



Model-free estimates that complement information obtained from the hazard ratio

Salil Vasudeo Deo^{1,2} · Vaishali Salil Deo¹ · Varun Sundaram^{1,2}

Received: 9 January 2021 / Revised: 12 February 2021 / Accepted: 16 February 2021 / Published online: 5 April 2021

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

Abstract

Learning objectives

1. Understand the limitations of the modeling of survival data, especially as pertains to the Cox proportional hazards model.
2. An introduction to model-free estimates of survival, namely, the restricted mean survival time/restricted mean lost time.
3. Use R (The R Foundation for Statistical Computing, Austria) or STATA® (The STATA Corp, College Station, TX, USA) to perform analyses and obtain these parameters from a dataset.

Keywords Biostatistics · Survival analysis · Median survival time · Restricted mean survival time · Kaplan and Meier method

Introduction

The Cox proportional hazards model is among the most frequent methods used to present results of survival analyses in clinical trials [1]. However, as for any model, results are based on certain assumptions that may not be fulfilled in every case. Therefore, in this brief paper, we present an overview of model-free estimates to report the results of survival analyses.

Limitations of the Cox model and some published examples

Randomized controlled trials (RCT) in cardiovascular medicine are designed to compare the clinical outcome of two treatment arms. Participants are enrolled; they are randomized to receive one of two treatments (study or control). Patients are then followed up, either for a specific time period or until a

predefined number of events have occurred. Trials are often designed to observe differences in a primary end point, which is often a composite of multiple clinical events. The Kaplan and Meier method or the cumulative failure function (inverse of Kaplan and Meier event-free survival) is often chosen to present time-related event rates. As both end points and statistical methods often need to be declared a priori, the Cox proportional hazards test (CPH) is most commonly used to calculate the hazard ratio (HR), i.e., the risk of observing the end point in the study arm versus the control arm.

However, as discussed earlier, HR is reliable only if the proportional hazard (PH) assumption is fulfilled. Yet, as demonstrated in Fig. 1, we often observe a changing relationship between the two study arms during the study period. In Fig. 1a, we have recreated the curve comparing the secondary end point (composite of death, stroke, myocardial infarction, and ischemia-driven re-intervention) between percutaneous intervention (PCI) and coronary artery bypass grafting (CABG) from the EXCEL trial [2]. The EXCEL trial was an RCT comparing clinical outcomes between CABG and PCI in patients with left main coronary artery stenosis. As demonstrated in Fig. 1a, the curves for CABG and PCI cross during follow-up. This situation violates the PH assumption. Investigators thus reported the results for this end point as odds ratio rather than the conventional HR. However, it is very clear from the graph that the relationship between CABG and PCI changes during the follow-up period. To present another example, the researchers

✉ Salil Vasudeo Deo
svd14@case.edu

¹ Louis Stokes VA Medical Center, Northeast Ohio Veteran Affairs Healthcare System, Cleveland, OH, USA

² Case School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA

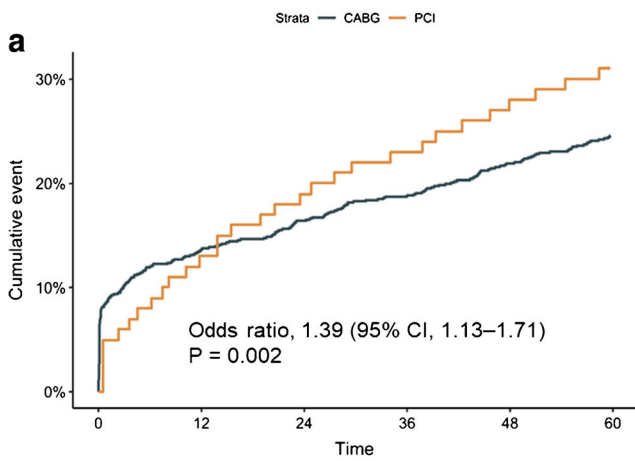
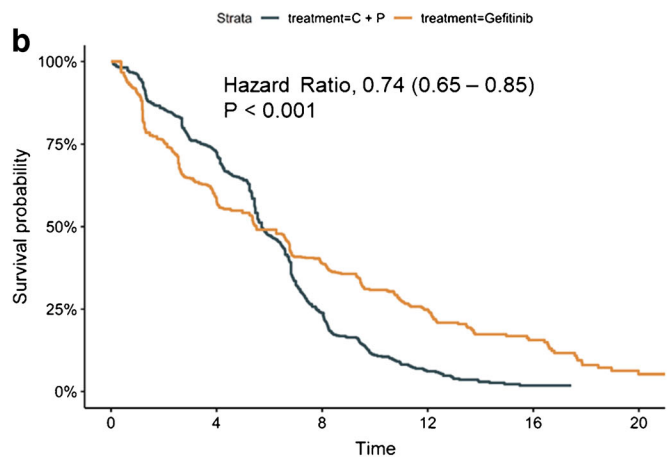


