An introduction to flexible parametric survival models and a discussion of the proportional hazards assumption

Paul W Dickman
Professor of Biostatistics
Karolinska Institutet

Talk at: Department of Statistics and Quantitative Methods
University of Milano-Bicocca
31 Mar 2023
`http://pauldickman.com/talk/`

Today's talk

- A non-technical introduction to flexible parametric survival models and why I like them.
- Overview of some of my research that extends and/or applies flexible parametric survival models.
- A general discussion of the proportional hazards assumption and discussion of the paper "Why Test for Proportional Hazards?" by Mats Stensrud & Miguel Hernán.
- Builds upon the two previous talks.

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 current department 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 net survival.
- General interest in statistical aspects of the design, analysis, and reporting of epidemiological studies.
- Epidemiology, with particular focus on cancer epidemiology and perinatal/reproductive epidemiology.
- Lots of administrative work.

About Karolinska Institutet and MEB

- Karolinska Institutet, founded in 1810, is a medical university;
  - 6500 undergraduate/masters students;
  - 2000 doctoral students;
  - 5000 FTE staff;
  - Ranked 41 (1st in Sweden) overall in Shanghai rankings.
- Since 1901, the Nobel Assembly at Karolinska Institutet has selected the Nobel laureates in Physiology or Medicine.
- Department of Medical Epidemiology and Biostatistics (MEB) one of 22 departments; 300 staff (including doctoral students).
- Focus on register-based epidemiology; especially strong in cancer epidemiology, psychiatric epidemiology, and biostatistics.

Paradise? [1]

A paradise for epidemiologists?

Hans-Olov Adami The Lancet 1996;2:588

For three reasons—the structure of its health system, the existence of nationwide registers, and the systematic use of national registration numbers—Sweden offers exceptional opportunities for epidemiological research.

- I would add 'willingness of the public to contribute to research' and 'outstanding clinical researchers'.

Some common survival models in epidemiology

- Commonly used models have the same basic formulation.

$$h_i(t) = h_0(t)\exp(\mathbf{x_i}\beta)$$

$$\ln(h_i(t)) = \ln(h_0(t)) + \mathbf{x_i}\beta$$

- Proportional hazards assumed by default (but can be relaxed).
- Primary difference is in specification of the baseline hazard:
  - Cox model: $h_0(t)$ an arbitrary function of time; not estimated.
  - Poisson regression model: $h_0(t)$ is a step function.
  - Weibull model: $h_0(t) = \lambda\gamma t^{\gamma-1}$
  - Flexible parametric model: $h_0(t)$ modelled using splines.

Why I use flexible parametric survival models

- I analyse large population-based datasets where
  - The proportional hazards assumption is rarely appropriate.
  - The hazard function is of interest.
  - A hazard ratio does not tell the whole story.
- I model excess mortality/net survival among cancer patients.
  - Not possible to fit the Cox model.
  - Proportional excess hazards assumption is rarely appropriate.
  - Quantities other than the excess hazard ratio are of interest.
- Quantification and presentation of absolute risks and rates.
  - Should be done more than it is.
  - Much easier with parametric estimate of the baseline hazard.
- Many useful extensions are much easier in a parametric setting.

Sex differences in bladder cancer survival [3]

European Journal of Cancer 95 (2018) 52–58

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Bladder cancer survival: Women better off in the long run

Bettina Kulle Andreassen [a,*], Tom Kristian Grimsrud [a], Erik Skaaheim Haug [b,c]

- See Radkiewicz *et al.* (2017) [2] for a similar Swedish study.

Time-varying excess hazard ratio [3]

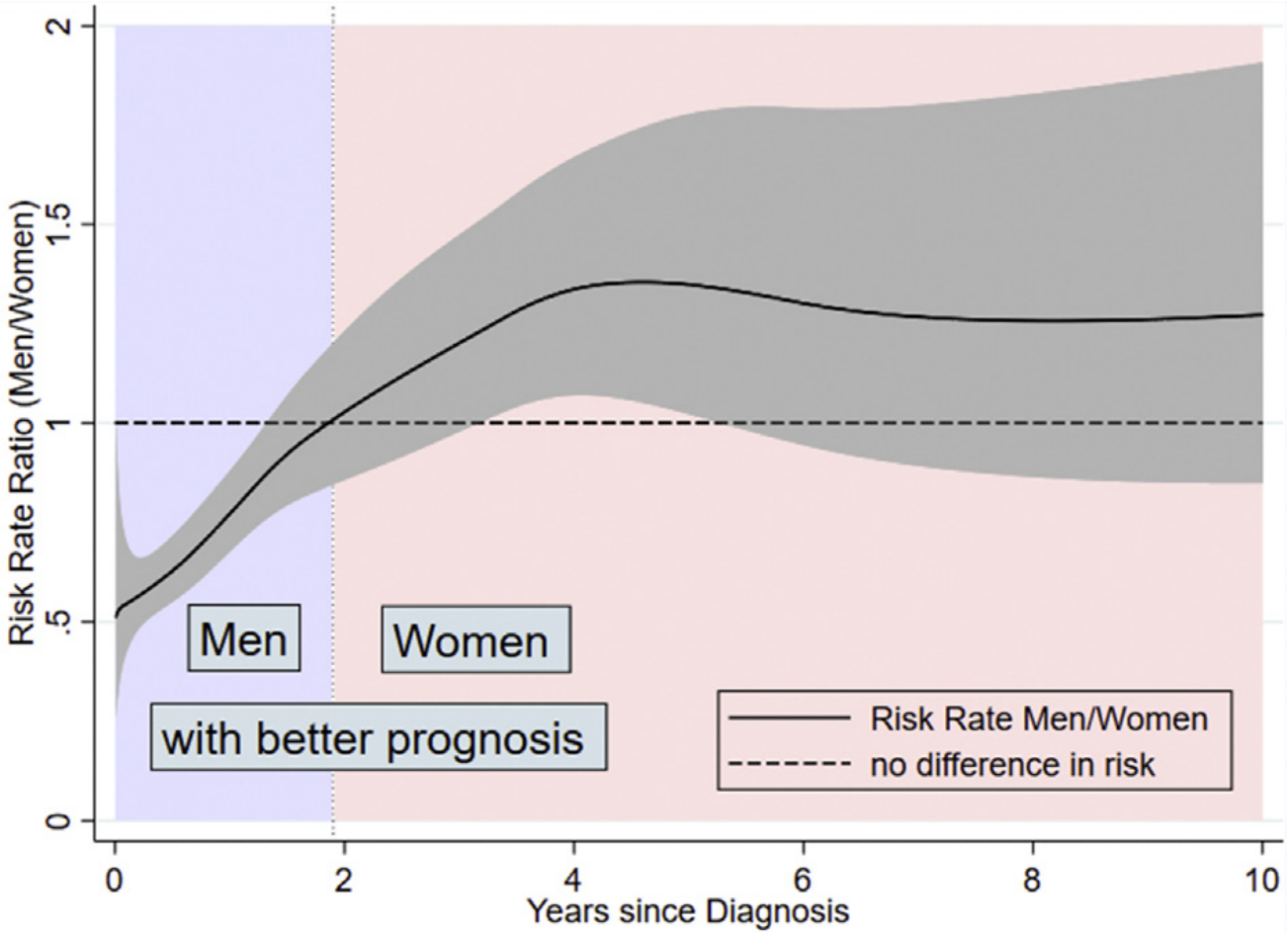
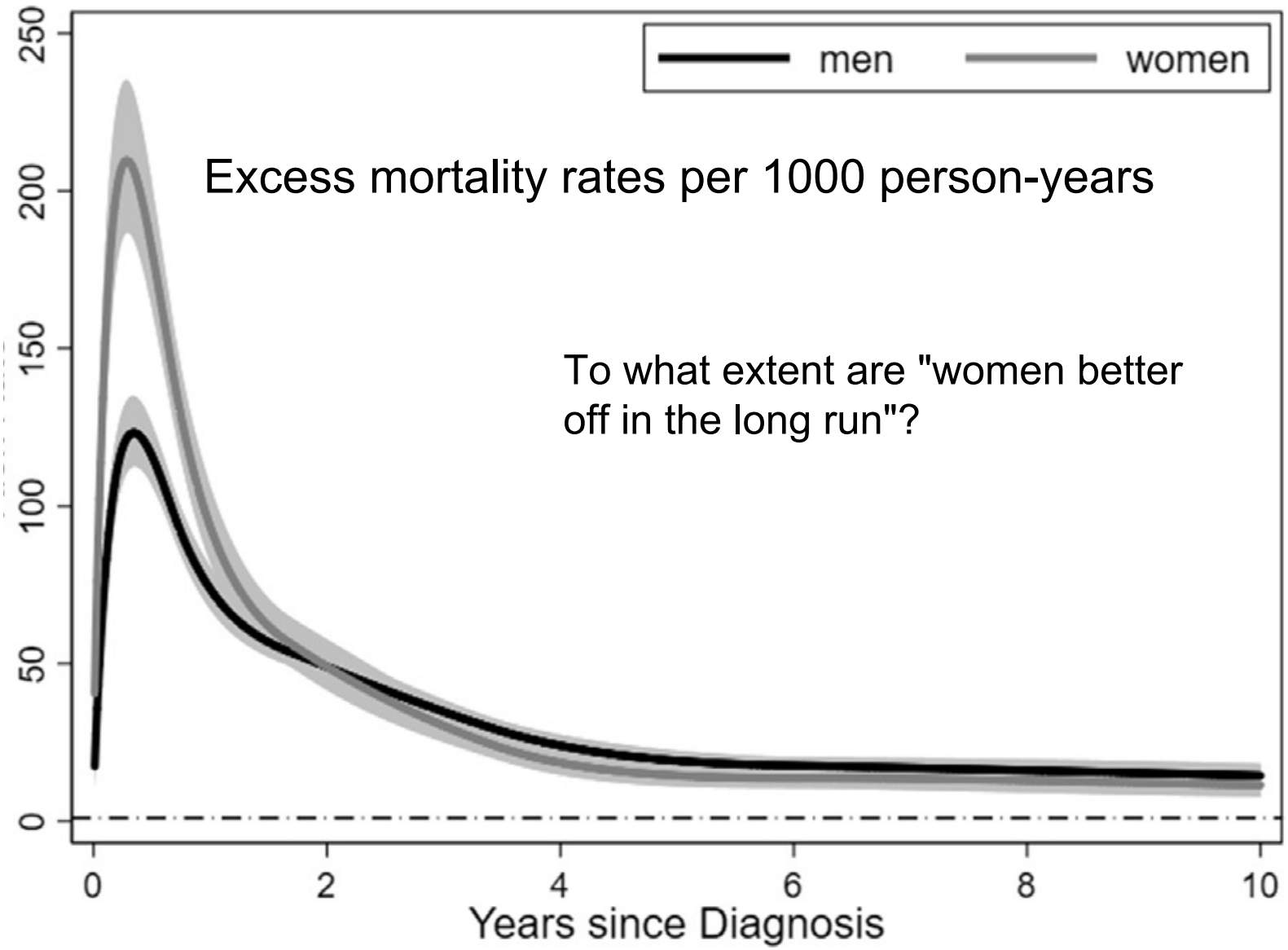


Fig. 2. Risk ratio (excess mortality rate ratio) including confidence intervals for men versus women with bladder cancer diagnosis. The

Baseline excess mortality rates [3]



Marginal and standardised survival [3, 4]

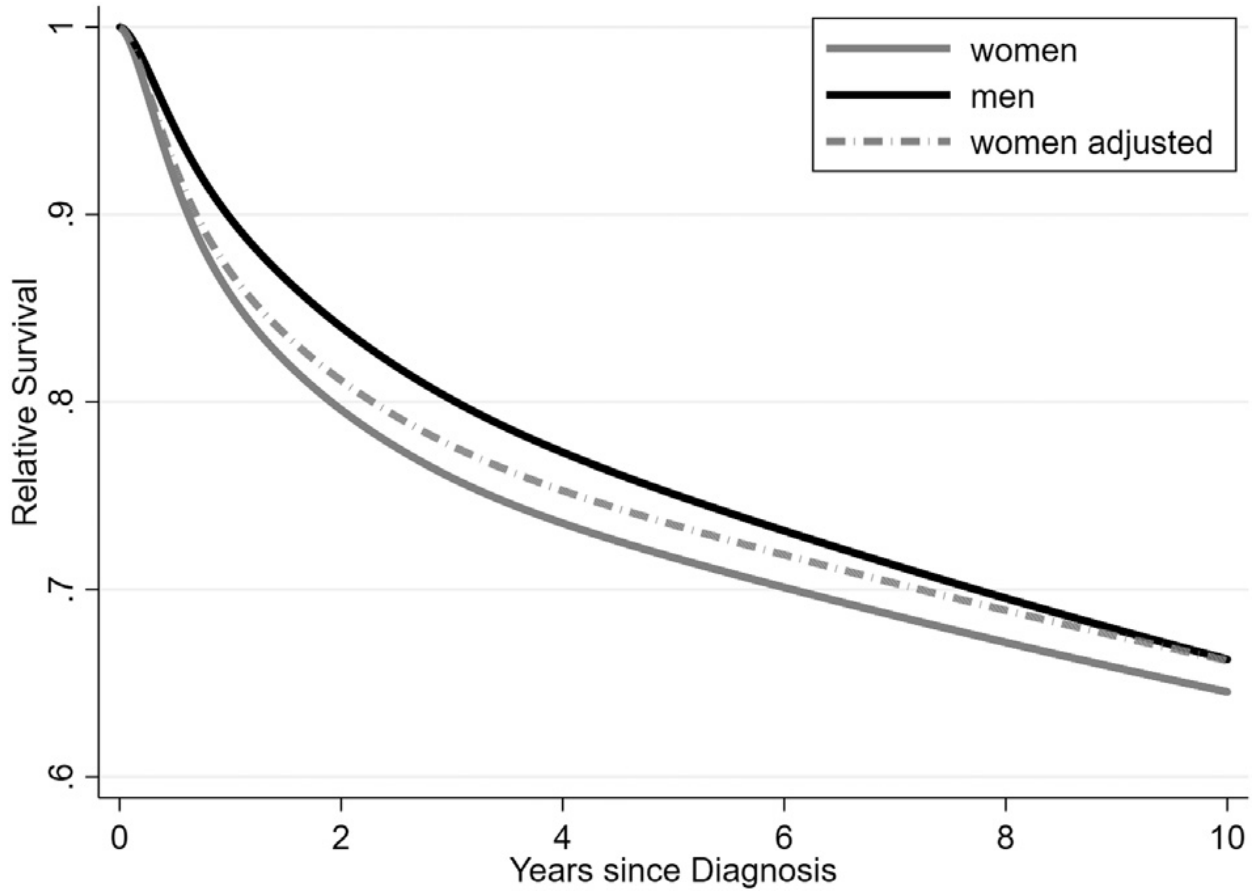
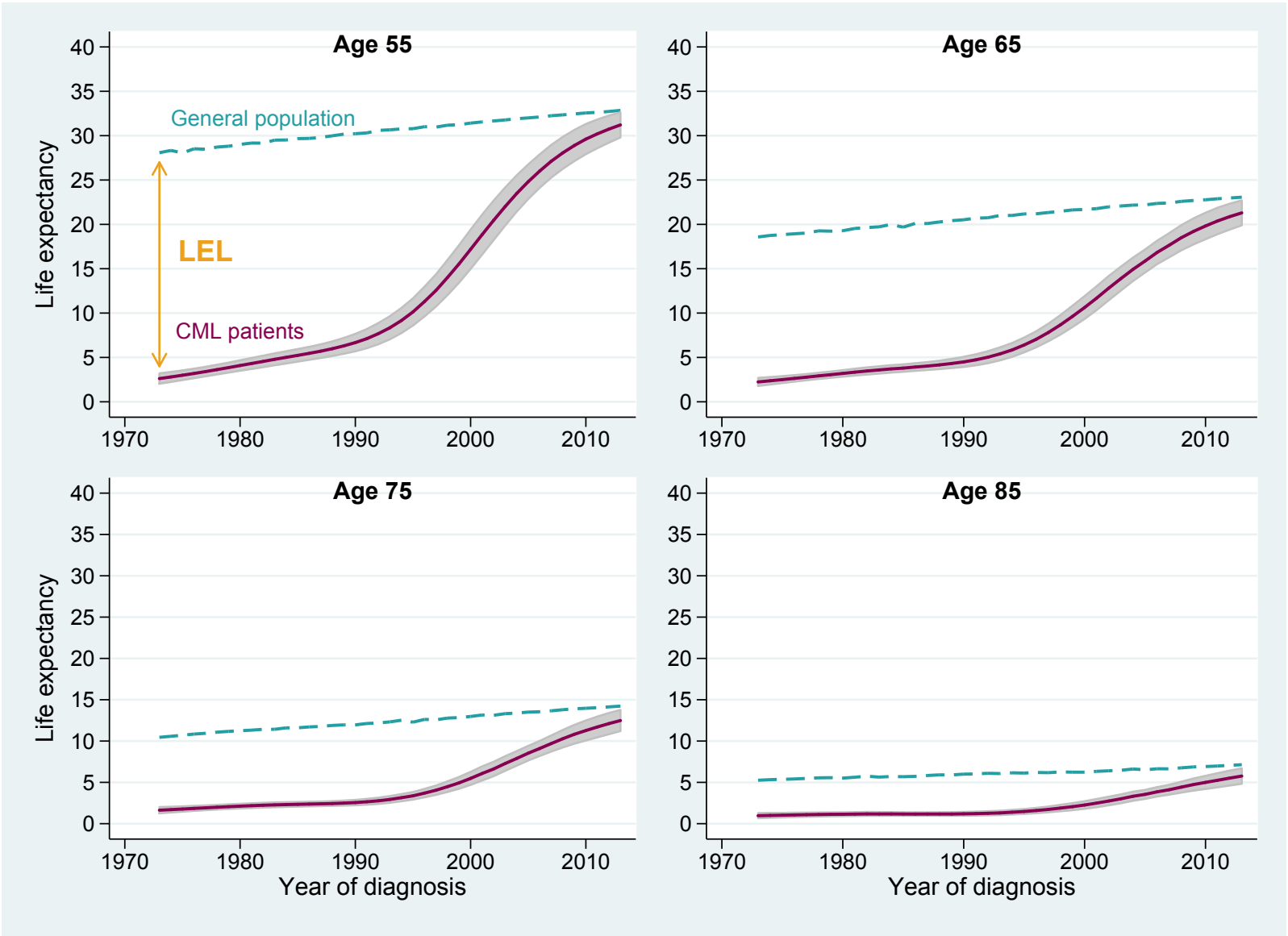


Fig. 3. Relative survival for men, women and women assuming the same T-stage distribution as men. Black (grey) lines: mean survival curve for men (women); Dashed grey line: survival curve for women when assuming men's covariate pattern.

