available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Platinum Opinion

Restricted Mean Survival Times to Improve Communication of Evidence from Cancer Randomized Trials and Observational Studies

Ludovic Trinquart [a], *Anna Bill-Axelson* [b], *Jennifer R. Rider* [c,*]

[a] *Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;* [b] *Department of Surgical Sciences, Uppsala University, Uppsala, Sweden;* [c] *Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA*

The hazard ratio (HR) and median survival time for each treatment/exposure group are the most commonly reported treatment effect measures for censored outcome data in randomized trials and observational cohort studies. The methodological literature is demonstrating renewed interest in another measure of the treatment effect, the difference in restricted mean survival time (RMST), which was described by Kaplan and Meier in 1958 [1]. In a randomized experiment, we found that presenting HR- or RMST-based measures influenced how clinicians judged treatment effects [2]. Nearly half of participants misinterpreted the HR, and presentation of HRs rather than differences in RMST led to inflated views of treatment benefits.

In this Platinum Opinion, we explain how differences in RMST can provide additional insights into HRs and aid in the interpretability of results. We illustrate these concepts using the SPCG-4 trial, which randomly allocated participants with localized prostate cancer to radical prostatectomy (RP) or watchful waiting (WW) [3]. We focus on two primary endpoints after 18 yr of follow-up: death from any cause and death from prostate cancer. Readers can find the statistical code used to generate the figures at https://osf.io/j6sm5/.

Kaplan-Meier curves for any cause of death and prostate cancer death in SPCG-4 through 18 yr of follow-up are shown in Figs. 1A and 2A. The HRs estimated using the Cox model compare the hazard of death in the RP group to that in the WW group over the follow-up period. Over 18 yr of follow-up, the HR for death was 0.74 (95% confidence interval [CI] 0.61–0.90). The HR for prostate cancer death was 0.56 (95% CI 0.41–0.78). Two features of HRs are important to note. First, HRs are calculated assuming that the hazard of death in the RP group is proportional to the hazard in the WW group at any point over 18 yr. In SPCG-4, there is no evidence of nonproportional hazards (Grambsch-Therneau test: $p = 0.42$ for death and $p = 0.61$ for prostate cancer death). In particular, Figs. 1B and 2B show that the HRs are relatively constant as evidence accumulated over time [4]. However, the proportional hazards assumption is commonly violated in cancer randomized trials, resulting in HRs that are misleading because they reflect a summary of effects that vary over time [5]. Second, at a given time point, the hazard gives the instantaneous potential for death to occur per unit of time, given that participants have survived up to this time. The hazard is a probability per unit of time; thus, it is not a probability but a rate. In particular, the hazard can take values larger than 1, depending on the choice of time unit as days, weeks, or months. HRs cannot be interpreted as risk ratios. Only when the outcome rate is low will HRs be numerically similar to risk ratios [2,6]. At 18 yr of follow up in SPCG-4, the survival probability was 43.48% in the RP and 31.15% in the WW arm. These probabilities yield a risk ratio of $(1 - 0.4345)/(1 - 0.3115) = 0.82$, with a 95% CI of 0.76–0.88, as compared to a HR of 0.74 for death from any cause.

The RMST is the mean survival time up to a prespecified time point, 18 yr in our example. In the SPCG-4 trial, patients in the WW arm lived, on average, for 12.8 yr (95% CI 12.3–13.4) after randomization. The difference in RMST gives the mean gain in survival time with RP as compared to WW over 18 yr. Patients in the RP group lived, on average,

* Corresponding author. Department of Epidemiology, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118, USA.
Tel. +1 617 3583991; Fax: +1 617 6384558.
E-mail address: rider@bu.edu (J.R. Rider).