Fig. 1 a This figure presents the recreated Kaplan–Meier graph comparing the secondary end point (composite of death, stroke, myocardial infarction, and ischemia-driven re-intervention) between percutaneous intervention (PCI) and coronary artery bypass grafting

studied relapse-free survival at 12 months in lung cancer patients treated with the study drug (gefitinib) versus standard therapy (carboplatin + paclitaxel) [3]. Gefitinib, an oral agent, had been proven effective in a different type of lung cancer in earlier studies. As seen in Fig. 1b, gefitinib patients have higher event rates in the earlier part of the trial and then the two arms cross at approximately 5 months after therapy. This is, again, a violation of the PH assumption. Researchers reported HR of 0.74 (0.65–0.85; $p < 0.001$) favoring gefitinib. Several researchers naturally raised this point in a letter to the journal [4]. In this trial, as in the Excel graph earlier, the relationship between arms changes during follow-up. Schemper et al. report that, when the PH is violated, the calculated HR can be considered an average treatment effect over the entire study period [5]. However, we believe that such results are difficult to interpret clinically. In Fig. 2a, we present results for all-cause mortality from the COMPANION trial [6]. The COMPANION trial was



(CABG) from the EXCEL trial. **b** This figure graphs the relapse-free survival at 12 months in lung cancer patients treated with the study drug (gefitinib) versus standard therapy (carboplatin + paclitaxel). C carboplatin, P paclitaxel

an RCT that compared clinical outcomes between medical therapy and cardiac resynchronization therapy (CRT) in patients with advanced heart failure. In this figure, survival in both arms is similar at least for 180 days after enrollment into the trial. Thereafter, the curves diverge and patients in the CRT arm fare better than those receiving medical therapy. These are two examples, among many, that presenting time-to-event results as a single HR may be an oversimplification. In fact, we calculated the ratio of logarithm transformed hazard rates for both arms throughout the study period (Fig. 2b). As demonstrated here, HR gradually changes. While it initially favors medical therapy, after approximately 100 days of follow-up, it appears to favor the CRT arm. We believe that such information is clinically relevant and can aid decision-making. Given the limitations of the CPH model discussed above, we present some parameters that may improve our understanding if presented along with HR. The subsequent material presented herein borrows heavily

