Statistics and epidemiology

Paul W Dickman

Professor of Biostatistics
Department of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm, Sweden

29 January 2020
Young Researchers in Hematology Symposium
Karolinska Institutet and Karolinska University Hospital

Overview: I'll touch on the following topics

- About me.
- Interpreting survival functions.
- Hazard functions and proportional hazards.
- Competing risks.
- Other measures of survival.
- Lots of pictures and very little math.
- Please interrupt!
- Slides: `http://pauldickman.com/#talks`

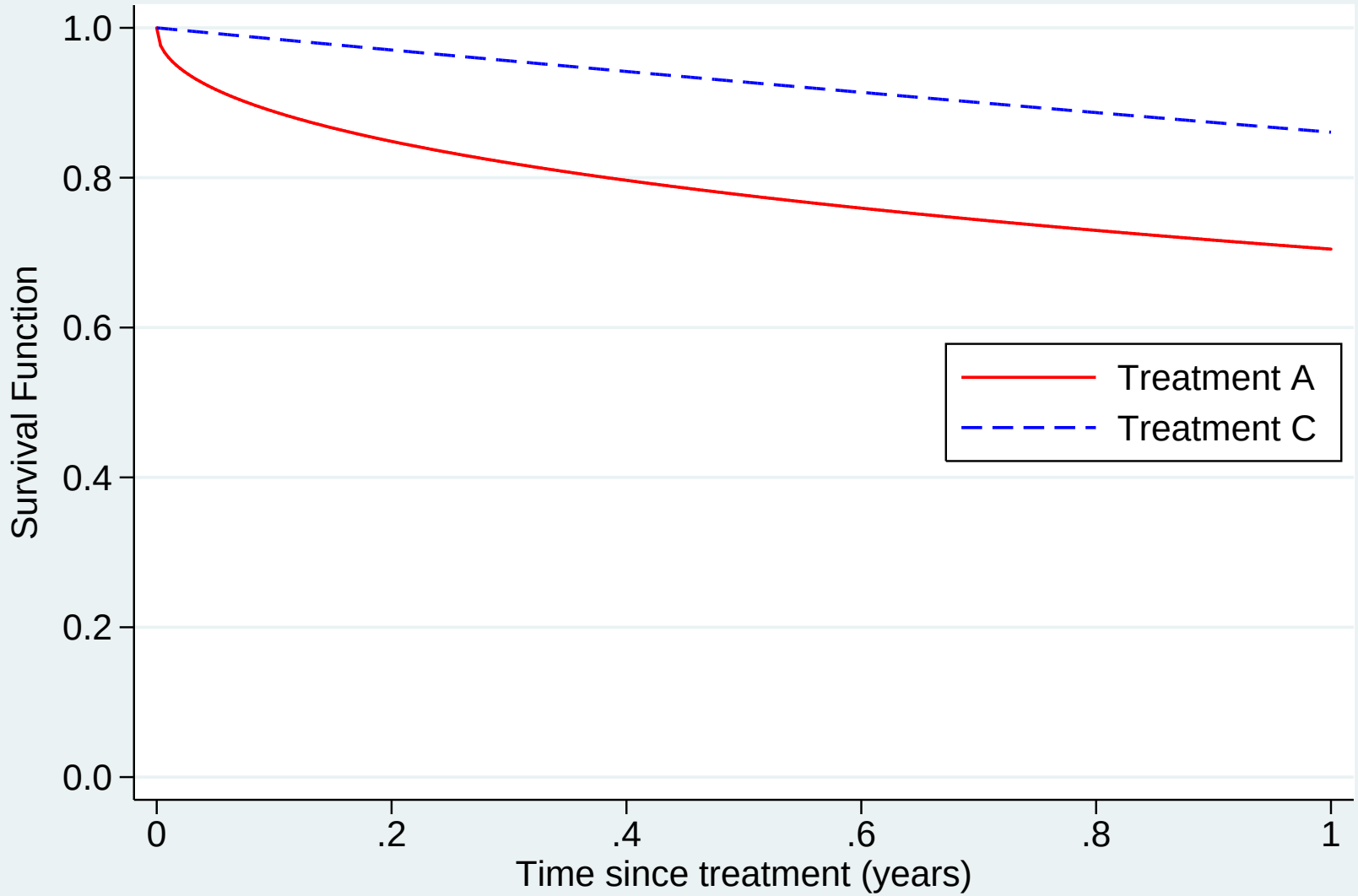
About me

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

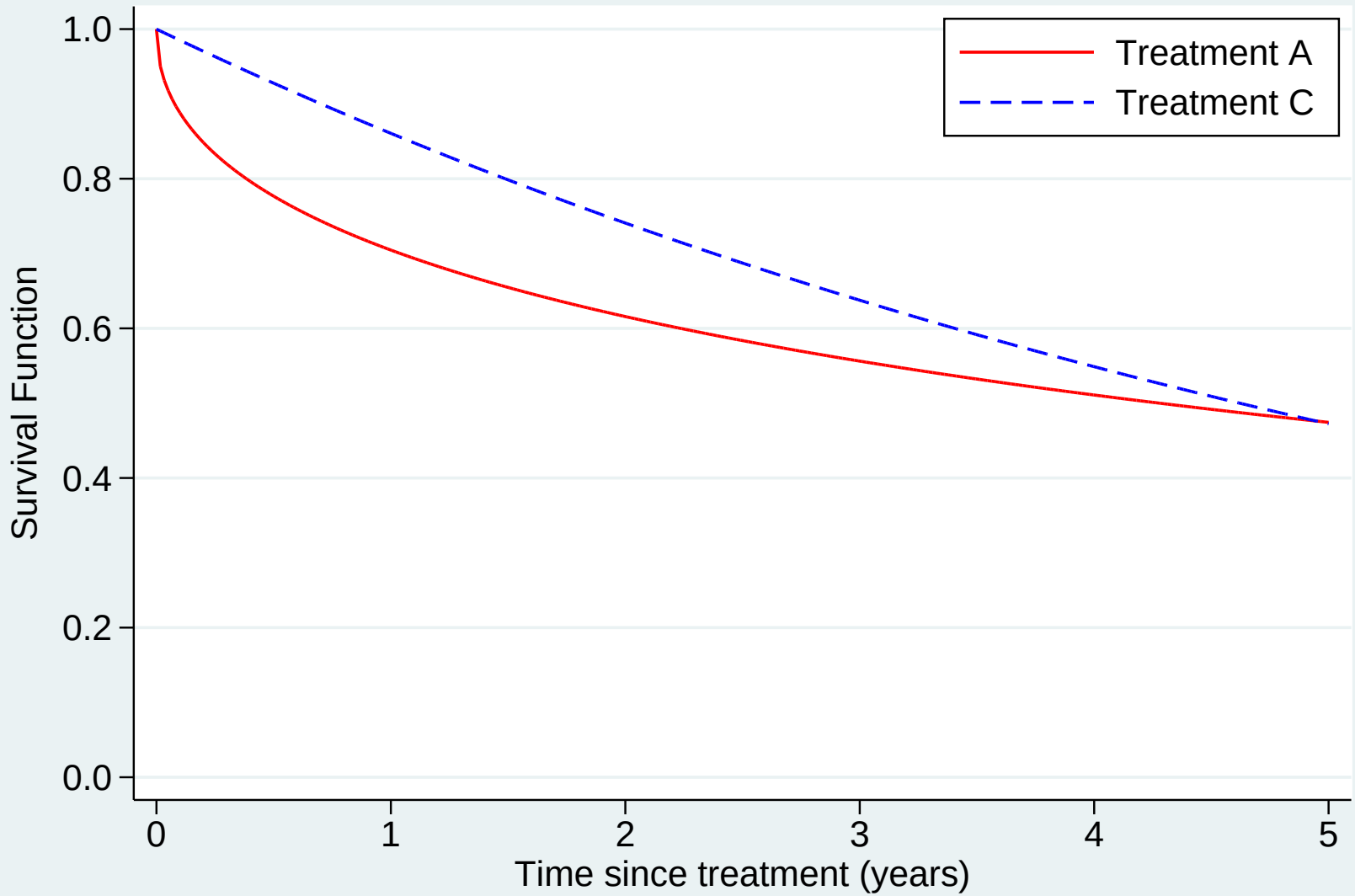
Current research interests

- Primary research interests are in the development and application of methods for population-based cancer survival analysis, particularly the estimation and modeling of relative survival.
- General interest in statistical aspects of the design, analysis, and reporting of epidemiological studies along with studies of disease aetiology, with particular focus on cancer epidemiology and perinatal/reproductive epidemiology.
- Collaborate closely with Paul Lambert (Biostatistician at University of Leicester) and Magnus Björkholm (Haematologist at KI/KS Solna).

