

JAMA Guide to Statistics and Methods

Why Test for Proportional Hazards?

Mats J. Stensrud, MD, DrPhilos; Miguel A. Hernán, MD, DrPH

The Cox proportional hazards model, introduced in 1972,¹ has become the default approach for survival analysis in randomized trials. The Cox model estimates the ratio of the hazard of the event or outcome of interest (eg, death) between 2 treatment groups. In



Supplemental content

formally, the hazard at any given time is the probability of experiencing the event of interest in the next interval among individuals who had not yet experienced the event by the start of the interval. Because the Cox model requires the hazards in both groups to be proportional, researchers are often asked to “test” whether hazards are proportional.

What Does It Mean That Hazards Are Proportional?

The hazards are proportional if the hazard ratio remains constant from day 1 of the study until the end of follow-up. In practice, this does not occur for most medical interventions. Three articles previously published in *JAMA* illustrate different scenarios regarding proportional hazards.

Scenario 1—No Immediate Effect

The Air Force/Texas Coronary Atherosclerosis Prevention Study² randomly assigned patients with atherosclerotic cardiovascular disease to either statin therapy or placebo. The hazard ratio of a major adverse cardiovascular event was 0.63 (95% CI, 0.50-0.79) for statin vs placebo. However, the cumulative incidences of major adverse cardiovascular event in the statin and placebo groups were almost identical during the first 6 months of follow-up and diverged thereafter. That is, the overall hazard ratio of 0.63 was a weighted average of the time-varying hazard ratios, which were close to 1 in the first months of follow-up and declined later.

Scenario 2—Immediate and Delayed Effects in Opposite Directions

The Norwegian Colorectal Cancer Prevention Trial³ randomly assigned individuals aged 50 to 64 years to flexible sigmoidoscopy screening or no screening. The hazard ratio of colorectal cancer was 0.80 (95% CI, 0.70-0.92) for screening vs no screening. However, the cumulative incidence was greater in the screening group until about 5 years of follow-up and lower after that time. That is, the hazard ratio of 0.80 was a weighted average of the time-varying hazard ratios, which were greater than 1 in the early follow-up and less than 1 in the later follow-up.

Scenario 3—Variations in Disease Susceptibility

A Women's Health Initiative study⁴ randomly assigned postmenopausal women to either estrogen plus progestin hormone therapy or placebo. The hazard ratio of coronary heart disease was 1.24 (95% CI, 1.00-1.54) for hormone therapy vs placebo. However, the hazard ratio was 1.8 during the first year and 0.70 after 5 years of follow-up. The overall hazard ratio of 1.24 was a weighted average of the time-varying hazard ratios throughout the follow-up.

Why Are Hazards Usually Not Proportional in Medical Studies?

Hazards are not proportional when the treatment effect changes over time. In scenario 1, the effect of statin therapy on cardiovascular events only became evident after 6 months or longer. In scenario 2, screening for colorectal cancer had both an immediate effect on the detection of undiagnosed cancers (hazard ratio of colorectal cancer greater than 1 in the early follow-up) and a delayed preventive effect due to the removal of cancer precursors (hazard ratio less than 1 later in the follow-up).⁵

Hazards may also not be proportional because disease susceptibility varies between individuals. Those with greater disease susceptibility are more likely to develop the disease earlier. In scenario 3, some women had a greater risk of coronary heart disease than others because of, for example, a genetic predisposition. Even if hormone therapy increased the risk of disease by a constant factor (eg, by 80%) at every single time of follow-up, it is still possible that the hazard ratio would have declined from 1.8 during the first year of follow-up to less than 1 in later years because the most susceptible women would have been diagnosed with coronary heart disease in the early follow-up.⁶ As a result, the most susceptible women would have been removed more rapidly from the treatment group than from the control group. After 5 years, women without coronary heart disease in the treatment group would have been, on average, less susceptible to develop the disease than those in the placebo group. That is, remaining disease-free for 5 years in the treatment group would have become a proxy for being resistant to the development of coronary heart disease. The hazard ratio after 5 years can be less than 1 even if hormone therapy did not prevent coronary heart disease in any women in the study.^{4,5}