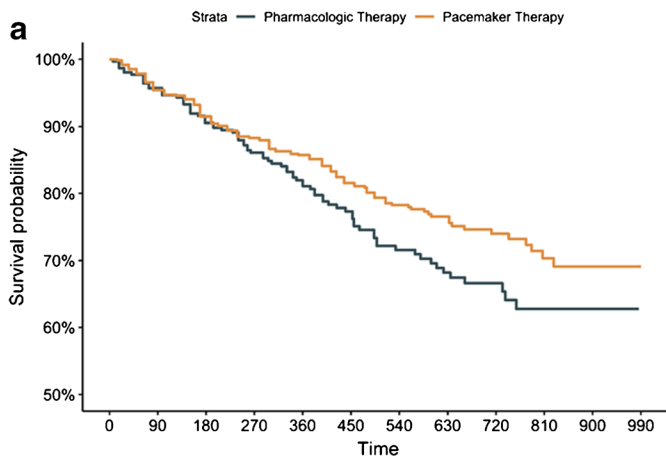
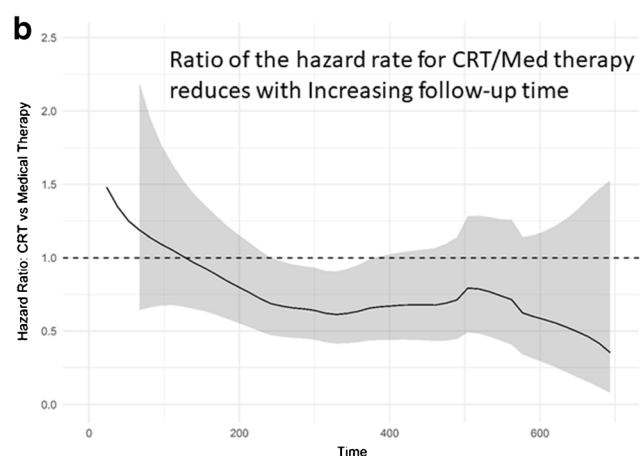


Fig. 2 a This Kaplan–Meier graph presents all-cause mortality from the COMPANION trial. The COMPANION trial was an RCT that compared clinical outcomes between medical therapy and cardiac resynchronization therapy (CRT) in patients with advanced heart failure. **b** This graph



presents the time varying hazard ratio between the CRT and medical therapy arms of the COMPANION trial. CRT cardiac resynchronization therapy

Table 1 This table presents the survival estimates at specific time points in the medical therapy and CRT arms

Time of follow-up (days)	Survival (medical therapy)	Survival (CRT therapy)	Survival ratio (CRT vs. pharmacologic therapy)	<i>p</i> value
180	0.9 (0.87–0.93)	0.91 (0.89–0.93)	0.98 (0.94–1.03)	0.64
360	0.81 (0.77–0.86)	0.85 (0.82–0.88)	0.95 (0.89–1.02)	0.18
540	0.71 (0.65–0.77)	0.78 (0.74–0.81)	0.91 (0.83–1.003)	0.05
720	0.66 (0.6–0.72)	0.73 (0.69–0.78)	0.9 (0.8–1.009)	0.07

Abbreviation: CRT cardiac resynchronization therapy

from the work of Uno et al. [7] and Royston et al. [8]. We attempt to explain their work in a non-mathematical manner and provide additional examples along the way. A supplemental file is also provided with the paper that will provide practical tips for conducting these analyses.

Model-free estimates

Ratio (or difference) in *t*-year survival rates

From the Kaplan and Meier estimates, it is easy to obtain the survival rate for both arms in the study at a specific time period (*t*). Results can then be presented as ratios of survival estimates with confidence intervals. Table 1 presents the survival estimates for medical therapy and CRT arms in the COMPANION trial. As we can see, the CRT arm survival gradually improves compared to medical therapy. While it does not attain statistical significance at $p < 0.05$, we can clearly see a meaningful difference in survival at 540 and 720 days of follow-up. These values correspond closely to the Kaplan Meier graph (Fig. 2a), where we see that the lines gradually diverge from each other after the initial time period. This estimate is easy to understand and interpret. As it is based on the Kaplan and Meier curve, it allows for right-censored

data. Presenting this estimate at regular intervals time points during the study period may provide readers with a better understanding regarding when the event rates have occurred.