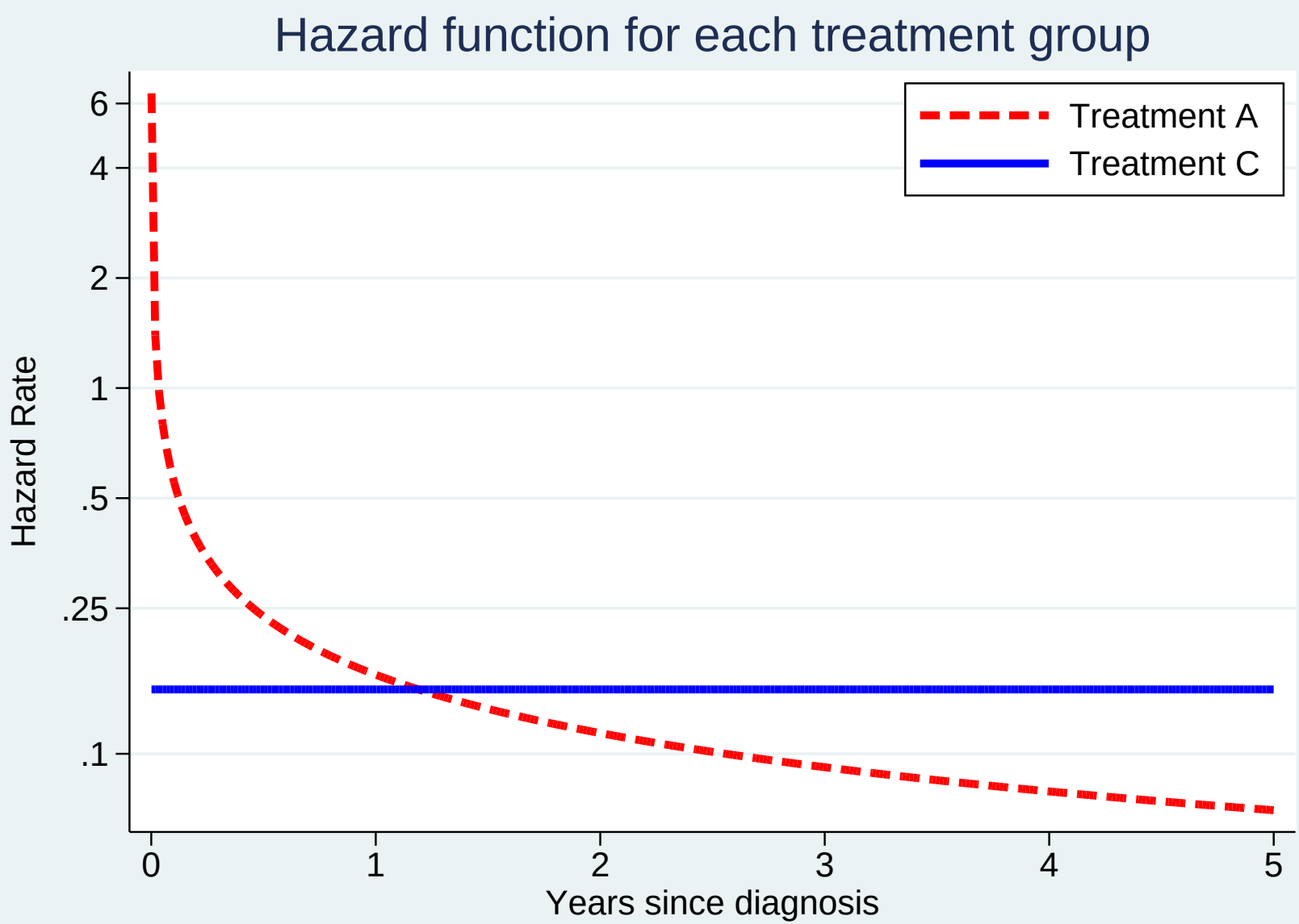
Which treatment (A or C) has the best survival?



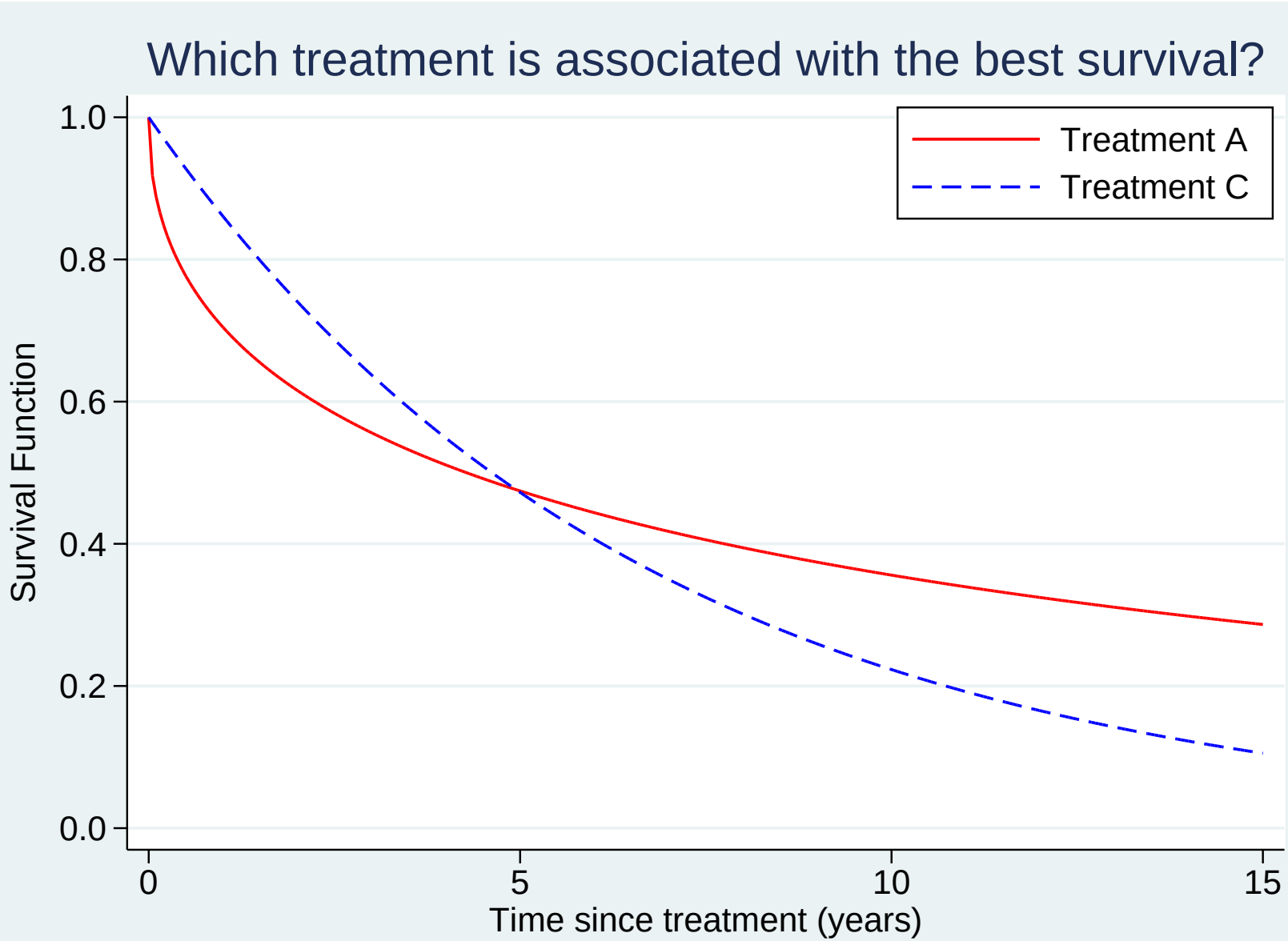
Which treatment (A or C) has the best survival?



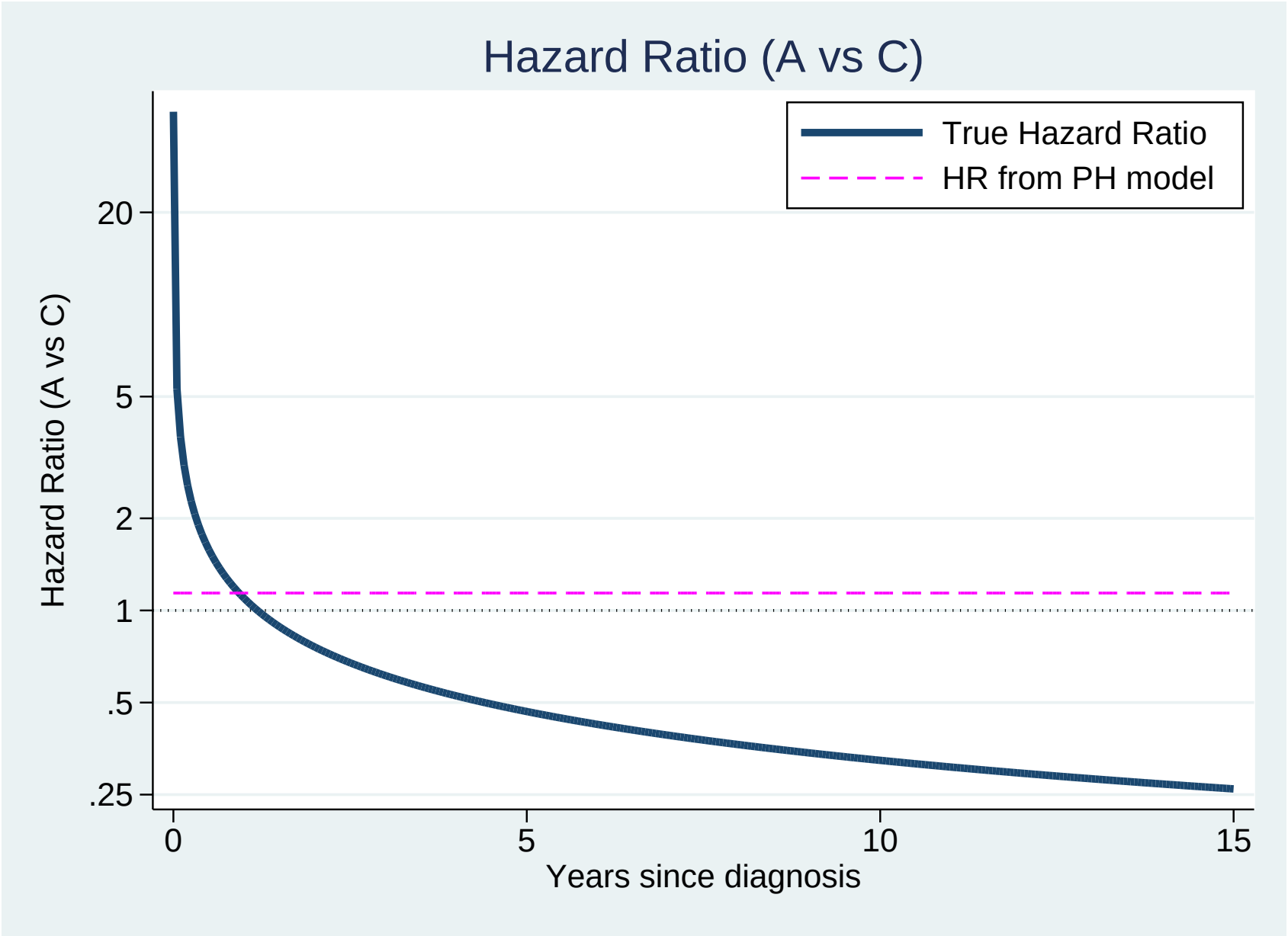
The two hazard functions



What about if we further extend the follow-up?