https://doi.org/10.1016/j.eururo.2019.04.002

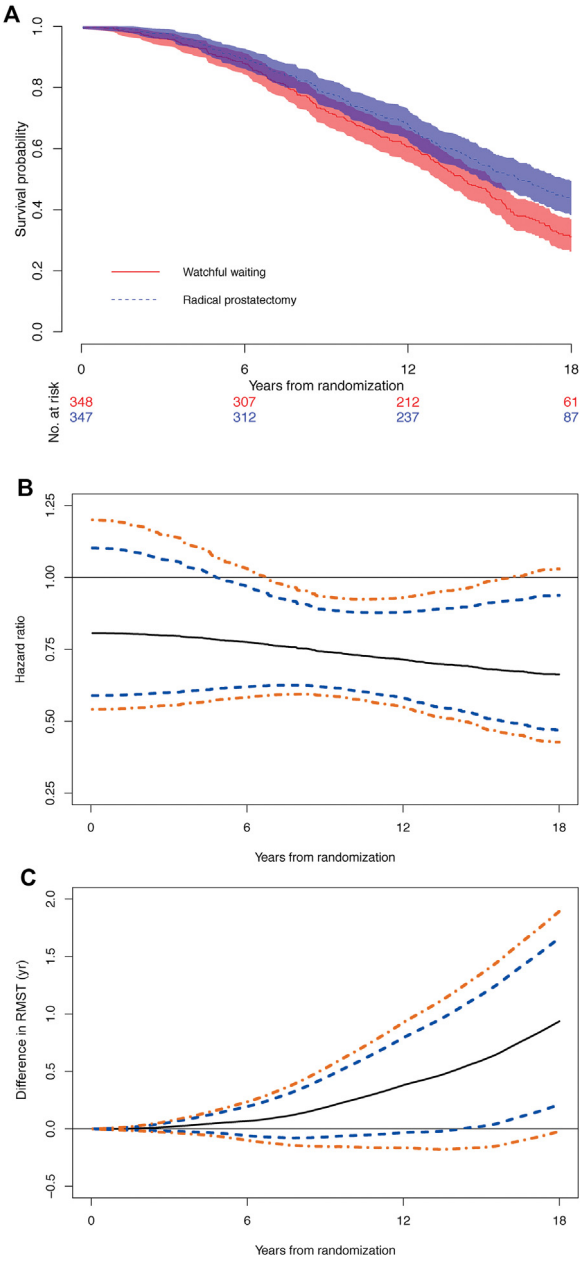

Fig. 1 – Results for death from any cause from the SPCG-4 trial over 18-yr follow-up. (A) Kaplan-Meier curves. (B) Hazard ratio curve. (C) Difference in restricted mean survival time (RMST) curve. Plots B and C show 95% pointwise and simultaneous confidence intervals in purple and green, respectively.