Loss in expectation of life: CML (Sweden) [5]



A sneak peek at my conclusions

- I use and advocate flexible parametric survival models. However,
- There is nothing wrong with using a Cox model.
- If you only want to estimate a hazard ratio and you 'know' you have proportional hazards then a Cox model is ideal.
- Can relax the PH assumption in the Cox model, and can estimate quantities other than HR.
- However, a parametric approach makes it easier to estimate quantities that provide more insight and may be more relevant to your research question.
- You will get the same hazard ratio, but a whole lot more.

An interview with Sir David Cox (Reid 1994 [6])

Reid "What do you think of the cottage industry that's grown up around [the Cox model]?"

Cox "In the light of further results one knows since, I think I would normally want to tackle the problem parametrically. . . . I'm not keen on non-parametric formulations normally."

Reid "So if you had a set of censored survival data today, you might rather fit a parametric model, even though there was a feeling among the medical statisticians that that wasn't quite right."

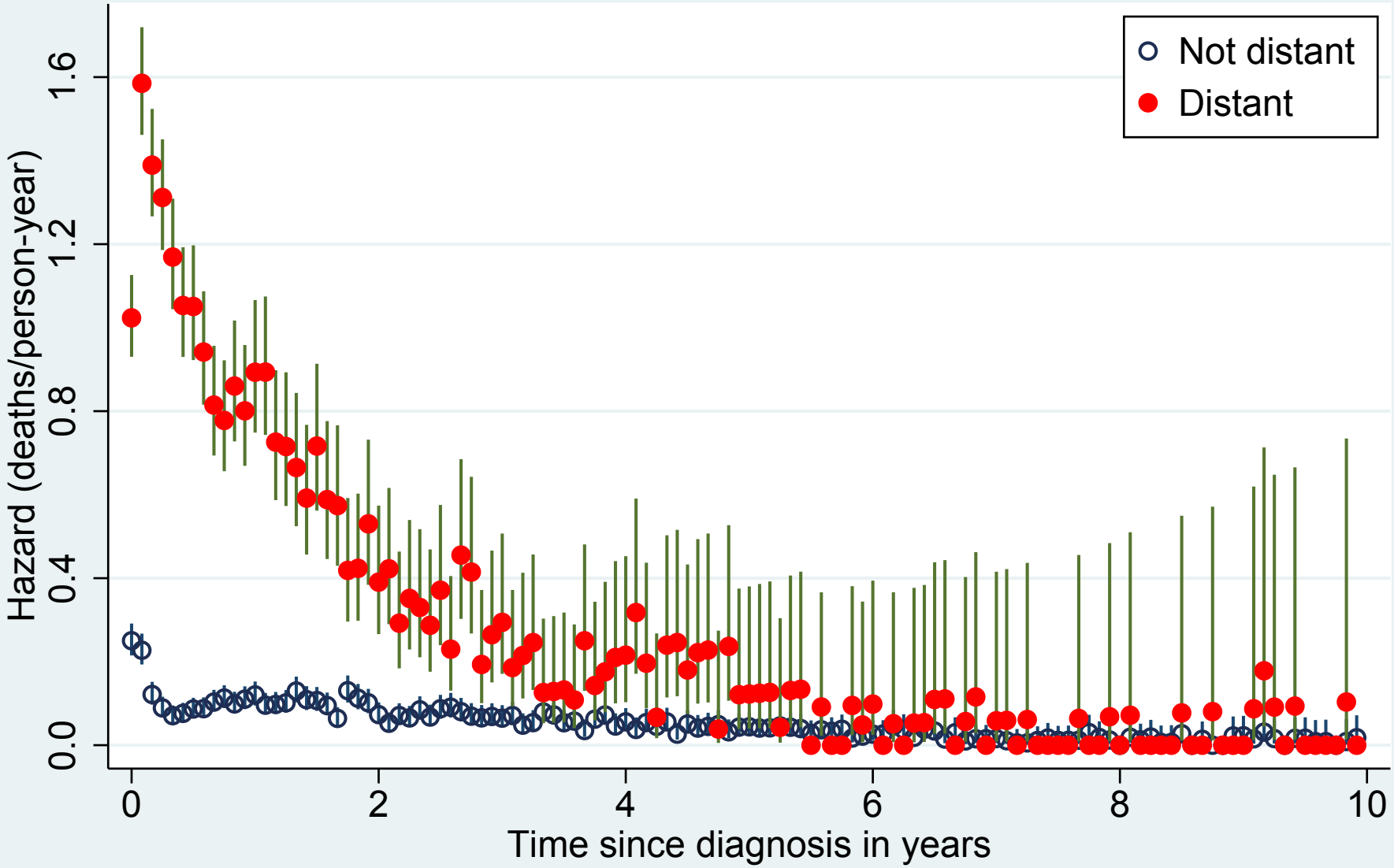
Cox "That's right, but since then various people have shown that the answers are very insensitive to the parametric formulation of the underlying distribution. And if you want to do things like predict the outcome for a particular patient, it's much more convenient to do that parametrically."

Example: survival of patients diagnosed with colon carcinoma

- I will use this dataset throughout the lecture.
- Patients diagnosed with colon carcinoma 1984–95. Potential follow-up to end of 1995; censored after 10 years.
- Outcome is death due to colon carcinoma.
- Interest is in the effect of clinical stage at diagnosis (distant metastases vs no distant metastases).
- How might we specify a statistical model for these data?

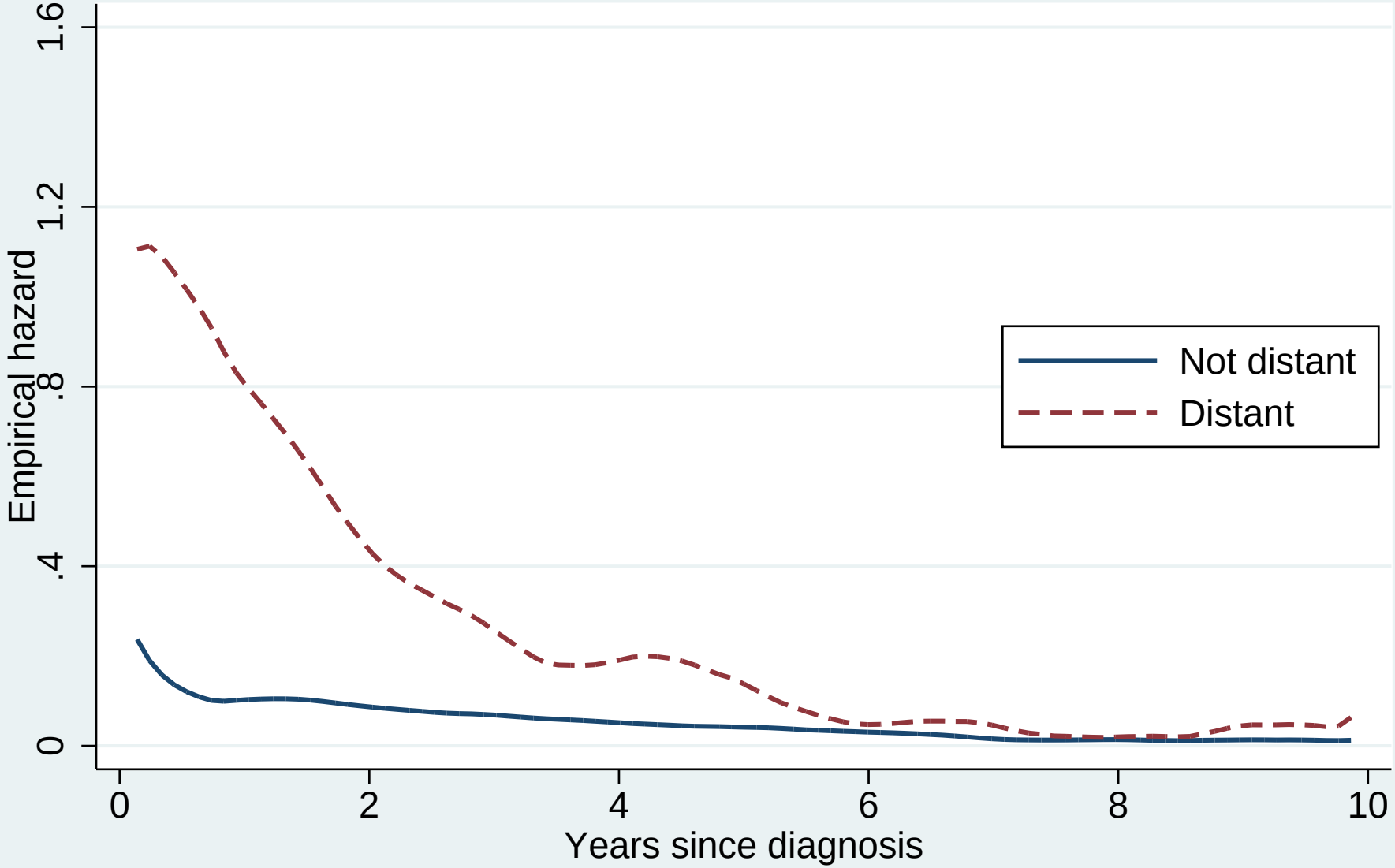
Empirical hazards by stage

strate fu distant, graph



Smoothed empirical hazards (cancer-specific mortality rates)

sts graph, by(distant) hazard kernel(epan2)



The Cox proportional hazards model

- The 'intercept' in the Cox model [7], the hazard (event rate) for individuals with all covariates x at the reference level, can be thought of as an arbitrary function of time[1], often called the baseline hazard and denoted by $h_0(t)$.
- The hazard at time t for individual with other covariate values is a multiple of the baseline

$$h(t|x) = h_0(t)\exp(x\beta).$$

- Alternatively

$$\ln[h(t|x)] = \ln[h_0(t)] + x\beta.$$

- Does not explicitly estimate $h_0(t)$ while estimating the log hazard ratios (β).

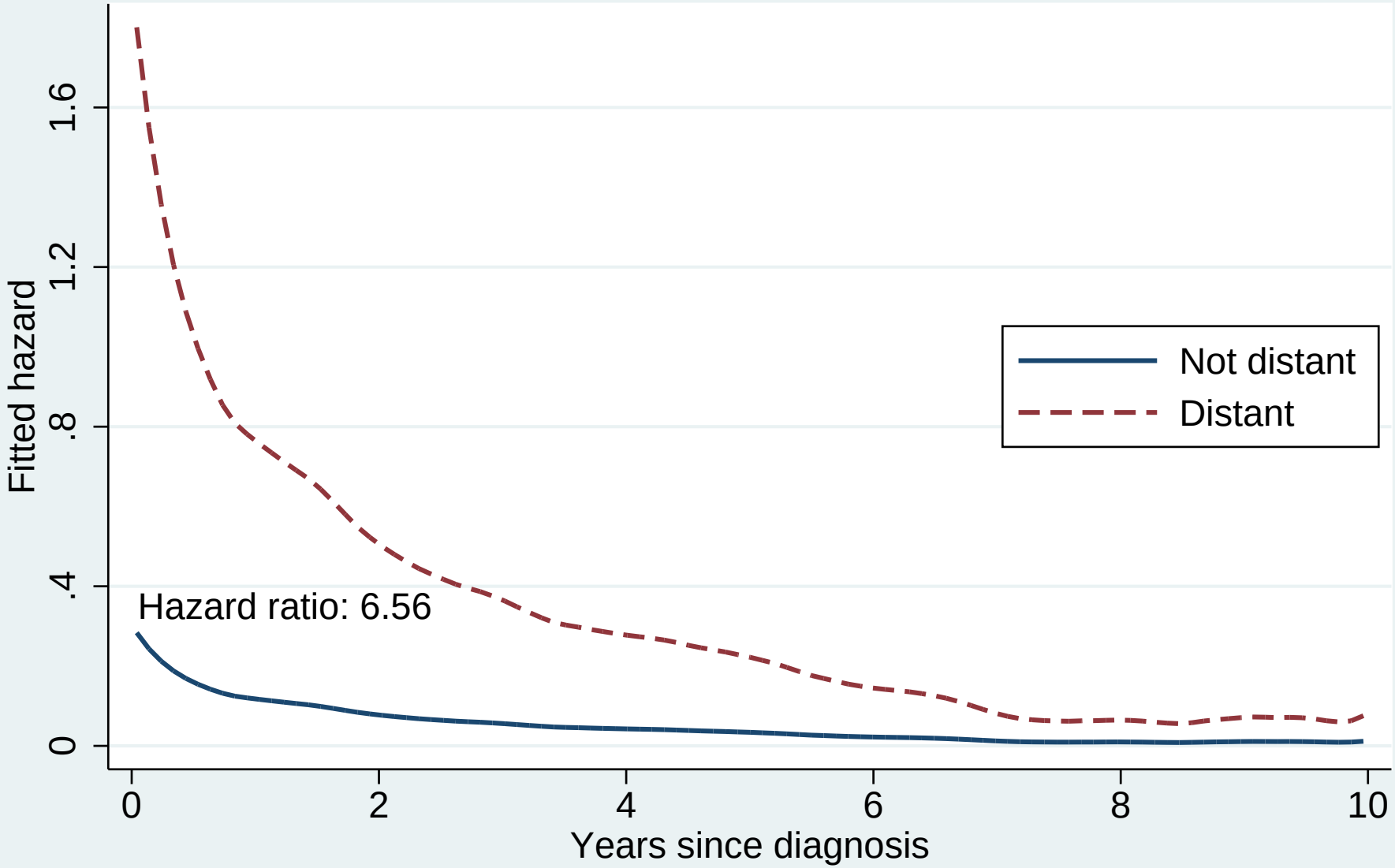
[1] time t can be defined in many ways, e.g., attained age, time-on-study, calendar time, etc.

Fit a Cox model to estimate the mortality rate ratio

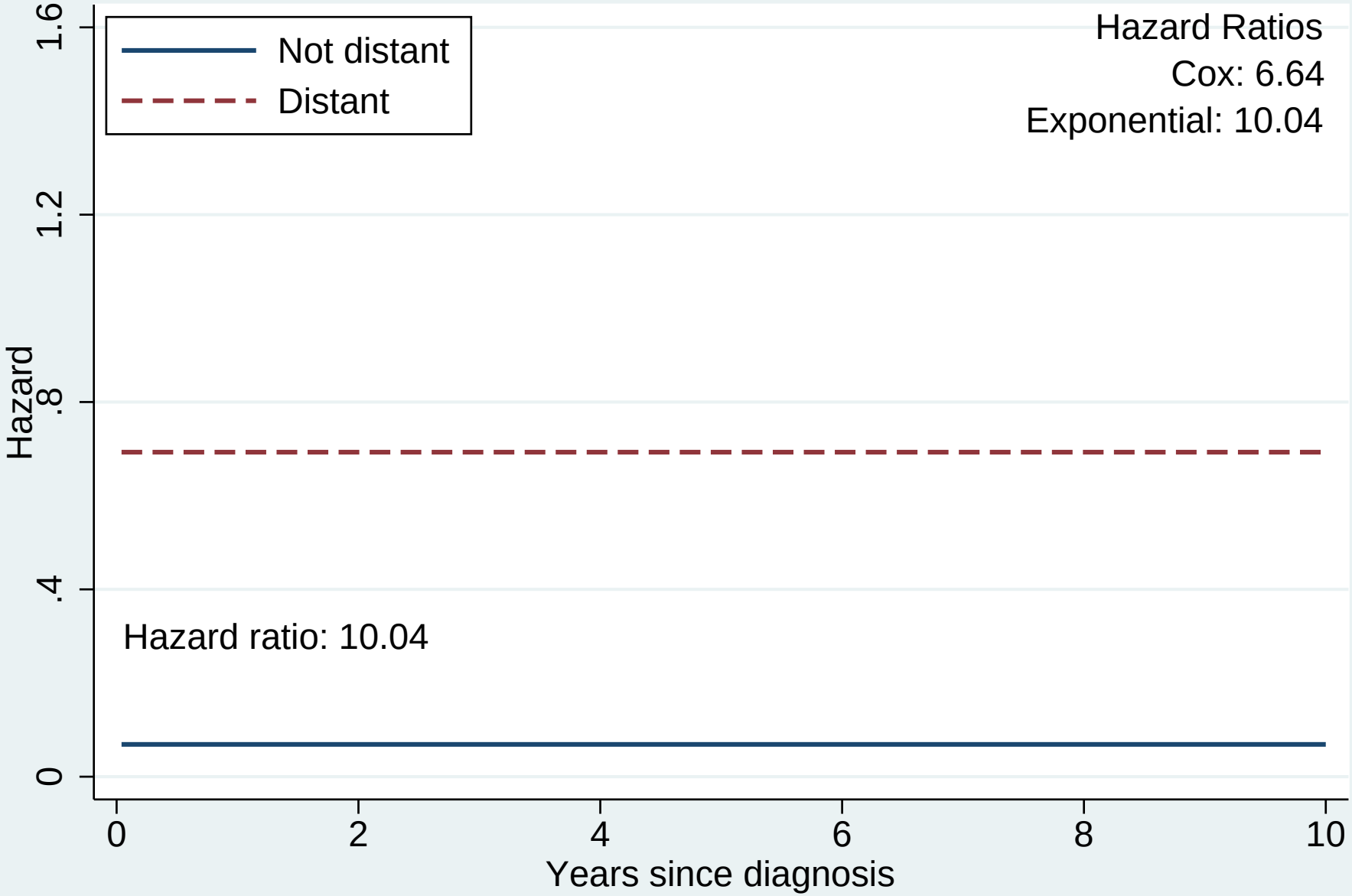
```
. stcox distant
         failure _d:  status == 1
   analysis time _t:  (exit-origin)/365.25
             origin:  time dx
               note:  time>10 trimmed

Cox regression -- Breslow method for ties

No. of subjects =        13208       Number of obs  =   13208
No. of failures =         7122
Time at risk    =  44013.26215
                                     LR chi2(1)     = 5544.65
Log likelihood  =   -61651.446       Prob > chi2    =  0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% C.I.]
-------------+----------------------------------------------------------------
distant |   6.557777   .1689328    73.00   0.000   6.235   6.897
------------------------------------------------------------------------------
```

