Parametric Survival Models

Paul C Lambert[1,2]

[1]Department of Health Sciences, University of Leicester, UK

[2]Medical Epidemiology & Biostatistics, Karolinska Institutet, Stockholm, Sweden

40+ years of the Cox model: 8/3/2013

University *of* Leicester

Karolinska Institutet

Why I use parametric models

- I analyse large population-based datasets where
  - The proportional hazards assumption is often not appropriate.
  - The hazard function is of interest.

Why I use parametric models

- I analyse large population-based datasets where
  - The proportional hazards assumption is often not appropriate.
  - The hazard function is of interest.
- I fit excess mortality/relative survival models in population-based cancer studies.
  - Was not an easy adaption for the Cox model.
  - Proportional excess hazards rarely true.
  - The excess hazard is of interest.

Why I use parametric models

- I analyse large population-based datasets where
  - The proportional hazards assumption is often not appropriate.
  - The hazard function is of interest.
- I fit excess mortality/relative survival models in population-based cancer studies.
  - Was not an easy adaption for the Cox model.
  - Proportional excess hazards rarely true.
  - The excess hazard is of interest.
- Quantification of absolute risks and rates.
  - I believe this should be done more than it is.
  - Much easier if you estimate the baseline.

The Cox model[1].

- Web of Science: over 26,938 citations (February 2013).
- Has an h-index of 13 from repeat mis-citations[1].

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

[1] http://occamstypewriter.org/boboh/2008/06/24/outdone_by_mis_prints/

The Cox model[1].

- Web of Science: over 26,938 citations (February 2013).
- Has an h-index of 13 from repeat mis-citations[1].

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

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

[1] http://occamstypewriter.org/boboh/2008/06/24/outdone_by_mis_prints/

The Cox model[1].

- Web of Science: over 26,938 citations (February 2013).
- Has an h-index of 13 from repeat mis-citations[1].

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

- 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.

[1] http://occamstypewriter.org/boboh/2008/06/24/outdone_by_mis_prints/

Quote from Sir David Cox (Reid 1994 [2])

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."

Quote from Sir David Cox (Reid 1994 [2])

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."

Quote from Sir David Cox (Reid 1994 [2])

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."

Proportional hazards models

- I will concentrate on models on the log cumulative hazard scale.

$$H_i(t) = H_0(t)\exp(\mathbf{x_i}\beta)$$

$$\ln(H_i(t)) = \ln(H_0(t)) + \mathbf{x_i}\beta$$

- Need to decide on functional form for $\ln(H_0(t))$.
- $\ln(H_0(t))$ can take standard parametric form (e.g. Weibull), but also more general non-linear function such as splines [3, 4].

Proportional hazards models

- I will concentrate on models on the log cumulative hazard scale.

$$H_i(t) = H_0(t)\exp(\mathbf{x_i}\beta)$$

$$\ln(H_i(t)) = \ln(H_0(t)) + \mathbf{x_i}\beta$$

- Need to decide on functional form for $\ln(H_0(t))$.
- $\ln(H_0(t))$ can take standard parametric form (e.g. Weibull), but also more general non-linear function such as splines [3, 4].
- Also can model on log hazard scale.

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

- Need numerical integration or Poisson regression for complex functions (e.g. splines) [5].

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.

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

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

- This is a linear function of $\ln(t)$

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.

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

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

- This 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 $\ln(t)$.

Flexible parametric models: incorporating splines

- We thus 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.

Flexible parametric models: incorporating splines

- We thus 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 with knots, $\mathbf{k_0}$, are used to model the log baseline cumulative hazard.

$$\ln[H(t|\mathbf{x}_i)] = \eta_i = s(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0) + \mathbf{x}_i\boldsymbol{\beta}$$

Flexible parametric models: incorporating splines

- We thus 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 with knots, $\mathbf{k_0}$, are used to model the log baseline cumulative hazard.

$$\ln[H(t|\mathbf{x}_i)] = \eta_i = s(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0) + \mathbf{x}_i\boldsymbol{\beta}$$

- For example, with 4 knots we can write

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

- We are fitting a linear predictor on the log cumulative hazard scale.

Flexible parametric models: incorporating splines

- We thus 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 with knots, $\mathbf{k_0}$, are used to model the log baseline cumulative hazard.

$$\ln[H(t|\mathbf{x}_i)] = \eta_i = s(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0) + \mathbf{x}_i\boldsymbol{\beta}$$

- For example, with 4 knots we can write

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

- We are fitting a linear predictor on the log cumulative hazard scale.

Survival and Hazard Functions

- We can transform to the survival function

$$S(t|\mathbf{x}_i) = \exp(-\exp(\eta_i))$$

Survival and Hazard Functions

- We can transform to the survival function

$$S(t|\mathbf{x}_i) = \exp(-\exp(\eta_i))$$

- The hazard function is a bit more complex.

$$h(t|\mathbf{x}_i) = \frac{ds\left(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0\right)}{dt}\exp(\eta_i)$$

- This involves the derivatives of the restricted cubic splines functions.
  - These are easy to calculate.

Survival and Hazard Functions

- We can transform to the survival function

$$S(t|\mathbf{x}_i) = \exp(-\exp(\eta_i))$$

- The hazard function is a bit more complex.

$$h(t|\mathbf{x}_i) = \frac{ds\left(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0\right)}{dt}\exp(\eta_i)$$

- This involves the derivatives of the restricted cubic splines functions.
  - These are easy to calculate.
- Survival and hazard function feed into the likelihood. No need for numerical integration or time-splitting.

Survival and Hazard Functions

- We can transform to the survival function

$$S(t|\mathbf{x}_i) = \exp(-\exp(\eta_i))$$

- The hazard function is a bit more complex.

$$h(t|\mathbf{x}_i) = \frac{ds\left(\ln(t)|\boldsymbol{\gamma}, \mathbf{k}_0\right)}{dt}\exp(\eta_i)$$

- This involves the derivatives of the restricted cubic splines functions.
  - These are easy to calculate.
- Survival and hazard function feed into the likelihood. No need for numerical integration or time-splitting.
- Royston and Parmar discuss other link functions including logit and probit[3].

Simulation Study (Rutherford et al.)[6]

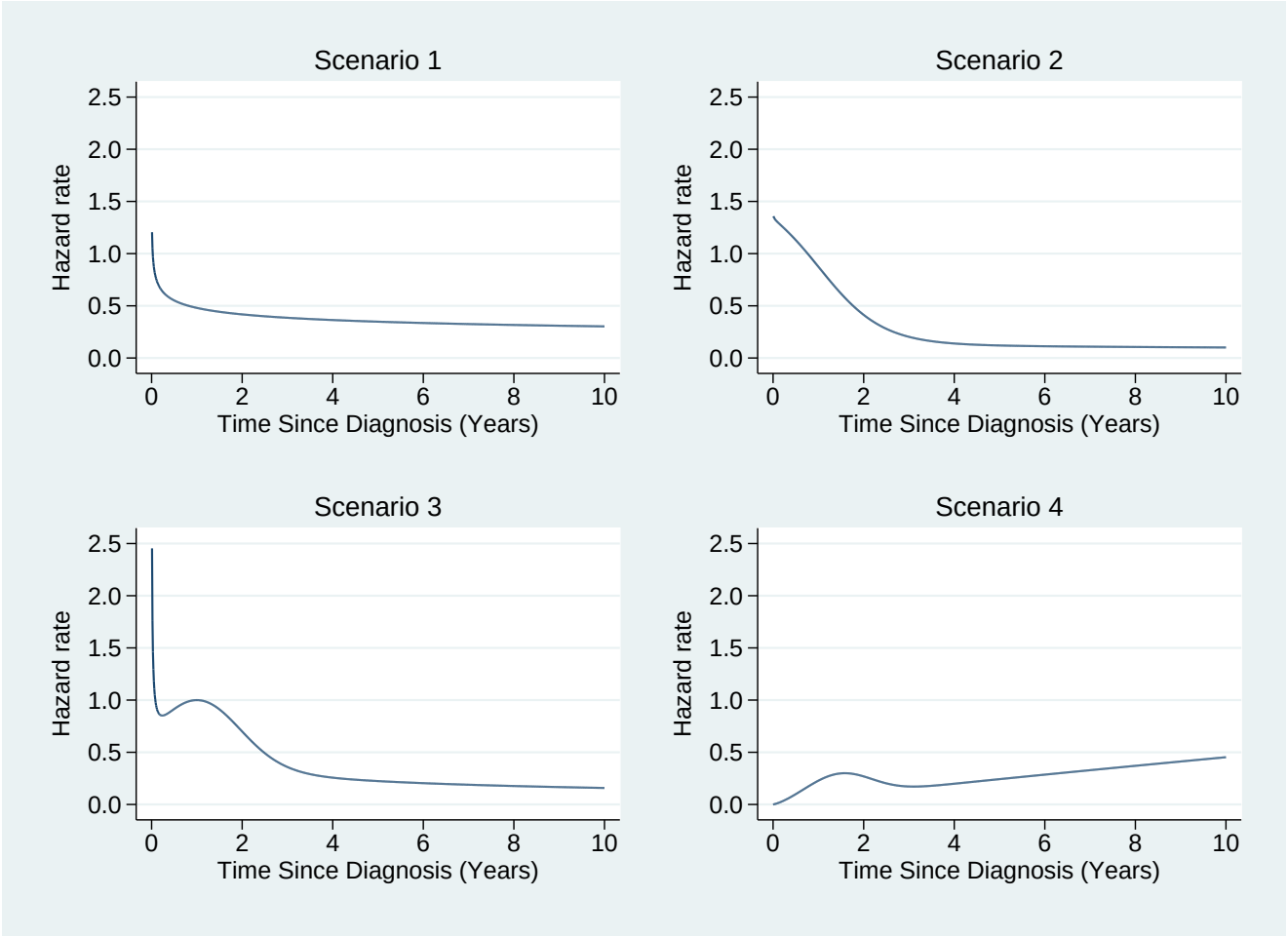
- Generate data assuming a mixture Weibull distribution,

$$S(t) = \pi \exp(-\lambda_1 t^{\gamma_1}) + (1 - \pi) \exp(-\lambda_2 t^{\gamma_2})$$

Simulation Study (Rutherford et al.)[6]

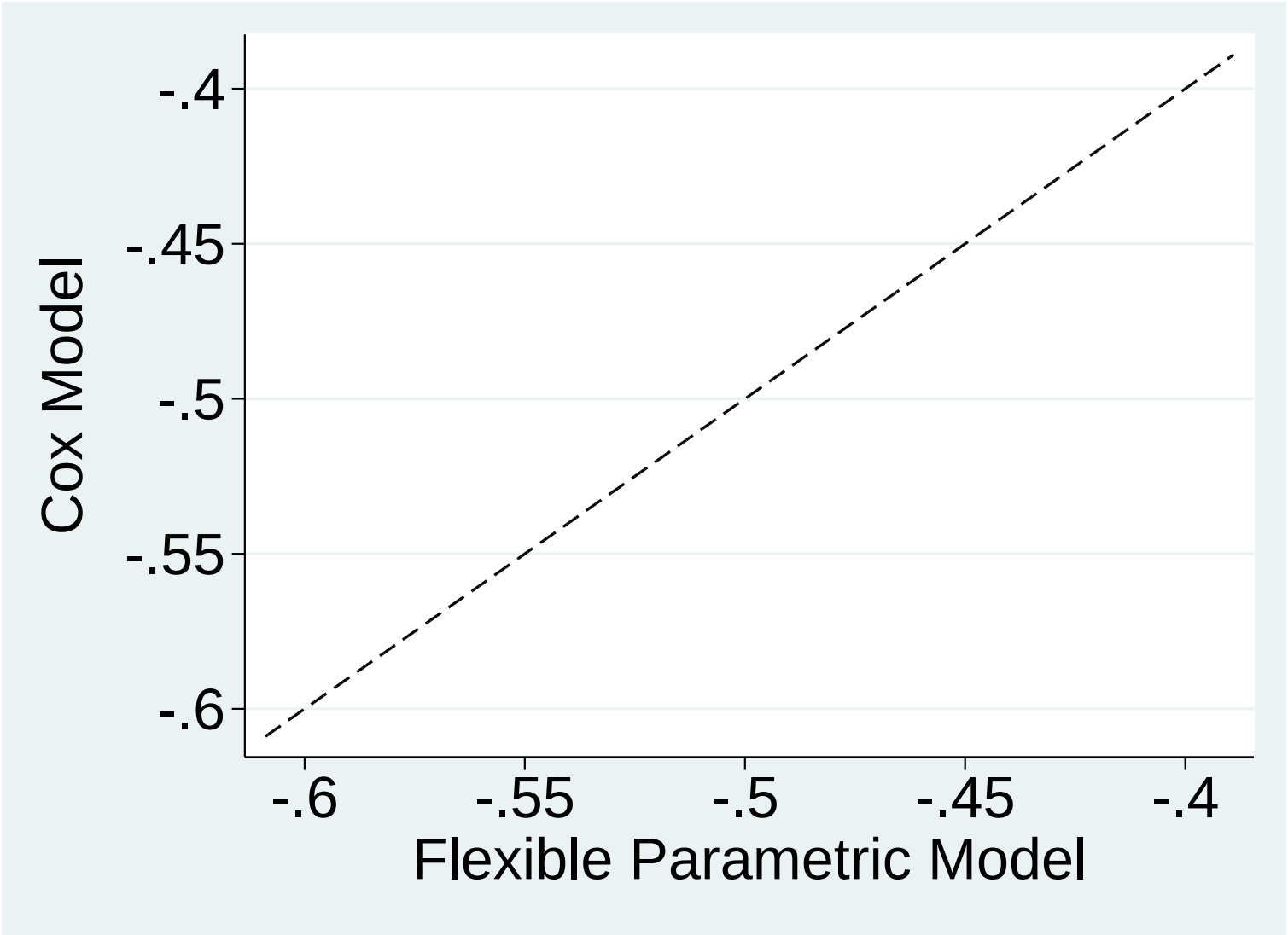
- Generate data assuming a mixture Weibull distribution,

$$S(t) = \pi \exp(-\lambda_1 t^{\gamma_1}) + (1-\pi)\exp(-\lambda_2 t^{\gamma_2})$$

