# An overview and some recent advances in statistical methods for population-based cancer survival analysis

Paul W Dickman[1] Paul C Lambert[1,2]

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

24 September 2014
Dept. of Statistics, University of Uppsala


## Overview of today's talk

- About me and my department.
- The state of epidemiology and biostatistics as research disciplines.
- Measures used in cancer control; why study patient survival.
- Intro to relative survival (excess mortality) and why it is the measure of choice for population-based cancer survival analysis.
- Flexible parametric models (Royston-Parmar models).



## 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 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.
- Not a mathematical statistician.



## I found paradise! [1]

**A paradise for epidemiologists?**

*Hans-Olov Adami* The Lancet 1996;2:588

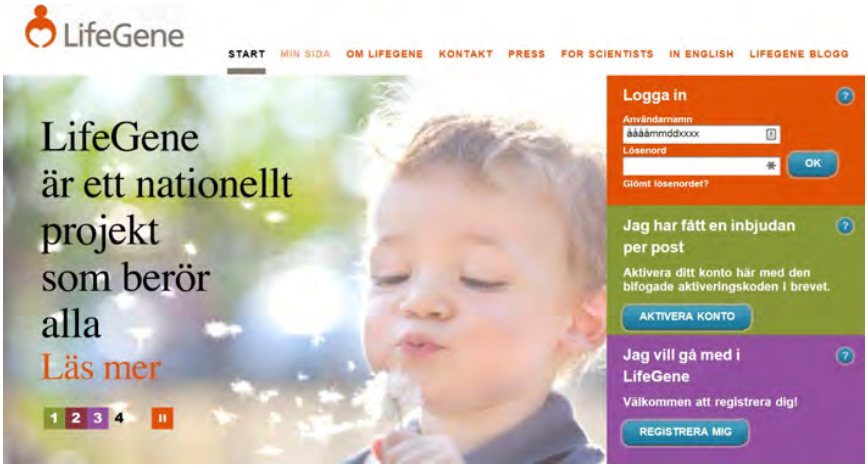
For three reasons—the structure of its health system, the existence of nationwide registers, and the systematic use of national registration numbers—Sweden offers exceptional opportunities for epidemiological research.

- I would add 'willingness of the public to contribute to research'.

## We have subsequently met some of the challenges

The final challenge is to initiate large-scale prospective studies. In this respect, Sweden has so far shown a lack of vision and foresight. We have no counterparts to the richly productive cohort studies initiated as early as the 1950s in the UK and later the USA, notably among doctors, nurses, and other health professionals. A large obstacle is the economic factor. In Sweden we have no funding mechanism to cover the costs of initiating a sizeable prospective study. Longstanding barriers need to be crossed—those between disciplines, between funding bodies, and between laboratory science and observational research. We eagerly await initiatives from the major funding agencies to take us beyond the tradition of small-scale thinking, provincial passivity, and self-sufficiency.








# MEB recruits two international top scientists

*Updated on 2014-06-27. Published on 2014-05-16* Denna sida på svenska

The Swedish Research Council has decided to grant a total of 522 million SEK over ten years for international recruitment of leading researchers to Karolinska Institutet. Two of these research centers of international top class will be located at the Department of Medical Epidemiology and Biostatistics (MEB) under the leadership of Patrick Sullivan, professor in psychiatric genetics and Cynthia Bulik, professor in eating disorders.



## Little progress in other areas

Evidently, Swedish epidemiologists have exceptional opportunities to add to knowledge in many areas of medicine. They do, however, face some important challenges. The large nationwide registers need to be better characterised—for example, with regard to completeness, agreement with hospital records, validity of diagnoses, and changes in registration over time. This sort of quality control will be much aided by the Epidemiology Centre recently established within the National Board of Health and Welfare, responsible for the maintenance of several registers.