Time-varying hazard ratio for A vs C



Relation between the survivor and hazard functions

$$
\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} \frac{\Pr(\text{event in } (t, t+\Delta t] \mid \text{alive at } t)}{\Delta t} \\
&= \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{S(t) \times \Delta t} \text{ where } F(t) = 1 - S(t) \\
&= \lim_{\Delta t \to 0} \frac{S(t+\Delta t) - S(t)}{\Delta t} \times \frac{-1}{S(t)} \\
&= \frac{\mathrm{d}S(t)}{\mathrm{d}t} \times \frac{-1}{S(t)} \text{ by definition of a derivative} \\
&= -\frac{\mathrm{d} \ln S(t)}{\mathrm{d}t} \text{ since } d/dx \ln(f(x)) = f'(x)/f(x)
\end{aligned}
$$

What does this mean in practice?

- $h(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \ln S(t)$
- In practical terms, this means that the event rate is proportional to the rate at which the survival function decreases.
- That is, if the survival function is decreasing sharply with time then the mortality rate is high (and vice versa).
- If the survival function is flat then the hazard is zero (and vice versa).
- The derivative of a function at a point is the slope of the [tangent to the] curve at that point. A curve that is decreasing (like the survival function) has a negative slope, hence the negative sign in the formula above.
- We can think of the hazard as being proportional to the rate of change of $S(t)$.

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

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

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

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

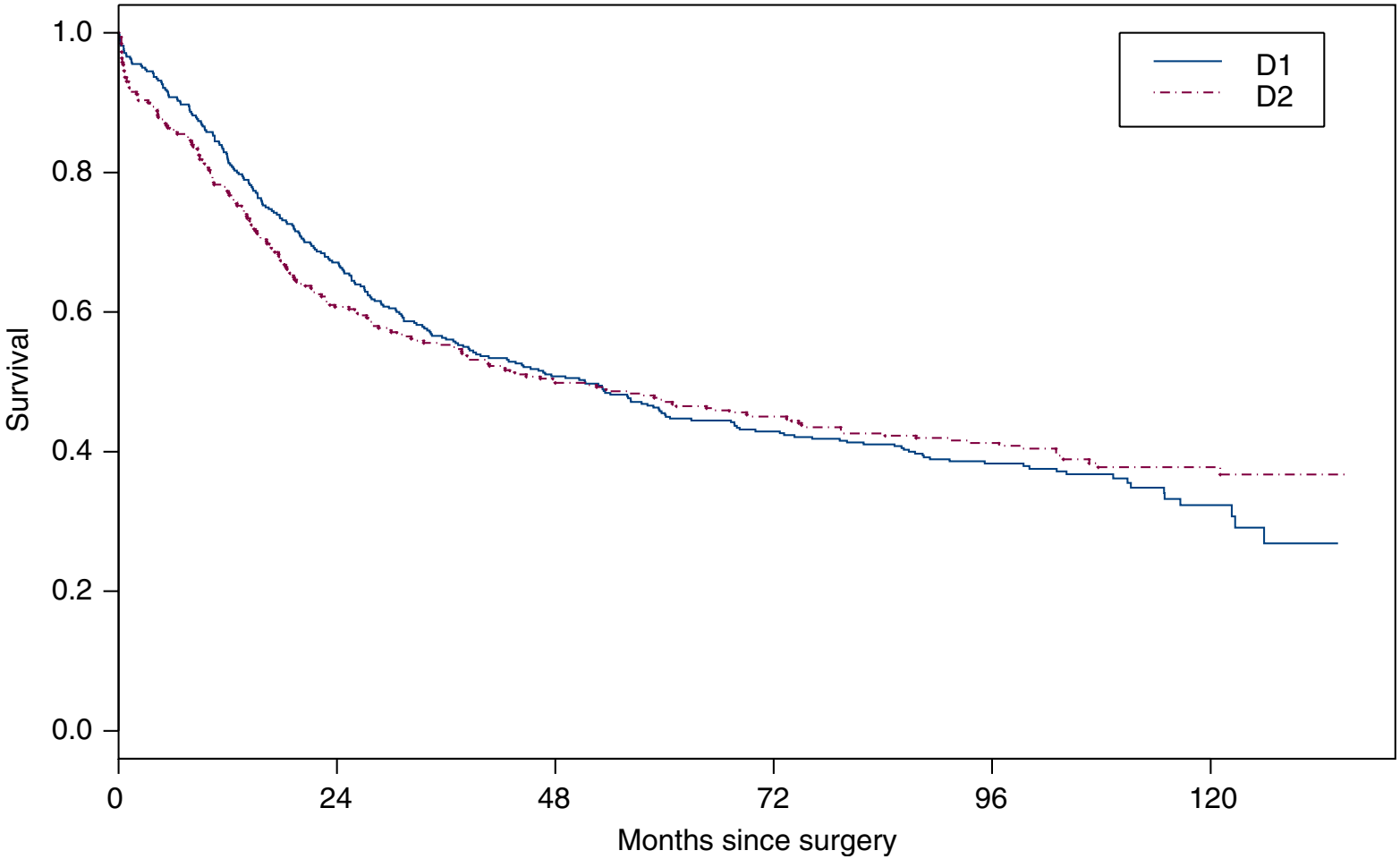


Figure 1. Kaplan–Meier plots of the survival curves for D1- and D2-dissection. The survival curves cross after 53 months.

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

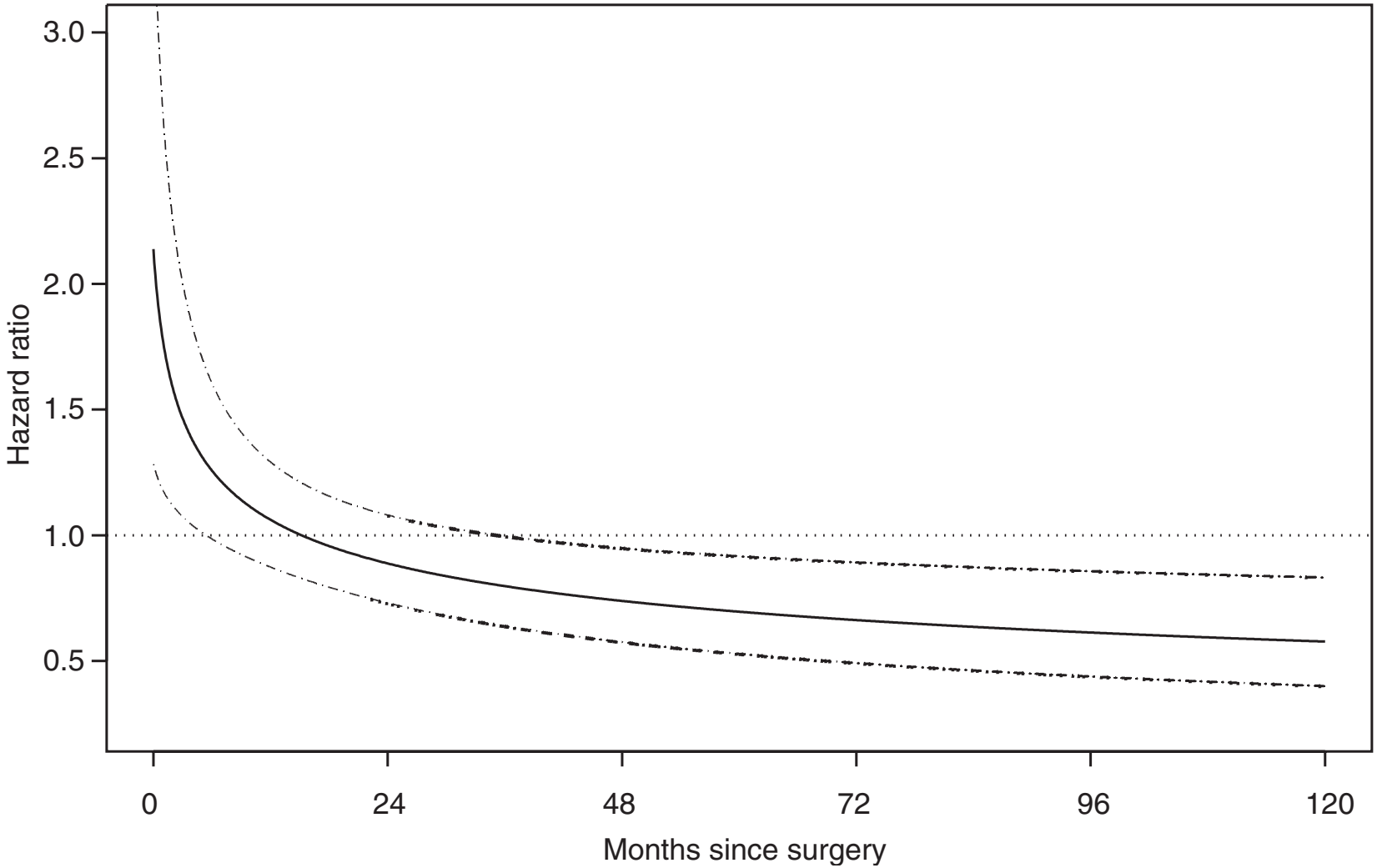
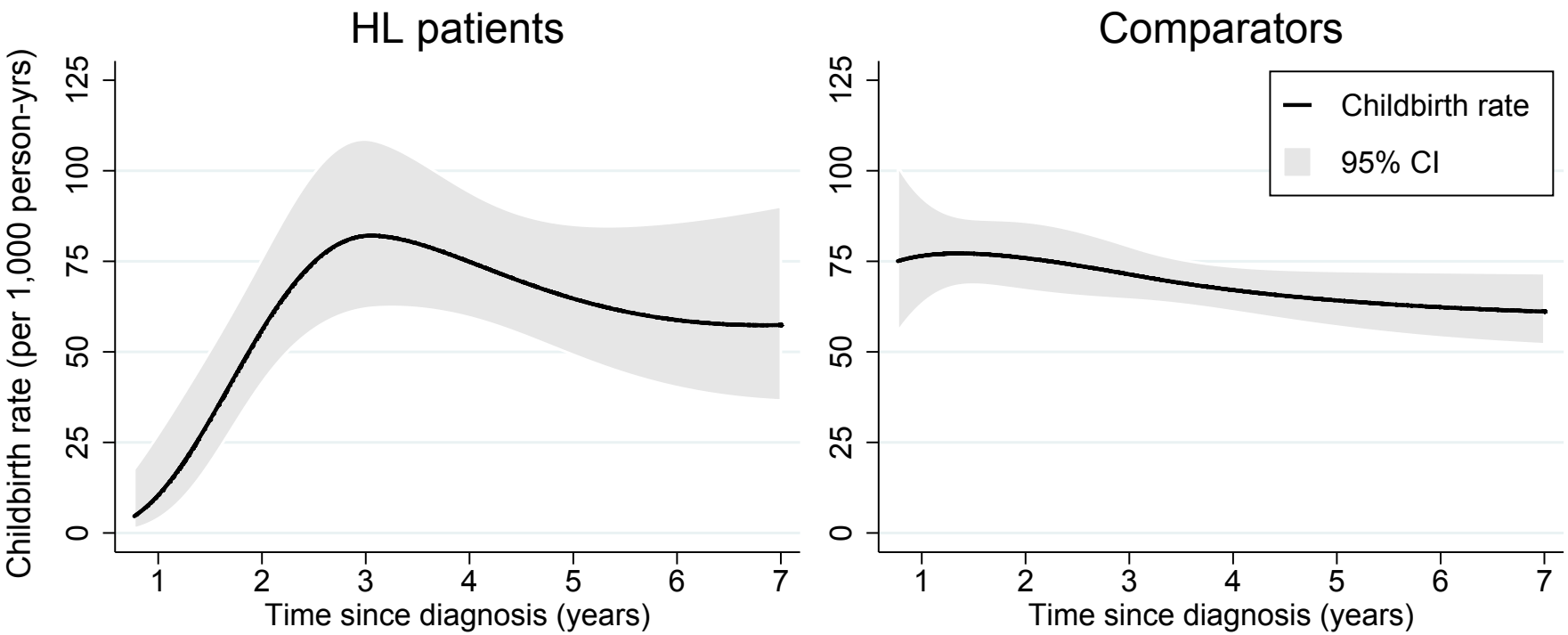