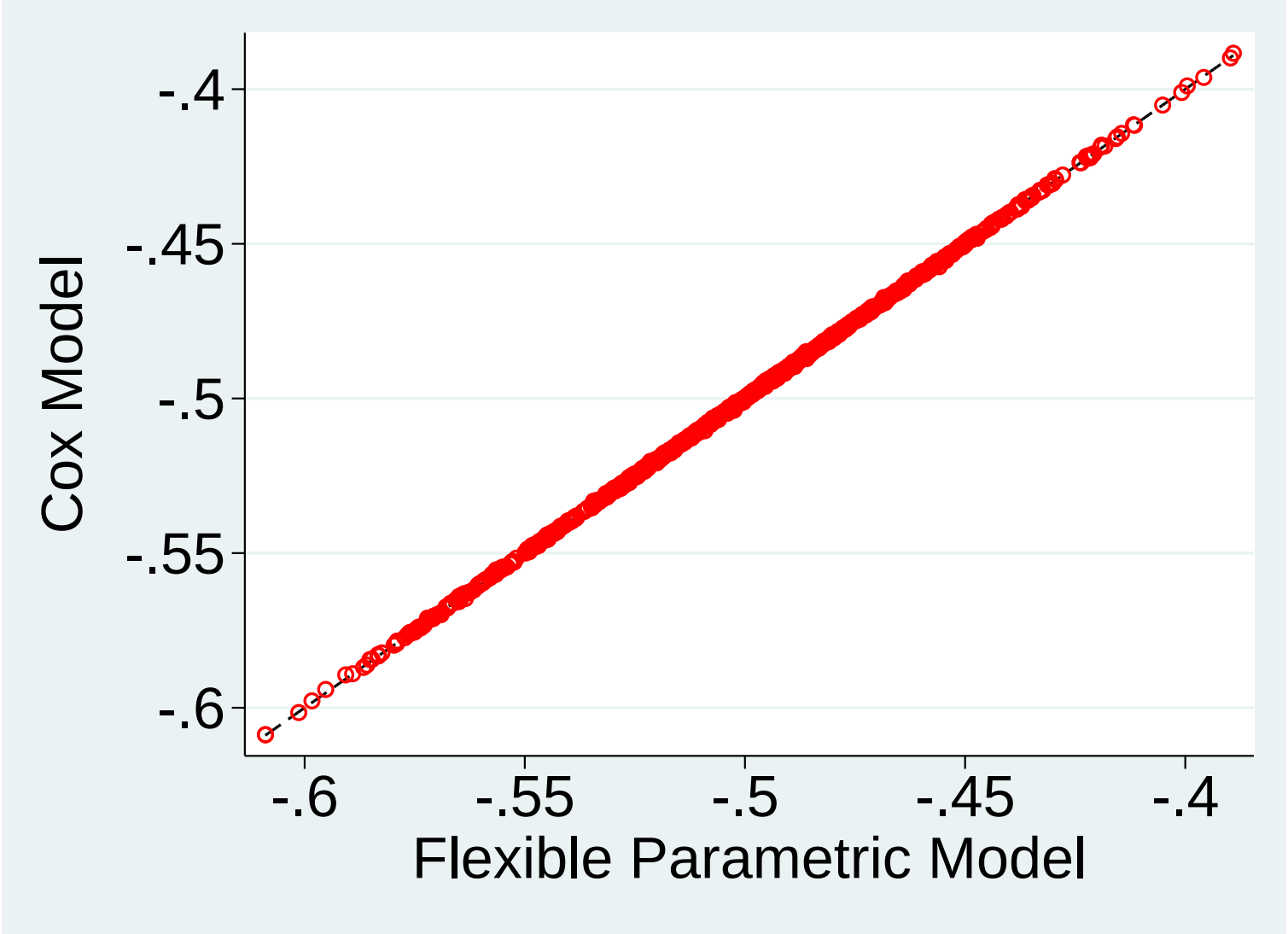


- Fit models using restricted cubic splines.

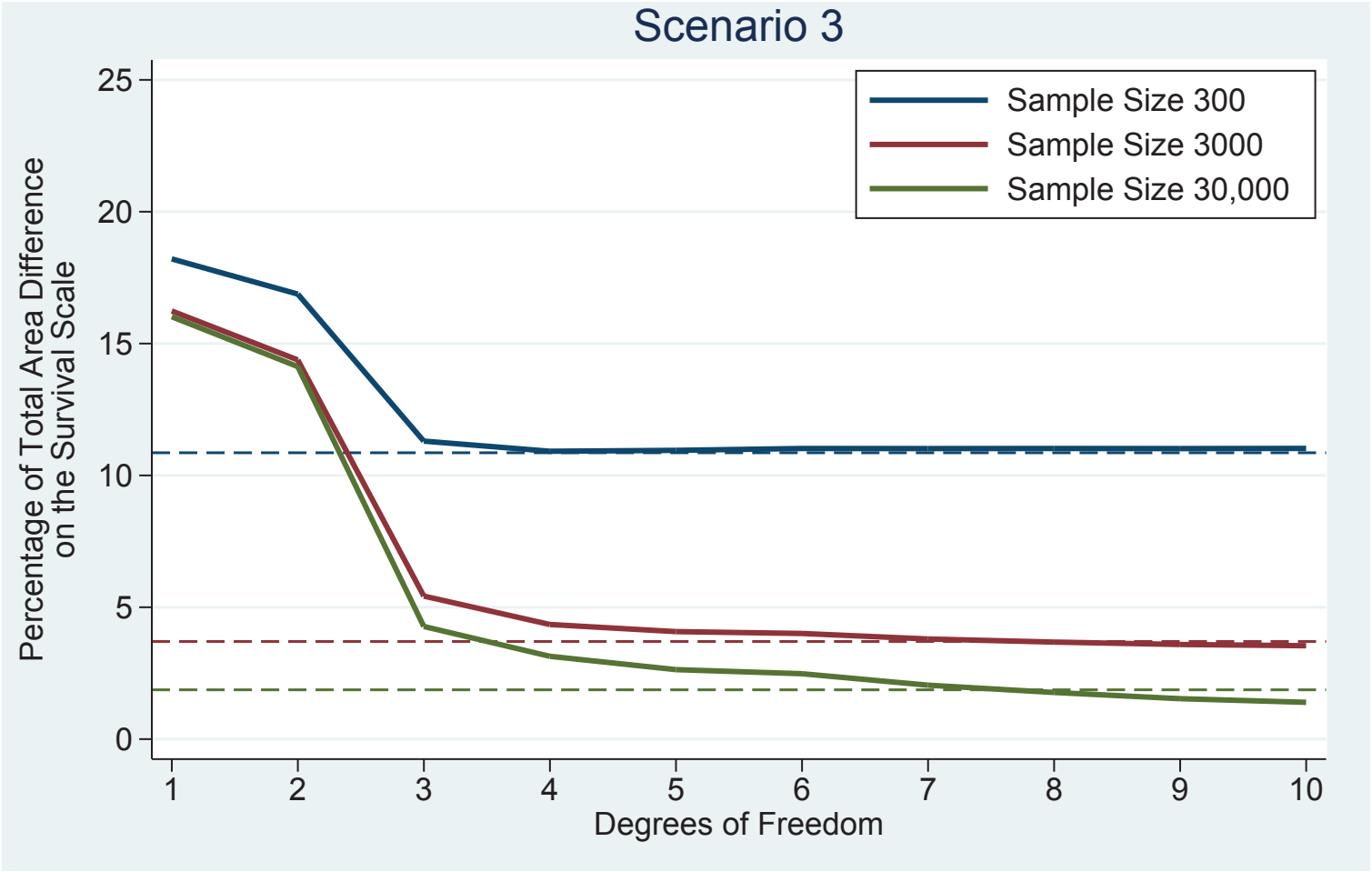
Scenario 3 comparison of Log Hazard Ratios



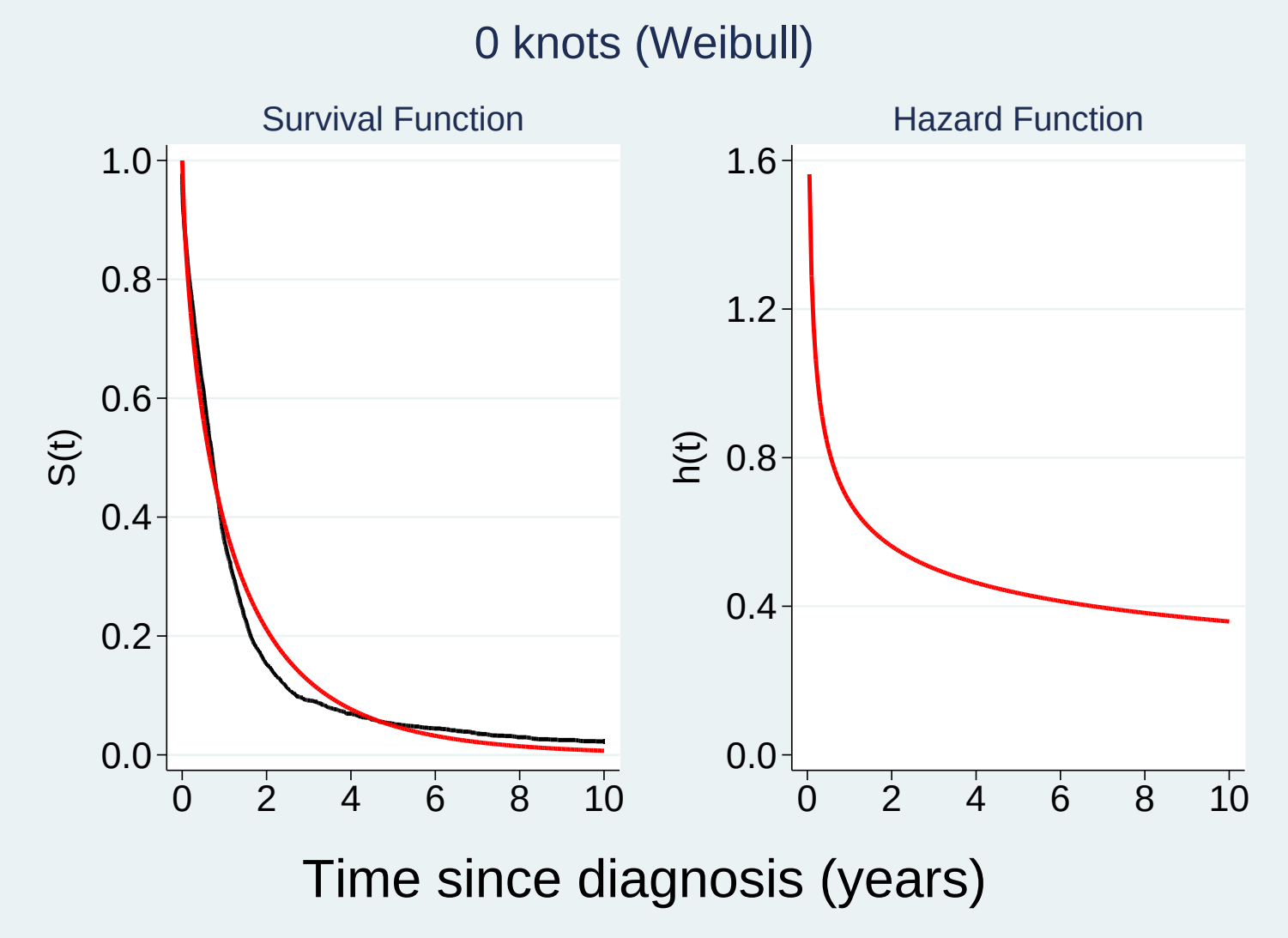
Scenario 3 comparison of Log Hazard Ratios