Fitted hazards from Cox model

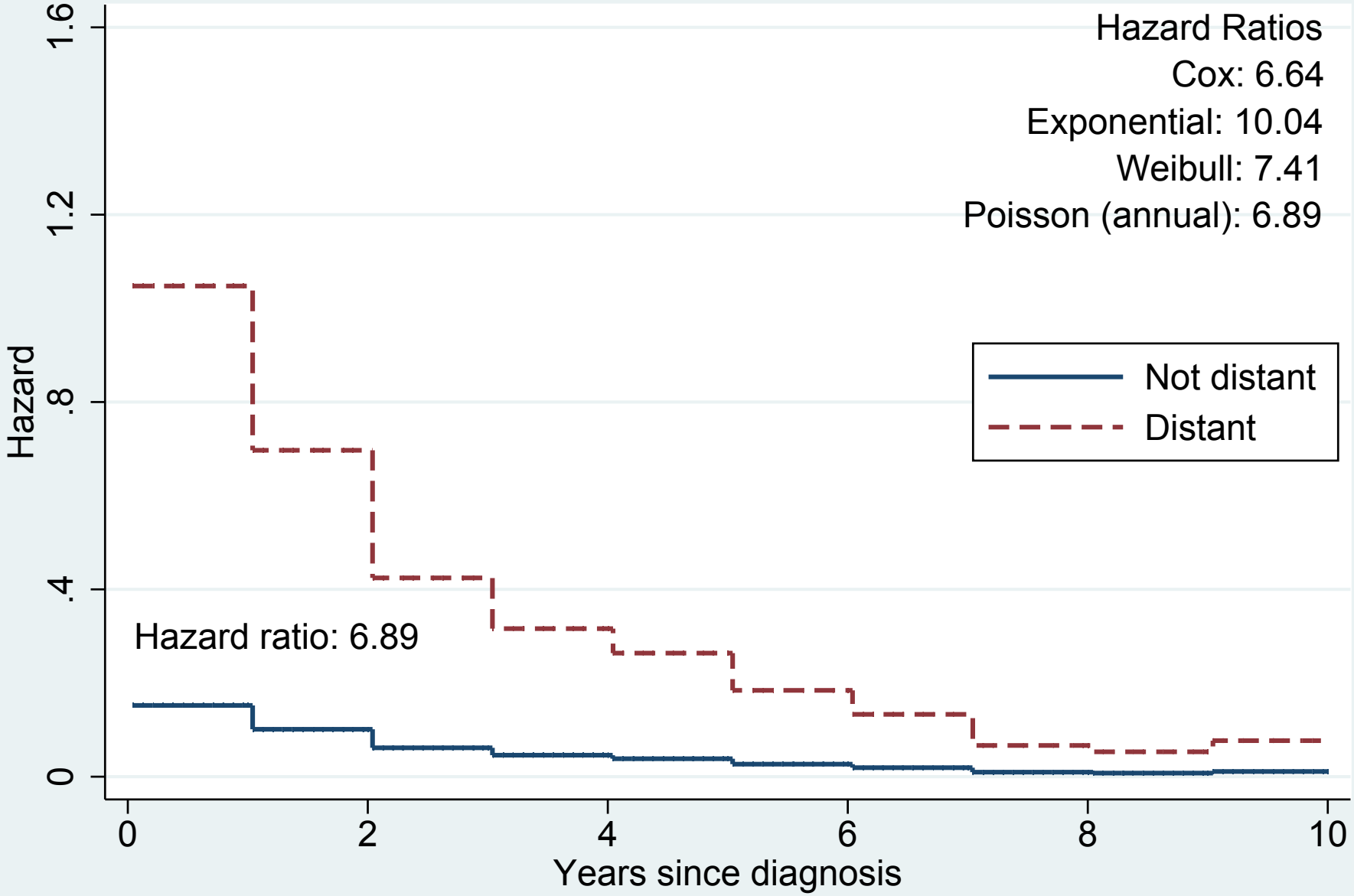
stcurve, hazard at1(distant=0) at2(distant=1) kernel(epan2)



Fitted hazards from parametric survival model (exponential)



Fitted hazards from Poisson model (yearly intervals)

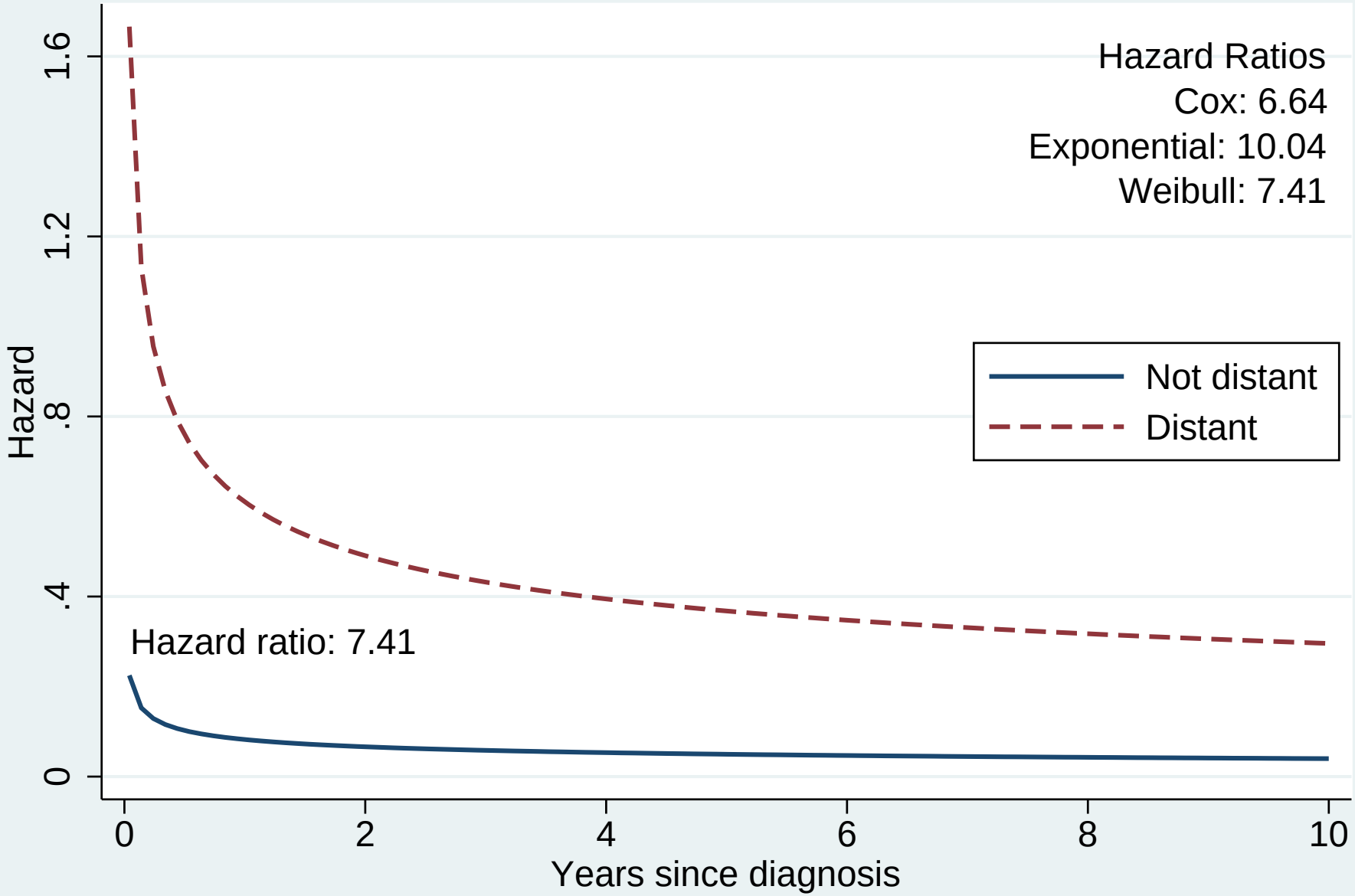


We can make Poisson regression more similar, and even equivalent to, Cox regression

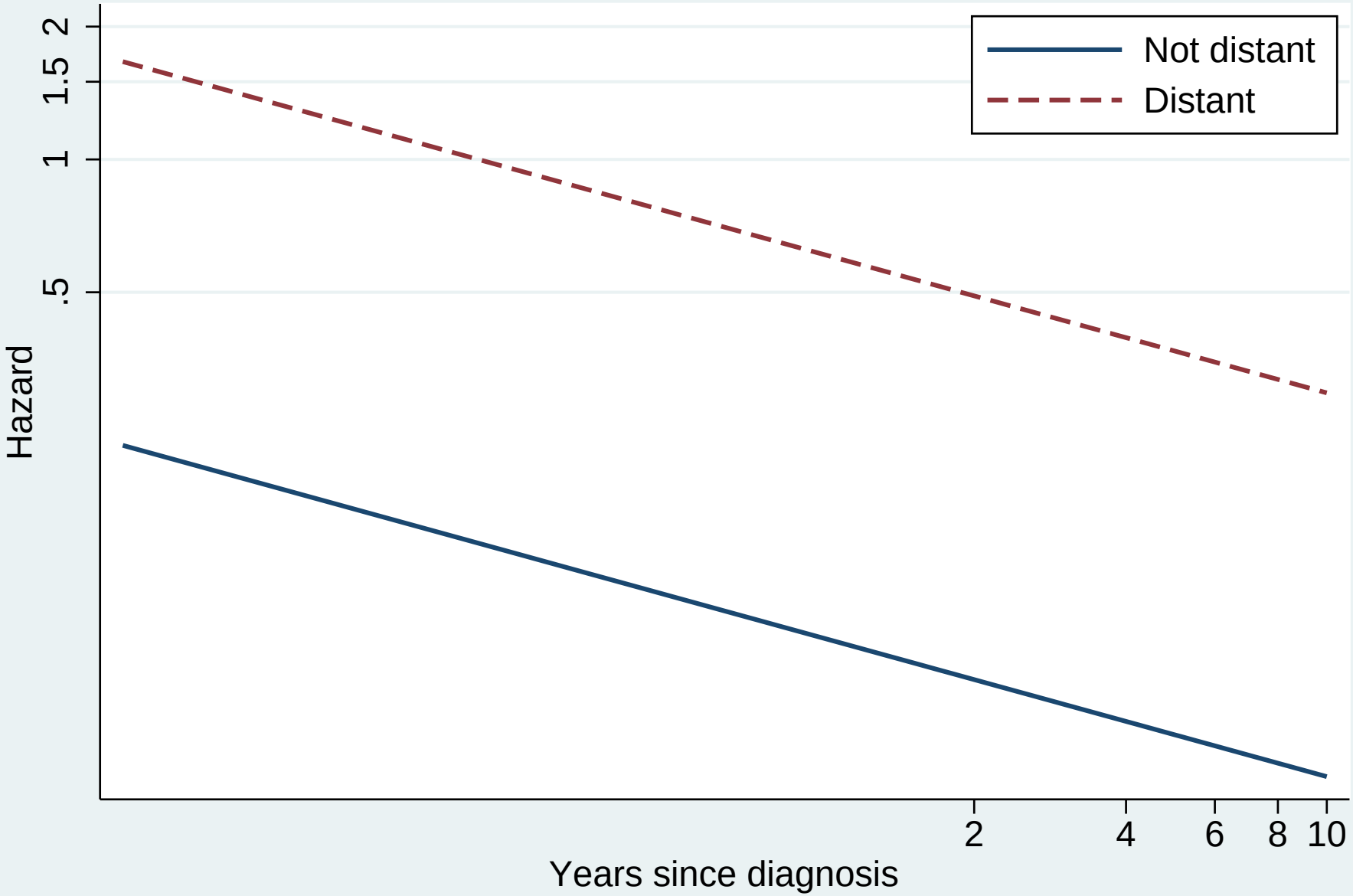
- We can make Poisson regression more similar to Cox regression by using a larger number of smaller intervals.
- If we split at each event time, then the estimates from Poisson regression are equivalent to those from Cox regression.

`www.pauldickman.com/software/stata/compare-cox-poisson/`

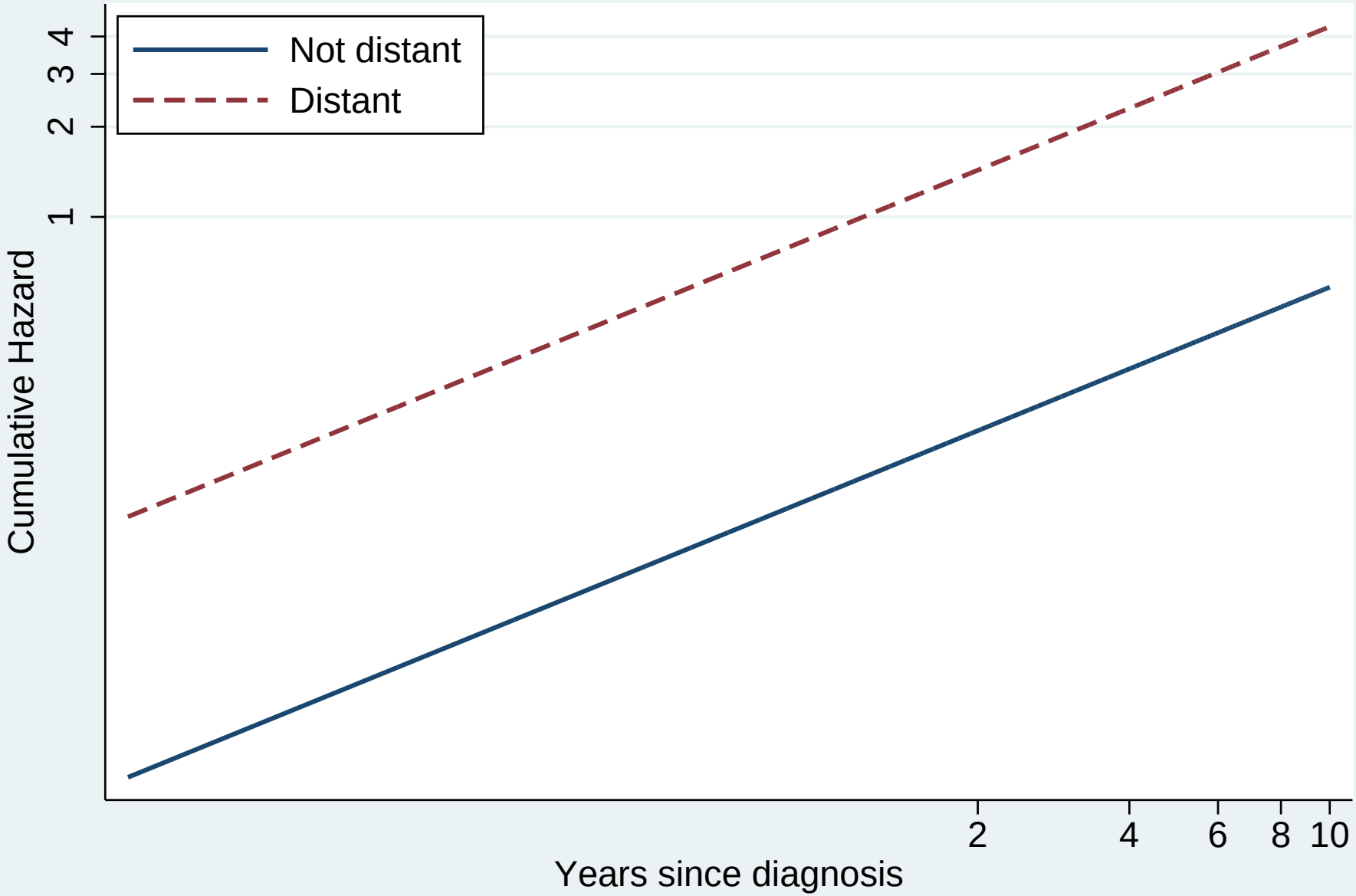
Fitted hazards from parametric survival model (Weibull)

