On the proportional hazards assumption in Cox regression

Paul Dickman

13 March 2024

http://pauldickman.com/talk/

Today's talk

- About me.
- Gentle introduction to the proportional hazards assumption.
- Views on statistical significance testing in epidemiology.
- Discussion of 'Why test for proportional hazards?' by Stensrud & Hernán [1].

Stensrud & Hernán [1]

'Statistical tests for proportional hazards are unnecessary'

- Shameless marketing.

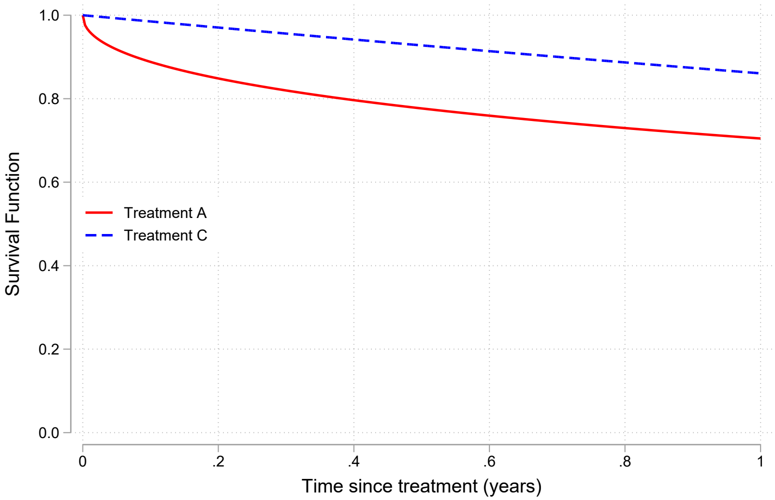
About me

- Born in Sydney Australia;
  studied mathematics and statistics in Newcastle (Australia).
- Worked in health services research;
  dabbled in industrial process control and quality improvement.
- Arrived in Sweden November 1993 for a 10 month visit to cancer epidemiology unit at KI. Stayed in Sweden for most of my PhD.
- Short Postdoc periods at Finnish Cancer Registry
  and Karolinska Institutet (cancer epidemiology).
- Joined MEB (MEP) in March 1999, attracted by the strong research environment and possibilities in register-based epidemiology.

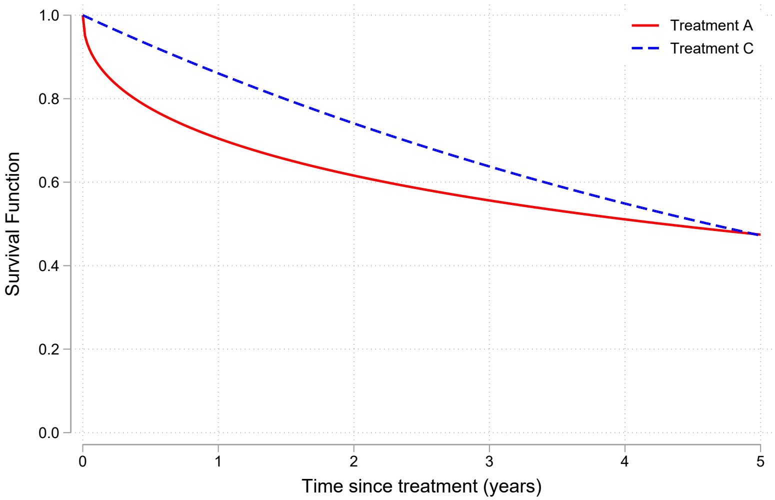
My research interests

- Development and application of methods for population-based cancer survival analysis, particularly the estimation and modeling of relative/net survival.
- General interest in statistical aspects of the design, analysis, and reporting of epidemiological studies.
- Epidemiology, with particular focus on cancer epidemiology.
- Lots of administrative work (deputy head of department and head of biostatistics group).
- Programme director for master's programme in biostatistics and data science (commences HT2024).