## New challenges arise

## New challenges arise

Eur J Epidemiol (2014) 29:227–230
DOI 10.1007/s10654-014-9909-0

COMMENTARY

**The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research**

**Olof Nyrén · Magnus Stenbeck · Henrik Grönberg**



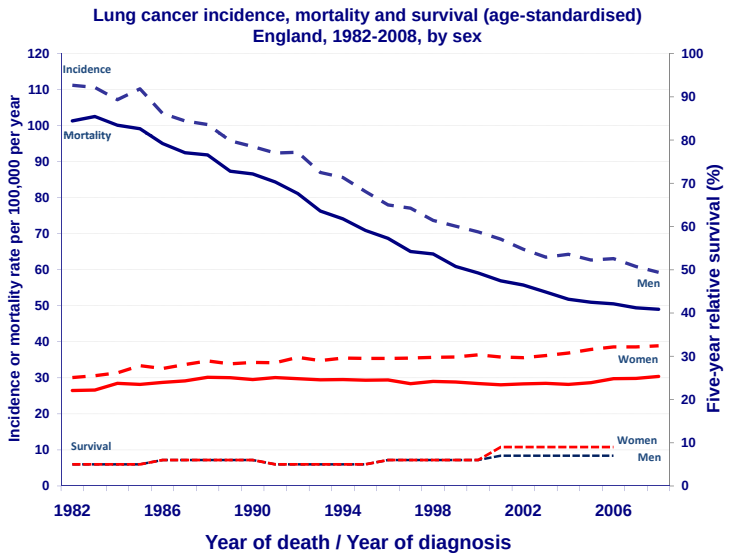
## About MEB (Dept. of Medical Epidemiology and Biostatistics)

- Established at KI 1997 as Medical Epidemiology (MEP) when Hans-Olov Adami moved the Department of Cancer Epidemiology from UU. Has grown from 40 to 300 FTE staff.
- Hans-Olov Adami was/is a very strong supporter of biostatistics and invested in developing biostatistics. Collaboration with Reinhold Bergström (and others) instilled Hans-Olov with a deep respect for biostatistics/biostatisticians.
- Today, 4 Professors and 3 Associate Professors (lektorer) of biostatistics at MEB (all foreign).
- Name change in 2002 (adding 'and Biostatistics').



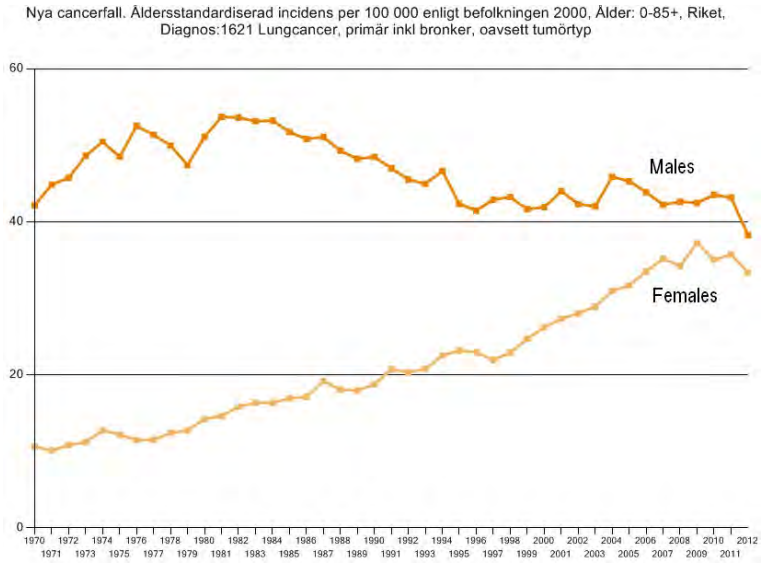
## My research interests

- 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.
- Recent interest has been in presenting information on patient survival in a manner relevant for patients and clinicians.
- 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).






# Lung cancer incidence in Sweden






# International comparisons of survival are hot!

## HOW EUROPE COMPARES: THE FIVE-YEAR SURVIVAL RATES

| | Women* | Men* | | Women* | Men* |
|---|---|---|---|---|---|
| Iceland | 58.2 | 48.5 | Spain | 55.3 | 44.9 |
| Sweden | 57.9 | 46.4 | Portugal | 54.9 | 45.6 |
| Italy | 57.5 | 47.6 | Netherlands | 54.8 | 45.7 |
| Finland | 56.9 | 46.2 | Denmark | 53.3 | 36.7 |
| Switzerland | 56.6 | 48.3 | Ireland | 51.4 | 42 |
| France | 56.6 | 45.5 | UK | 51.4 | 41.4 |
| Belgium | 56.3 | 48.1 | Poland | 49.8 | 39.4 |
| Norway | 55.8 | 43.2 | Czech Rep. | 49.7 | 37.2 |
| Austria | 55.7 | 47.6 | Slovenia | 49.4 | 36.5 |
| Germany | 55.5 | 47.4 | AVERAGE | 54.6 | 44.8 |