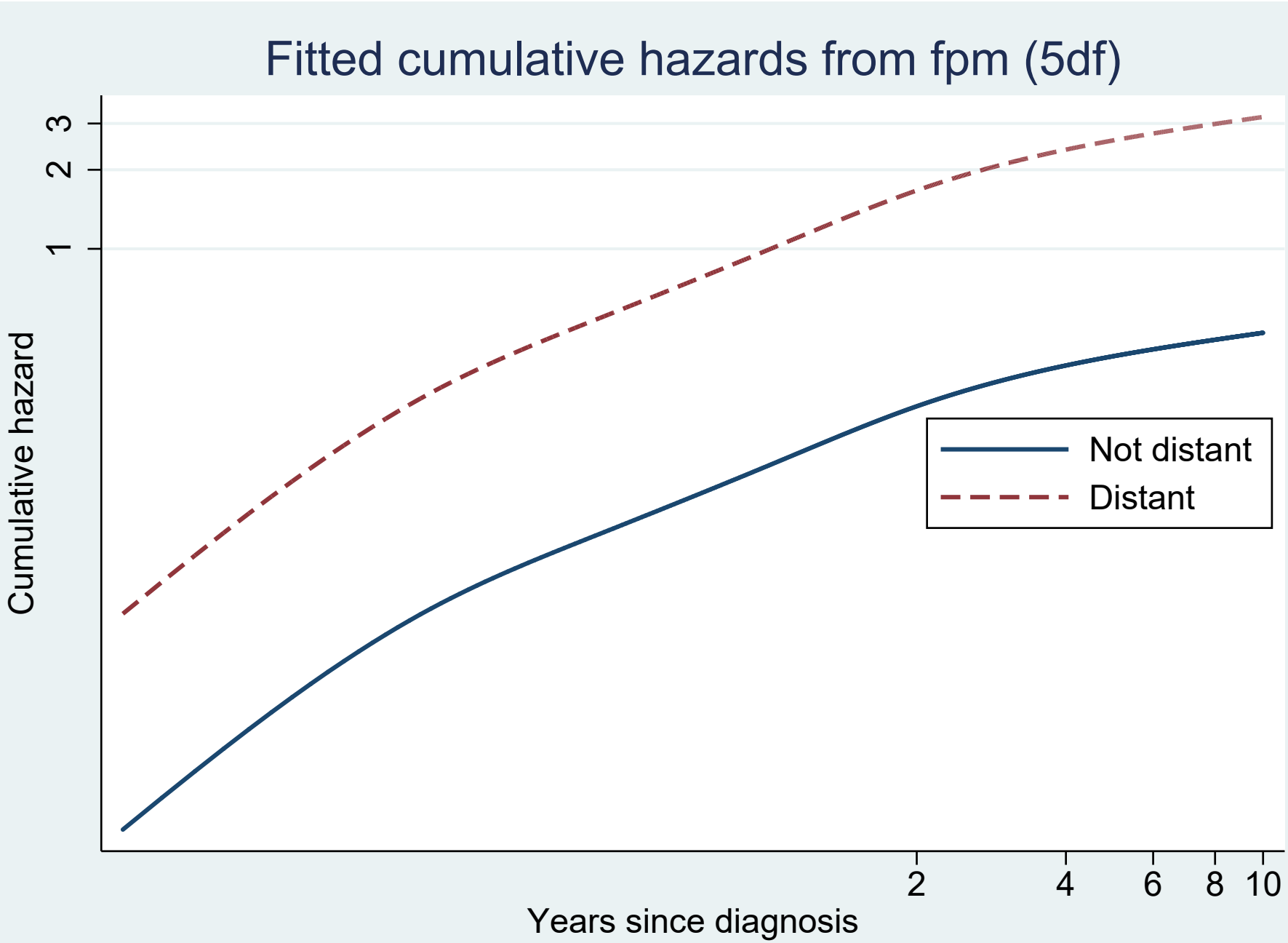


Fitted hazards from parametric survival model (Weibull)

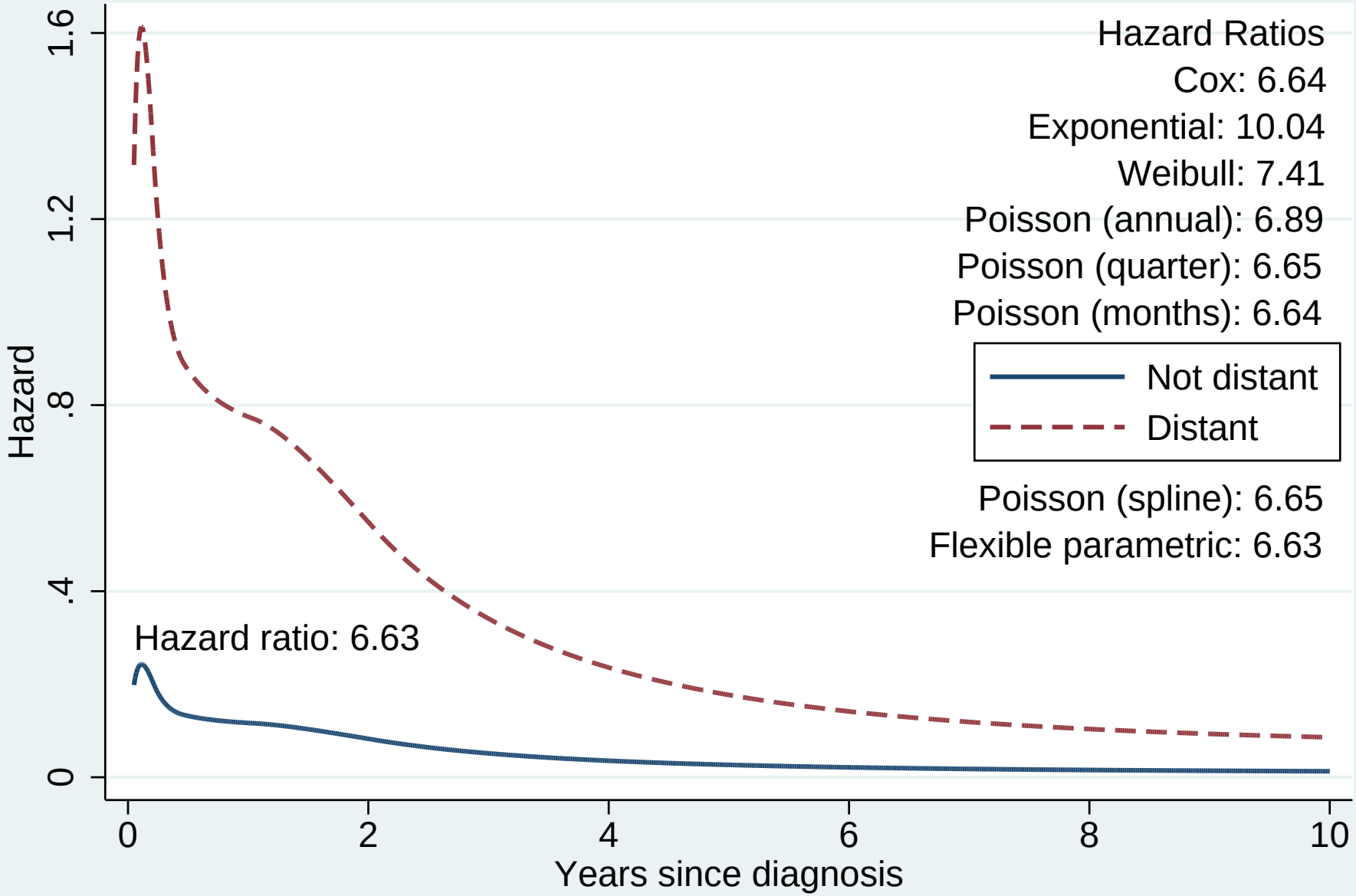


Fitted cumulative hazards from Weibull model




Fitted cumulative hazards from fpm (5df)
Cumulative hazard
1
2
3
Not distant
Distant
2
4
6
8
10
Years since diagnosis

Fitted hazards from flexible parametric model (5df)



Flexible Parametric Survival Models [8, 11, 12]

- First introduced by Royston and Parmar (2002) [8].
- Parametric estimate of the baseline hazard without the usual restrictions on the shape (i.e., flexible).
- Applicable for 'standard' and relative survival models.
- Can fit relative survival cure models (Andersson 2011) [9].
- Once we have a parametric expression for the baseline hazard we derive other quantities of interest (e.g., survival, hazard ratio, hazard differences, expectation of life).
- Can be fitted in Stata (`stpm2`) and R (`rstpm2` or `flexsurv`).
- Can also be estimated on the log-hazard scale [10]

The Cox model [7]

$$h_i(t|\mathbf{x}_i, \beta) = h_0(t) \exp(\mathbf{x}_i \beta)$$

- Advantage: The baseline hazard, $h_0(t)$ is not directly estimated from a Cox model.
- Disadvantage: The baseline hazard, $h_0(t)$ is not directly estimated from a Cox model.

Flexible Parametric Models: Basic Idea

- Consider a Weibull survival curve.

$$S(t) = \exp\left(-\lambda t^{\gamma}\right)$$

- If we transform to the log cumulative hazard scale.

$$\ln\left[H(t)\right] = \ln[-\ln(S(t))]$$

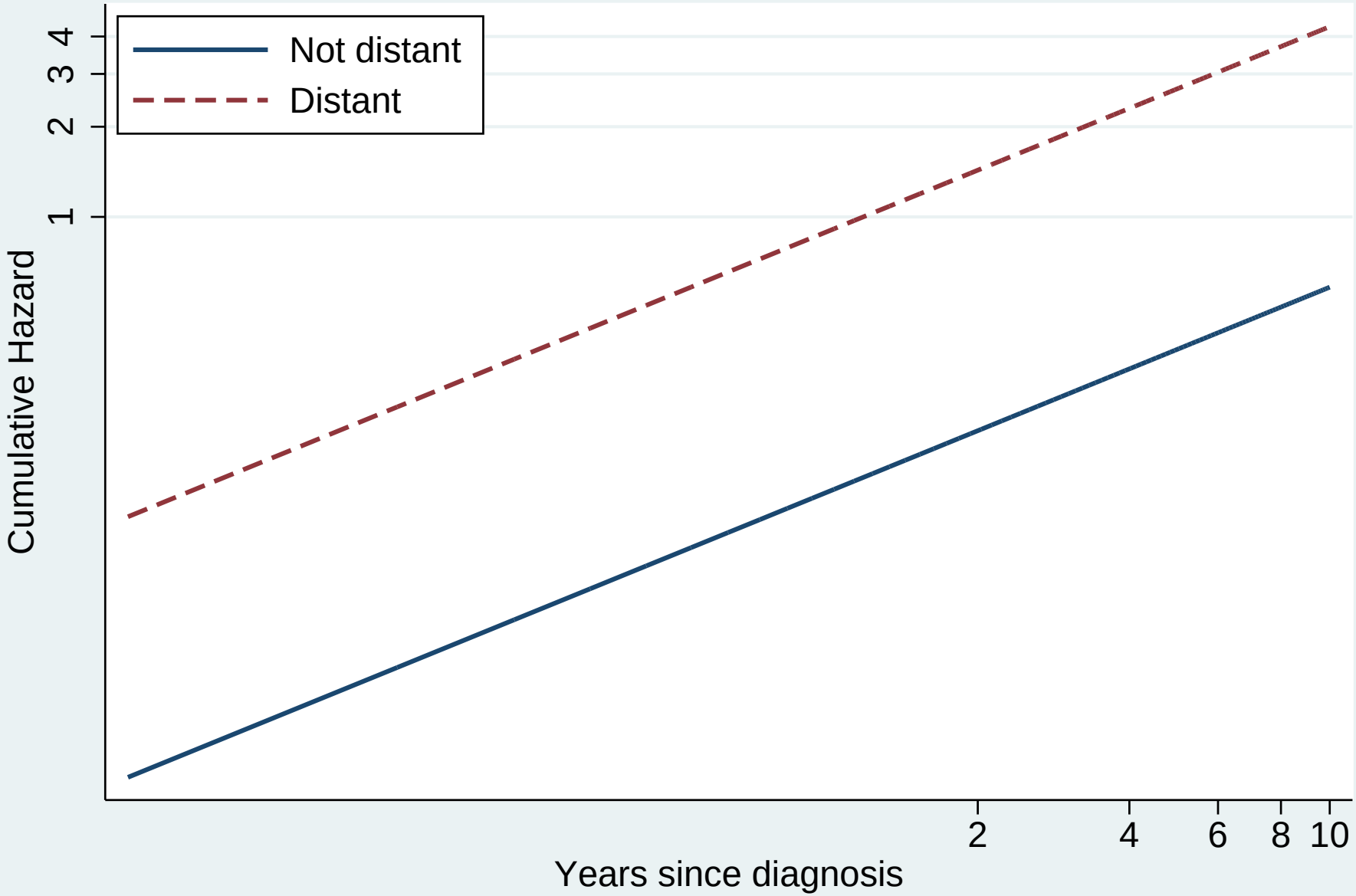
$$\ln\left[H(t)\right] = \ln(\lambda) + \gamma \ln(t)$$

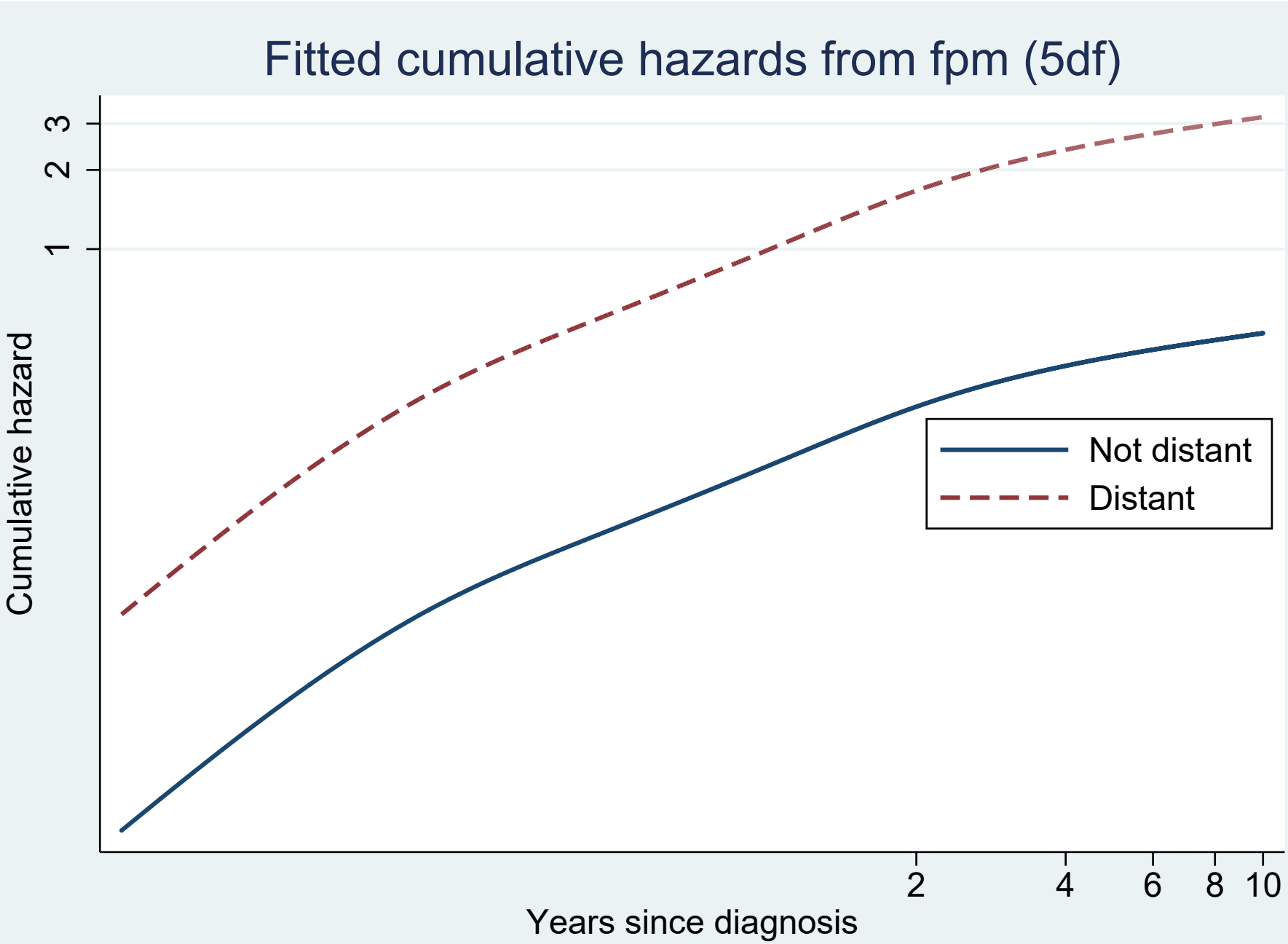
- The log cumulative hazard is a linear function of $\ln(t)$
- Introducing covariates gives

$$\ln\left[H(t|\mathbf{x}_i)\right] = \ln(\lambda) + \gamma \ln(t) + \mathbf{x}_i\boldsymbol{\beta}$$

- Rather than assuming linearity with $\ln(t)$ flexible parametric models use restricted cubic splines for $\ln(t)$.

