeneca, Celgene, Pfizer Inc., Helsinn, Eli-Lilly, Bristol-Myers Squibb, Novartis and Roche. GD reports personal fees from Bei Gene, non-financial support from Roche, outside the submitted work. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Chen TT. Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors. *J Natl Cancer Inst* 2015;107.
2. 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.
3. Ben-Aharon O, Magnezi R, Leshno M, et al. Median

Survival or Mean Survival: Which Measure Is the Most Appropriate for Patients, Physicians, and Policymakers? *Oncologist* 2019;24:1469-78.

Cite this article as: Daniele G, Giannarelli D, Bria E. Looking for a better measure of the benefit in clinical trials: a never-ending journey. *Ann Transl Med* 2020;8(14):893. doi: 10.21037/atm.2020.03.118