Ratio (or difference) of the median survival time

The percentiles of survival can be obtained from the Kaplan–Meier estimates for each arm. In the presence of right censoring, survival time intervals are often highly skewed, resulting in the median being a more reliable distribution parameter than the mean. The median survival time is defined as that follow-up time when 50% of the initial cohort is estimated to be alive. Hence, reporting the median survival time in each arm is an easy estimate to calculate and understand. If survival in either arm of the study does not fall below 50%, instead of the median, another quantile value can be reported instead. In Fig. 3, we present the Kaplan–Meier survival plot for male and female lung cancer patients. The dotted line corresponds to the median survival time (in days) for each group. For men and women, the median survival times were 270 (212–310) and 426 (345–524) days, respectively. Hence, in this study, the difference between median survival times for males versus females is -156 (-279 to -32 ; $p = 0.01$). From this estimate, we can infer that, in this study, survival for males was significantly lower than that for females.

Fig. 3 This graphs the Kaplan–Meier survival for male and female lung cancer patients. The dotted lines correspond to the median survival time (in days) for each group. For men and women, the median survival times were 270 (212–310) and 426 (345–524) days, respectively. Hence, in this study, the difference between median survival times for males versus females is -156 (95% confidence interval = -279 to -32); $p = 0.01$)

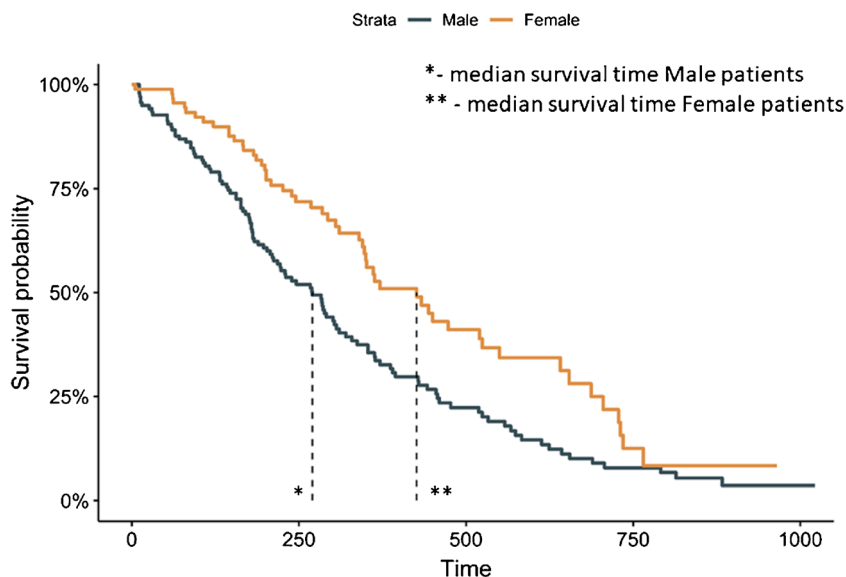
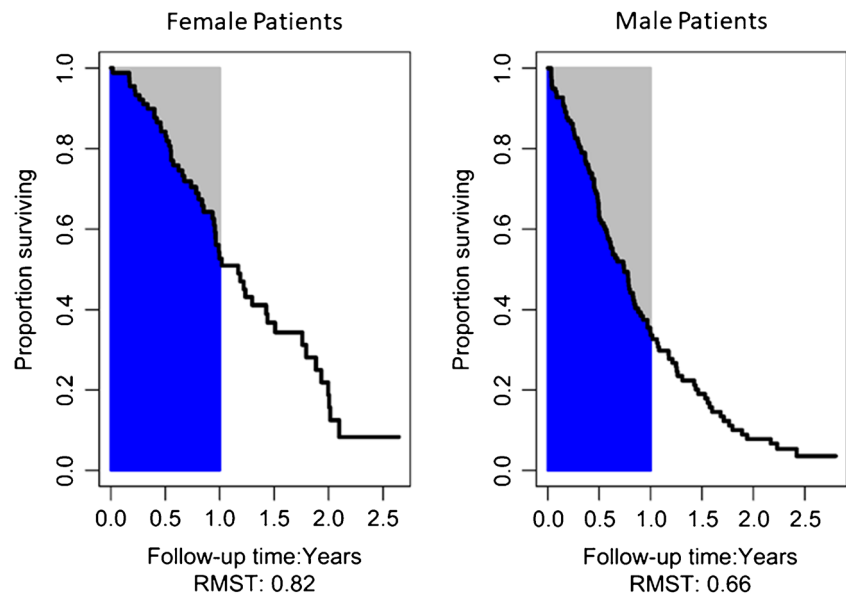