Fitted cumulative hazards from Weibull model




Fitted cumulative hazards from fpm (5df)
Cumulative hazard
1
2
3
Not distant
Distant
2
4
6
8
10
Years since diagnosis

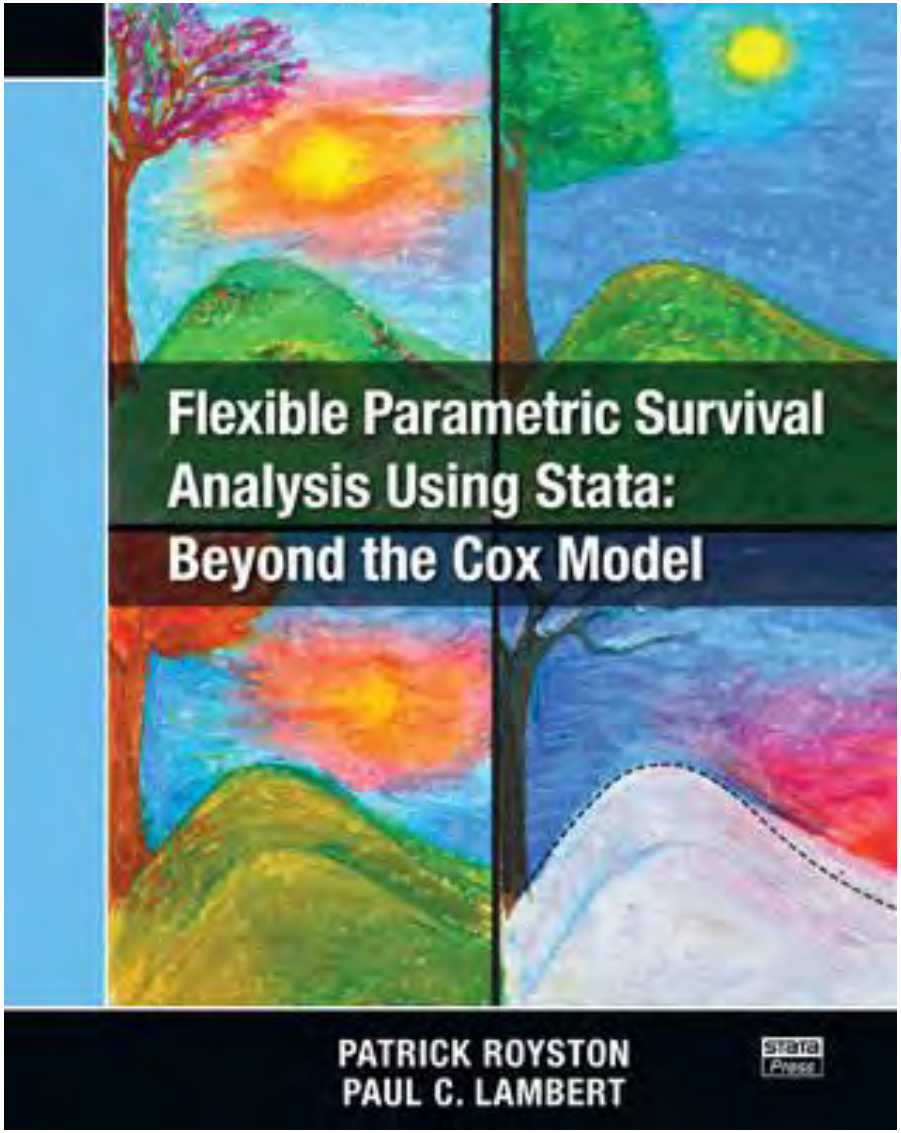
Flexible parametric models: Incorporating splines

- We model on the log cumulative hazard scale.

$$\ln[H(t|\mathbf{x}_i)] = \ln[H_0(t)] + \mathbf{x}_i\boldsymbol{\beta}$$

- This is a proportional hazards model.
- Restricted cubic splines are used to model the log baseline cumulative hazard.
- For example, with 4 knots we can write

$$\ln[H(t|\mathbf{x}_i)] = \eta_i = \underbrace{\gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i}}_{\text{log baseline cumulative hazard}} + \underbrace{\mathbf{x}_i\boldsymbol{\beta}}_{\text{log hazard ratios}}$$



Course in Italy, 5–10 June 2023, `http://cansurv.net/`
Dickman, Lambert, Rutherford, Andersson, Syriopoulou

Sensitivity to choice of knots; Simulation study by Rutherford et al. (2013) [13]

- 'Through the use of simulation we show that, provided a sufficient number of knots are used, the approximated hazard functions given by restricted cubic splines fit closely to the true function for a range of complex hazard shapes.'
- 'The simulation results also highlight the insensitivity of the estimated relative effects (hazard ratios) to the correct specification of the baseline hazard.'

Sensitivity analysis by Syriopoulou et al. (2019) [14]

- 'Although estimates do not depend heavily on the number of knots, too few knots should be avoided, as they can result in a poor fit.'
- 'Interactive graphs engage researchers in assessing model sensitivity to a wide range of scenarios and their use is highly encouraged.'

Implementation in Stata [11]

stpm2 available from SSC

```
ssc install stpm2
```

All-cause or cause-specific survival

```
stpm2 distant, scale(hazard) df(5)
```

Relative survival (excess mortality)

```
stpm2 distant, scale(hazard) df(5) bhazard(rate)
```

Time-dependent effects

```
stpm2 distant, sc(hazard) df(5) bh(rate) tvc(distant) dftvc(3)
```

Cure model

```
stpm2 distant, sc(hazard) df(5) bh(rate) tvc(distant) dftvc(3) cure
```

Continuing with the colon carcinoma example

- Patients diagnosed with colon carcinoma 1984–95. Potential follow-up to end of 1995; censored after 10 years.
- Outcome is death due to colon carcinoma.
- **We have restricted to patients with localised stage**.
- This example will be used for the remainder of the lecture.

Fitting proportional hazards models

- I will start with PH models to illustrate basic concepts and will show later how to relax the PH assumption.

Proportional hazards models

```
. stcox male agegrp2 agegrp3 agegrp4
. stpm2 male agegrp2 agegrp3 agegrp4, scale(hazard) df(5)
```

- The `scale(hazard)` option requests the model be fitted on the log cumulative hazard scale.
- The `df(5)` option implies using 4 internal knots and 2 boundary knots for the baseline cumulative hazard.

Cox proportional hazards model

```
. stcox male agegrp2 agegrp3 agegrp4

Cox regression with Breslow method for ties

No. of subjects =       6,274                 Number of obs =  6,274
No. of failures =       1,687
Time at risk    = 30,962.0616
                                              LR chi2(4)    = 155.93
Log likelihood = -14073.066                   Prob > chi2   = 0.0000

------------------------------------------------------------------------------
          _t | Haz. ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        male |   1.098541   .0548618     1.88   0.060     .9961089    1.211507
     agegrp2 |   .9006346   .1257767    -0.75   0.454     .6849762    1.184191
     agegrp3 |   1.216113   .1539427     1.55   0.122     .9489076    1.558562
     agegrp4 |   2.030934   .2567928     5.60   0.000     1.585146    2.602091
------------------------------------------------------------------------------
```

- The above estimates are adjusted for the baseline hazard (i.e., that mortality may depend on time since diagnosis) but the baseline hazard is not estimated along with the other parameters.

Flexible parametric proportional hazards model

```
. stpm2 male agegrp2 agegrp3 agegrp4, scale(hazard) df(5) eform

Log likelihood = -5898.9448                     Number of obs = 6,274

------------------------------------------------------------------------------
             |     exp(b)   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
xb           |
        male |   1.101218   .0549999     1.93   0.054      .998528    1.214468
     agegrp2 |   .9029138   .1260949    -0.73   0.465     .6867097    1.187188
     agegrp3 |   1.223325   .1548601     1.59   0.111     .9545278    1.567816
     agegrp4 |   2.059039   .2603789     5.71   0.000     1.607032    2.638181
       _rcs1 |   2.324953   .0454633    43.15   0.000     2.237532    2.415789
       _rcs2 |   1.052631   .0142623     3.79   0.000     1.025045    1.080959
       _rcs3 |   1.010869   .0075236     1.45   0.146     .9962299    1.025723
       _rcs4 |   1.081719   .0070315    12.08   0.000     1.068025    1.095589
       _rcs5 |   1.004954   .0046494     1.07   0.285     .9958821    1.014108
       _cons |   .1460823   .0181255   -15.50   0.000     .1145467    .1862998
------------------------------------------------------------------------------
```

- The `eform` option requests exponentiated parameter estimates (i.e., hazard ratios).
- The `_rcs` parameters are the spline basis vectors; the estimates do not have a simple interpretation.

Comparison of estimates, PH models

Variable	cox	stpm2
male	1.0985	1.1012
	0.0549	0.0550
agegrp2	0.9006	0.9029
	0.1258	0.1261
agegrp3	1.2161	1.2233
	0.1539	0.1549
agegrp4	2.0309	2.0590
	0.2568	0.2604

Legend: HR/se

- The hazard ratios and standard errors are similar.
- I have yet to find an example of a proportional hazards model where there is a large difference in the estimated hazard ratios.

Simple predictions

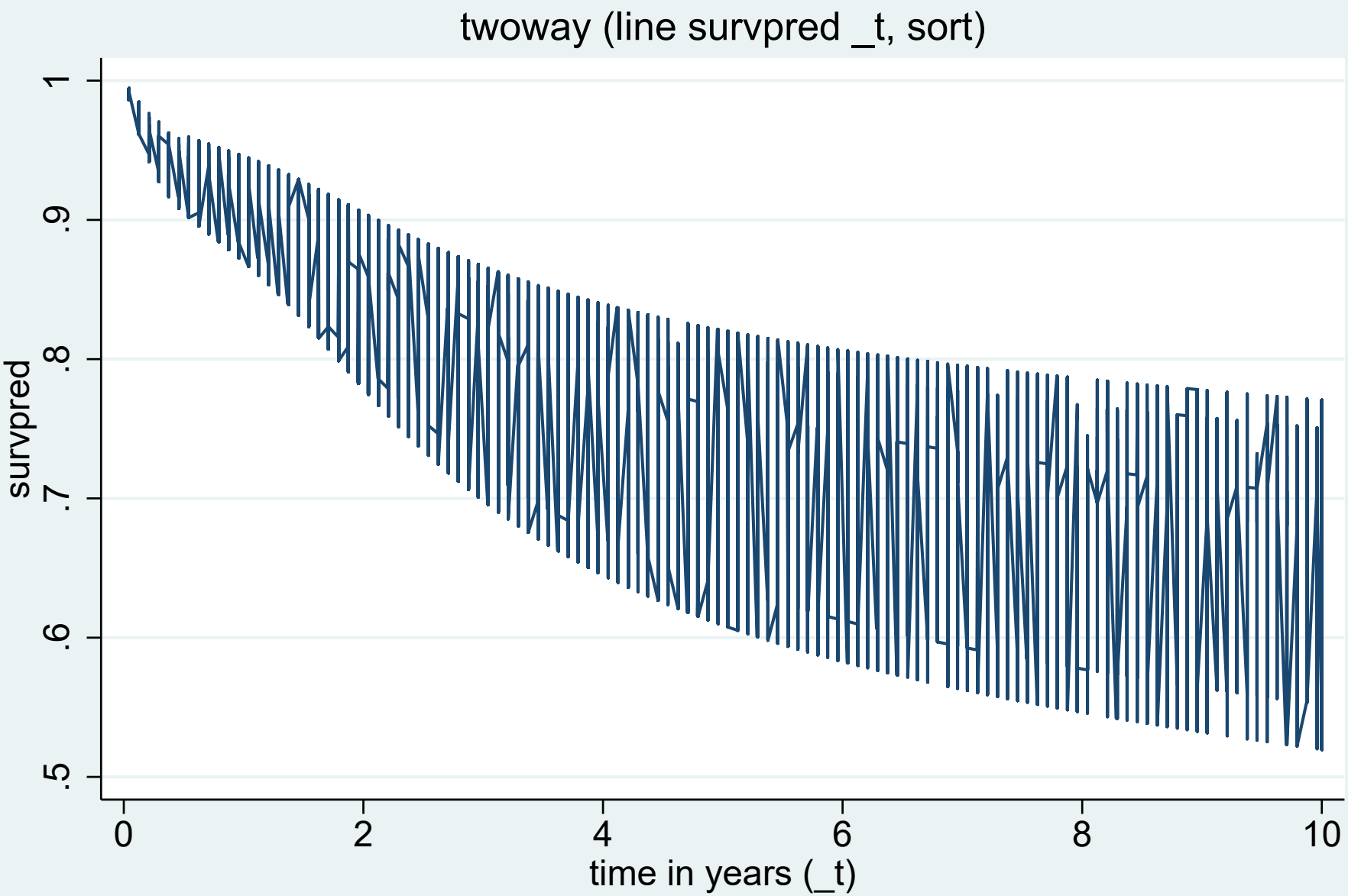
- `stpm2` has a very powerful postestimation command, `predict`, for model-based predictions.

Predicting the survival and hazard functions

```
. predict survpred, survival
. predict hazpred, hazard
```

- For confidence intervals, include the `ci` option.
- Model-based prediction is very powerful, but should be performed with caution.
- Following is a plot of `survpred` (predicted cause-specific survival) against time ($\_t$) for the model we just fitted.

Predicted survival, but probably not as we had hoped



Survival predictions in Stata – technical details

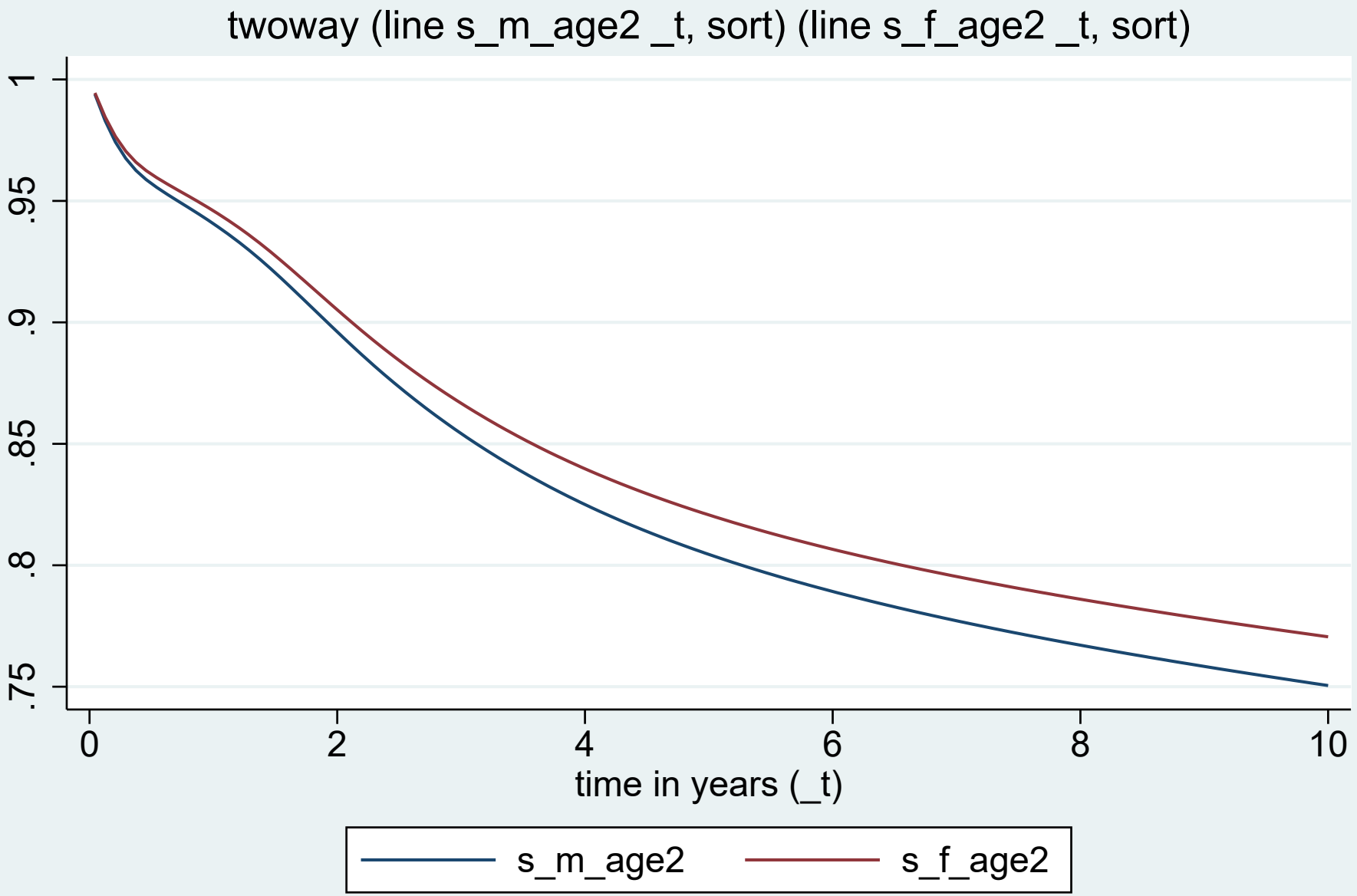
- For each observation, Stata predicts the requested quantities at the value of $\_t$ (exit time).
- For each value of $\_t$ there are 8 possible predicted values of the survival function (one for each combination of age and sex).
- Use the `at()` option to predict for a specified covariate pattern.

Predicted survival for males and females in age group 2

```
. predict s_m_age2, survival at(male 1 agegrp2 1) zeros
. predict s_f_age2, survival at(male 0 agegrp2 1) zeros
```

- The `zeros` option sets all covariates not in `at()` to zero.

Predicted survival by sex, ages 60-74, from PH model



Survival predictions from PH model

- Predictions on the previous slide are based on a PH model, which may or may not be appropriate.
- On the next slide we will see how to relax the PH assumption.
- These are conditional (rather than marginal) estimates. That is, estimates of survival for an individual with specified values of sex and age group.
- I will show later how to obtain marginal (population-averaged) estimates.