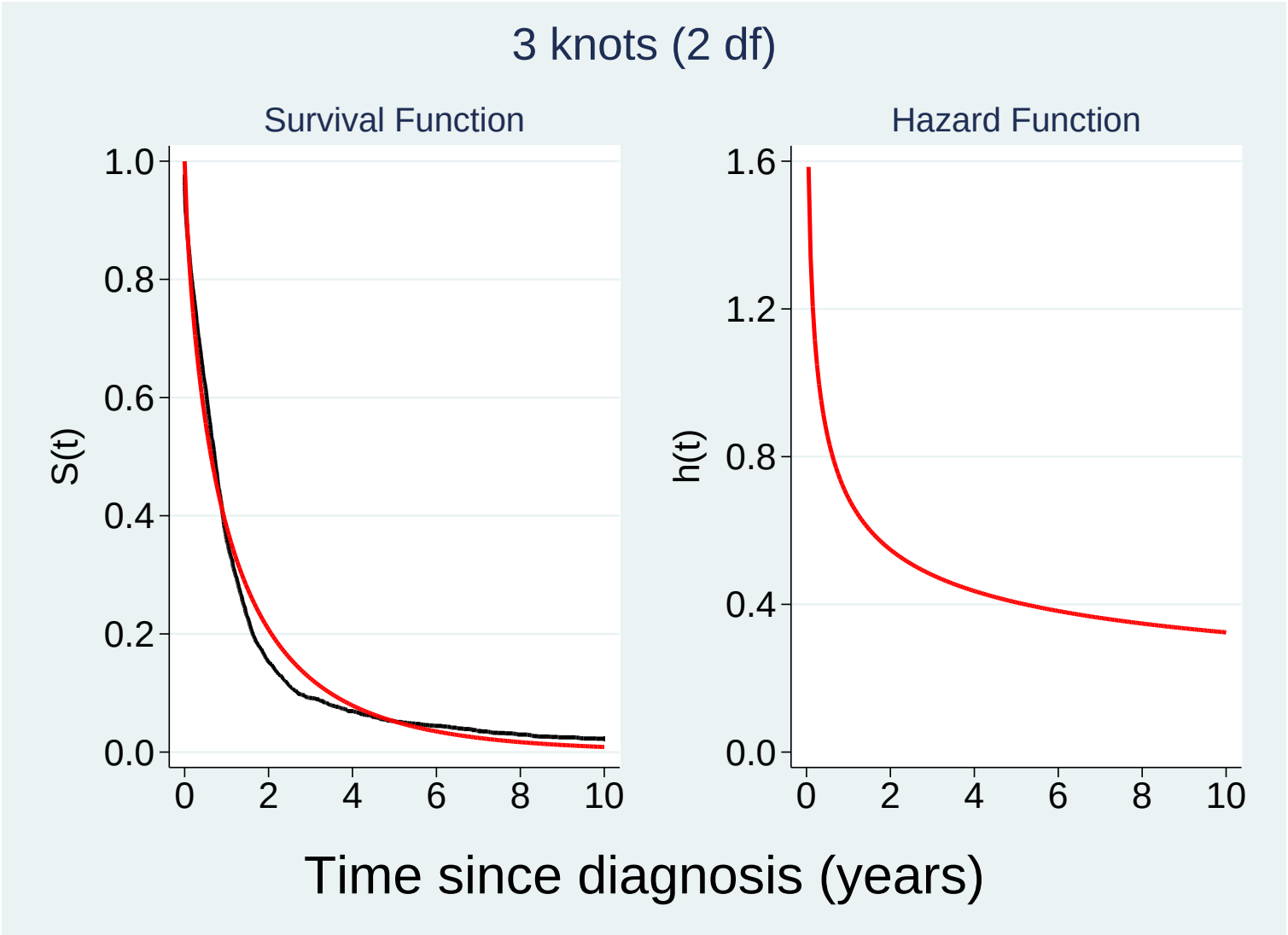
Restricted cubic splines vs true model



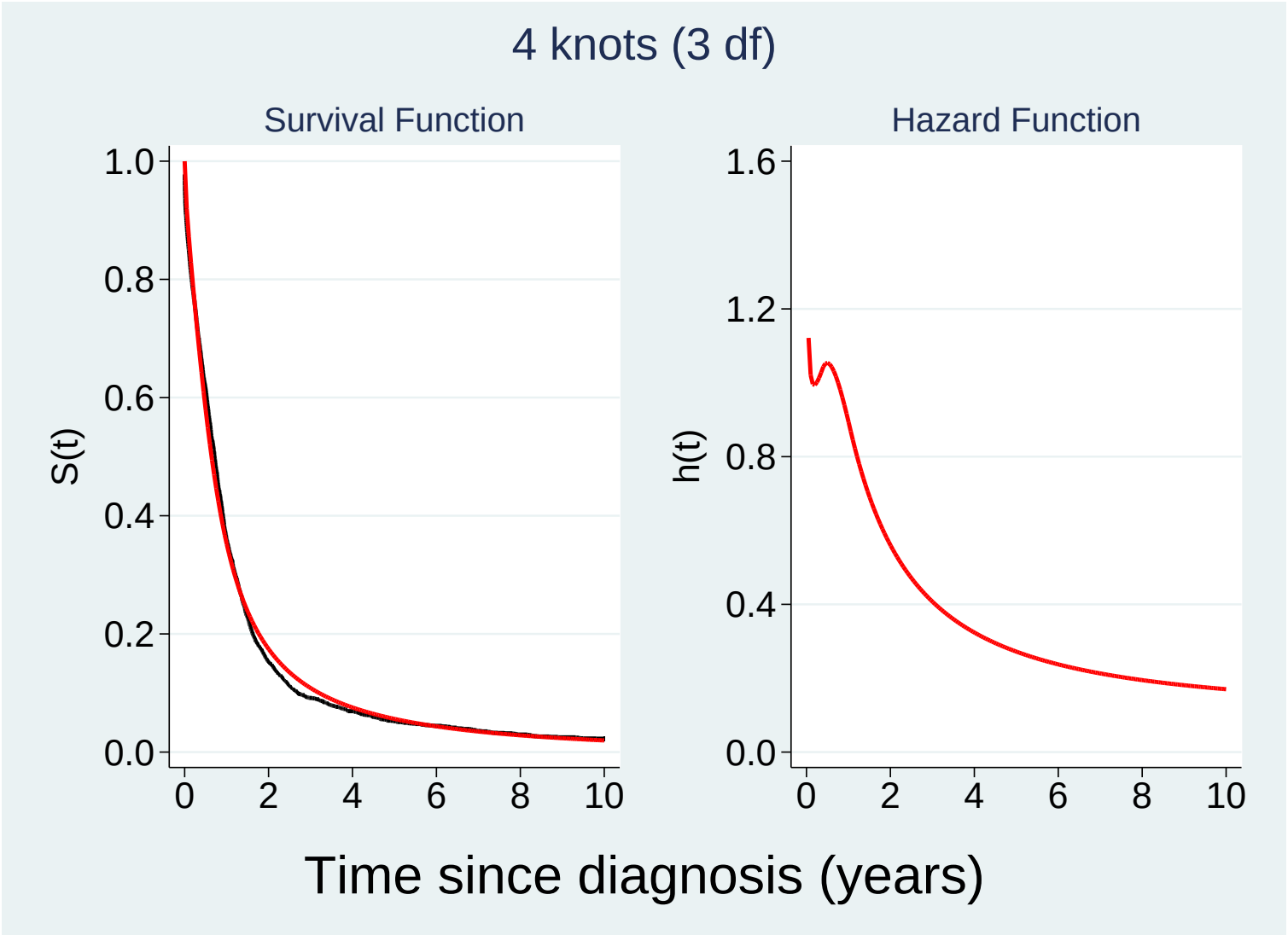
Choice of knots: Scenario 3



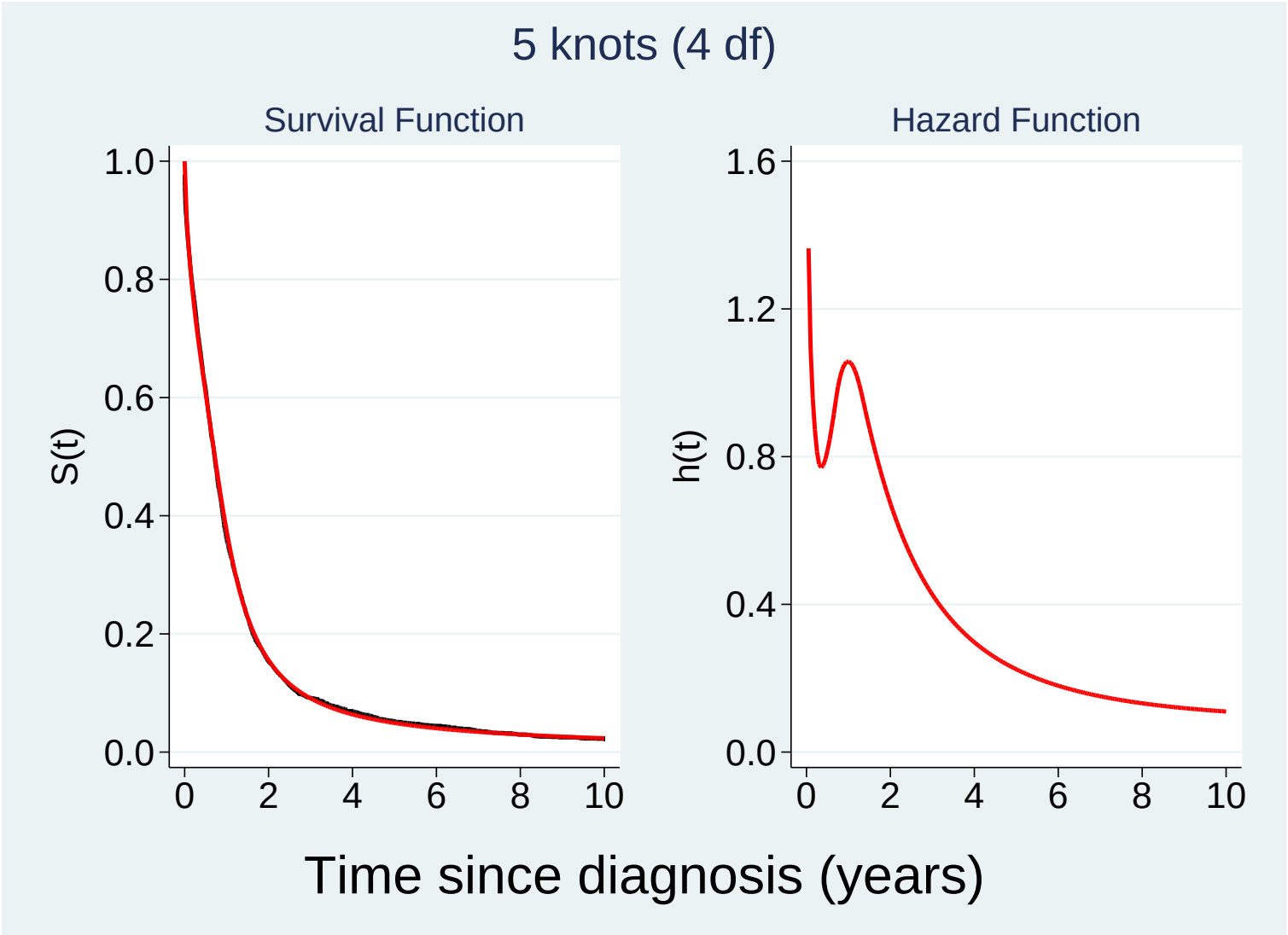
Choice of knots: Scenario 3



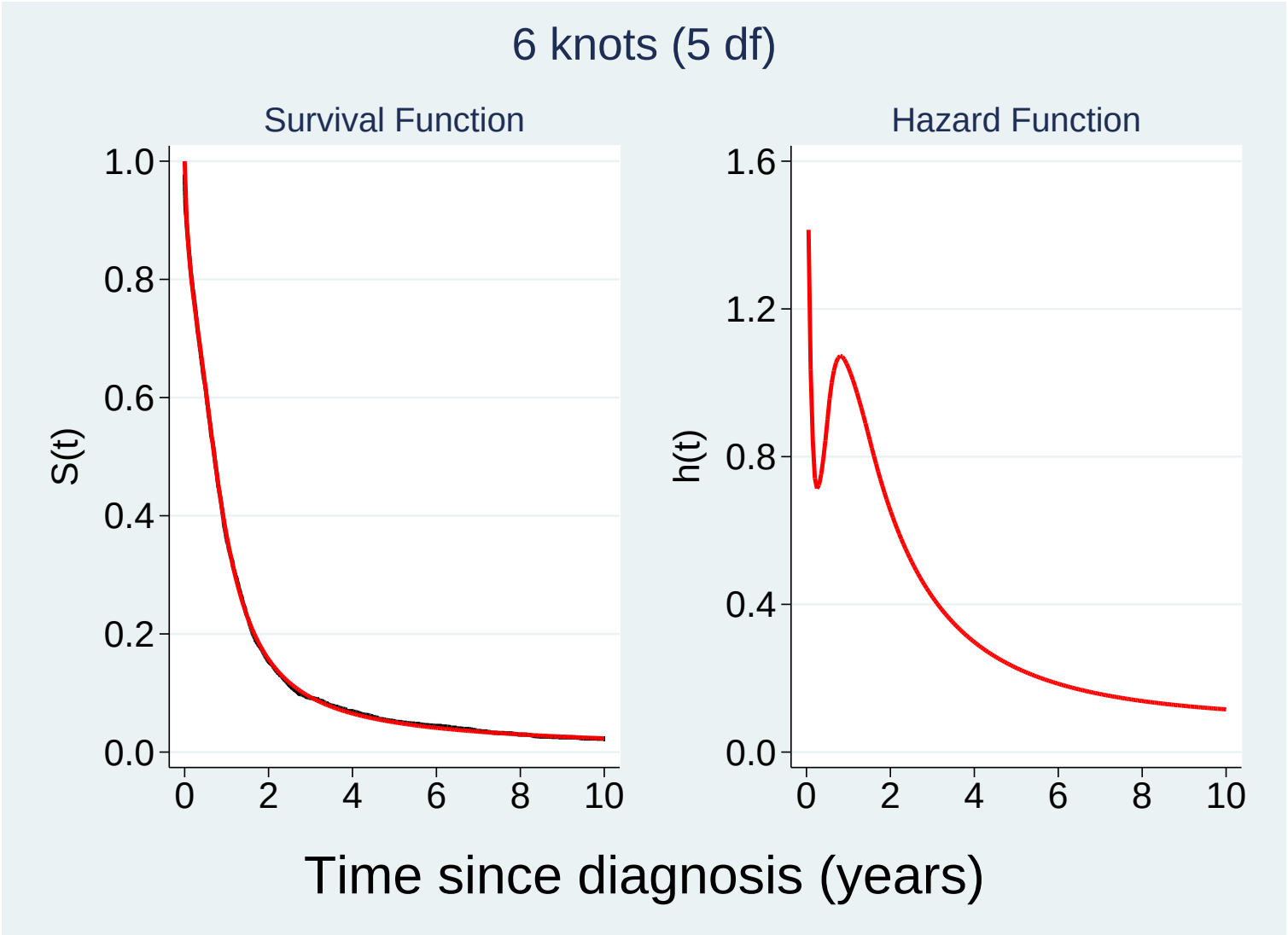
Choice of knots: Scenario 3



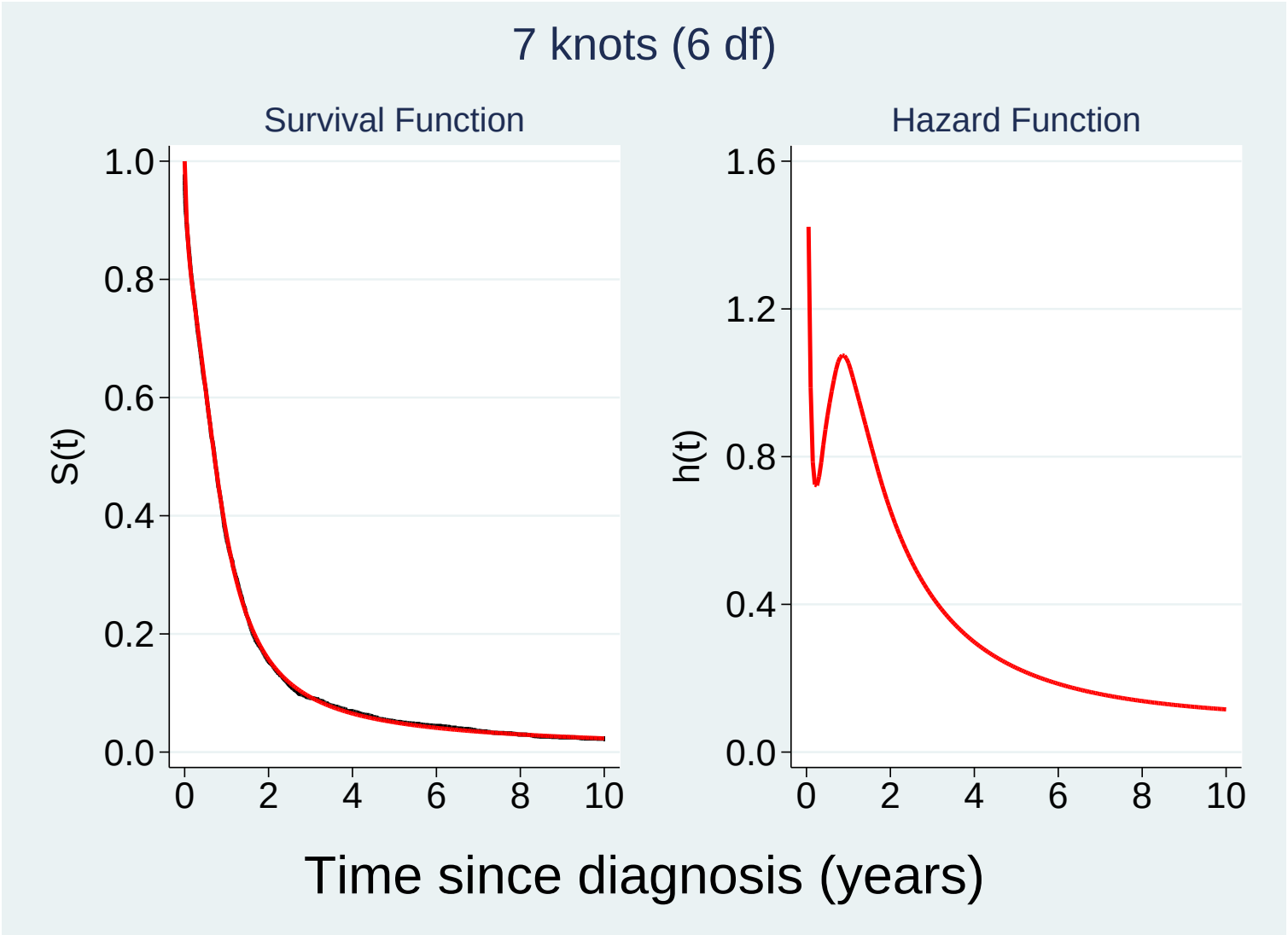
Choice of knots: Scenario 3



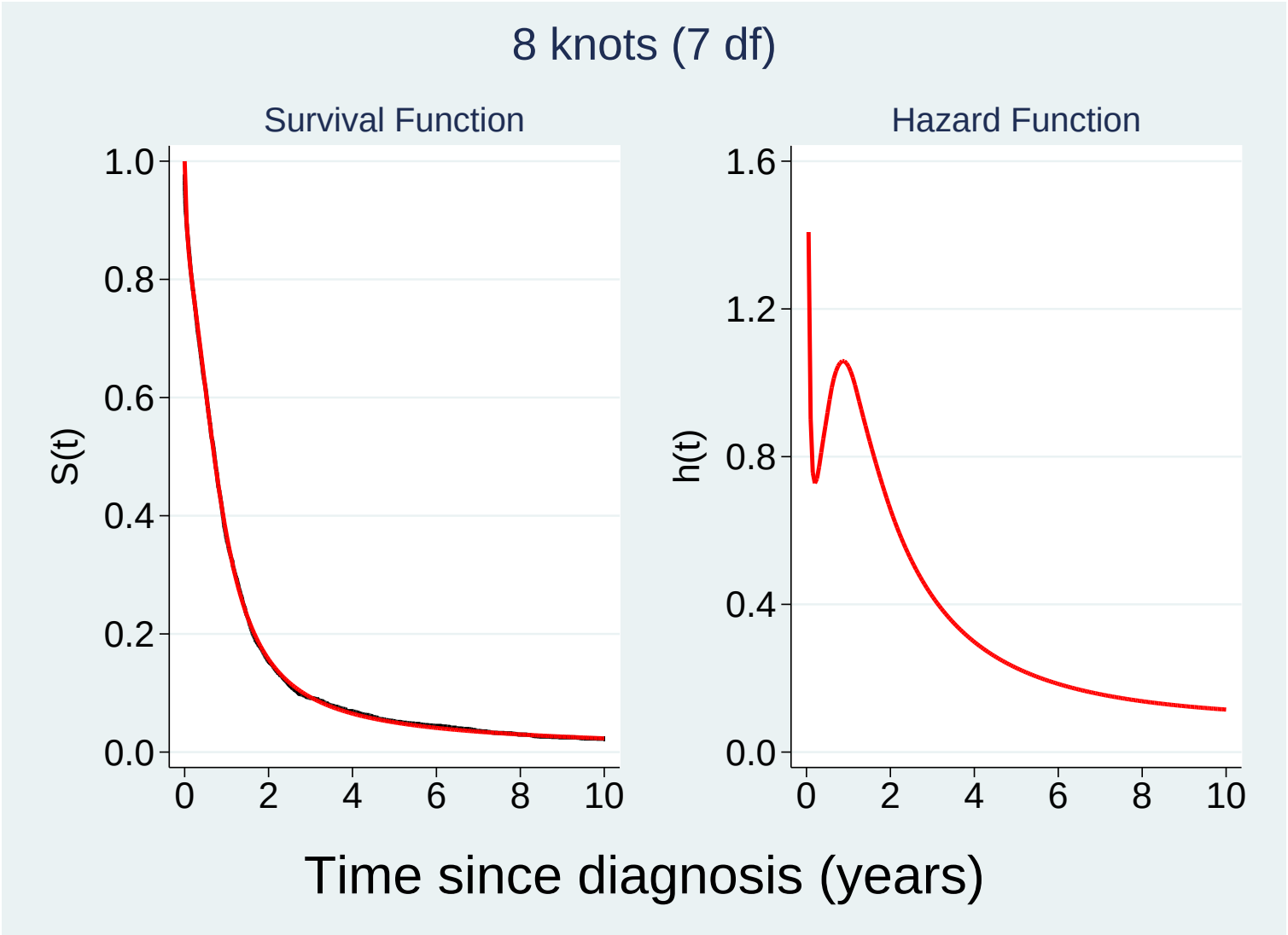
Choice of knots: Scenario 3



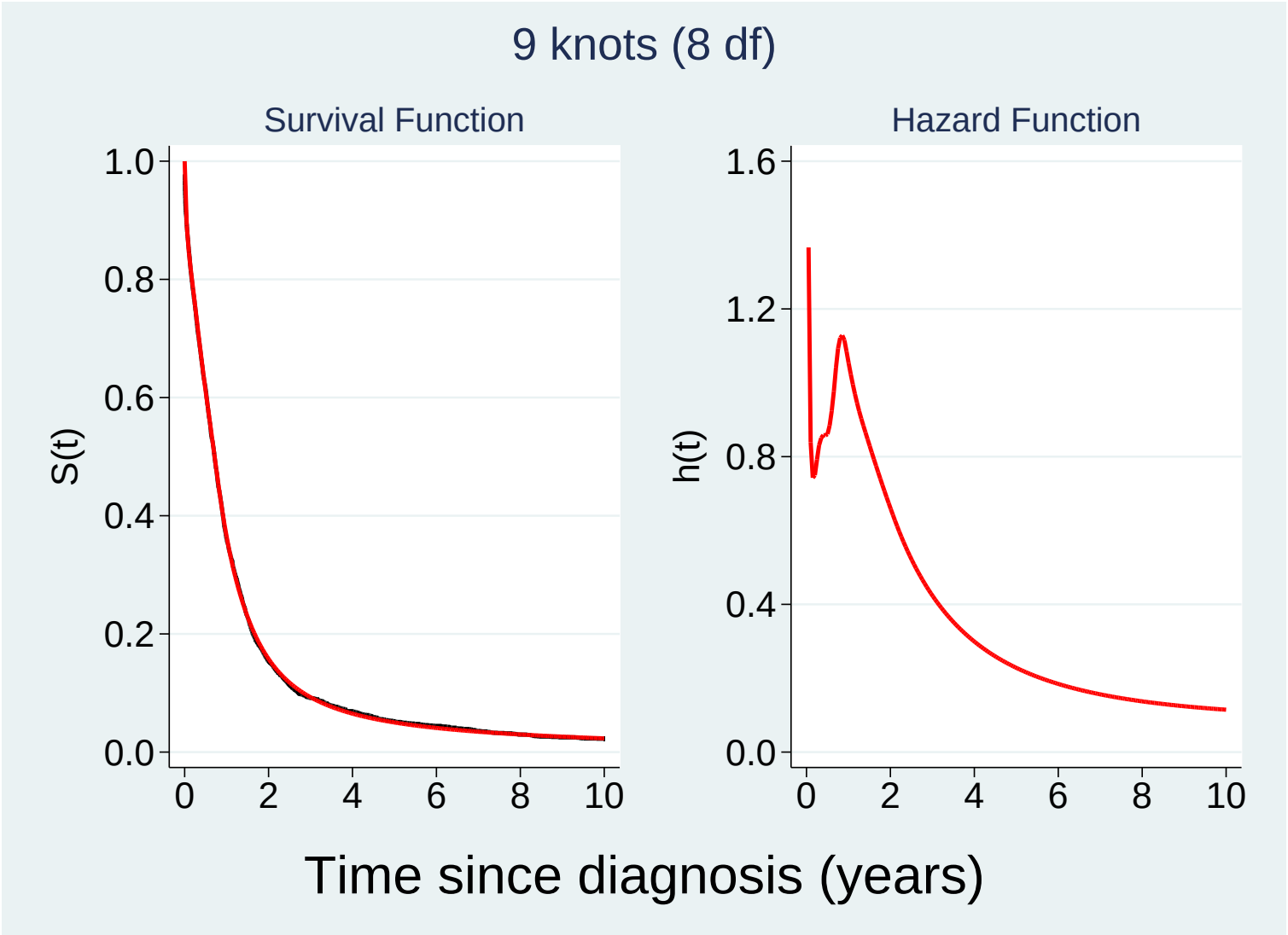
Choice of knots: Scenario 3



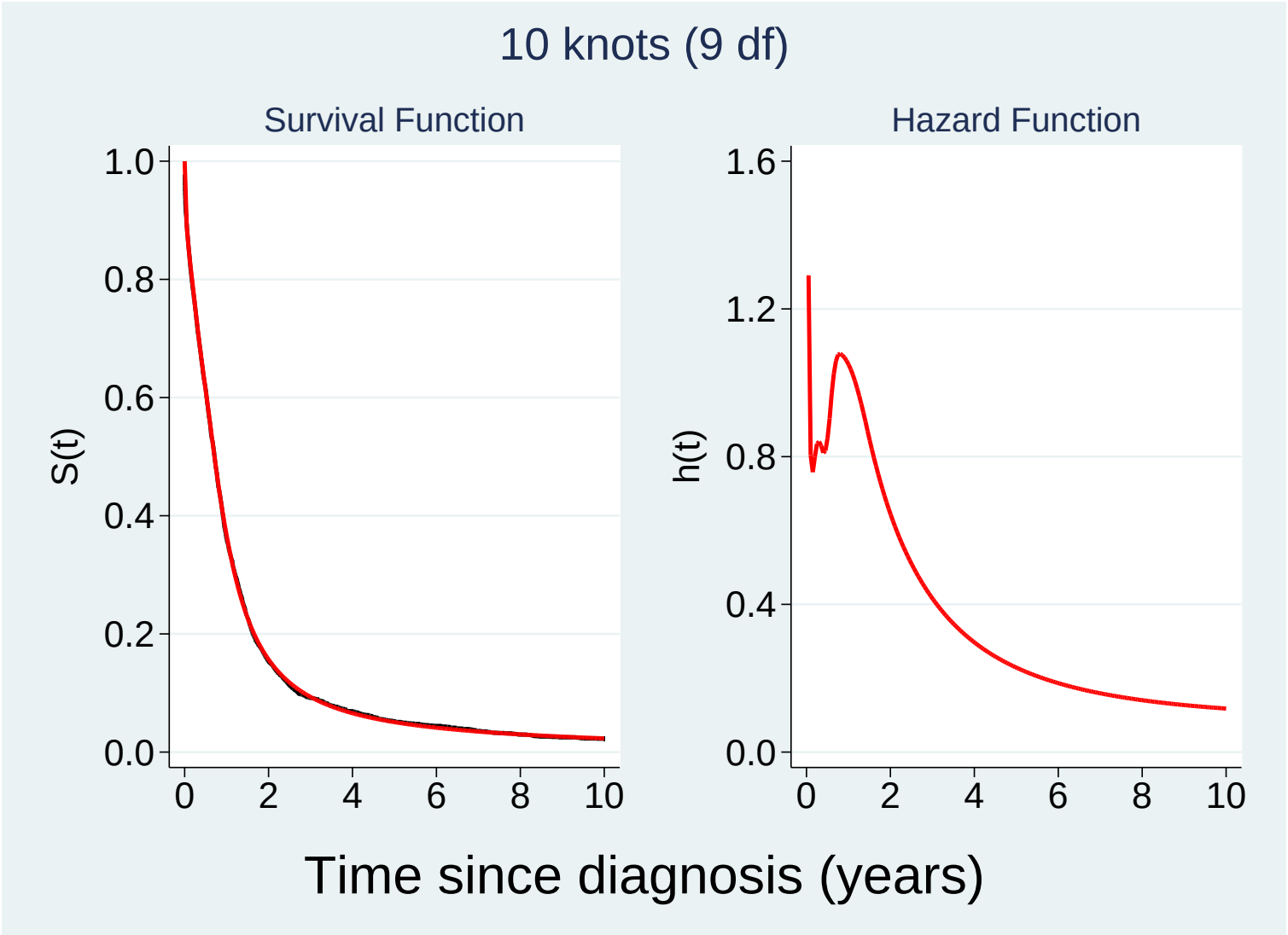
Choice of knots: Scenario 3



Choice of knots: Scenario 3



Choice of knots: Scenario 3



Model Selection

- Estimated hazard and survival functions fairly insensitive to knot location.
- AIC and BIC can be used as rough guides to choose models.
- Not crucial (within reason) to inferences from model.
  - We often present a sensitivity analysis to show this..
- Could treat number of knots and their locations as unknowns..
- However, it is an area where more work is still required.

Modelling time-dependent effects

- In studies I am involved in we frequently have non-proportional hazards.
- Non-proportional effects can be introduced.
- With D covariates with time-dependent effects.

$$\ln\left[H_i(t|\mathbf{x}_i)\right] = s\left(\ln(t)|\gamma, \mathbf{k}_0\right) + \sum_{j=1}^{D} s\left(\ln(t)|\delta_j, \mathbf{k}_j\right)x_{ij} + \mathbf{x}_i\beta$$

- Interaction between covariate and time (spline function).
- Generally have fewer knots for interaction term than for baseline.
- Need some caution with interpretation with multiple time-dependent effects.

Example of Attained Age as the Time-scale

- Study from Sweden[7] comparing incidence of hip fracture of,
  - 17,731 men diagnosed with prostate cancer treated with bilateral orchiectomy.
  - 43,230 men diagnosed with prostate cancer not treated with bilateral orchiectomy.
  - 362,354 men randomly selected from the general population.
- Study entry is 6 months post diagnosis.
- Outcome is femoral neck fracture.
- Risk of fracture varies by age.
- Attained age is used as the main time-scale.
- Alternative way of "adjusting" for age.
- Gives the age specific incidence rates.
- Actually, two timescales, but will initially ignore time from diagnosis.