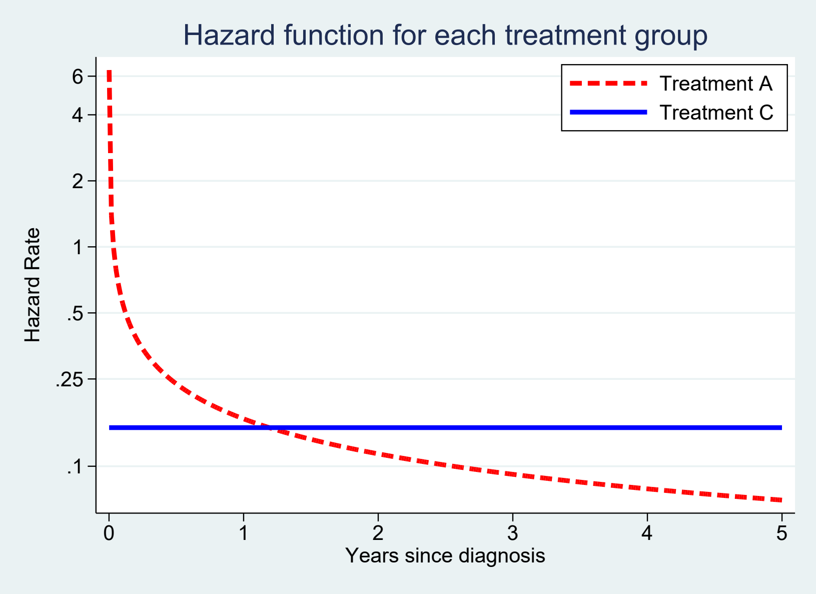
Which treatment (A or C) has the best survival?



Which treatment (A or C) has the best survival?



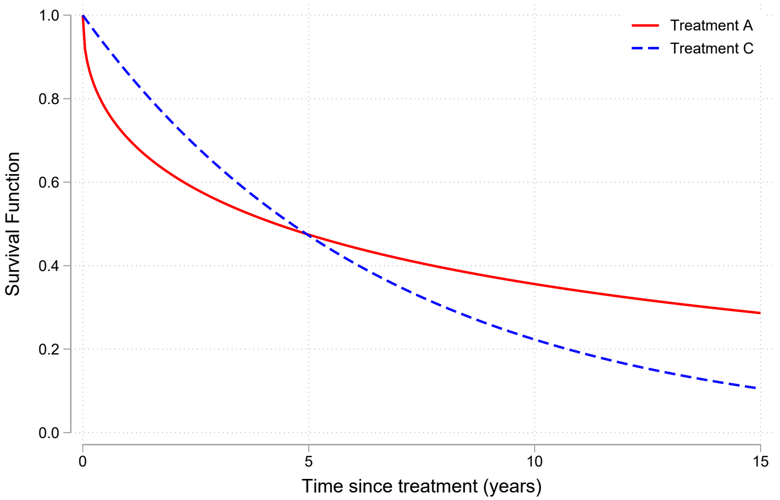
The two hazard functions



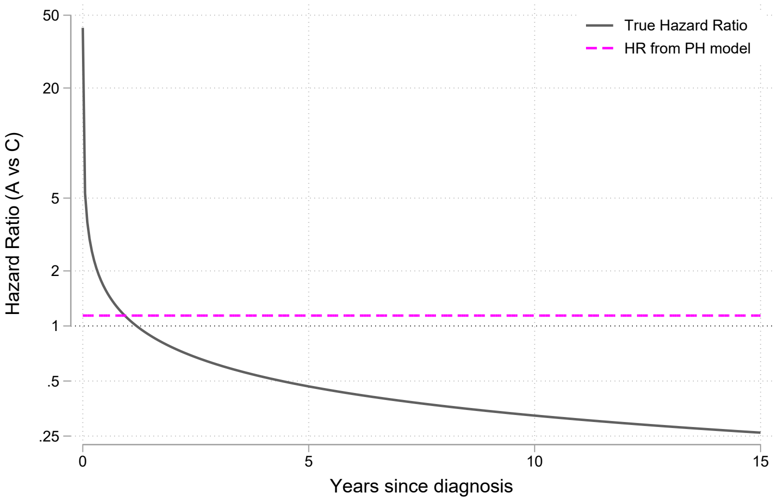
The proportional hazards assumption

- Hazard functions for any two covariate patters are proportional.
- Equivalently, log hazard functions have constant difference.
- Equivalently, hazard ratio is constant over time.
- Equivalently, no interaction between covariates and time.
- Can relax the PH assumption by modelling covariate by time interactions.

What about if we further extend the follow-up?