Figure 4. The estimated hazard ratio with 95 per cent confidence intervals based on Cox regression with treatment as time-dependent effect. A hazard ratio of one indicates equality of the hazard rates of D1 and D2.

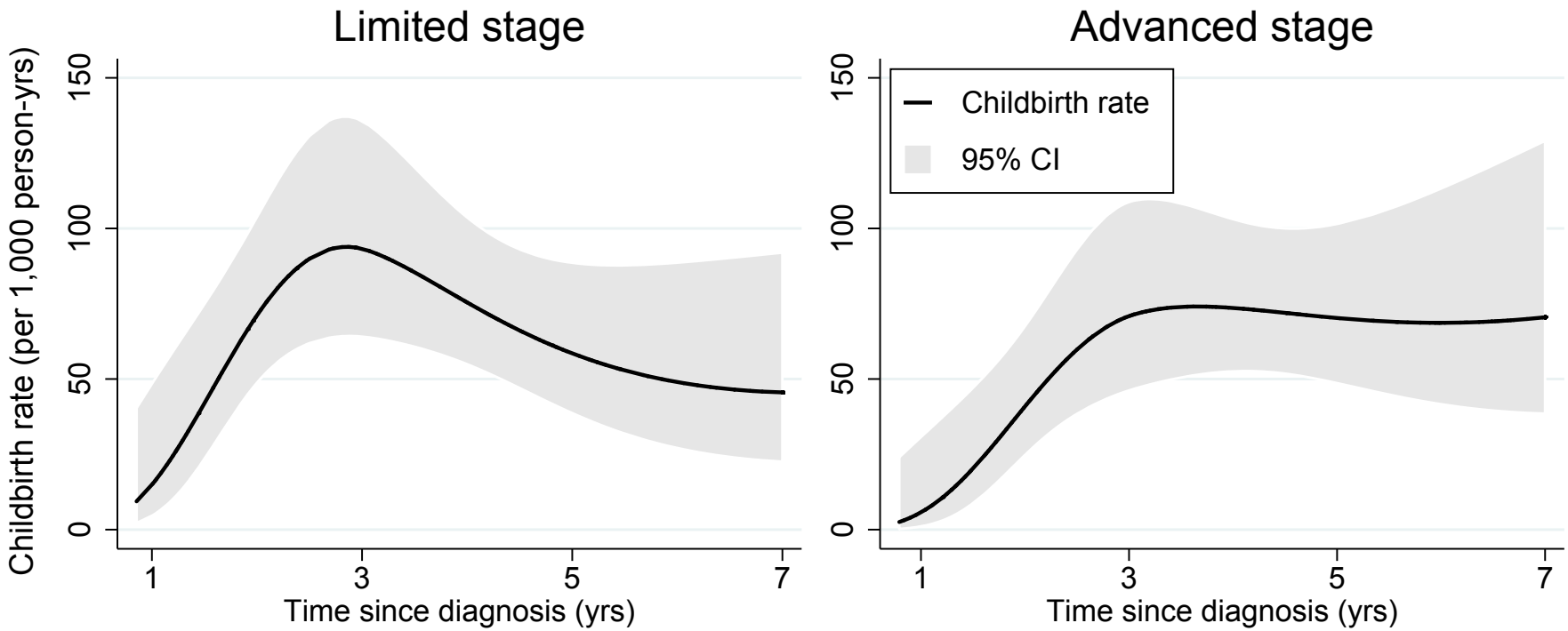
Childbirth rates among Hodgkin lymphoma survivors in Sweden (Weibull *et al.* 2018 [1])

Sometimes the hazard is a useful descriptive measure



Childbirth rates among Hodgkin lymphoma survivors in Sweden (Weibull *et al.* 2018 [1])

Sometimes the hazard is a useful descriptive measure



Chai-Adisaksopha *et al. BMC Hematology* (2016) 16:17
DOI 10.1186/s12878-016-0055-7

BMC Hematology

RESEARCH ARTICLE Open Access

A systematic review of using and reporting survival analyses in acute lymphoblastic leukemia literature

Chatree Chai-Adisaksopha[1,2], Alfonso Iorio[1,2*], Christopher Hillis[2,3], Wendy Lim[1] and Mark Crowther[1,2]

Conclusions: The use and reporting of survival analysis in adult ALL patients undergoing allo-SCT have significant limitations. Notably, the finding of crossing survival curves was common and none of the studies assessed for the proportional hazards assumption. We encourage authors, reviewers and editors to improve the quality of the use and reporting of survival analysis in the hematology literature.

What to do if you have non-proportional hazards

- Non-PH means you have different estimates of the HR at different points of time.
- Simply report the HR at selected time points (e.g., in a table) or a graph of the HR as a function of time.
- Disclaimer: assumes the HR is a sensible measure for your study design and research questiuon, you have fitted an appropriate model, and the differences in the HR are substantial (clinically and/or statistically).

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.

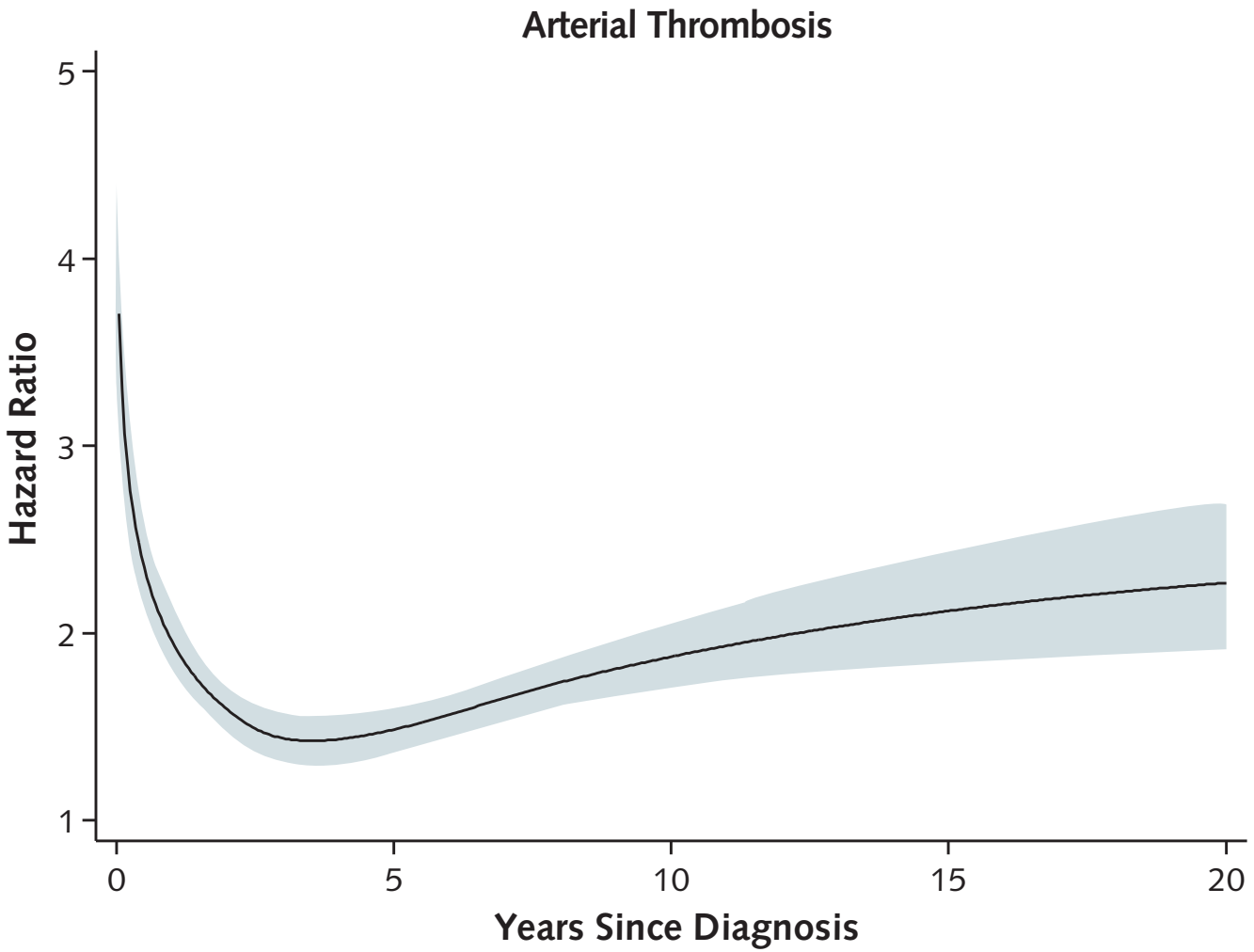
Ann Intern Med. 2018;168:317-325. doi:10.7326/M17-0028 **Annals.org**

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.