Estimates from a proportional hazards model

Cox Model

Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.09 (1.93 to 2.27)

Estimates from a proportional hazards model

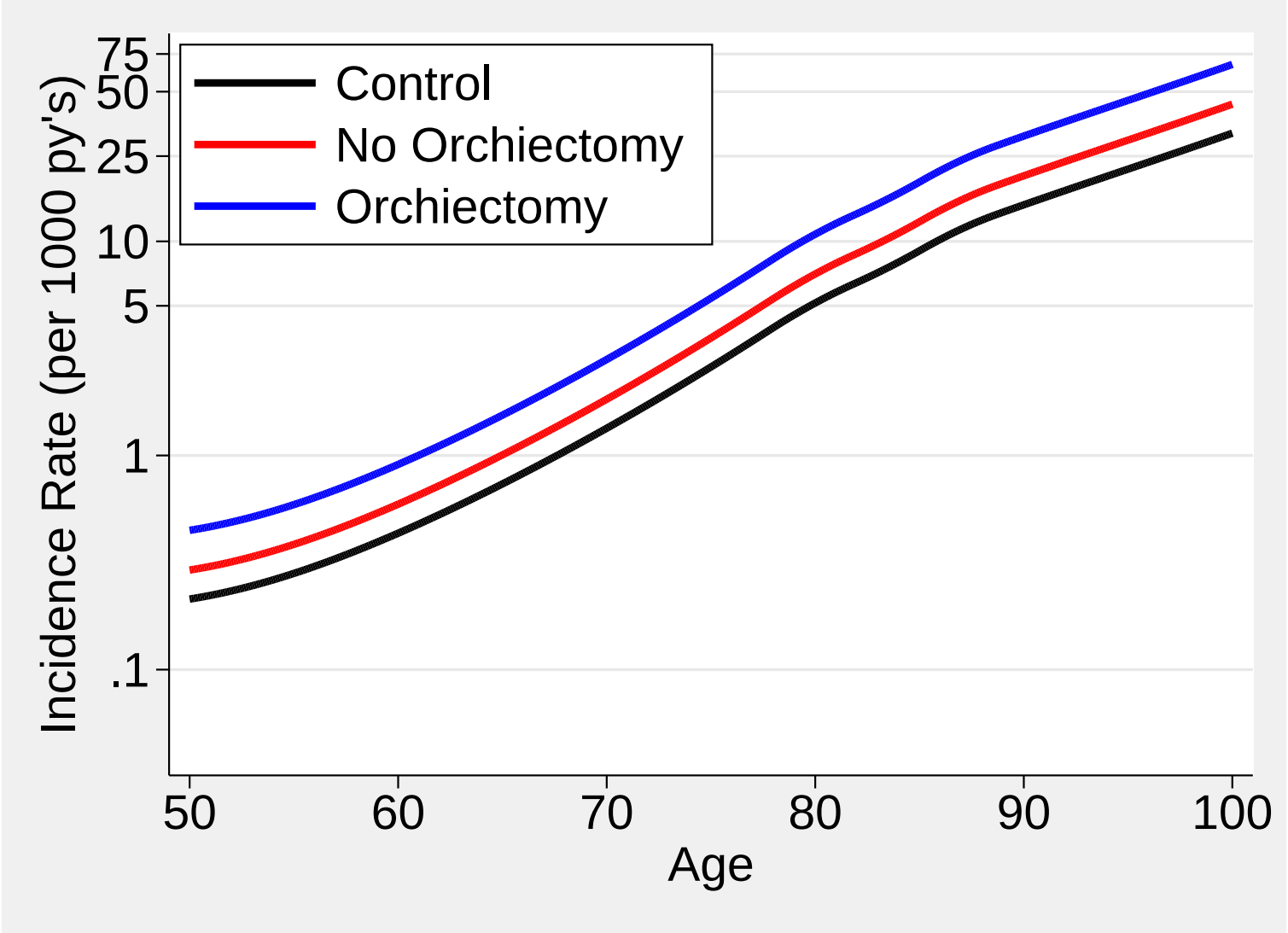
Cox Model

Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.09 (1.93 to 2.27)

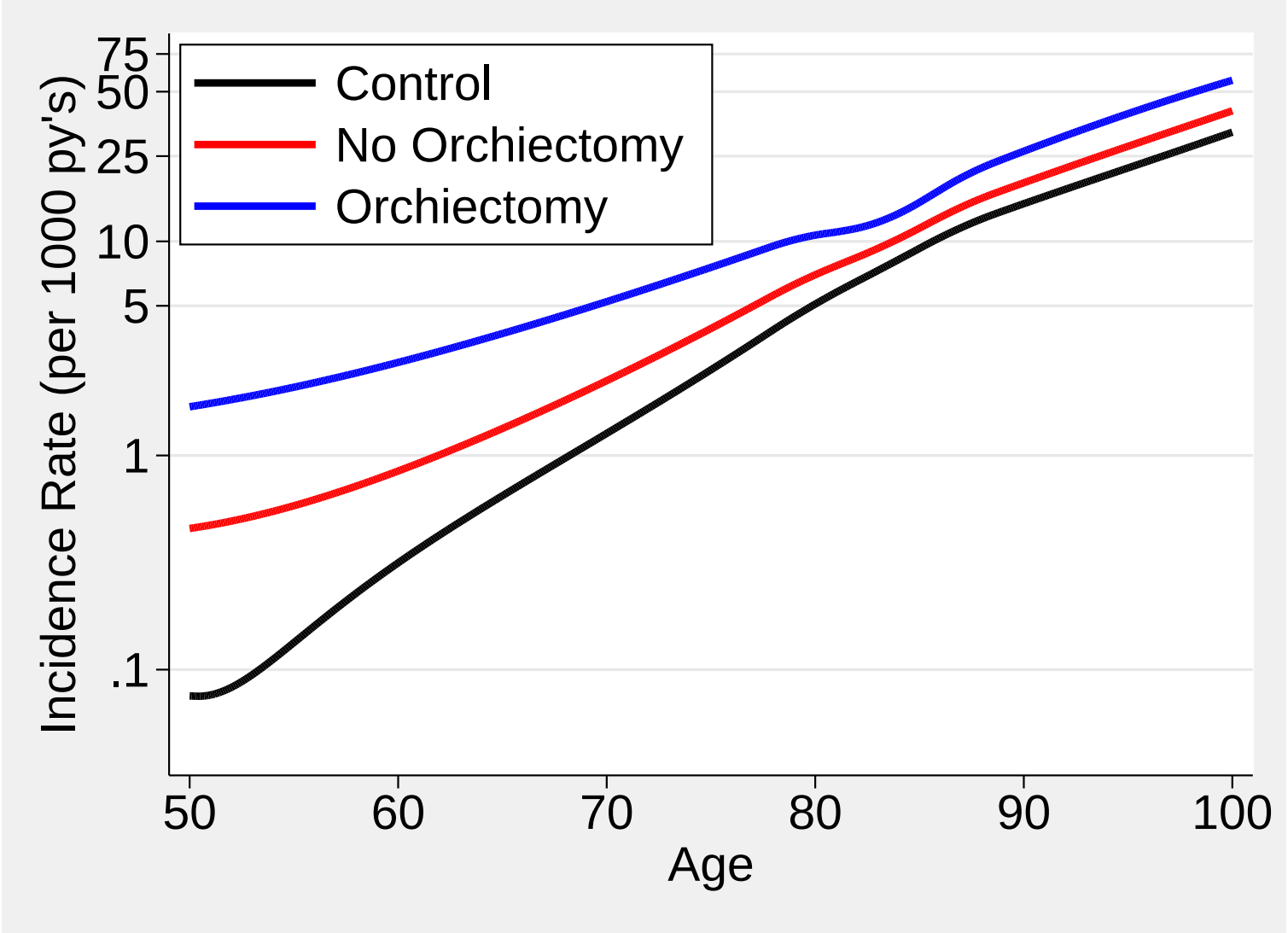
Flexible Parametric Model

Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.09 (1.93 to 2.27)

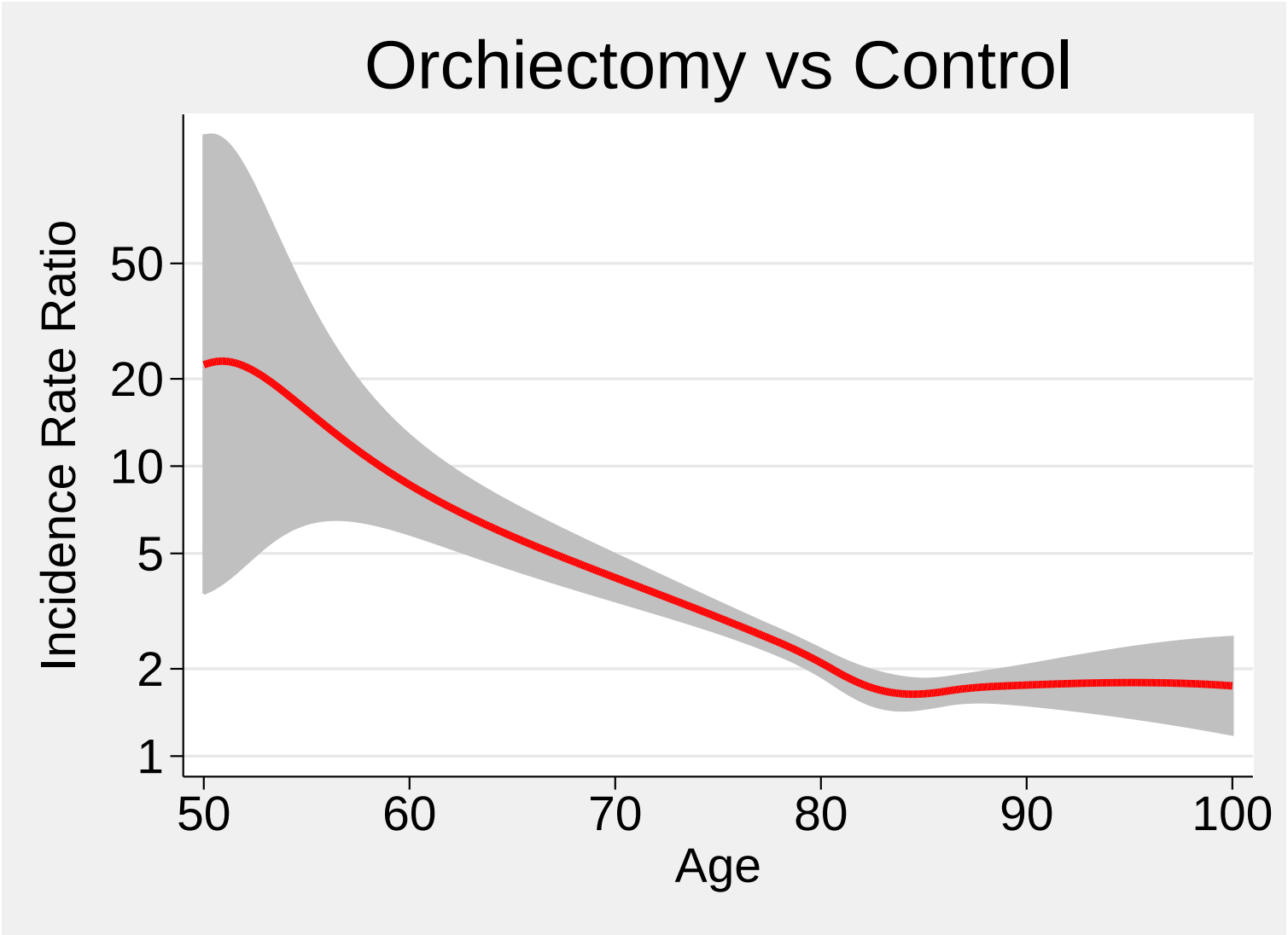
Proportional Hazards



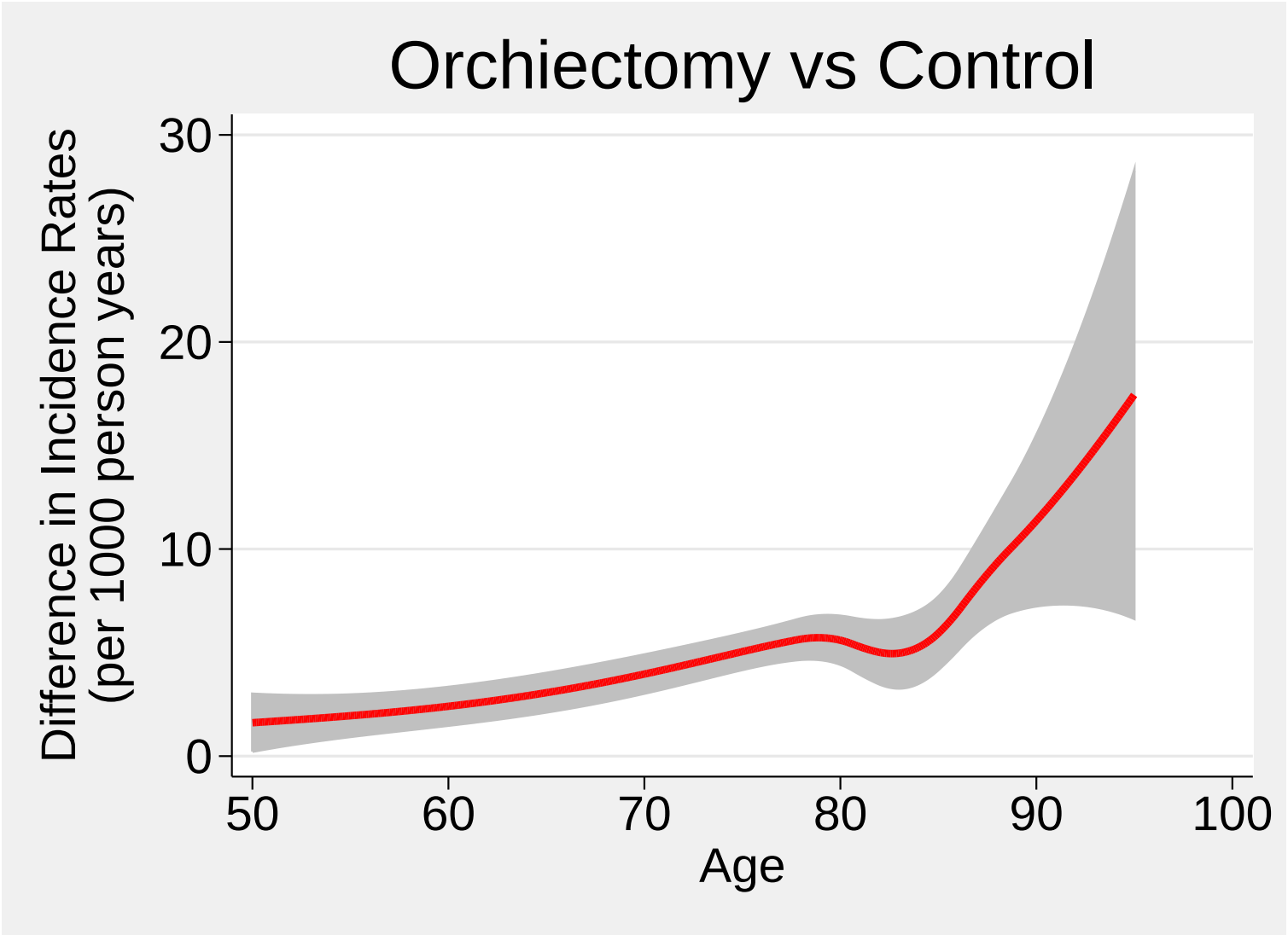
Non Proportional Hazards



Incidence Rate Ratio



Incidence Rate Difference



Multiple Time-scales

- Both attained age and time since diagnosis can be modelled simultaneously, i.e. two time-scales[7]. Main time-scale is age.
- Better to use hazard scale.
- Model for PH, but can be extended to time-dependent effects.