Fig. 4 This graph presents the separate curves for female and male patients as presented in Fig. 3. The shaded blue area in each graph presents the restricted mean survival time (RMST) for that group at 1 year from surgery. The RMST value for each group is also presented below the graphs



The restricted mean survival time

The restricted mean survival time (RMST) is a model-free parameter that represents the difference in the areas subtended by the two arms of the study for a specific time. To provide a simple example, we calculated the post-operative survival of males and females with advanced lung cancer. We decided to observe the RMST for each group at 1 year from surgery. In Fig. 4, the RMST for each group is shaded blue. Hence, the difference in RMST between groups is interpreted as the average delay in end points between arms. When considering this situation between two drugs, say a study and control drug, the difference in RMST can provide an understanding of clinical benefit achieved by the treatment over control. In the above example, on average, when compared to males, mortality was delayed by 0.15 (0.07–0.23) years (p value < 0.01) in females.

The RMST is a model-free measure that is intuitive and easy to understand. It can be calculated at different time periods; hence, unlike the fixed HR, it provides a better understanding of time-related changes in events between the two arms in the study. More importantly, it is a value that may aid clinical decision making.

As recommended by Uno et al. [7] and Kloecker et al. [9], we believe that the parameters presented above are excellent supplements to the HR obtained from the CPH model. We believe that researchers present more than one estimate when they report results of time-to-event analyses.

Software available

A brief search of literature demonstrates that both R 4.0.2 (The R Foundation of Statistical Computing, Austria) and

STATA® (StataCorp, College Station, TX) have user-written commands for calculating RMST and other parameters presented in our paper (Stata and R (R Foundation for Statistical Computing)). STATA commands available are *standsurv*, *strmst2*, *strmst*, and *stpmean*. In R, the packages “survRM2” can be used to calculate RMST while “surv2sampleComp” provides a range of other model-free parameters. The supplemental section provides an overview of both R and STATA commands. The supplemental section consists of an example dataset (Supplemental file 1) and two supplemental scripts (Supplemental files 2 and 3) that provide code for use with R or STATA®.

Conclusion

We have presented a brief overview of model-free parameters. We believe that they complement HR, allow readers to gain a better understanding of results, and may aid clinical decision making. We encourage the wider adoption of these methods in clinical research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12055-021-01167-4>.

Acknowledgements This material is the result of work supported with resources and use of facilities at the Louis Stokes Cleveland VA Medical Center, VA Northeast Ohio Healthcare System. The views expressed in this article are those of the authors. They do not represent the position or policy of the Department of Veteran Affairs or the United States Government.

Funding None.

Declarations

IRB approval Not required as this is an editorial. The data used in the study is publicly available.

Patient consent Not required as this is an editorial.

Conflict of interest The authors declare no competing interests.

References

1. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol*. 2014;21:686–94.
2. Stone GW, Kappetein AP, Sabik JF, et al. Five-year outcomes after PCI or CABG for left main coronary disease. *N Engl J Med*. 2019;381:1820–30.
3. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med*. 2009;361:947–57.
4. Ballatori E, Fatigoni S, Roila F. Treatment of lung cancer. *N Engl J Med*. 2009;361:2485.
5. Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Stat Med*. 2009;28:2473–89.
6. Bristow MR, Saxon LA, Boehmer J, et al. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med*. 2004;350:2140–50.
7. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol*. 2014;32:2380–5.
8. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med*. 2011;30:2409–21.
9. Kloecker DE, Davies MJ, Khunti K, Zaccardi F. Uses and limitations of the restricted mean survival time: illustrative examples from cardiovascular outcomes and mortality trials in Type 2 diabetes. *Ann Intern Med*. 2020;172:541–52.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.