**Percentage of patients who survived these cancers after five years****

| LUNG | | PROSTATE | |
|---|---|---|---|
| England | 8.4 | England | 69.7 |
| Euro average | 12.2 | Euro avg | 77.7 |
| **BOWEL** | | **BREAST** | |
| England | 49.9 | England | 77.3 |
| Euro avg | 54.3 | Euro avg | 81.6 |

All Figures %. *Countries ranked on female survival rates **Figures are for England only

# Not everyone is a fan of cancer survival [2]

## UK cancer survival statistics

### Are misleading and make survival look worse than it is

**RESEARCH, p 335**

**Valerie Beral** professor of epidemiology, Cancer Epidemiology Unit, University of Oxford, Oxford OX3 7LF
**pa.valerie.beral@ceu.ox.ac.uk**
**Richard Peto** professor of medical statistics and epidemiology, Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU), University of Oxford, Oxford OX3 7LF

**Competing interests:** Both authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from either

In the linked article, Autier and colleagues report that (population based) breast cancer mortality rates have fallen over the past two decades in many European countries, with a greater decline in the United Kingdom than in any other large country.[1] That the UK is leading Europe in the speed with which national breast cancer mortality rates are falling is in stark contrast to, and at first sight difficult to reconcile with, claims that survival after breast cancer onset is worse in the UK than elsewhere in western Europe.[2]

The unpromising UK cancer survival estimates are, however, misleading. In contrast, population based mortality trends are reasonably reliable (at least in middle age, for example, people aged 35-69 years) because a death certificate is legally required before someone can be buried

vival calculations based on re
survival rates seem significantl

Information in cancer regis
is virtually complete because
mentions cancer is automatica
registries that, between them,
then registered, and further info
successfully) from medical rec
for decades played an importan
identify people with cancer. Wi
tion, many such cancers could
it, many people who die of can
than the death certificate ever t
certificate only" cases) or may h



# June 2011 – Response to Beral & Peto by Coleman *et al.* (including Dickman & Lambert)

- 'This editorial is unfounded, untenable and inconsistent. It has elicited critical responses (refs) but the BMJ editor reports the authors were too busy to defend it. In fact, the editorial is indefensible. We suggest you withdraw it.'
- 'The editorial is unfounded. The provocative title "UK cancer survival statistics are misleading and make survival look worse than it is" is pure conjecture. The article slides directly from assertion to conclusion, with no evidence in between.'
- 'The editorial thus undermines research designed to explain the UK cancer survival deficit, as well as policy designed to reduce the deficit. That is a disservice to cancer patients in the UK.'



# Beral & Peto arguments debunked [3]

BJC

**British Journal of Cancer** (2013) 108, 691–698 | doi: 10.1038/bjc.2013.12

**Keywords:** cancer registration; registration errors; relative survival; population-based data; simulation

## A comprehensive assessment of the impact of errors in the cancer registration process on 1- and 5-year relative survival estimates

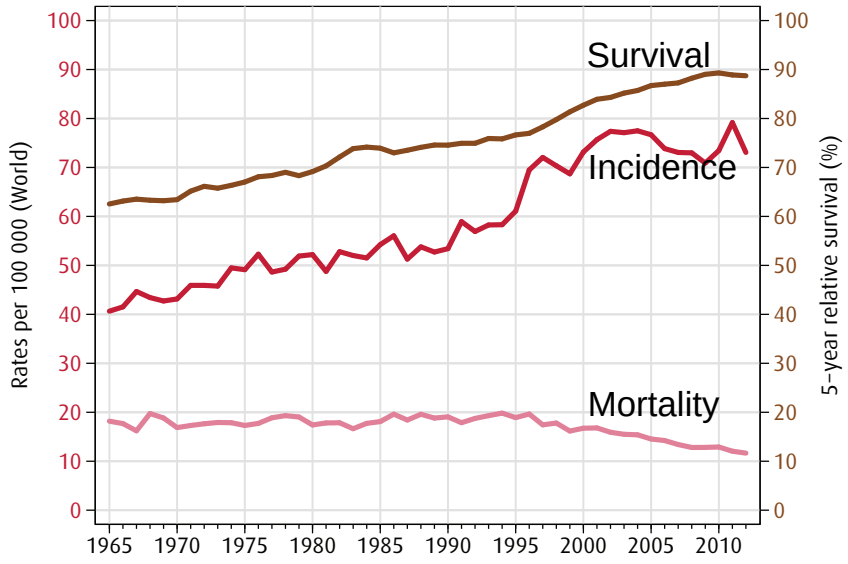
M J Rutherford*[,1], H Møller[2] and P C Lambert[1,3]

## From Dickman & Adami (2006) [4] 'Interpreting trends in cancer patient survival'

- Until primary prevention programmes succeed to the point of eradicating cancer, doctors must effectively diagnose and treat the cancers that arise and require a means of measuring progress in this specific area.
- Patient survival rates provide such a measure whereas population mortality rates may not as they also reflect changes in incidence.
- For example, lung cancer mortality rates are decreasing in many countries, not because we have become better at diagnosing and treating those individuals that develop lung cancer but because successful primary prevention has reduced lung cancer incidence.