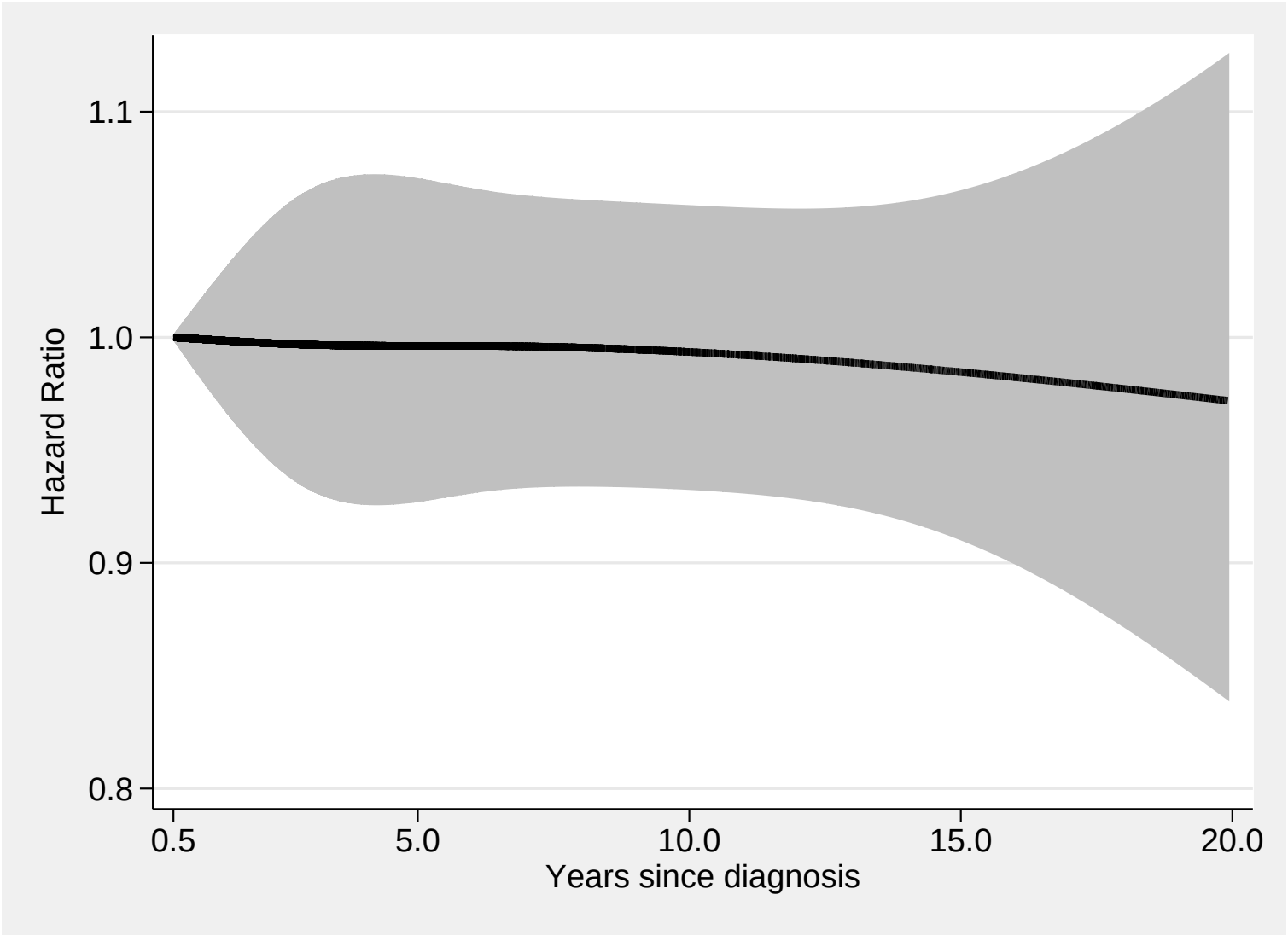
$$\ln[h(a|\mathbf{x}_i, a_{0i})] = s\left(a|\boldsymbol{\gamma_0}, \mathbf{k}_0\right) + \mathbf{x}_i\boldsymbol{\beta} + s\left(a - a_{0i}|\boldsymbol{\gamma_1}, \mathbf{k}_1\right)$$

- a_{0i} is age at diagnosis
- Numerical integration required to obtain cumulative hazard for each individual at each iteration.

$$\ln L_i = d_i \ln[h(t_i)] - \int_{t_{0i}}^{t_i} h(u)du$$

- For orchiectomy data (N=423,315) takes about 15 minutes on my 4 year old laptop using `stgenreg` in Stata[8]

Hazard Ratio for 2nd time scale (reference 0.5)



Quantification of Differences

- Time-dependent hazard ratios.
- Differences in hazard rates.
- Differences in survival function (also NNT).
- Avoidable Deaths.
  - E.g. For a cohort of women diagnosed in one year in England, we estimated that there would be 1020 fewer deaths by 5 years if England could achieve the same excess mortality rates as Norway [9].
- Loss in expectation of life.
  - Requires extrapolation.

Extrapolation

- Often interest in extrapolation of survival curves beyond range of follow-up.
  - For example, in economic evaluations.
  - Often a standard parametric model is fitted (e.g. Weibull) and use estimated parameters for long-term extrapolation.
- Our interest arose form wanting to estimate the loss in expectation of life and related measures in population-based cancer studies.

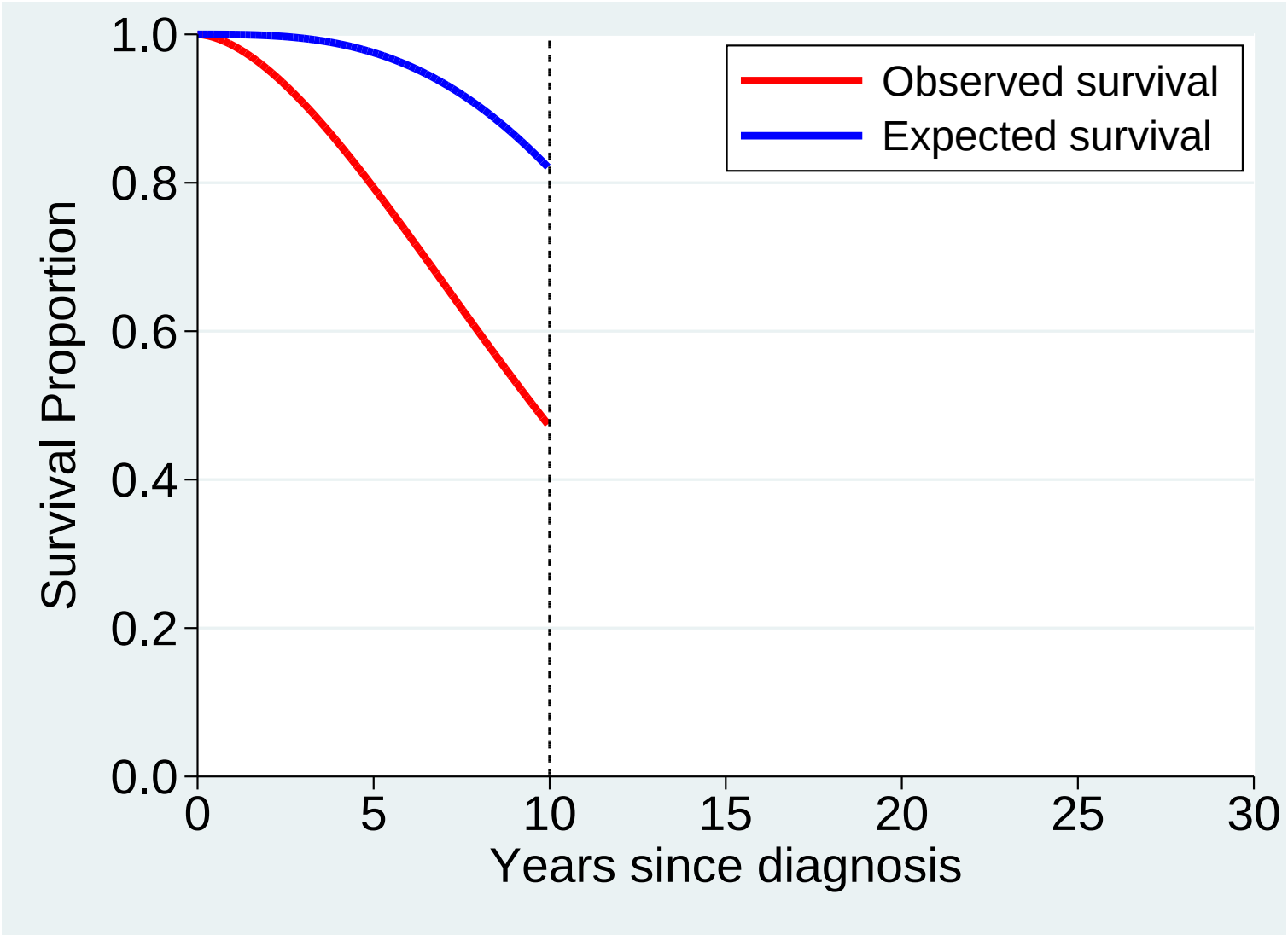
Extrapolation

- Often interest in extrapolation of survival curves beyond range of follow-up.
  - For example, in economic evaluations.
  - Often a standard parametric model is fitted (e.g. Weibull) and use estimated parameters for long-term extrapolation.
- Our interest arose form wanting to estimate the loss in expectation of life and related measures in population-based cancer studies.

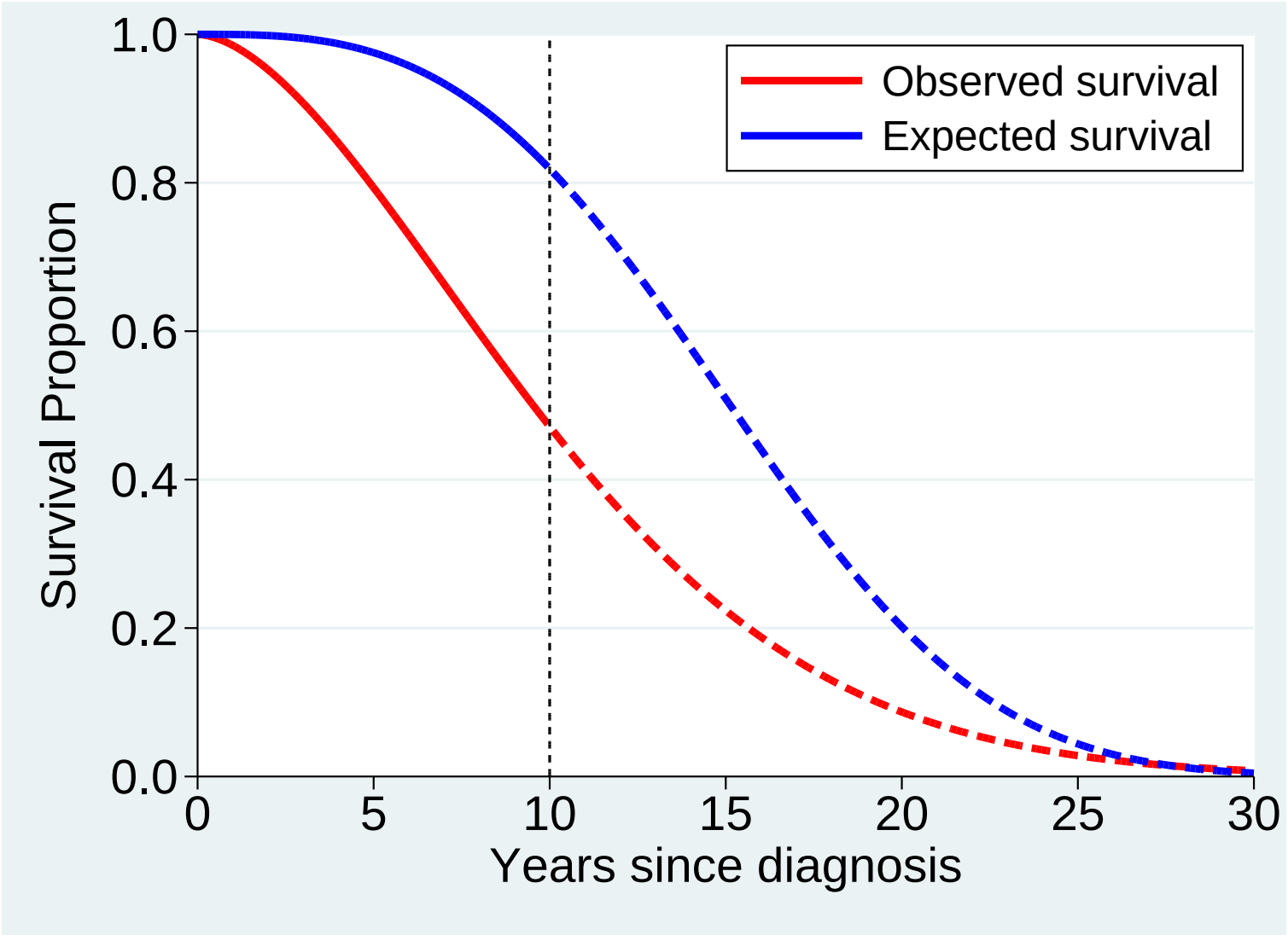
- How many life-years are lost due to a particular cancer?
- How many life-years are lost due to differences in cancer patient survival between social-economic groups?
- How many life-years would be gained if England had the same cancer patient survival as Sweden?

- Need to extrapolate survival curve.

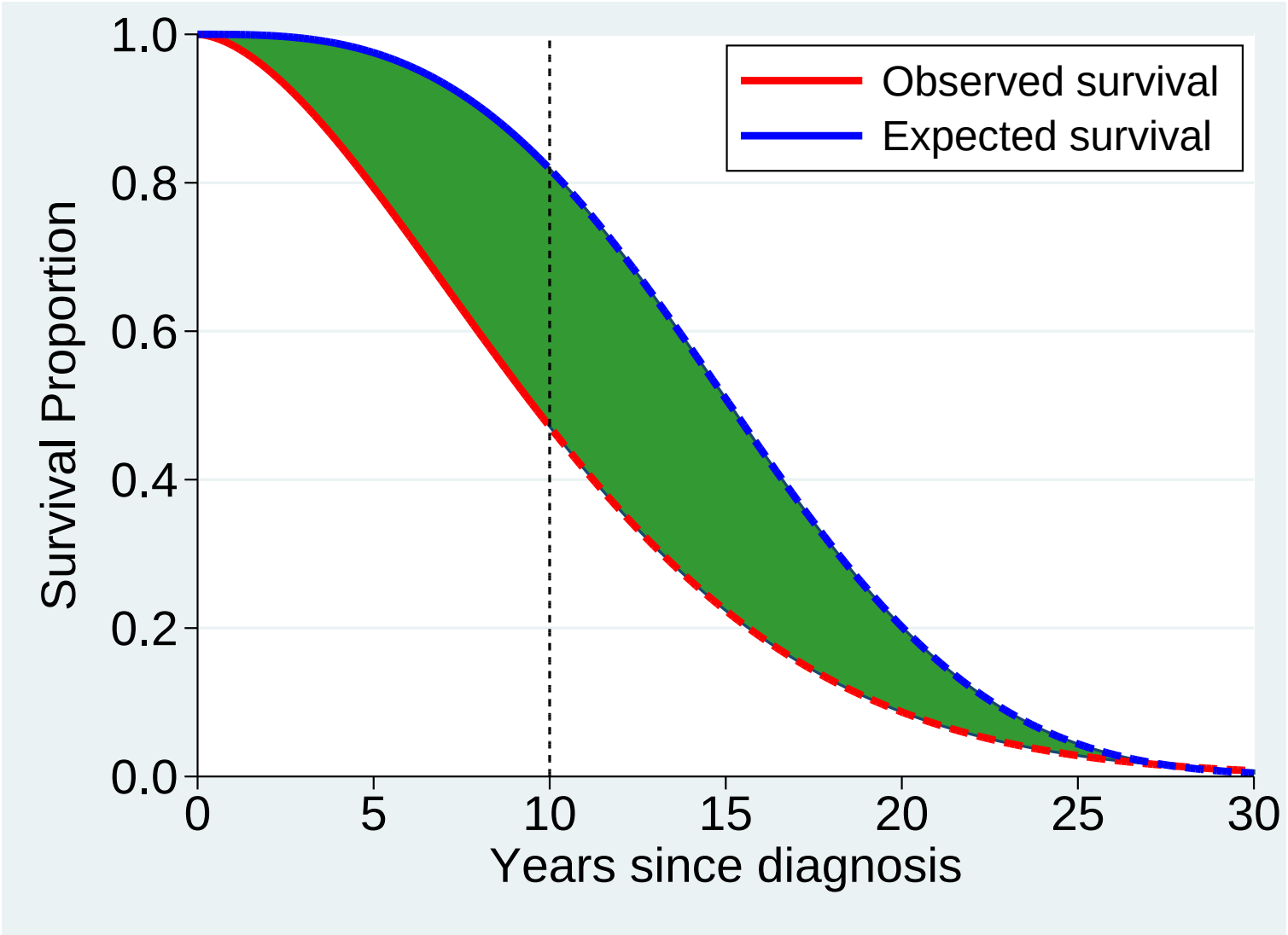
Loss in expectation of life



Loss in expectation of life



Loss in expectation of life



Excess Mortality Models

- Common way of modelling population-based cancer data.
- Interested in mortality associated with a diagnosis of cancer without the need for cause of death information [10].
- All-cause mortality, $h(t)$ made up of two components, expected, $h^*(t)$, and excess, $\lambda(t)$, mortality.

$$h(t) = h^*(t) + \lambda(t)$$

- $\lambda(t)$ usually modelled parametrically[7, 11, 12].
- $h^*(t)$ assumed known - obtained from routine data sources.