Introduction to competing risks

- Vital status might be coded as follows:

```
Freq.   Numeric  Label
 4642          0  Alive
 8369          1  Dead: disease/condition of interest
 2549          2  Dead: other causes
```

- We are typically interested in the probability of dying (or not dying) due to a specific disease/condition.
- Other events are known as 'competing events' or 'competing risks'.
- Based on the research question, we choose between one of two quantities to estimate:
  1. *Eliminate* the competing events (estimate net survival)
  2. *Accommodate* the competing events (estimate crude survival)

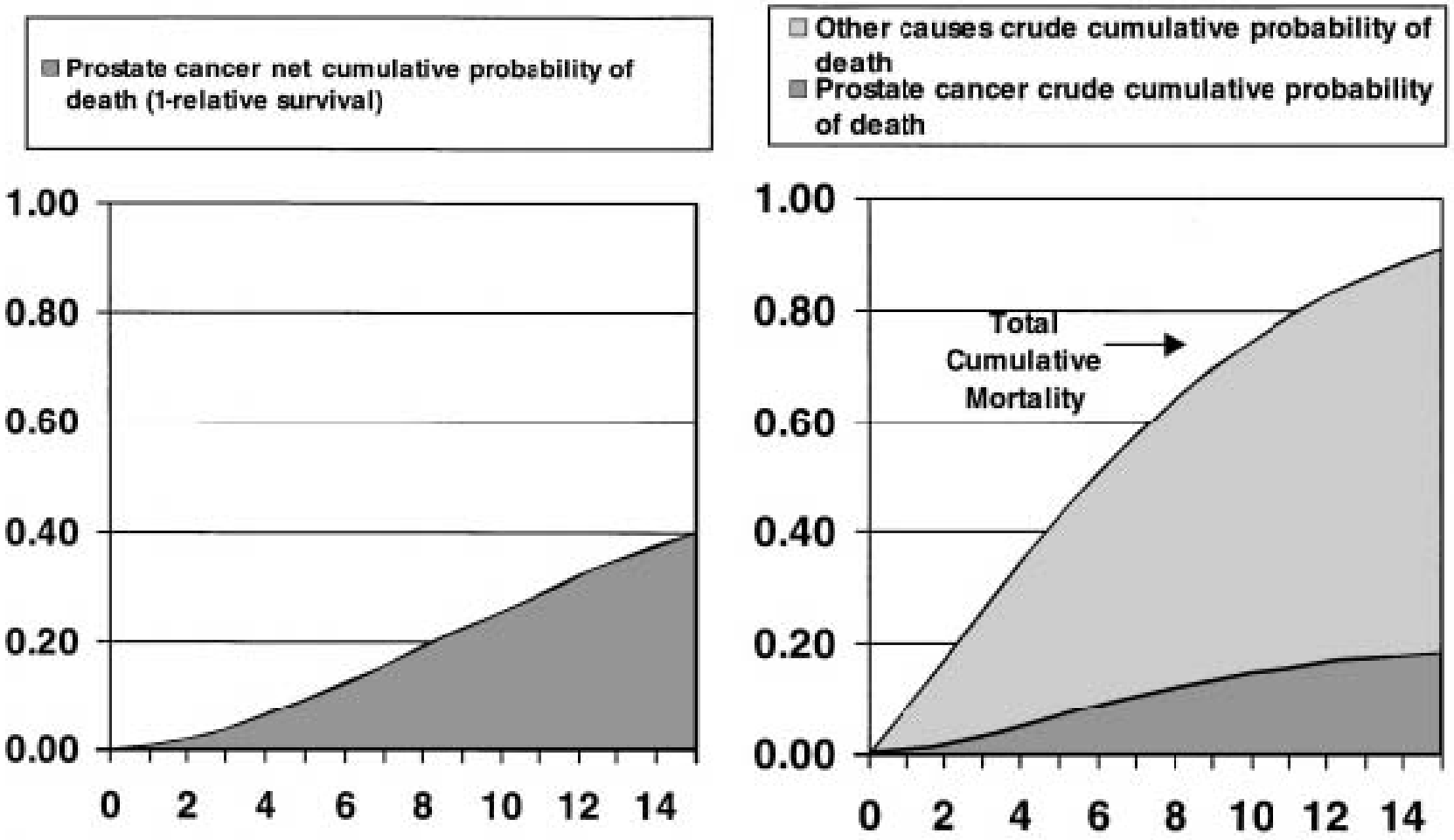
We have a choice of two measures

Net probability of death due to cancer	=	Probability of death in a hypothetical world where the cancer under study is the only possible cause of death

Crude probability of death due to cancer	=	Probability of death in the real world where you may die of other causes before the cancer kills you

- Net probability also known as the marginal probability.
- Crude probability also known as cumulative incidence function.

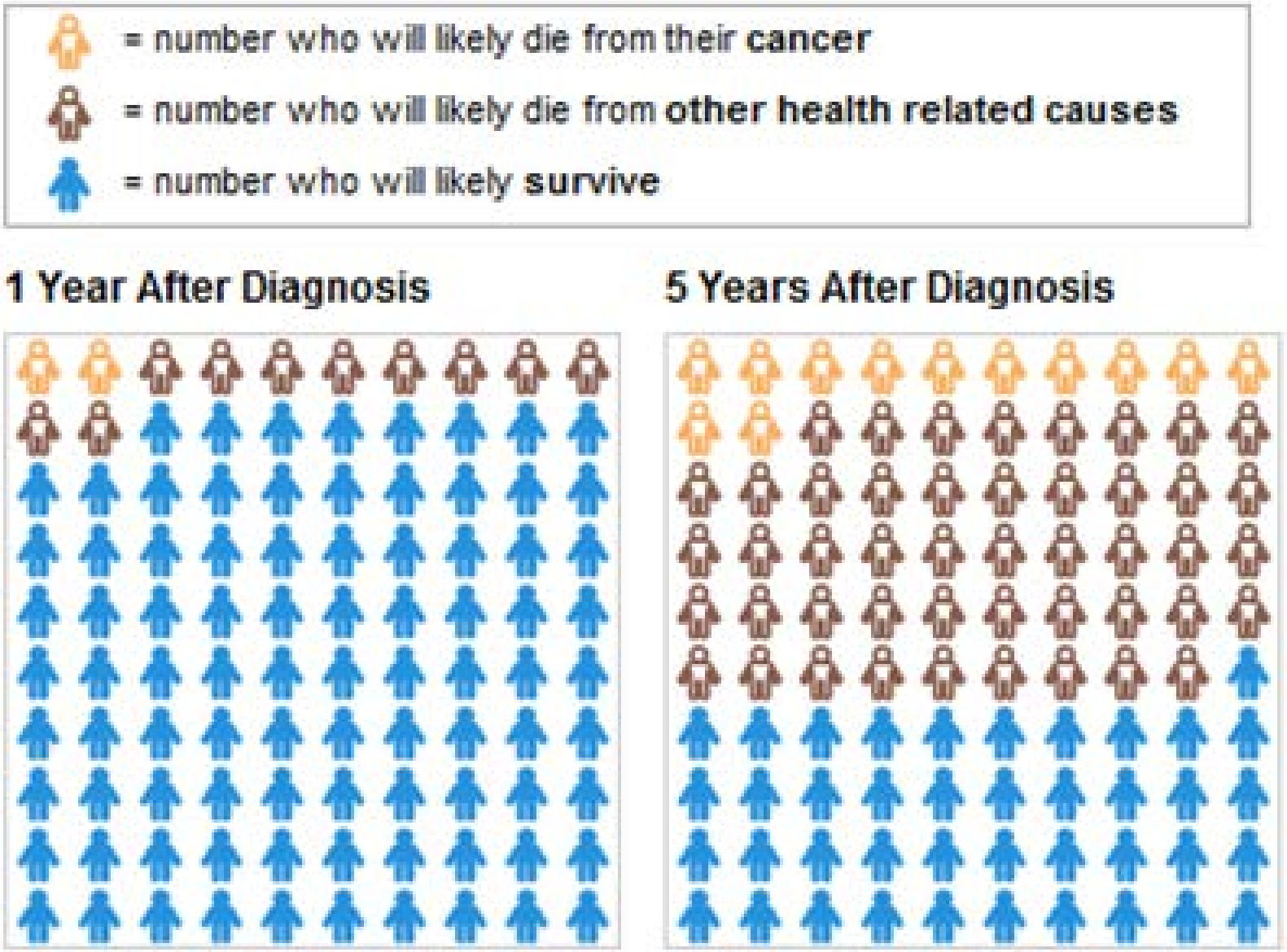
Net (left) and crude (right) probabilities of death in men with localized prostate cancer aged 70+ at diagnosis (Cronin and Feuer [2])



Choose the measure most appropriate for your research question!

- Is survival of cancer patients improving over time (due to better management)? Hypothetical world.
- How many patients will survive and require continued care? Real world.
- How does cancer patient survival in Sweden compare to other countries (with a view to comparing effectiveness of healthcare systems)?
- Do beta blockers causally affect the probability of surviving cancer? Observational study comparing the survival of patients with a cancer diagnosis who use beta blockers compared to those who do not.