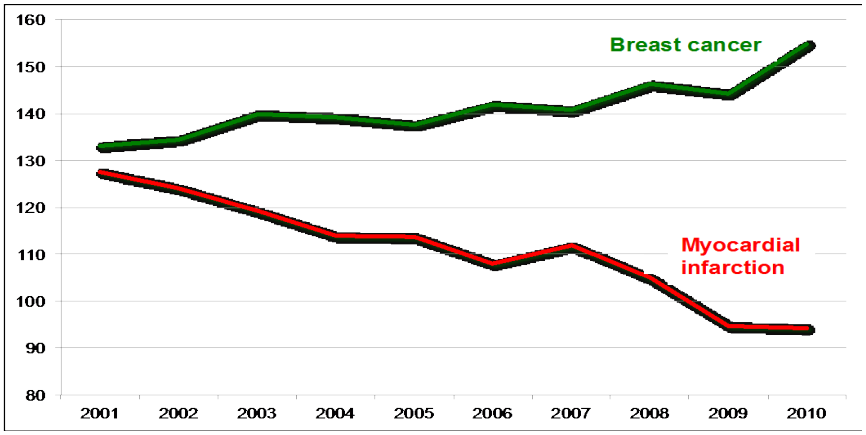


## Breast cancer in Norway [5]






## Trends in incidence in Sweden



## How might we measure the prognosis of cancer patients?

- Total mortality (among the patients).
- Our interest is typically in net mortality (mortality associated with a diagnosis of cancer).
- Cause-specific mortality provides an estimate of net mortality (under certain assumptions).
- When estimating cause-specific mortality only those deaths which can be attributed to the cancer in question are considered to be events.

$$\text{cause-specific mortality} = \frac{\text{number of deaths due to cancer}}{\text{person-time at risk}}$$

The survival times of patients who die of causes other than cancer are censored.



## Need to consider competing risks [6]

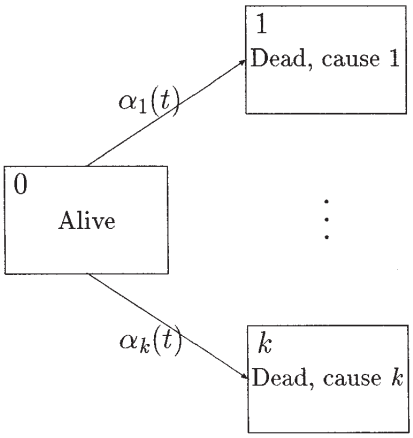



**Figure 1** The competing risks multi-state model.



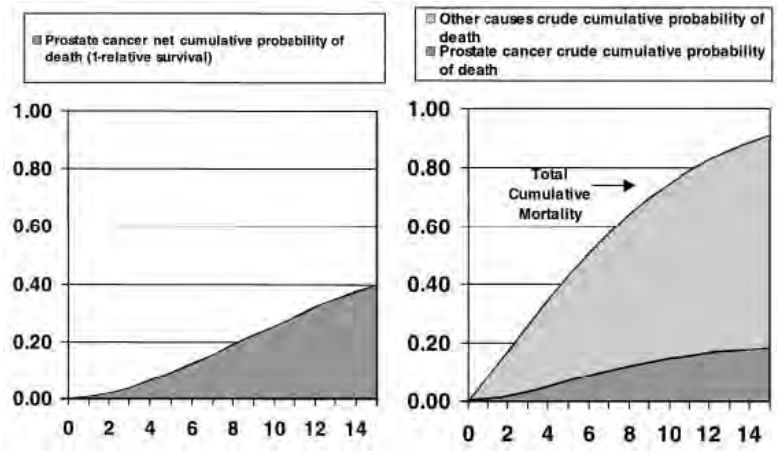
## Many synonyms for the same concept

| | | |
|---|---|---|
| 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 the cause-specific cumulative incidence function (Geskus) or the cumulative incidence function.