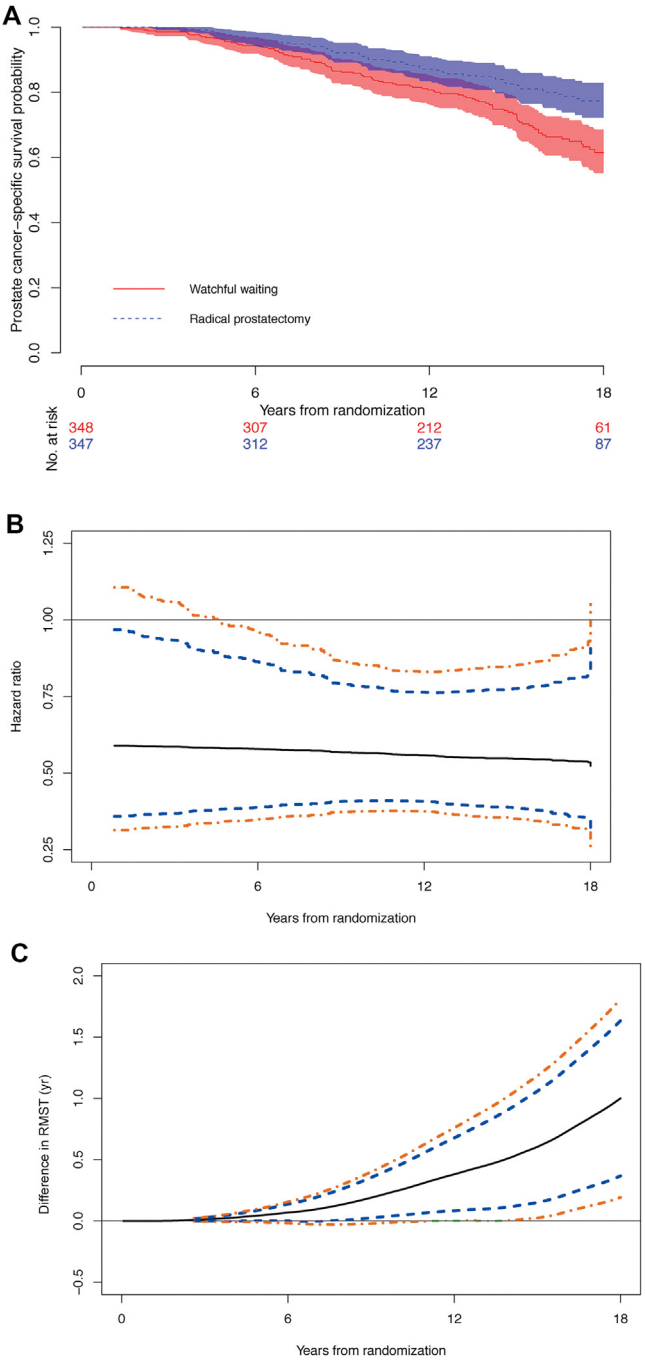

Fig. 2 – Results for death from prostate cancer from the SPCG-4 trial over 18-yr follow-up. (A) Kaplan-Meier curves. (B) Hazard ratio curve. (C) Difference in restricted mean survival time (RMST) curve. Plots B and C show 95% pointwise and simultaneous confidence intervals in purple and green, respectively.

for an additional 11.2 mo (95% CI 2.2–20.2) in 18 yr. For prostate cancer–specific survival, the gain was also considerable, with a difference in RMST of 12.0 mo (95% CI 4.5–19.5). Similarly to the HR, the difference in RMST captures the entirety of the survival curves up to the prespecified time point, by taking the area between the two curves. In addition, measuring treatment effects in the time domain through differences in RMST provides an intuitive interpretation of the magnitude of treatment effects. In particular, the RMST in the control group gives a background against which one can compare the gain in mean survival time. HRs do not provide such information. Moreover, the difference in RMST is valid whether the proportional hazards assumption is met or not. In many trials, large HRs correspond to small, potentially clinically insignificant differences in RMST [5]. Differences in RMST also help to highlight treatments resulting in a considerable gain in life expectancy, as in our current example. In observational

studies, covariate adjustment is possible when estimating differences in RMST to compare exposure groups [7].

Absolute treatment-effect measures generally convey trial results in a way that improves patient understanding as compared to HR. The difference between survival probabilities at a fixed time point is an alternative to the difference in RMST that may be particularly advantageous for measuring long-term effects [8]. However, the difference in RMST captures the survival pattern over the whole follow-up duration without loss of information. One potential caveat with the difference in RMST is that conclusions can vary depending on the choice of time point, contrary to the HR, which does not change over time when hazards are proportional. In this regard, the difference in RMST curves in Figs. 1C and 2C are helpful for examining if the experimental treatment loses or gains potency in comparison to the control over the follow-up period [9]. The simultaneous confidence band quantifies the uncertainty for the whole curve, and thus there is no need to prespecify a single time point. For both endpoints, the gain in restricted mean lifetime with RP compared to WW increases over time. In the most recent analysis from the SPCG-4 study, at 23-yr follow-up, a mean of 2.9 yr of life was gained with RP compared to WW [10]. Finally, the difference in RMST is restricted to the study follow-up duration, and is not to be interpreted as a gain or loss over a lifespan.

In conclusion, the choice of treatment effect presented influences the clinical take-home message. This phenomenon was eloquently described by the late Stephen Jay Gould in his essay, "The median isn't the message," which was inspired by his own reaction to statistics for median survival times following a diagnosis of mesothelioma [11]. Reporting multiple treatment-effect measures will foster more effective communication of study results. Framing the clinical message around differences in RMST may benefit shared decision-making.

Conflicts of interest: The authors have nothing to disclose.

References

[1] Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958;53:457–81.
[2] Weir IR, Marshall GD, Schneider JI, et al. Interpretation of time-to-event outcomes in randomized trials: an online randomized experiment. Ann Oncol 2019;30:96–102.
[3] Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932–42.
[4] Yang S, Prentice RL. Estimation of the 2-sample hazard ratio function using a semiparametric model. Biostatistics 2011;12:354–68.
[5] Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. J Clin Oncol 2016;34:1813–9.
[6] Sutradhar R, Austin PC. Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Ann Epidemiol 2018;28:54–7.
[7] Trinquart L, Jacot J, Conner SC, Porcher R. Reply to K.H. Eng et al.. J Clin Oncol 2017;35:466–7.
[8] Chen TT. Milestone survival: a potential intermediate endpoint for immune checkpoint inhibitors. J Natl Cancer Inst 2015;107:djv156.
[9] Zhao L, Claggett B, Tian L, et al. On the restricted mean survival time curve in survival analysis. Biometrics 2016;72:215–21.
[10] Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in prostate cancer—29-year follow-up. N Engl J Med 2018;379:2319–29.
[11] Gould SJ. The median isn't the message. AMA J Ethics 2013;15:77–81.