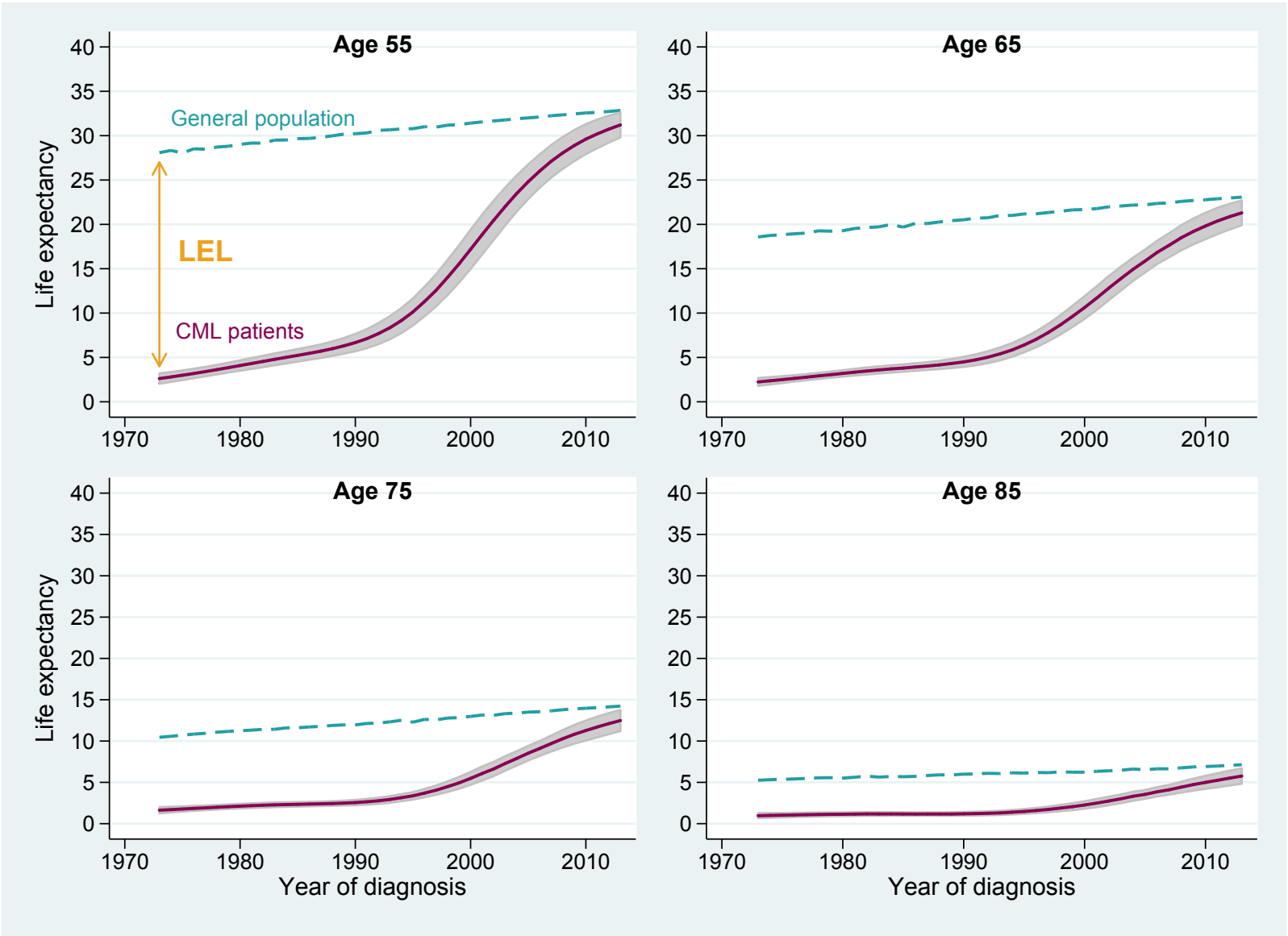
Natural frequencies presented using infographics



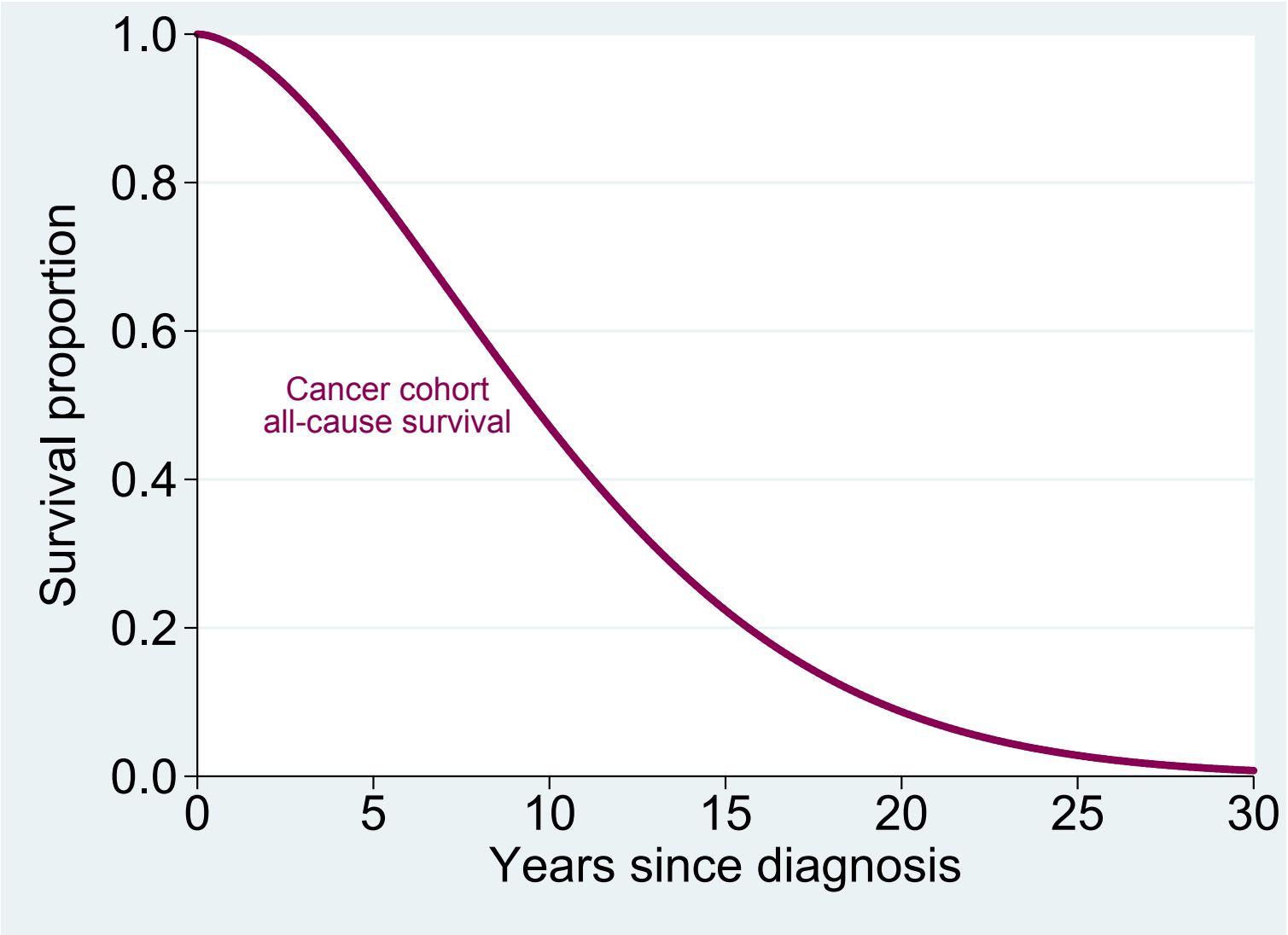
Impact of a cancer diagnosis on life expectancy

- We have promoted a number of alternative metrics in population-based cancer studies. One of which is reduction in life expectancy associated with a diagnosis of cancer.
- Extrapolation to end of life is needed for this.
- We know a lot about how mortality rates vary by demographic factors. We can utilise this external information to help with our extrapolation.

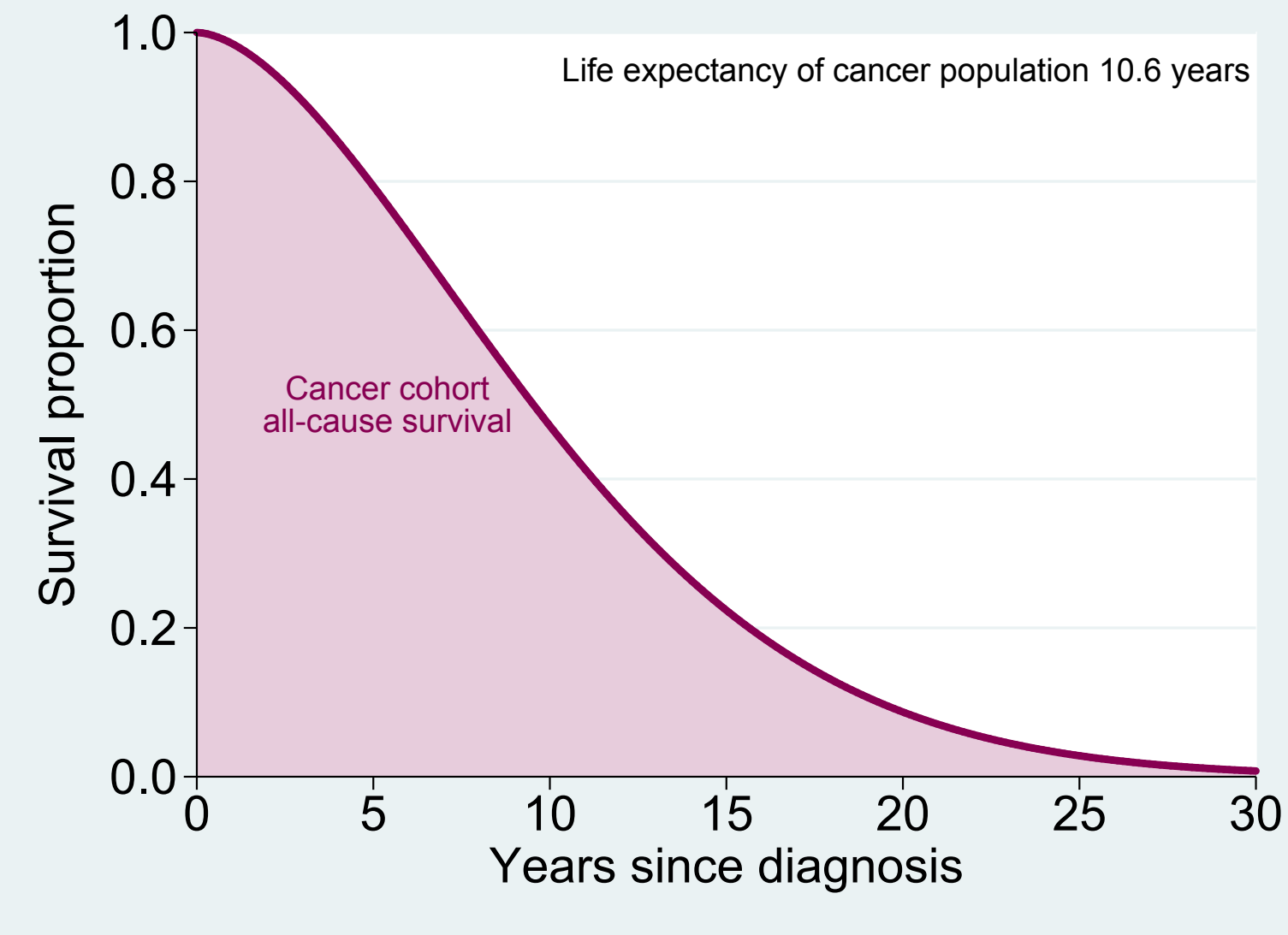
Loss in expectation of life: CML (Sweden) [3]