Time varying hazard ratio for A vs C



Limited (D1) vs. extended (D2) lymph node dissection for gastric cancer

STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:2807–2821
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2143

Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial

H. Putter[1,*,†], M. Sasako[2], H. H. Hartgrink[3], C. J. H. van de Velde[3] and J. C. van Houwelingen[1]

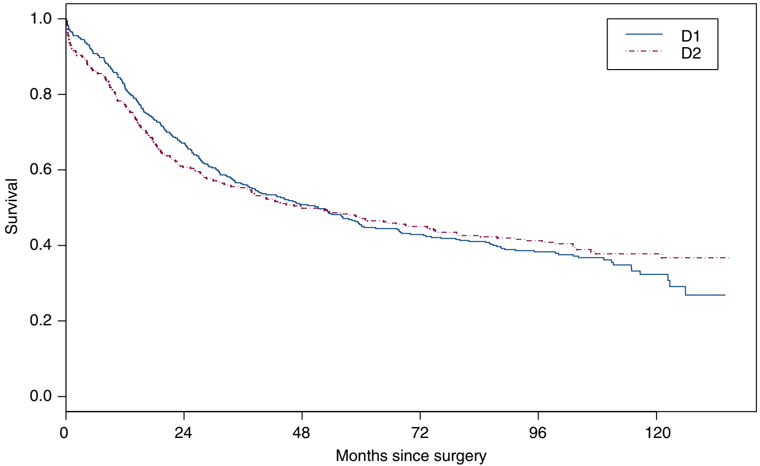


Figure 1. Kaplan–Meier plots of the survival curves for D1- and D2-dissection. The survival curves cross after 53 months.

The Cox regression with only randomization as a time-fixed effect gives an estimated hazard ratio of 0.97 of D2 dissection compared to D1-dissection, with a *p*-value of 0.73. The survival

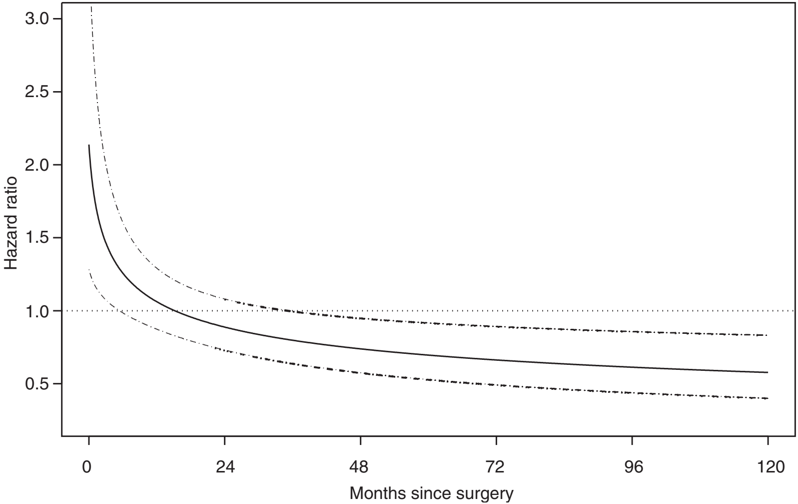


Figure 4. The estimated hazard ratio with 95 per cent confidence intervals based on Cox regression with treatment as time-dependent effect. A hazard ratio of one indicates equality of the hazard rates of D1 and D2.

P-values are not central in epidemiology

Estimation rather than testing

Epidemiology (the journal) has a longstanding policy of discouraging the use of statistical significance testing, that practice that judges study results according to whether a P-value exceeds or does not exceed a standard yet arbitrary cutoff value. (Lang et al. 1998) [2]

- 'Causal analyses of existing databases: no power calculations required' (Hernán 2021) [3]
- 'Why Stating Hypotheses in Grant Applications Is Unnecessary' (Hernán and Greenland 2024) [4]
- For causal inference, focus is on identifying an appropriate estimand and quantifying the effect as unbiasedly and precisely as possible.

JAMA Guide to Statistics and Methods

March 13, 2020

Why Test for Proportional Hazards?

Mats J. Stensrud, MD, DrPhilos[1,2]; Miguel A. Hernán, MD, DrPH[1,3,4]

» Author Affiliations | Article Information

JAMA. 2020;323(14):1401-1402. doi:10.1001/jama.2020.1267

Overview of my thoughts on the paper [1]

- Our comments in Sjölander and Dickman (2024) [5]
- Nice paper; I agree with essentially everything.
- The statement that 'statistical tests for proportional hazards are unnecessary' is potentially controversial, but I agree.
- I am concerned that the statement may be (mis)interpreted by some as 'assessing proportional hazards is unnecessary'.
- Researchers should understand the concept of proportional hazards, to which this paper makes a valuable contribution.
- Researchers should consider the time-varying nature of hazard ratios in the design and reporting of their studies and should assess the proportional hazards assumption in the analysis.
- Do formal tests have any value in assessing PH?
- Does the 'tests are unnecessary' claim apply to all effect modifiers and to other models?

Why Are Hazards Usually Not Proportional?