Time-dependent effects (non-proportional hazards)

- Fitting time-dependent effects is done using the `tvc()` and `dftvc()` options.

stpm2 with non-PH

```
stpm2 male agegrp2-agegrp4, scale(hazard) df(5) ///
            tvc(male agegrp2-agegrp4) dftvc(2) eform
```

Cox model with non-PH

```
stcox male agegrp2-agegrp4, tvc(male agegrp2-agegrp4) texp(_t)
```

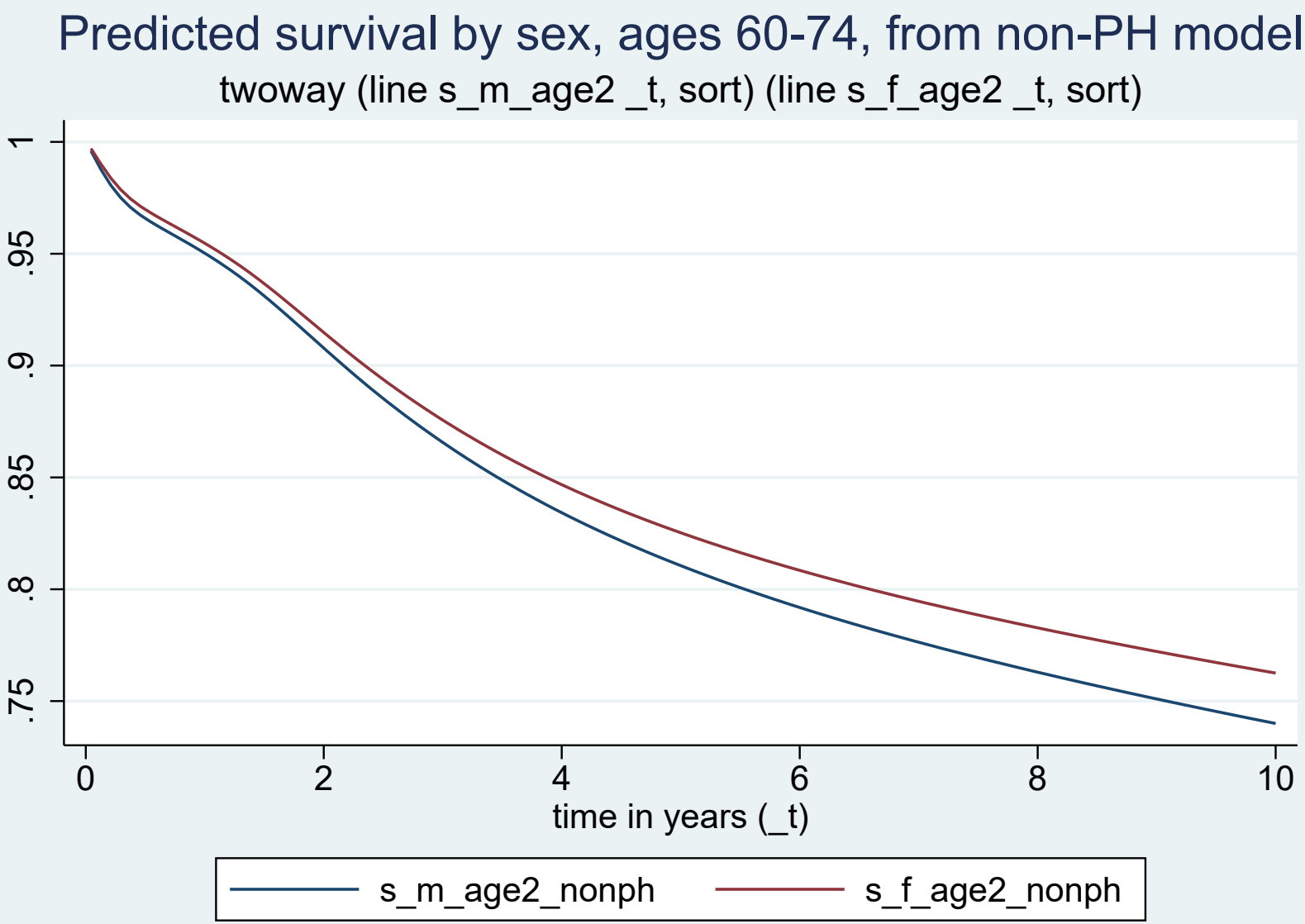
- We are considering time-varying *coefficients*, not time-varying *covariates*.

Predictions from non-proportional hazards models

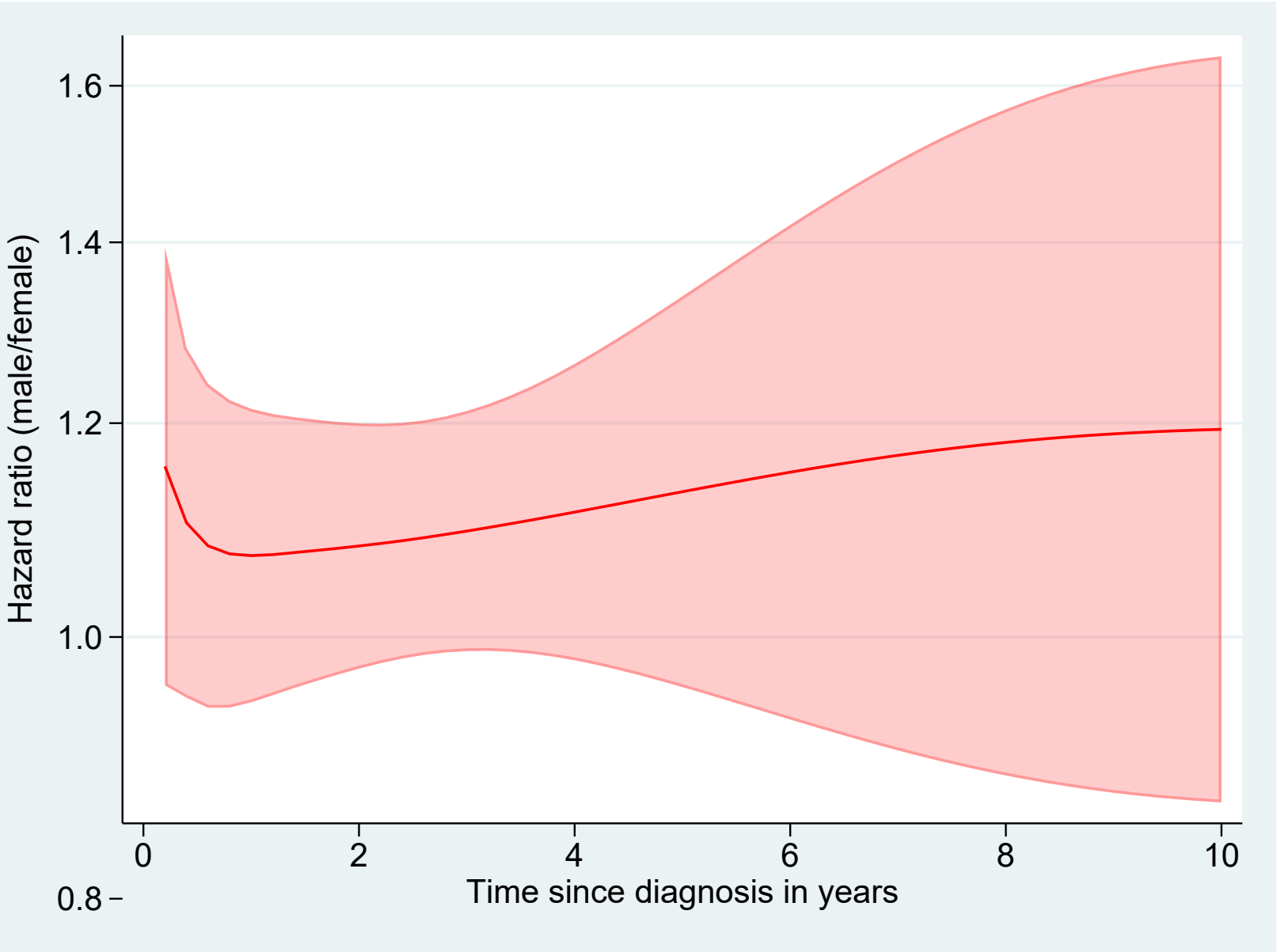
- Syntax for predict is same as with PH model, but we now have the option of estimating time-varying hazard ratios using the `hrnumerator()` and `hrdenominator()` options.

```
predict s_m_age2_nonph, survival at(male 1 agegrp2 1) zeros
predict s_f_age2_nonph, survival at(male 0 agegrp2 1) zeros

// predict time-varying excess hazard ratio (males/females)
predict hr_sex, hrnumerator(male 1) hrdenominator(male 0) ci
```

Predicted hazards (non-PH model)



Predicted hazard ratio for males/females



Time-varying excess hazard ratio [3]

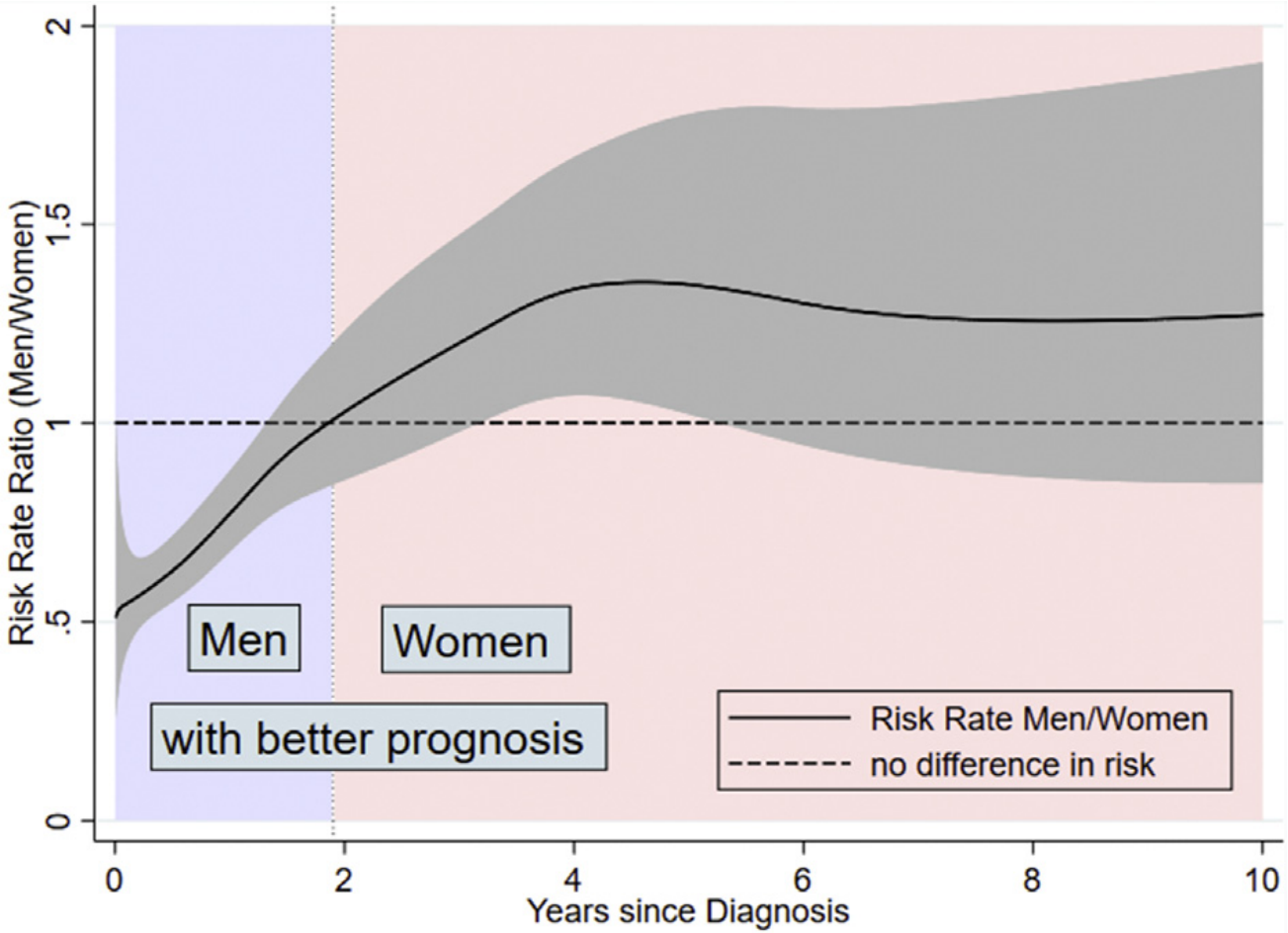


Fig. 2. Risk ratio (excess mortality rate ratio) including confidence intervals for men versus women with bladder cancer diagnosis. The

The predict command is extremely powerful!

Syntax of predict and how to access the help file

```
. predict newvar [if] [in] [, statistic ]
. help stpm2 postestimation
```

- Statistics for predict include:

sdiff	difference in survival functions
hdiff	difference in hazard functions
rmst	restricted mean survival time
lifelost	loss in expectation of life (after a relative survival model)
cure	cure proportion (after fitting a cure model)
uncured	survival function for uncured (after fitting a cure model)
meansurv	population averaged (marginal) survival

Marginal (population-averaged) survival curves: average of individual predictions

- The predicted survival function for individual i is

$$\widehat{S}_i(t) = \exp\left(-H_0(t)\exp\left(\mathbf{x}_i\beta\right)\right)$$

- We average over all predicted survival functions

$$\widehat{S}^P(t) = \frac{1}{N}\sum_{i=1}^{N}\widehat{S}_i(t)$$

- The model can be as complex as required (continuous covariates, interactions, non-linear functions, non-proportional hazards).
- We are predicting a function, not $S(t)$ at a single time point.

Software for marginal measures and regression standardisation

- With Stata `stpm2`, the `meansurv` option to `predict` produces an average of predicted survival curves for each observation.
- `standsurv` is much faster and has more features, see: `https://pclambert.net/software/standsurv/`.
- R users can use the stdReg package (Arvid Sjölander).

Marginal survival curves with stpm2, predict meansurv

```
// Fit model, allowing non-proportional hazards
stpm2 male agegrp2-agegrp4, scale(hazard) df(5) bhazard(rate) ///
            tvc(male agegrp2-agegrp4) dftvc(2) eform nolog

// Marginal survival for entire cohort
predict s_marginal, meansurv timevar(temptime)

// Marginal survival for each sex
predict s_m_marginal if male==1, meansurv timevar(temptime)
predict s_f_marginal if male==0, meansurv timevar(temptime)
```

- `s_marginal` is the average of all 6,274 predicted curves.
- `s_m_marginal` is the average of the 2,620 curves for males.
- `s_f_marginal` is the average of all 3,654 curves for females.
- `s_m_marginal` and `s_f_marginal` are not comparable, but we have estimates for the entire population (i.e., not conditional).

'The hazard ratio has a built-in selection bias' [15]

The Hazards of Hazard Ratios

Miguel A. Hernán

The hazard ratio (HR) is the main, and often the only, effect measure reported in many epidemiologic studies. For dichotomous, non–time-varying exposures, the HR is defined as the hazard in the exposed groups divided by the hazard in the unexposed groups. For all practical purposes, hazards can be thought of as incidence rates and thus the HR can be roughly interpreted as the incidence rate ratio. The HR is commonly and conveniently estimated via a Cox proportional hazards model, which can include potential confounders as covariates.

Unfortunately, the use of the HR for causal inference is not straightforward even in the absence of unmeasured confounding, measurement error, and model misspecification. Endowing a HR with a causal interpretation is risky for 2 key reasons: the HR may change over time, and the HR has a built-in selection bias. Here I review these 2 problems and

Standardised survival curves

- Marginal (population-averaged) survival curves, but 'comparable' (standardised).

```
// Standardised survival (using entire cohort as standard)
predict s_m_std, at(male 1) meansurv timevar(temptime)
predict s_f_std, at(male 0) meansurv timevar(temptime)

// Standardised survival (using males as the standard)
predict s_m_std_m if male==1, at(male 1) meansurv timevar(temptime)
predict s_f_std_m if male==1, at(male 0) meansurv timevar(temptime)
```

- If the model is appropriate and there are no unmeasured confounders, the difference in standardised survival probabilities is an estimate of the causal effect of treatment on survival.
- Under assumptions, the difference between the bottom two estimates is the causal effect among the exposed (if being male is the exposure).