Excess Mortality Models

- Common way of modelling population-based cancer data.
- Interested in mortality associated with a diagnosis of cancer without the need for cause of death information [10].
- All-cause mortality, $h(t)$ made up of two components, expected, $h^*(t)$, and excess, $\lambda(t)$, mortality.

$$h(t) = h^*(t) + \lambda(t)$$

- $\lambda(t)$ usually modelled parametrically[7, 11, 12].
- $h^*(t)$ assumed known - obtained from routine data sources.
- Extensions to Cox model: Sasieni[13] and Pohar-Perme *et al.*[14].
  - Estimation needs non-parametric smoother of baseline excess hazard function (updated at each iteration).
- We use models on the log *cumulative* excess hazard scale[15].
- Also known as relative survival models, $R(t) = exp(-\int_0^t \lambda(u)du)$

Extrapolation (Andersson *et al.*)[16]

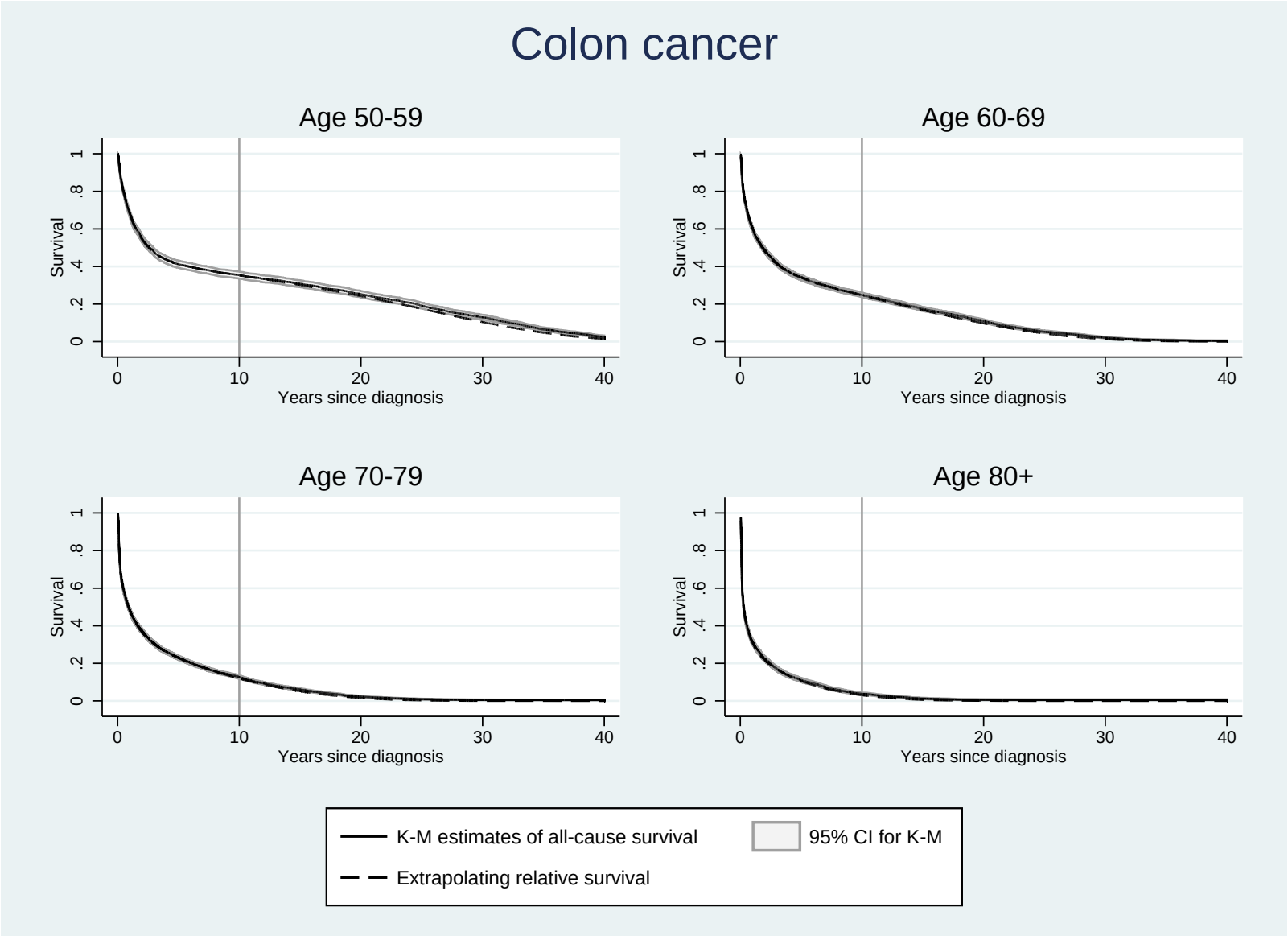
$$h(t) = h^*(t) + \lambda(t) \qquad S(t) = S^*(t)R(t)$$

- As follow-up time increases $h^*(t)$ tends to dominate.
- Make simple assumptions about excess mortality after end of follow-up.
- Derive relative survival, $R(t)$. Combine with expected survival, $S^*(t)$ to obtain all cause survival, $S(t)$.
- Loss in expectation of life, $\int S^*(t) - \int S(t)$.

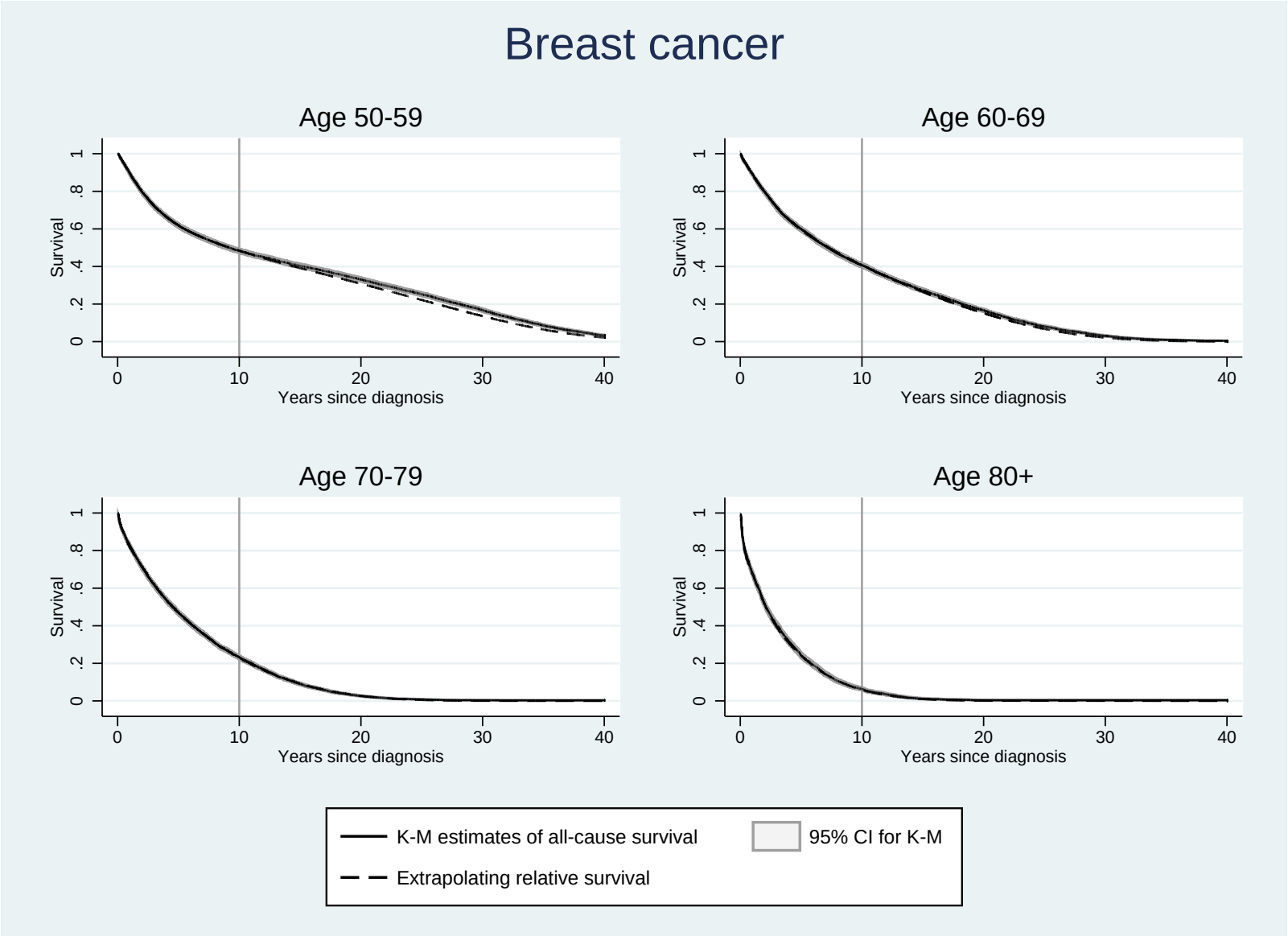
Three possible assumptions when extrapolating

1. No excess mortality after certain point in time (Population cure).
2. Constant excess mortality after certain point in time.
3. Excess mortality estimated from final linear component.