## Net (left) and crude (right) probabilities of death in men with localized prostate cancer aged 70+ at diagnosis (Cronin and Feuer [7])






## Should we use crude or net survival?

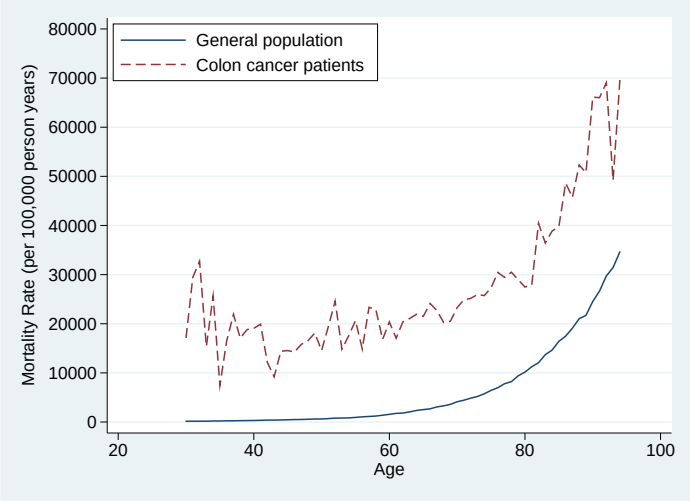
1. Comparing patient survival between countries.
2. Studying temporal trends in patient survival.
3. Communicating prognosis to patients.



## Cause-specific survival can estimate net survival (assuming conditional independence)

- Using cause-specific methods requires that reliably coded information on cause of death is available.
- Even when cause of death information is available to the cancer registry via death certificates, it is often vague and difficult to determine whether or not cancer is the primary cause of death.
- How do we classify, for example, deaths due to treatment complications?
- Consider a patient treated with radiation therapy and chemotherapy who dies of cardiovascular disease. Do we classify this death as 'due entirely to cancer' or 'due entirely to other causes'?

## All-cause mortality for males with colon cancer and Finnish population






## Relative survival aims to estimate net survival (still need conditional independence)

- We estimate excess mortality: the difference between observed (all-cause) and expected mortality.

$$\begin{array}{ccccc} \text{excess} & = & \text{observed} & - & \text{expected} \\ \text{mortality} & & \text{mortality} & & \text{mortality} \end{array}$$

- Relative survival is the survival analog of excess mortality — the relative survival ratio is defined as the observed survival in the patient group divided by the expected survival of a comparable group from the general population.

$$\text{relative survival ratio} = \frac{\text{observed survival proportion}}{\text{expected survival proportion}}$$



## Cervical cancer in New Zealand 1994 – 2001 Life table estimates of patient survival

Women diagnosed 1994 - 2001 with follow-up to the end of 2002

| I | N | D | W | Effective number at risk | Interval-specific observed survival | Interval-specific relative survival | Cumulative observed survival | Cumulative expected survival | Cumulative relative survival |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1559 | 209 | 0 | 1559.0 | 0.86594 | 0.87472 | 0.86594 | 0.98996 | 0.87472 |
| 2 | 1350 | 125 | 177 | 1261.5 | 0.90091 | 0.90829 | 0.78014 | 0.98192 | 0.79450 |
| 3 | 1048 | 58 | 172 | 962.0 | 0.93971 | 0.94772 | 0.73310 | 0.97362 | 0.75296 |
| 4 | 818 | 32 | 155 | 740.5 | 0.95679 | 0.96459 | 0.70142 | 0.96574 | 0.72630 |
| 5 | 631 | 23 | 148 | 557.0 | 0.95871 | 0.96679 | 0.67246 | 0.95766 | 0.70218 |
| 6 | 460 | 10 | 130 | 395.0 | 0.97468 | 0.98284 | 0.65543 | 0.94972 | 0.69013 |
| 7 | 320 | 5 | 129 | 255.5 | 0.98043 | 0.98848 | 0.64261 | 0.94198 | 0.68219 |
| 8 | 186 | 3 | 134 | 119.0 | 0.97479 | 0.98405 | 0.62641 | 0.93312 | 0.67130 |
| 9 | 49 | 1 | 48 | 25.0 | 0.96000 | 0.97508 | 0.60135 | 0.91869 | 0.65457 |

## Relative survival example (skin melanoma)

Table 1: Number of cases (N) and 5-year observed ($p$), expected ($p^*$), and relative ($r$) survival for males diagnosed with localised skin melanoma in Finland during 1985–1994.