Quotes from Stensrud & Hernán [1]

1. Hazards are not proportional when the treatment effect changes over time.
2. Hazards may also not be proportional because disease susceptibility varies between individuals [6].

- (1) is just the familiar assumption of constancy of effect, often called no interaction or no effect modification, where the potential effect modifier in this case is time.
- (1) applies to other covariates in the Cox model and to other regression models whereas (2) is specific to time.
- Does this mean we should never perform statistical tests for effect modification?

‘Statistical tests for PH are unnecessary’

Because it is expected that the hazard ratio will vary over the follow-up period, tests of proportional hazards yielding high P values are probably underpowered.

- I agree, but am concerned that the ‘tests are unnecessary’ statement may be interpreted by some as ‘assessing PH is unnecessary’ or ‘it’s fine to just report the HR from a PH model’.
- Researchers should consider the time-varying nature of hazard ratios in the design and reporting of their studies and should assess the proportional hazards assumption in the analysis.
- Another issue is that there is no omnibus test of PH.
- Arguably the most common test, based on scaled Schoenfeld residuals, tests the null of PH against the alternative that the HR changes as a linear or log-linear function of time.

Alternative measures

Quote from Stensrud & Hernán [1]

Reports of hazard ratios should be supplemented with reports of effect measures directly calculated from absolute risks, such as the survival differences or the restricted mean survival difference, at times prespecified in the study protocol. These measures are arguably more helpful for clinical decision-making and more easily understood by patients.

- I very much agree.

Estimating the HR from a PH model

Quote from Stensrud & Hernán [1]

Another limitation is that the magnitude of the Cox HR depends on the distribution of losses to follow-up (censoring), even if the losses occur at random. This limitation can be overcome by estimating an inverse probability-weighted hazard ratio.

- The statement is indisputably true, but how much difference does it make in practice?
- The authors show using simulations (see next slide taken from supplementary material) that differences can be considerable.
- Those three scenarios, however, concern large departures from PH and I would not consider reporting the HR from a PH model.
- How large is the ‘bias’ when a PH model is reasonable?

Table from supplementary material

Table. Simulated trials under the 3 scenarios described in the Figure in the main text. Each trial included 50,000 individuals and was analyzed first including all individuals and then after randomly censoring individuals such that about 20% of the events were unmeasured. The magnitude of the Cox hazard ratio depends on the censoring proportion even though the survival difference does not change.

Scenario	Censoring	Hazard ratio (95% CI), Cox proportional hazards model	3-year survival difference, % (95% CI), Kaplan-Meier estimator
1	No	0.69 (0.66 to 0.72)	3.2 (2.6 to 3.8)
	Yes	0.71 (0.67 to 0.74)	3.1 (2.5 to 3.8)
2	No	0.51 (0.48 to 0.54)	3.6 (3.1 to 4.1)
	Yes	0.62 (0.58 to 0.66)	3.6 (3.0 to 4.1)
3	No	1.27 (1.22 to 1.32)	–5.2 (–5.8 to –4.5)
	Yes	1.34 (1.28 to 1.40)	–5.2 (–5.9 to –4.5)

Conclusion from Sjölander and Dickman (2024) [5]

Conclusion

Statistical inference is built upon assumptions. While we note that not all assumptions are equally realistic, and not all assumptions are necessary for inference, we also note that the proportional hazards assumption is similar to other assumptions commonly made in statistical modelling. Formal statistical tests of proportional hazards may be unnecessary, but analysts should assess the appropriateness of the assumption for their data and research question. Thus, analysts must understand the assumption, how and why it might be violated, and how one interprets estimated hazard ratios from a proportional hazards model; the tutorial by SH is an excellent resource for gaining such understanding.

Annals of Internal Medicine

ORIGINAL RESEARCH

Risk for Arterial and Venous Thrombosis in Patients With Myeloproliferative Neoplasms

A Population-Based Cohort Study

Malin Hultcrantz, MD, PhD; Magnus Björkholm, MD, PhD; Paul W. Dickman, MSc, PhD; Ola Landgren, MD, PhD; Åsa R. Derolf, MD, PhD; Sigurdur Y. Kristinsson, MD, PhD*; and Therese M.L. Andersson, MSc, PhD*

Background: Patients with myeloproliferative neoplasms (MPNs) are reported to be at increased risk for thrombotic events. However, no population-based study has estimated this excess risk compared with matched control participants.

Objective: To assess risk for arterial and venous thrombosis in patients with MPNs compared with matched control participants.

Design: Matched cohort study.

Setting: Population-based setting in Sweden from 1987 to 2009, with follow-up to 2010.