Marginal and standardised survival [3, 4]

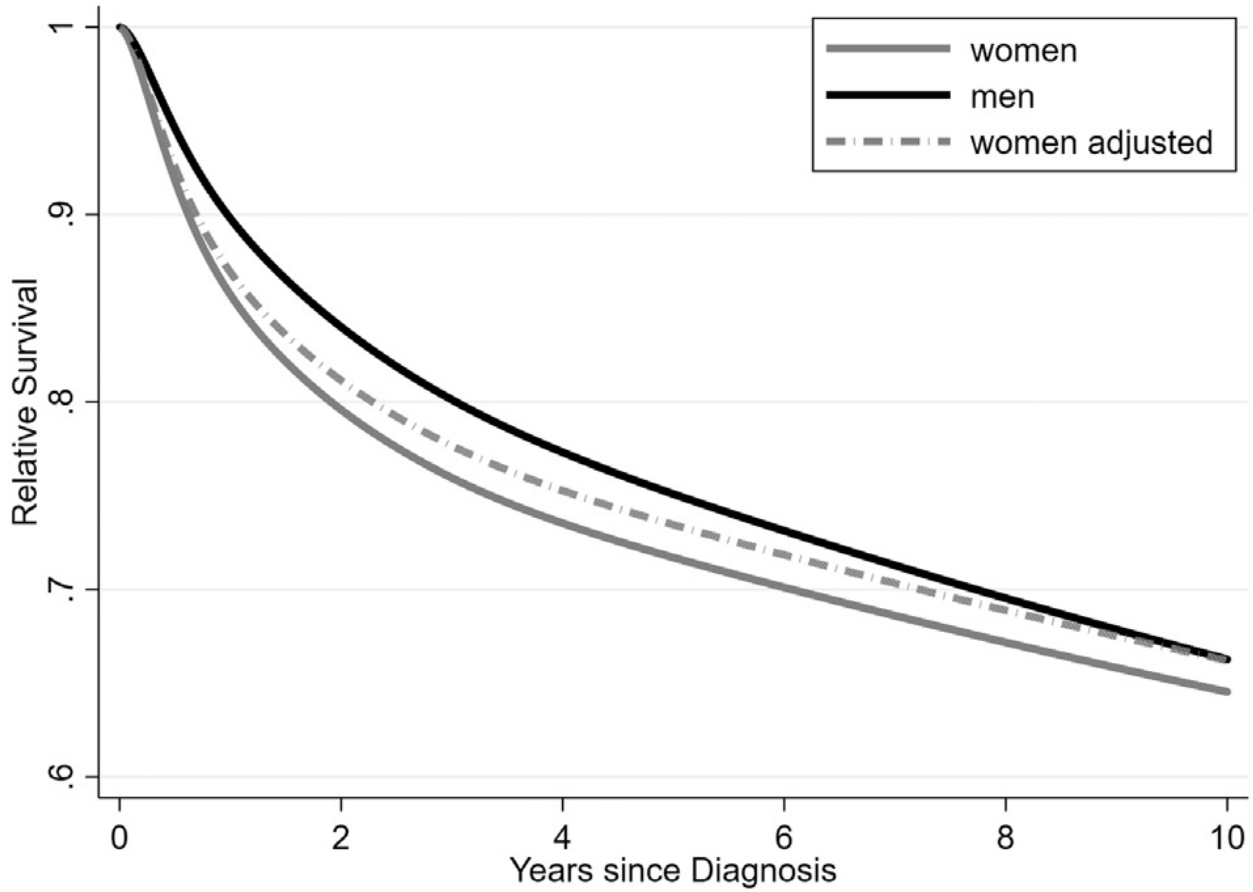


Fig. 3. Relative survival for men, women and women assuming the same T-stage distribution as men. Black (grey) lines: mean survival curve for men (women); Dashed grey line: survival curve for women when assuming men's covariate pattern.

Standardised survival curves (with some math)

- When interest lies in comparing the survival of (two) exposure groups we can standardize to the same covariate distribution.
- Let X be the exposure of interest (e.g., male sex).
- Let Z denote the set of measured covariates (age group).

$$\widehat{R}^P(t|X=x,Z) = \frac{1}{N}\sum_{i=1}^{N}\widehat{R}_i\left(t|X=x,Z=z_i\right)$$

- Note that the average is over the marginal distribution of Z, not over the conditional distribution of Z among those with $X=x$.
- We are forcing the same covariate distribution on both exposure groups.

Standardised survival curves

- We first predict a relative survival curve for all 6,274 patients under the assumption they are male, and average these curves.

$$\widehat{R}^P(t|X = male, Z) = \frac{1}{N}\sum_{i=1}^{N}\widehat{R}_i\left(t|X = male, Z = z_i\right)$$

- We then predict a relative survival curve for all 6,274 patients under the assumption they are female, and average these curves.

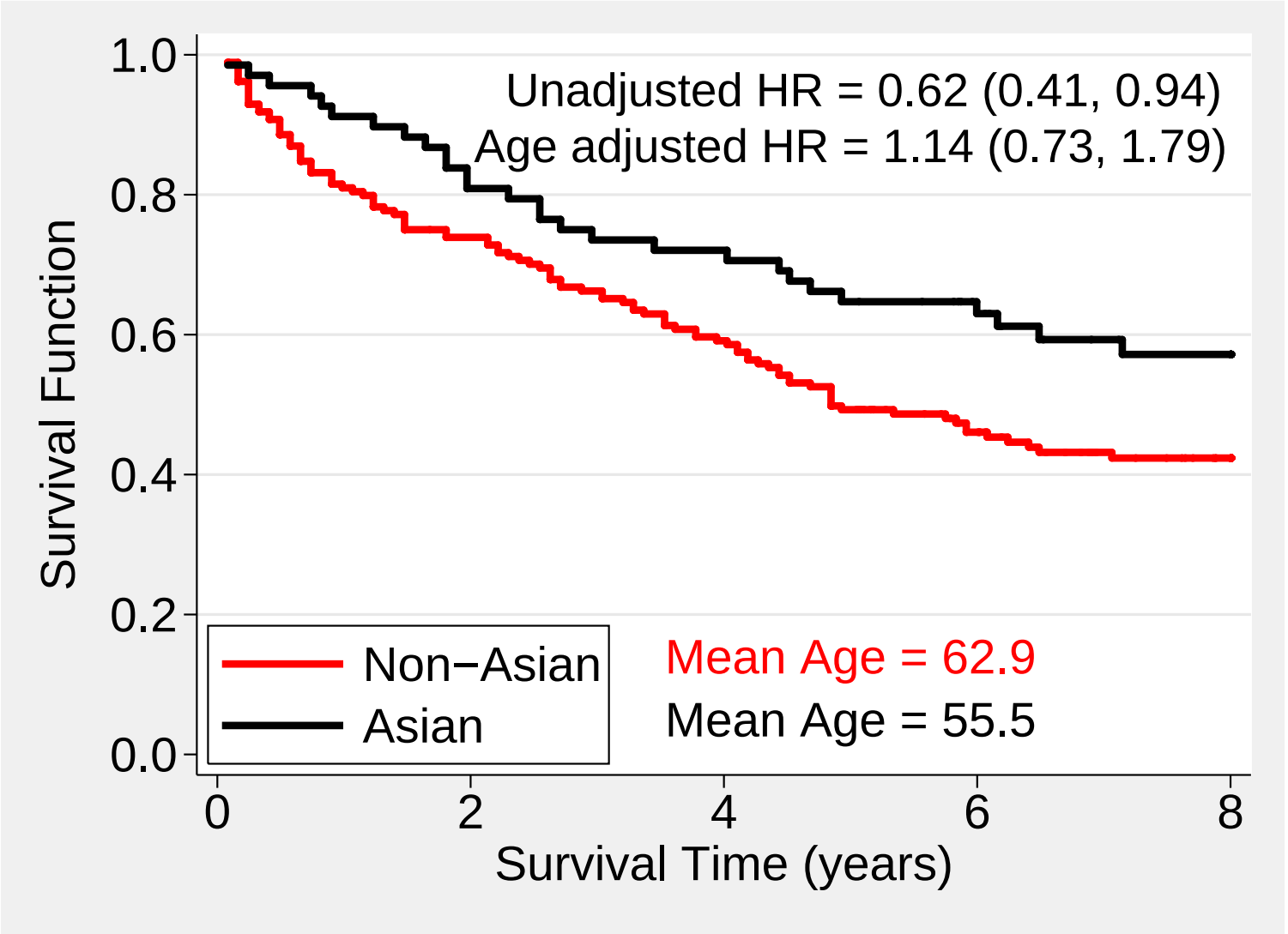
$$\widehat{R}^P(t|X = female, Z) = \frac{1}{N}\sum_{i=1}^{N}\widehat{R}_i\left(t|X = female, Z = z_i\right)$$

- Both resulting marginal relative survival curves are averaged over the same covariate distribution (age distribution in the entire population). The two curves have been age-standardised and are comparable (with respect to confounding by age).

Example: Renal dialysis

- 252 patients entering a renal dialysis program in Leicestershire, England 1982-1991 with follow-up to the end of 1994.
- Interest in difference in survival by ethnicity (Non-South Asian vs South Asian).
- At the time of the study, approximately 25% of the population were of South Asian origin.

Kaplan-Meier Curves - Renal Replacement Therapy



Predictions for Standardised Survival Curves

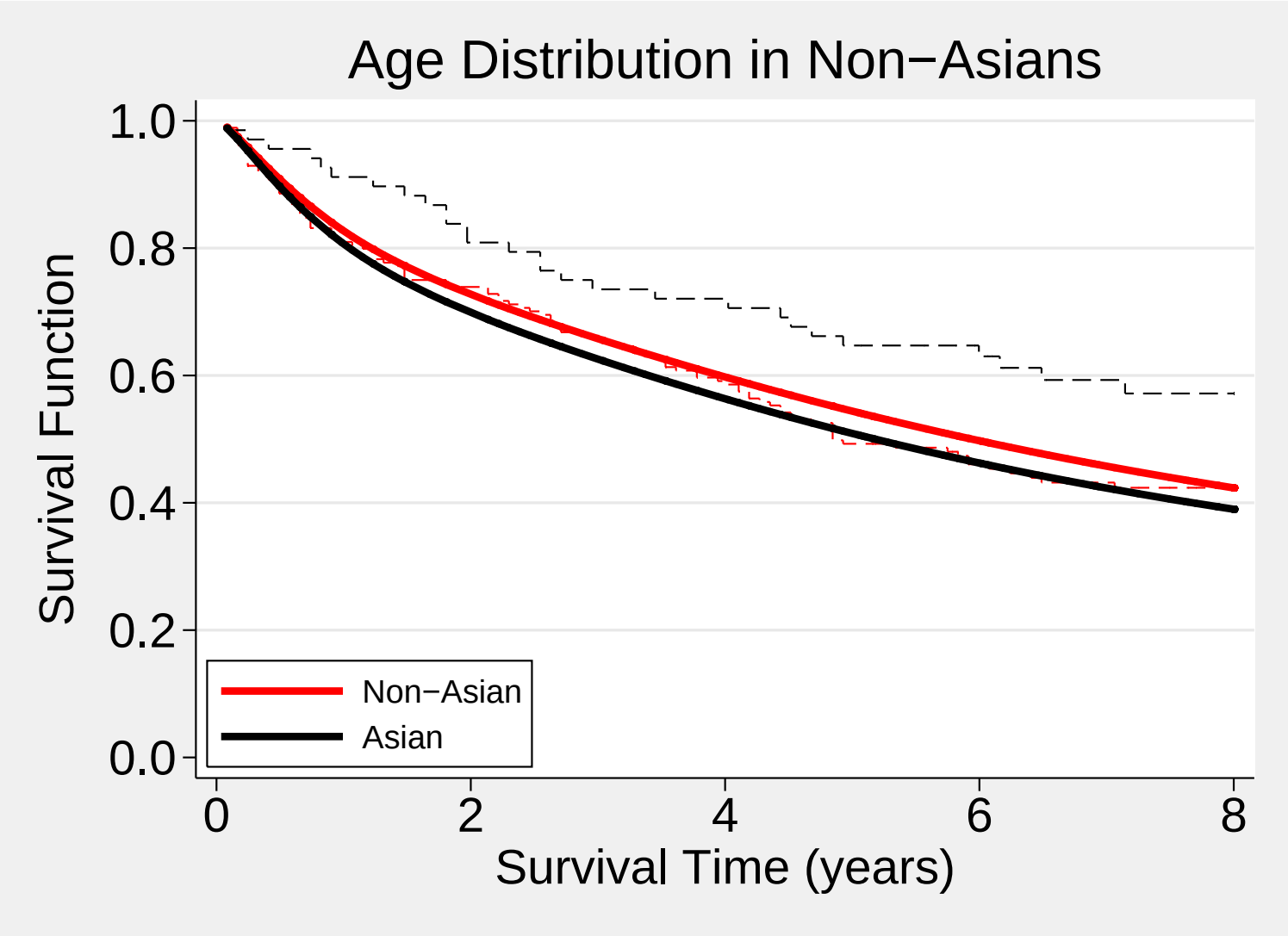
The meansurv option

```
stpm2 asian age, df(3) scale(hazard)
/* Age distribution for study population as a whole */
predict meansurv_pop0, meansurv at(asian 0)
predict meansurv_pop1, meansurv at(asian 1)

/* Age distribution for non-asians */
predict meansurv_pop0b if asian == 0, meansurv at(asian 0)
predict meansurv_pop1b if asian == 0, meansurv at(asian 1)

/* Age distribution for asians */
predict meansurv_pop0c if asian == 1, meansurv at(asian 0)
predict meansurv_pop1c if asian == 1, meansurv at(asian 1)
```

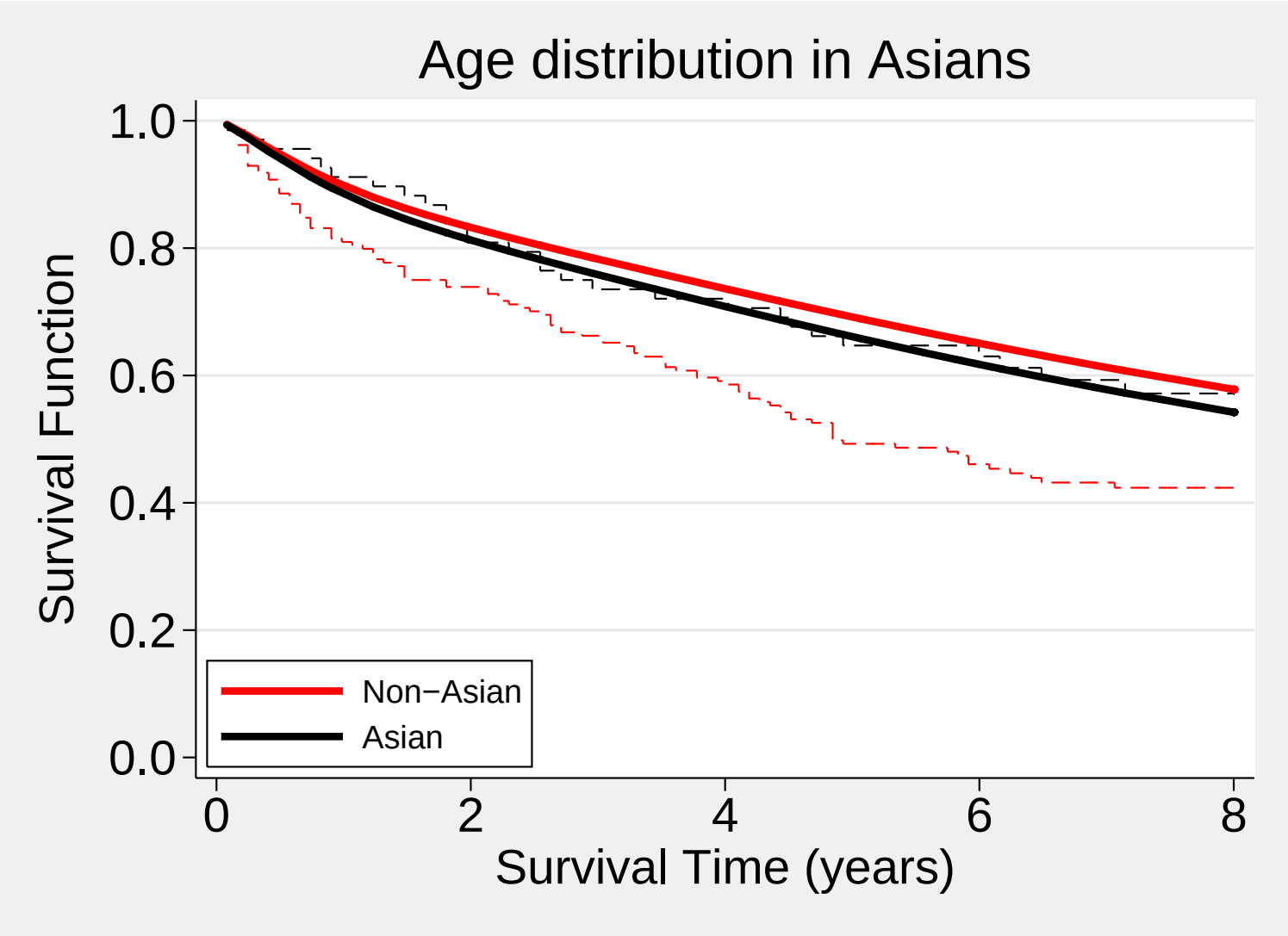
- $S(t)$ calculated for each subject in the study population and averaged.

Standardized Survival Curve 1

Standardized Survival Curve 2



Standardized Survival Curve 3



A non-technical overview (no mathematics) [16]

Standardised survival probabilities: a useful and informative tool for reporting regression models for survival data

Elisavet Syriopoulou[1]✉, Tove Wästerlid[2,3], Paul C. Lambert[1,4] and Therese M.-L. Andersson[1]

BACKGROUND: When interested in studying the effect of a treatment (or other exposure) on a time-to-event outcome, the most popular approach is to estimate survival probabilities using the Kaplan–Meier estimator. In the presence of confounding, regression models are fitted, and results are often summarised as hazard ratios. However, despite their broad use, hazard ratios are frequently misinterpreted as relative risks instead of relative rates.
METHODS: We discuss measures for summarising the analysis from a regression model that overcome some of the limitations associated with hazard ratios. Such measures are the standardised survival probabilities for treated and untreated: survival probabilities if everyone in the population received treatment and if everyone did not. The difference between treatment arms can be calculated to provide a measure for the treatment effect.
RESULTS: Using publicly available data on breast cancer, we demonstrated the usefulness of standardised survival probabilities for comparing the experience between treated and untreated after adjusting for confounding. We also showed that additional important research questions can be addressed by standardising among subgroups of the total population.
DISCUSSION: Standardised survival probabilities are a useful way to report the treatment effect while adjusting for confounding and have an informative interpretation in terms of risk.