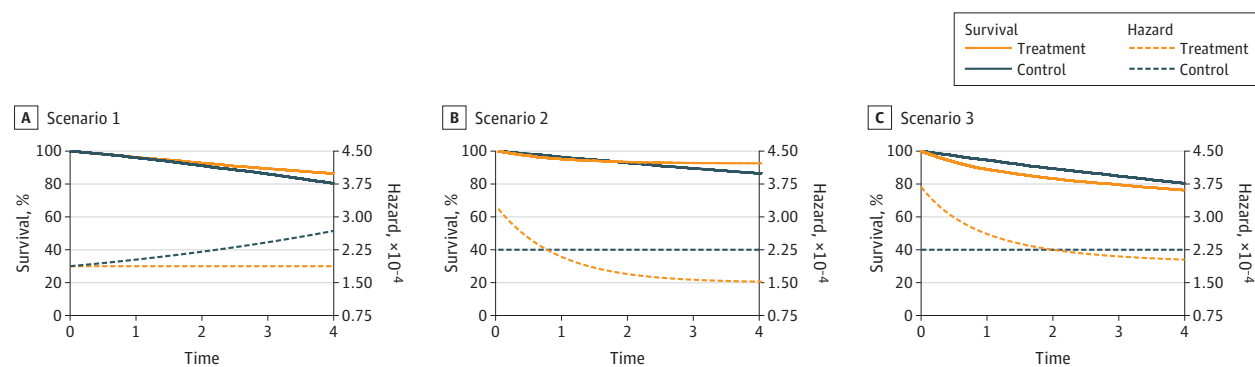
The Figure depicts the 3 scenarios described above. These examples illustrate why hazards are not expected to be proportional in almost any clinical study. The exception is when the treatment has no effect—then the hazard ratio is constant at 1 throughout the follow-up.

What Are the Problems of Using Hazard Ratios From Proportional Hazards Models?

A mortality hazard ratio estimate of, for instance, 0.7 for treatment vs placebo cannot be interpreted as a constant 30% mortality decrease in the treatment group at all times during the follow-up period. Rather, a hazard ratio of 0.7 means that, on average, treatment decreases mortality during the follow-up period. The magnitude of the cumulative benefit at a particular time can only be conveyed by a comparison of the survival (proportion of individuals alive) in each group.

One limitation of using Cox regression models when the hazard ratio is not constant during the follow-up period is reporting an incorrect standard variance estimator when the statistical model includes covariates other than the treatment group indicator.⁷ This limitation

Figure. Nonproportional Hazards and Survival Curves in 3 Hypothetical Trials Comparing a Treatment vs a Control



In scenario 1, both the hazards (dotted lines) and the survival curves (solid lines) gradually diverge (ie, the hazard ratio is not constant but is always greater than 1). In scenario 2, both the hazards and the survival curves cross (ie, the hazard ratio goes from greater than 1 to less than 1). In scenario 3, the hazards cross

because of depletion of susceptibles in the treatment group, but the survival curves do not cross. The hazards would have been proportional if the dotted lines were straight and horizontal.

can be overcome, and valid 95% confidence intervals can be estimated, by using bootstrapping methods. Another limitation is that the magnitude of the Cox hazard ratio 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 (eAppendix in the Supplement).

How Should Hazard Ratios Be Interpreted?

As a weighted average of the time-varying hazard ratios, the hazard ratio estimate from a Cox proportional hazards model is often used as a convenient summary of the treatment effect during the follow-up. However, a hazard ratio from a Cox model needs to be interpreted as a weighted average of the true hazard ratios over the

entire follow-up period. The 95% confidence interval should be estimated using a valid method such as bootstrapping and also using inverse probability weighting to adjust for losses to follow-up.

An implication is that statistical tests for proportional hazards 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.

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,⁸ at times prespecified in the study protocol. These measures are arguably more helpful for clinical decision-making and more easily understood by patients.

ARTICLE INFORMATION

Author Affiliations: Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Stensrud, Hernán); Department of Biostatistics, Oslo Centre for Biostatistics and Epidemiology, University of Oslo, Oslo, Norway (Stensrud); Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Hernán); Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts (Hernán).

Corresponding Author: Mats J. Stensrud, MD, DrPhilos, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115 (mstensrud@hsph.harvard.edu).

Section Editors: Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA; and Edward H. Livingston, MD, Deputy Editor, JAMA.

Published Online: March 13, 2020.
doi:10.1001/jama.2020.1267

Conflict of Interest Disclosures: Dr Stensrud reported receiving grants from the Research

Council of Norway (NFR239956/F20) and the ASISA Fellowship. Dr Hernán reported receiving a grant from the National Institutes of Health (R37 AI02634).

Additional Information: The simulated data and R code are available at <https://github.com/CausalInference>.

REFERENCES

- Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol*. 1972;34:187-202.
- Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. *JAMA*. 1998;279(20):1615-1622. doi:10.1001/jama.279.20.1615
- Holme Ø, Løberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized clinical trial. *JAMA*. 2014;312(6):606-615. doi:10.1001/jama.2014.8266
- Rossouw JE, Anderson GL, Prentice RL, et al; Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288(3):321-333. doi:10.1001/jama.288.3.321
- Manson JE, Hsia J, Johnson KC, et al; Women's Health Initiative Investigators. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med*. 2003;349(6):523-534. doi:10.1056/NEJMoa030808
- Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13-15. doi:10.1097/EDE.0b013e3181c1ea43
- DiRienzo A, Lagakos S. Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *J R Stat Soc Series B Stat Methodol*. 2001;63:745-757. doi:10.1111/1467-9868.00310
- Pak K, Uno H, Kim DH, et al. Interpretability of cancer clinical trial results using restricted mean survival time as an alternative to the hazard ratio. *JAMA Oncol*. 2017;3(12):1692-1696. doi:10.1001/jamaoncol.2017.2797