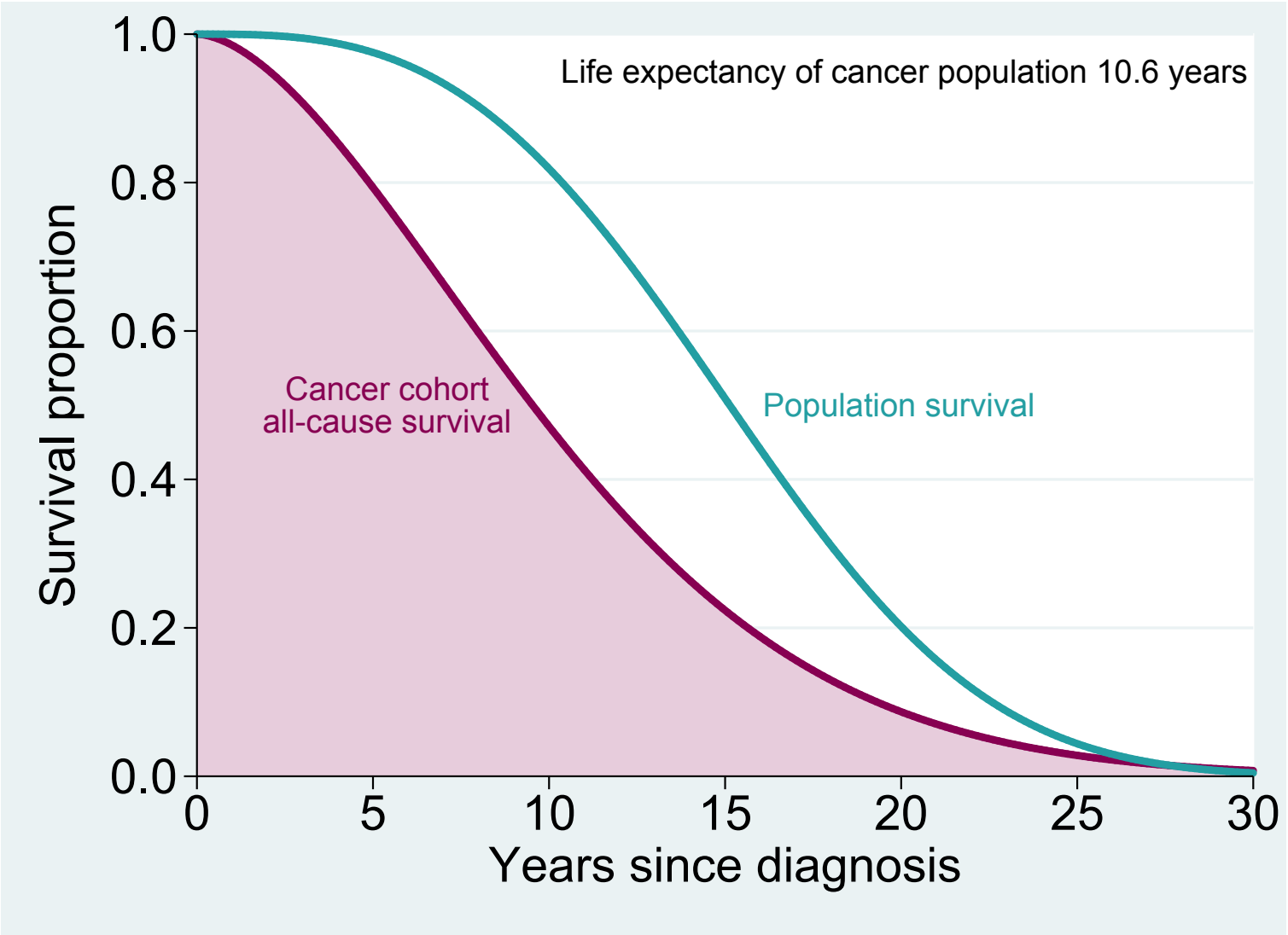
Estimating expectation of life



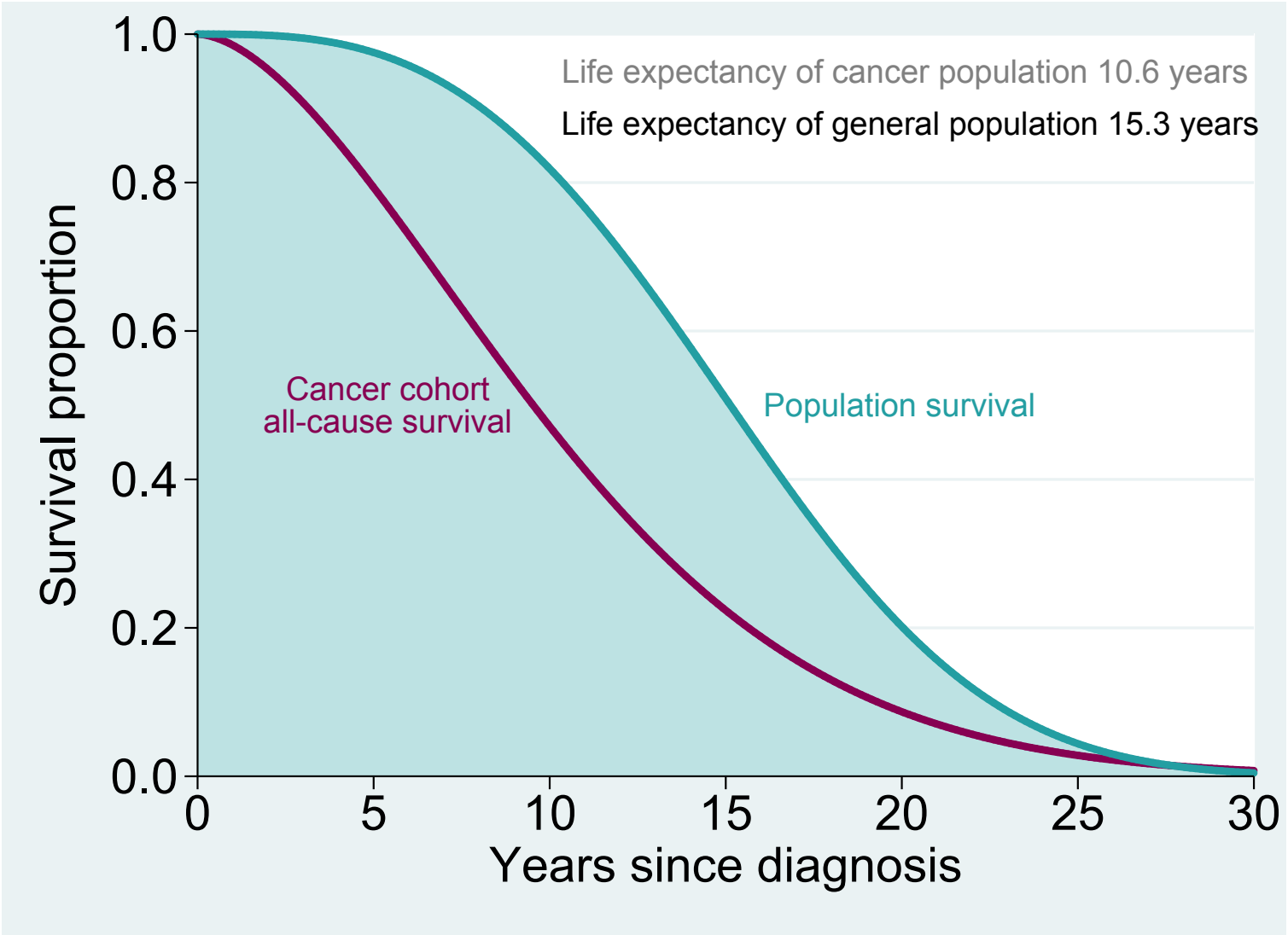
Estimating expectation of life



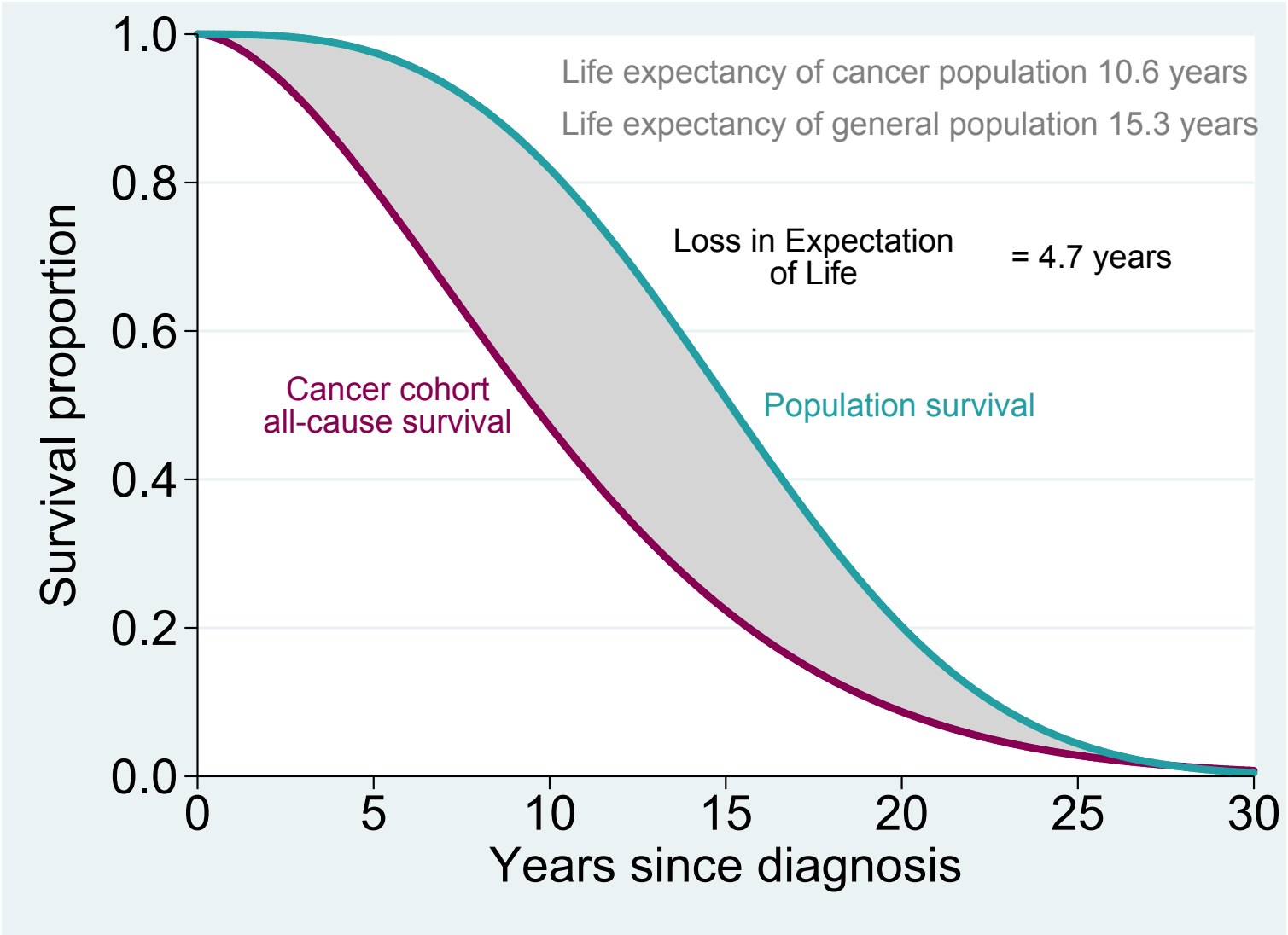
Estimating expectation of life



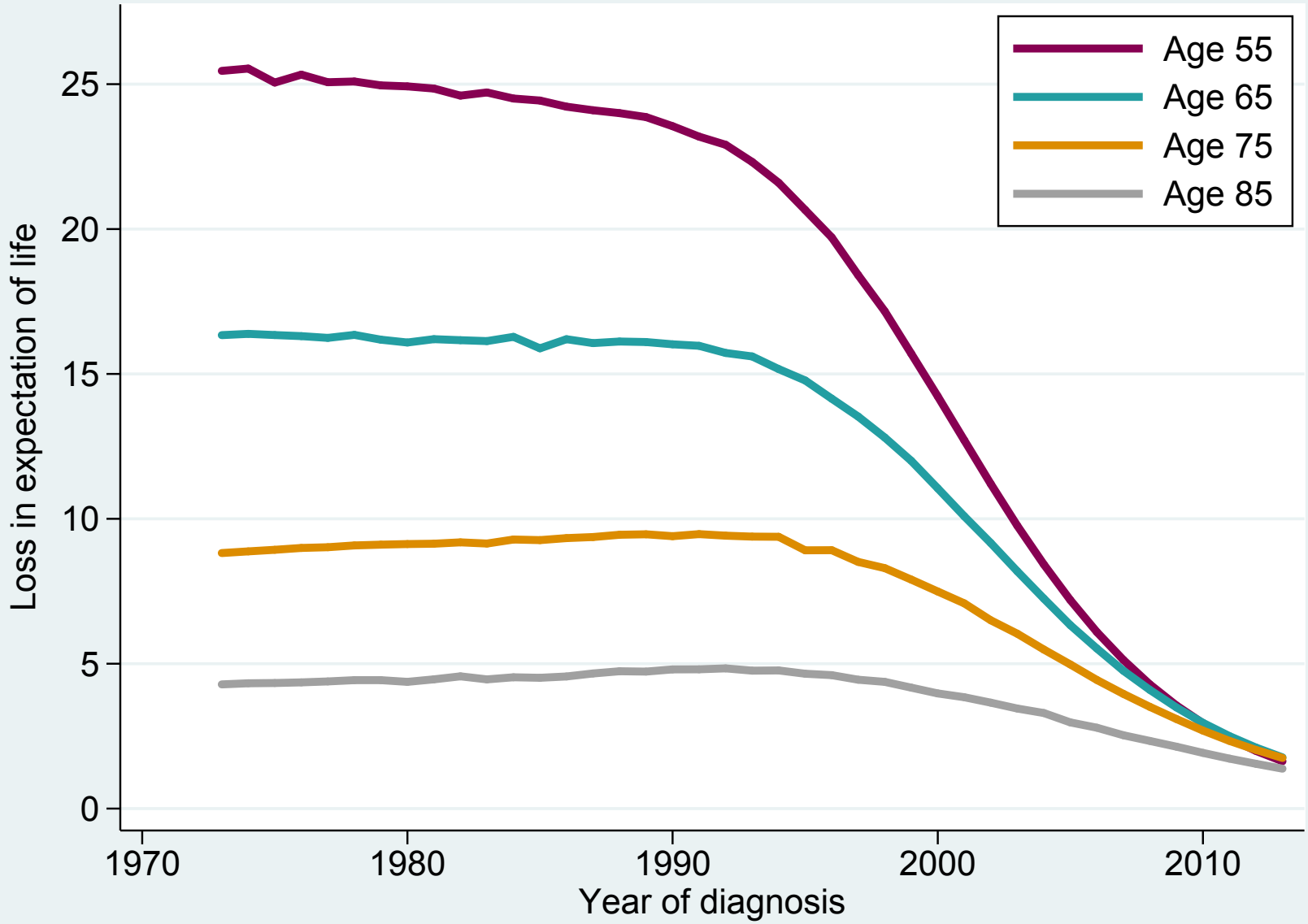
Estimating expectation of life



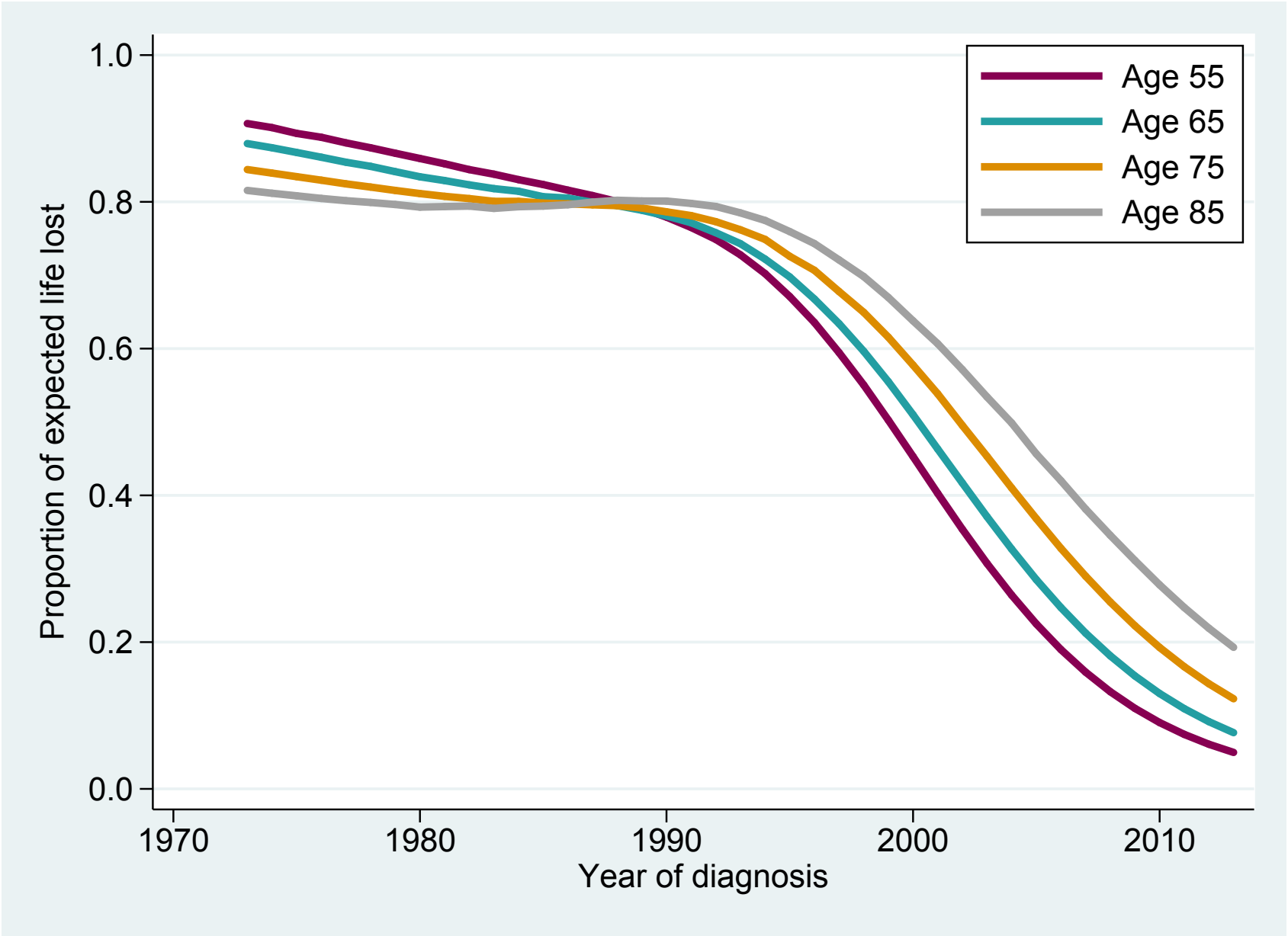
Loss in expectation of life



Loss in expectation of life: CML (Sweden)



Proportion of expected life lost: CML (Sweden)



Course at summer school in Italy, 1–6 June 2020

References

[1] Weibull CE, Johansson ALV, Eloranta S, Smedby KE, Björkholm M, Lambert PC, *et al.*. Contemporarily treated patients with Hodgkin lymphoma have childbearing potential in line with matched comparators. *Journal of Clinical Oncology* 2018;**36**:2718–2725.

[2] Cronin KA, Feuer EJ. Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Statistics in Medicine* 2000; **19**:1729–1740.

[3] 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.