Patients: 9429 patients with MPNs and 35 820 matched control participants.

Measurements: The primary outcomes were rates of arterial and venous thrombosis. Flexible parametric models were used to calculate hazard ratios (HRs) and cumulative incidence with 95% CIs.

Results: The HRs for arterial thrombosis among patients with MPNs compared with control participants at 3 months, 1 year, and 5 years were 3.0 (95% CI, 2.7 to 3.4), 2.0 (CI, 1.8 to 2.2), and 1.5 (CI, 1.4 to 1.6), respectively. The corresponding HRs for venous thrombosis were 9.7 (CI, 7.8 to 12.0), 4.7 (CI, 4.0 to 5.4), and 3.2 (CI, 2.9 to 3.6). The rate was significantly elevated across all age groups and was similar among MPN subtypes. The 5-year cumulative incidence of thrombosis in patients with MPNs showed an initial rapid increase followed by gentler increases during follow-up. The HR for venous thrombosis decreased during more recent calendar periods.

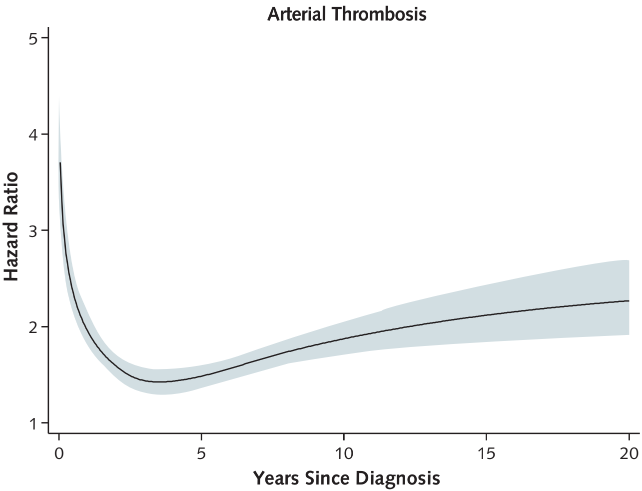
Limitation: No information on individual laboratory results or treatment.

Conclusion: Patients with MPNs across all age groups have a significantly increased rate of arterial and venous thrombosis compared with matched control participants, with the highest rates at and shortly after diagnosis. Decreases in the rate of venous thrombosis over time likely reflect advances in clinical management.

Primary Funding Source: The Cancer Research Foundations of Radiumhemmet, Blodcancerfonden, the Swedish Research Council, the regional agreement on medical training and clinical research between Stockholm County Council and Karolinska Institutet, the Adolf H. Lundin Charitable Foundation, and Memorial Sloan Kettering Cancer Center.

Ann Intern Med. 2018;168:317-325. doi:10.7326/M17-0028 **Annals.org**
For author affiliations, see end of text.
This article was published at Annals.org on 16 January 2018.
* Drs. Kristinsson and Andersson contributed equally to this work.

Figure 1. Arterial (*top*) and venous (*bottom*) thrombosis during follow-up in patients with MPNs versus matched control participants.



https://news.ki.se/calendar/biostatistics-network-meeting-biostatistics-meets-artificial-intelligence

CONFERENCES AND SYMPOSIUMS

Biostatistics network meeting: Biostatistics meets artificial intelligence

04-04-2024
Add to iCal

4:00 pm - 8:30 pm

Campus Solna
Lecture hall Erling Persson, Aula Medica

MSc Biostatistics and Data Science

- Interdisciplinary programme designed for students with a background in mathematics, statistics, and computer science.
- Aims to train qualified data analysts and researchers in the fields of biology, medicine, and public health.
- Designed to meet the data-driven future of the life sciences.
- The first and only programme of its kind in the Nordics.
- Starting in the fall of 2024!

Karolinska Institutet KTH Stockholm University

References

[1] Stensrud MJ, Hernán MA. Why test for proportional hazards? *JAMA* 2020;**323**:1401–1402.

[2] Lang JM, Rothman KJ, Cann CI. That confounded p-value. *Epidemiology* 1998;**9**:7–8.

[3] Hernán MA. Causal analyses of existing databases: no power calculations required. *Journal of clinical epidemiology* 2022;**144**:203–205.

[4] Hernán MA, Greenland S. Why stating hypotheses in grant applications is unnecessary. *JAMA* 2024;**331**:285–286.

[5] Sjölander A, Dickman P. Why test for proportional hazards - or any other model assumptions? *American journal of epidemiology* 2024;.

[6] Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010;**21**:13–15.