British Journal of Cancer (2022) 127:1808–1815; https://doi.org/10.1038/s41416-022-01949-6

This paper is more technical [4]

International Epidemiological Association

International Journal of Epidemiology, 2020, 1–10
doi: 10.1093/ije/dyz268
Original article

Original article

Marginal measures and causal effects using the relative survival framework

Elisavet Syriopoulou ,[1]* Mark J Rutherford[1] and Paul C Lambert[1,2]

[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK and [2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

*Corresponding author. Biostatistics Research Group, Department of Health Sciences, University of Leicester University Road, Leicester LE1 7RH, UK. E-mail: e.syriopoulou@leicester.ac.uk

Editorial decision 28 November 2019; Accepted 3 December 2019

Even more

- Cure models [9].
- Random effect models [17].
- Joint models [18].
- Multi-state models
- Competing Risks
  - Cause-specific models [19]
  - Direct modelling (subhazards) [20, 20].
- Restricted mean survival time [21].
- Prognostic modelling.

Conclusion

- There is nothing wrong with using a Cox model.
- If you only want to estimate a hazard ratio and that you 'know' you have proportional hazards then a Cox model is ideal.
- Can relax the PH assumption in the Cox model, and can estimate quantities other than HR.
- However, a parametric approach makes it easier to estimate quantities that provide more insight and may be more relevant to your research question.
- You will get the same hazard ratio, but a whole lot more.

Why test for PH? A discussion.

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 [22]

- 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?

How Should Hazard Ratios Be Interpreted?

Quote from Stensrud & Hernán [22]

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.

- I agree with the intepretation (second sentence) but I'm not sure I understand the distinction between what they claim is often done (first sentence) and what should be done.
- Similar to how we interpret effect estimates in any model?

Why Are Hazards Usually Not Proportional?

Quotes from Stensrud & Hernán [22]

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 [15].

- (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. Or is it?
- Does this mean we should never perform statistical tests for effect modification?

'Statistical tests for PH are unnecessary'

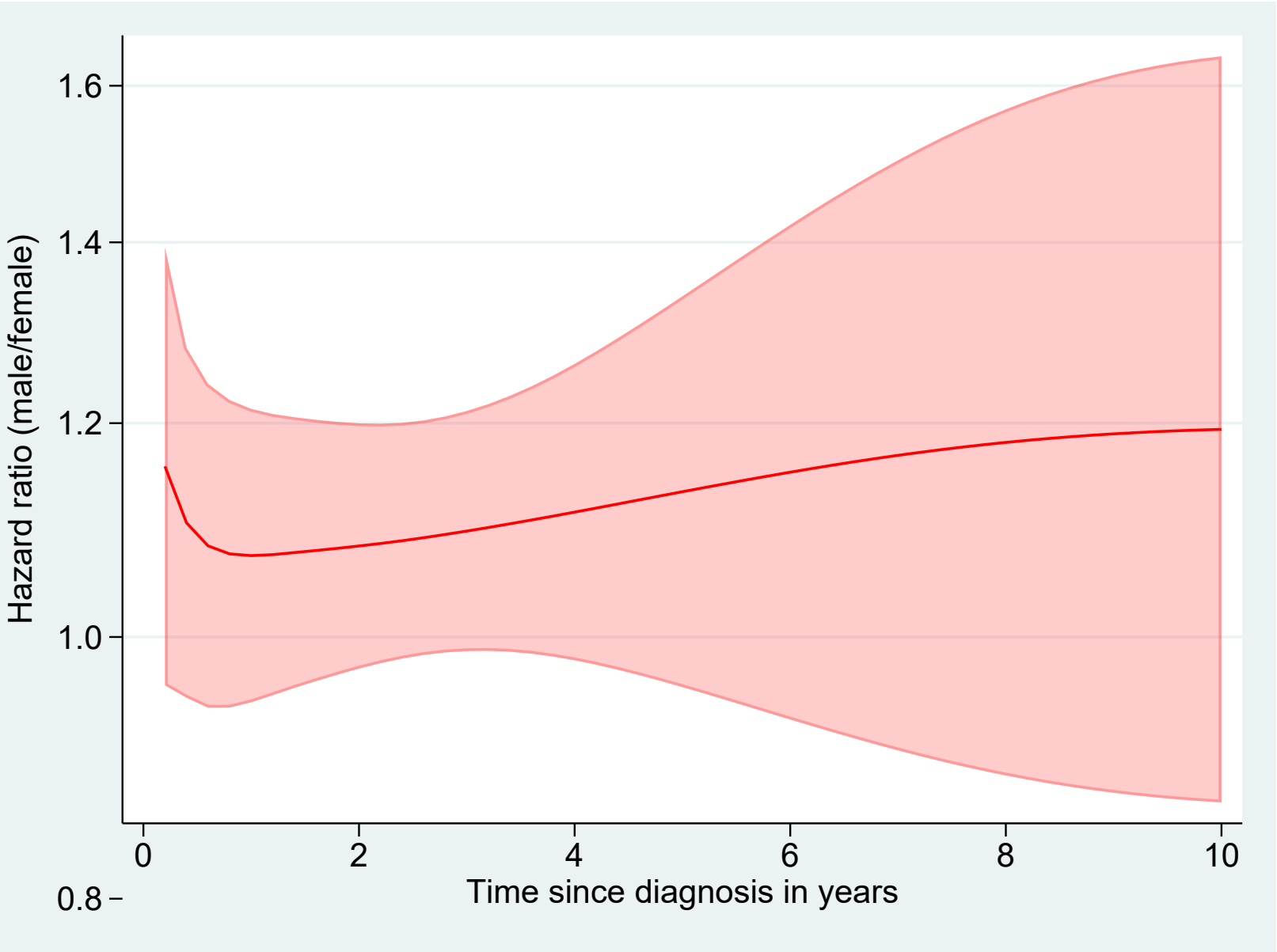
'Because it is expected that the hazard ratio will vary over the follow-up period, tests of proportional hazards yielding high P values are probably underpowered.'

- 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'.
- PH is not the only H_0 that is never true. Should we abandon null hypothesis significance testing?

'Statistical tests for PH are unnecessary' 2

- I argue that 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.
- How should we report the effect of sex for the colon and bladder examples? What guides our decision making?

Predicted hazard ratio for males/females



Predicted excess hazard ratio for males/females [3]

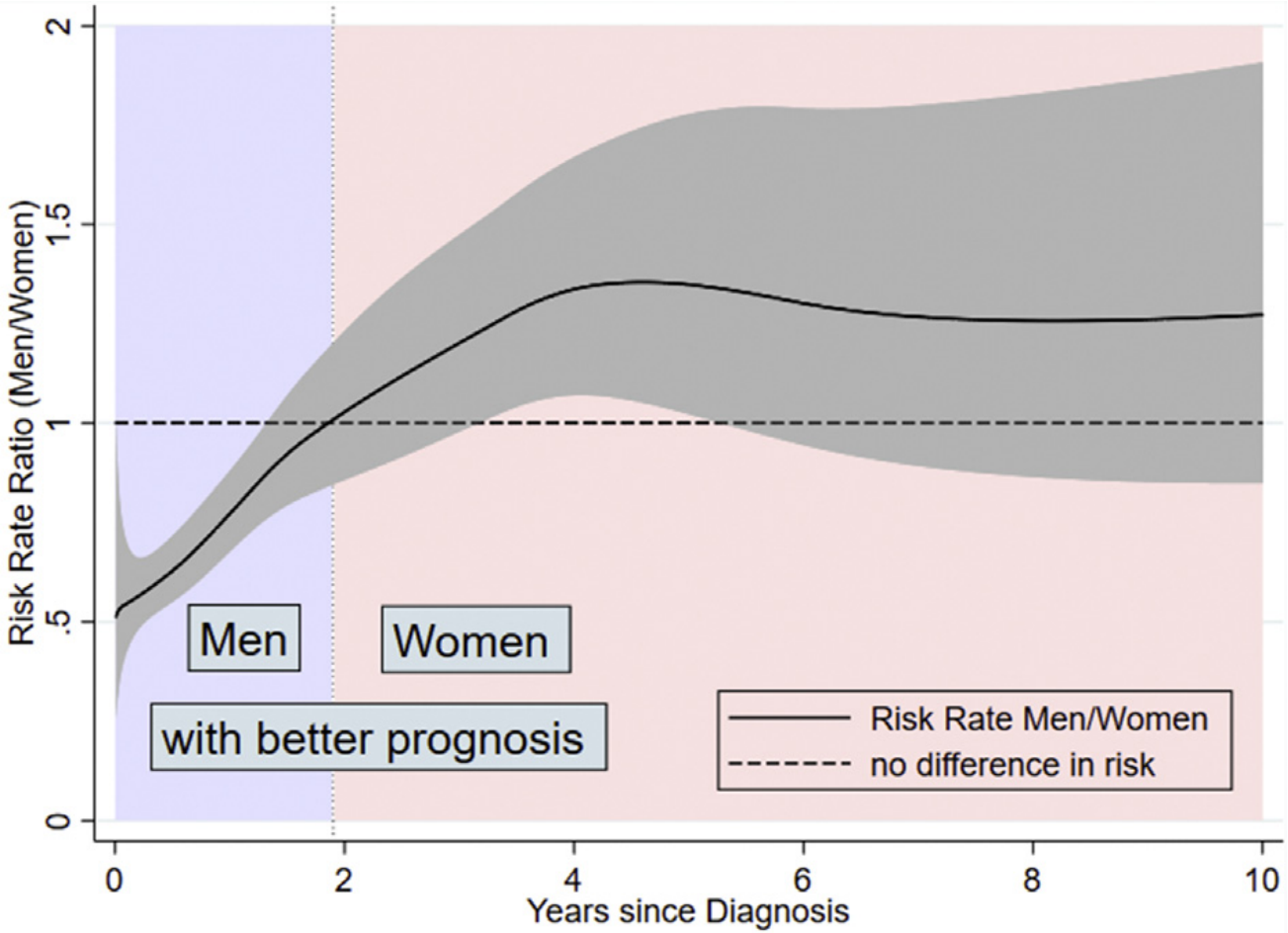


Fig. 2. Risk ratio (excess mortality rate ratio) including confidence intervals for men versus women with bladder cancer diagnosis. The

Alternative measures

Quote from Stensrud & Hernán [22]

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. Such measures are easy to obtain when using flexible parametric models.

Estimating the HR from a PH model

Quote from Stensrud & Hernán [22]

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)

Estimating the HR from a PH model

Quote from Stensrud & Hernán [22]

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 [23]. This limitation can be overcome, and valid 95% confidence intervals can be estimated, by using bootstrapping methods.

- The statement is indisputably true, but how much difference does it make in practice?
- How many of you do this?

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.

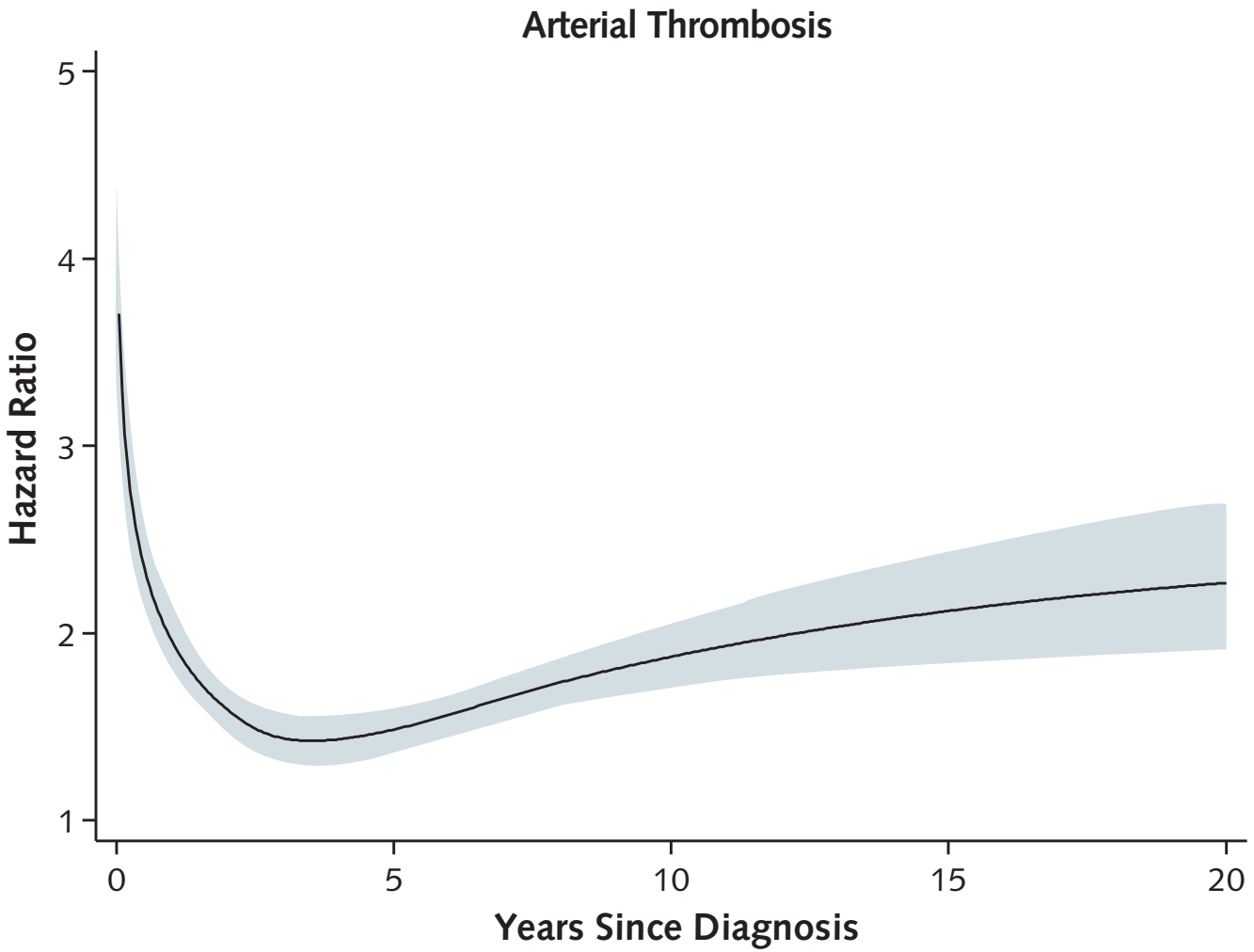
Ann Intern Med. 2018;168:317-325. doi:10.7326/M17-0028 **Annals.org**

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.



References

[1] Adami HO. Sweden: A paradise for epidemiologists? *Lancet* 1996;**347**:588–589.

[2] Radkiewicz C, Johansson AL, Dickman PW, Lambe M, Edgren G. Sex differences in cancer risk and survival: A Swedish cohort study. *European Journal of Cancer* 2017;**84**:130–140.

[3] Andreassen BK, Grimsrud TK, Haug ES. Bladder cancer survival: Women better off in the long run. *European Journal of Cancer* 2018;**95**:52–58.

[4] Syriopoulou E, Rutherford M, Lambert P. Marginal measures and causal effects using the relative survival framework. *International Journal of Epidemiology* 2020;**49**:619–628.

[5] Bower H, Björkholm M, Dickman PW, Höglund M, Lambert PC, Andersson TML. Life expectancy of chronic myeloid leukemia patients is approaching the life expectancy of the general population. *Journal of Clinical Oncology* 2016;**34**:2851–2857.

[6] Reid N. A conversation with Sir David Cox. *Statistical Science* 1994;**9**:439–455.

[7] Cox DR. Regression models and life-tables (with discussion). *JRSSB* 1972;**34**:187–220.

[8] Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002;**21**:2175–2197.

References 2

[9] Andersson TML, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Medical Research Methodology* 2011;**11**:96.

[10] Bower H, Crowther MJ, Lambert PC. Strcs: A command for fitting flexible parametric survival models on the log-hazard scale. *The Stata Journal* 2016;**16**:989–1012.

[11] Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *The Stata Journal* 2009;**9**:265–290.

[12] Royston P, Lambert PC. *Flexible parametric survival analysis in Stata: Beyond the Cox model*. Stata Press, 2011.

[13] Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2015;**85**:777–793.

[14] Syriopoulou E, Mozumder SI, Rutherford MJ, Lambert PC. Robustness of individual and marginal model-based estimates: A sensitivity analysis of flexible parametric models. *Cancer Epidemiology* 2018;**58**:17–24.

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

References 3

[16] Syriopoulou E, Wästerlid T, Lambert PC, Andersson TML. Standardised survival probabilities: a useful and informative tool for reporting regression models for survival data. *British journal of cancer* 2022;**127**:1808–1815.

[17] Crowther MJ, Look MP, Riley RD. Multilevel mixed effects parametric survival models using adaptive gauss-hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med* 2014;**33**:3844–3858.

[18] Crowther MJ, Abrams KR, Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine* 2012;**31**:4456–4471.

[19] Hinchliffe SR, Lambert PC. Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions. *BMC Medical Research Methodology* 2013;**13**:13.

[20] Lambert PC, Wilkes SR, Crowther M. Flexible parametric modelling of the cause-specific cumulative incidence function. *Statistics in Medicine* 2017;**36**:1429–1446.

[21] Royston P, Parmar MKB. 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 medical research methodology* 2013;**13**:152.

[22] Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? *JAMA* 2020; **323**:1401–1402.

References 4

[23] DiRienzo AG, Lagakos SW. Effects of model misspecification on tests of no randomized treatment effect arising from Cox's proportional hazards model. *Journal of the Royal Statistical Society Series B Statistical Methodology* 2001;**63**:745–757.