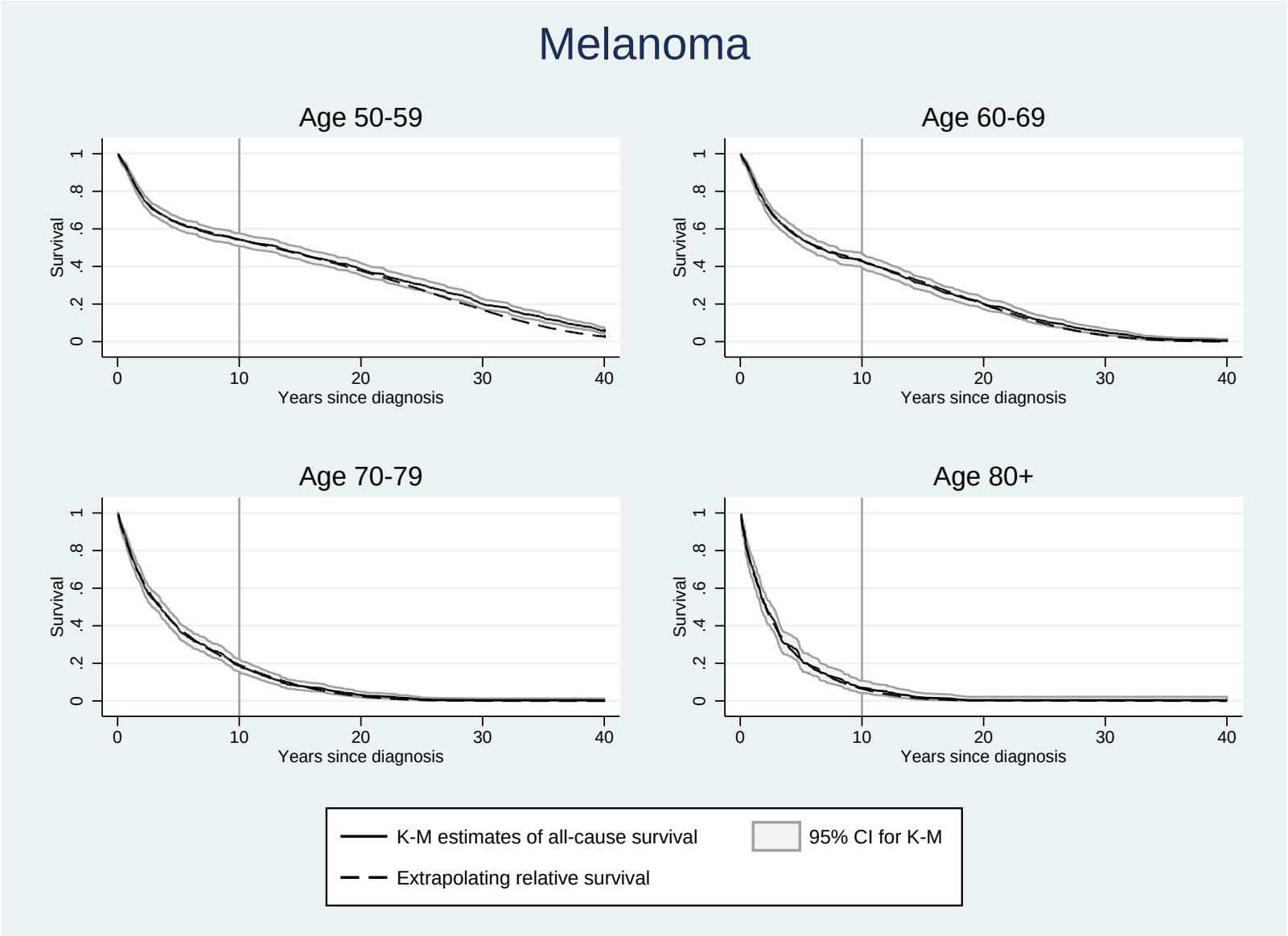
Colon Cancer



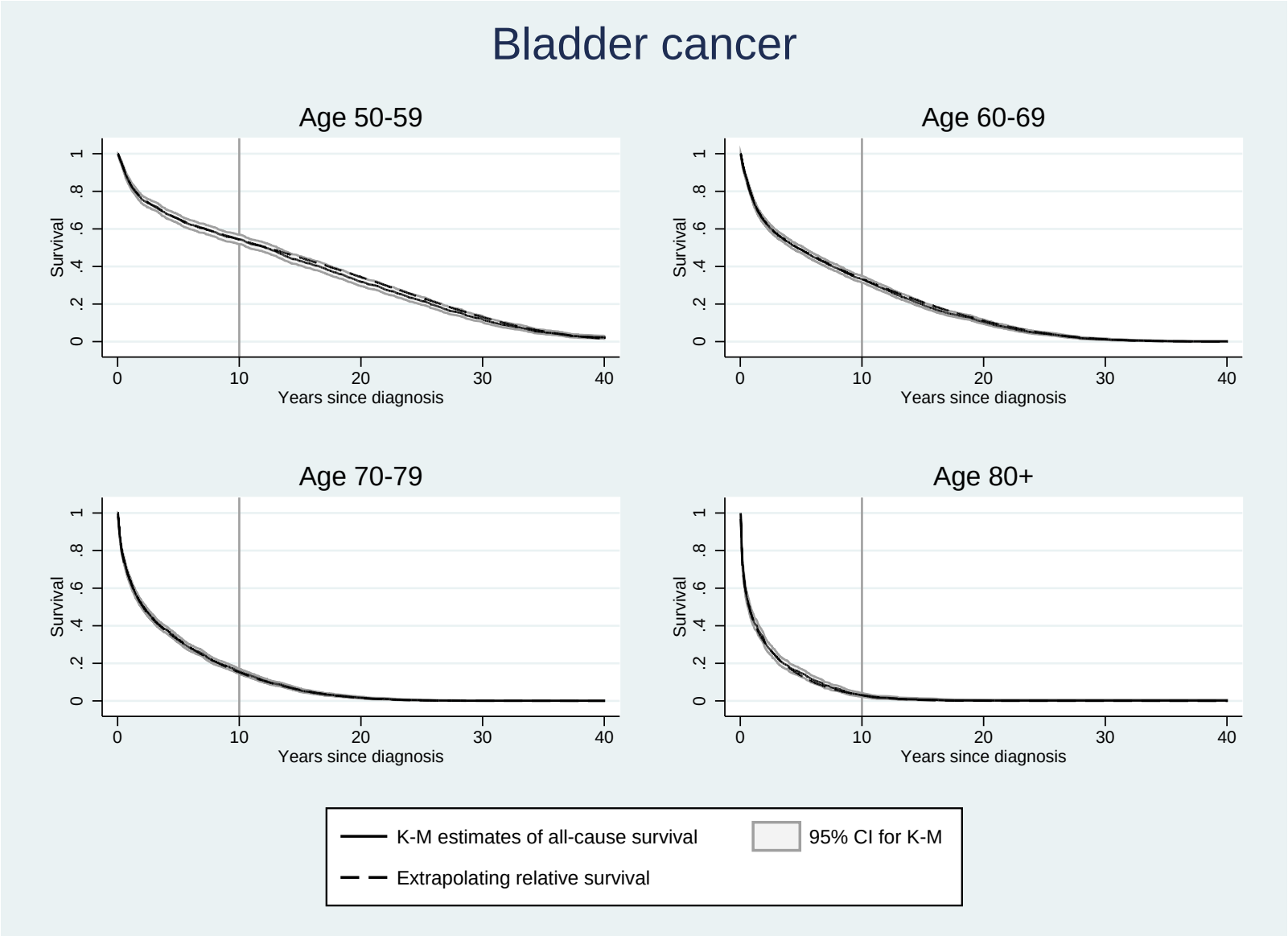
Breast Cancer



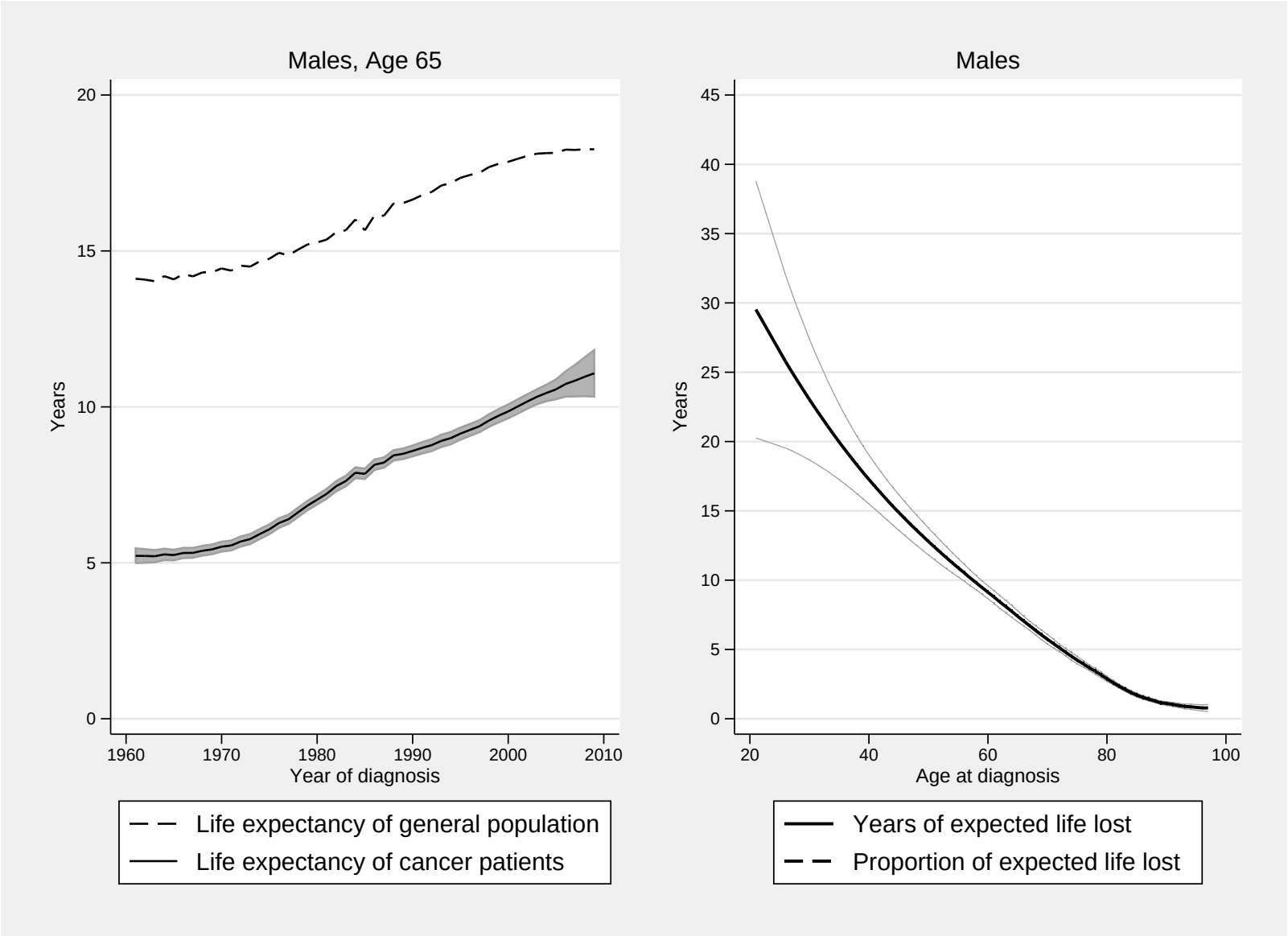
Melanoma


Melanoma
Age 50-59
Age 60-69
Age 70-79
Age 80+
Survival
0
.2
.4
.6
.8
1
0
10
20
30
40
Years since diagnosis
K-M estimates of all-cause survival
95% CI for K-M
Extrapolating relative survival

Bladder Cancer



Loss in Expectation of Life: Colon Cancer Sweden

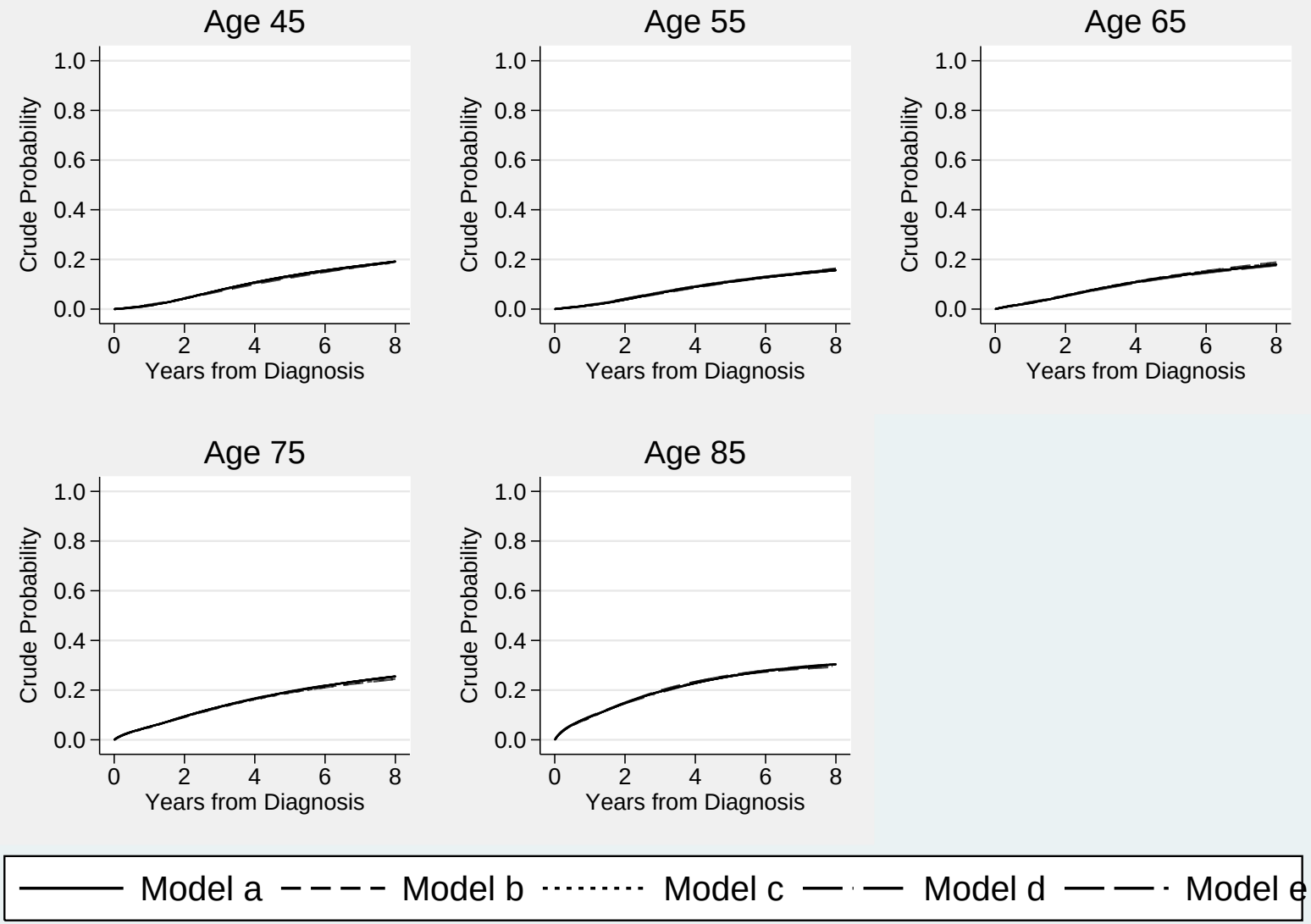


Sensitivity to the number of knots

- A potential criticism of these models is sensitivity to the number and location of the knots.
- We generally perform sensitivity analyses to convince ourselves and others that are conclusions are not sensitive to these choices.
- Below is a range an example of a range of models fitted in a large study (N=328,567) comparing breast cancer survival in England and Norway[9].

Model	Baseline df_b	Time-dependent df_t	age df_a	No. of Parameters	BIC
Model (a)	6	3	4	31	304640.8
Model (b)	5	2	3	21	304872.5
Model (c)	8	4	3	32	304461.6
Model (d)	8	5	8	57	304672.1
Model (e)	7	4	5	43	304607.6

Knot sensitivity analysis



Software

Log cumulative hazard scale

Stata - `stpm2`[17]

R - `Rstpm2`[a], `flexsurv`[b]

[a] `http://rstpm2.r-forge.r-project.org/`

[b] `http://cran.r-project.org/web/packages/flexsurv`

Log hazard scale

Stata - `stgenreg`[8]

Summary

- I use the Cox model and will continue to use it.
- However, I find certain problems easier to tackle parametrically.
- We need to improve the way we quantify what our model parameters mean at both the population and individual level. Generally need estimates of absolute rates/risks for this.
- Most of my applications are with fairly large data sets.
- 'Reasonable' choices of knots lead to very similar fitted values.
- Still some issues round model choice to be resolved.
- Parametric models particularly useful for extrapolation.

Acknowledgements

Paul Dickman	-	Karolinska Institutet
Patrick Royston	-	MRC Clinial Trials Unit, London
Therese Andersson	-	Karolinska Institutet
Michael Crowther	-	University of Leicester
Sandra Eloranta	-	Karolinska Institutet
Sally Hinchliffe	-	University of Leicester
Chris Nelson	-	University of Leicester
Mark Rutherford	-	University of Leicester
Hannah Twitchell	-	University of Leicester

References I

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

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

[3] 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.

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

[5] Carstensen B. Demography and epidemiology: Practical use of the Lexis diagram in the computer age or: Who needs the Cox model anyway? Tech. rep., Department of Biostatistics, University of Copenhagen, 2006.

[6] Rutherford MJ, Crowther MJ, Lambert PC. The use of restriced cubic splines to appoximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2013;**(Submitted)**.

[7] Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Statistics in Medicine* 2004;**23**:51–64.

[8] Crowther MJ, Lambert PC. stgenreg: A stata package for general parametric survival analysis. *Journal of Statistical Software* 2013;**(in press)**.

References II

[9] Lambert PC, Holmberg L, Sandin F, Bray F, Linklater KM, Purushotham A, *et al.*. Quantifying differences in breast cancer survival between England and Norway. *Cancer Epidemiology* 2011;**35**:526–533.

[10] Dickman PW, Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006;**260**:103–117.

[11] Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine* 2007; **26**:5486–5498.

[12] Bolard P, Quantin C, Esteve J, Faivre J, Abrahamowicz M. Modelling time-dependent hazard ratios in relative survival: application to colon cancer. *Journal of Clinical Epidemiology* 2001;**54**:986–996.

[13] Sasieni PD. Proportional excess hazards. *Biometrika* 1996;**83**:127–141.

[14] Pohar Perme M, Henderson R, Stare J. An approach to estimation in relative survival regression. *Biostatistics* 2009;**10**:136–146.

[15] Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine* 2010;**29**:885 – 895.

References III

[16] Andersson TL, Dickman P, Eloranta S, Lambe M, Lambert P. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine submitted* 2013;.

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