| Age | N | $p$ | $p^*$ | $r$ |
|---|---|---|---|---|
| 15–29 | 67 | 0.947 | 0.993 | 0.954 |
| 30–44 | 273 | 0.856 | 0.982 | 0.872 |
| 45–59 | 503 | 0.824 | 0.943 | 0.874 |
| 60–74 | 449 | 0.679 | 0.815 | 0.833 |
| 75+ | 200 | 0.396 | 0.505 | 0.784 |

- Relative survival controls for the fact that expected mortality depends on demographic characteristics (age, sex, etc.).
- In addition, relative survival may, and usually does, depend on such factors.



## Breakthrough paper in 2012; An unbiased estimator for net survival [8]

**On Estimation in Relative Survival**

**Maja Pohar Perme,[1,*] Janez Stare,[1] and Jacques Estève[2]**

[1]Department of Biostatistics and Medical Informatics, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia

[2]Université Claude Bernard, Hospices Civils de Lyon, Service de Biostatistique, 162, Avenue Lacassagne 69424 Lyon Cedex 03, France

[*]*email:* maja.pohar@mf.uni-lj.si

SUMMARY. Estimation of relative survival has become the first and the most basic step when reporting cancer survival statistics. Standard estimators are in routine use by all cancer registries. However, it has been recently noted that these estimators do not provide information on cancer mortality that is independent of the national general population mortality. Thus they are not suitable for comparison between countries. Furthermore, the commonly used interpretation of the relative survival curve is vague and misleading. The present article attempts to remedy these basic problems. The population quantities of the traditional estimators are carefully described and their interpretation discussed. We then propose a new estimator of net survival probability that enables the desired comparability between countries. The new estimator requires no modeling and is accompanied with a straightforward variance estimate. The methods are described on real as well as simulated data.

KEY WORDS: Age standardization; Cancer registry data; Competing risks; Net survival; Relative survival; Survival analysis.



## In the relative survival setting, which approach should one use for estimating net survival?

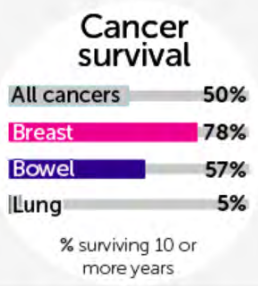
- A hot topic in our field and a focus of our current research.
- I won't bore you with the details.

## http://www.cancerresearchuk.org/cancer-info/cancerstats/survival/

Relative survival was estimated to be 50%.

### Cancer survival statistics

- 50% of adult cancer patients diagnosed in 2010-2011 in England and Wales are predicted to survive 10 or more years.
- 46% of men and 54% of women cancer patients diagnosed in 2010-2011 in England and Wales are predicted to survive 10 or more years.
- Cancer survival rates in the UK have doubled in the last 40 years.






## What does a relative survival of 50% mean? 10-year probabilities of death

| Measure | Age 40 | Age 60 | Age 80 |
|---|---|---|---|
| Net prob. of death (1-rel surv) | 0.50 | 0.50 | 0.50 |
| Crude (actual): cancer death | 0.49 | 0.48 | 0.42 |
| Crude (actual): non-cancer death | 0.02 | 0.08 | 0.42 |
| Crude (actual): any cause death | 0.51 | 0.57 | 0.84 |



ELSEVIER

European Journal of Cancer 40 (2004) 326–335

European Journal of Cancer
www.ejconline.com

Review

## Period analysis for 'up-to-date' cancer survival data: theory, empirical evaluation, computational realisation and applications

H. Brenner[a,*], O. Gefeller[b], T. Hakulinen[c,d]

[a]*Department of Epidemiology, German Centre for Research on Ageing, Bergheimer Str. 20, D-69115 Heidelberg, Germany*
[b]*Department of Medical Informatics, Biometry and Epidemiology, University of Erlangen-Nuremberg, Waldstrasse 6, D-91054 Erlangen, Germany*
[c]*Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland*
[d]*Department of Public Health, University of Helsinki, FIN-00170 Helsinki, Finland*

## Modelling excess mortality

### Relative Survival Model

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

$$\begin{array}{ccccc} \text{Observed} & & \text{Expected} & & \textit{Excess} \\ \text{Mortality Rate} & = & \text{Mortality Rate} & + & \textit{Mortality Rate} \end{array}$$

- Cox model cannot be applied to model a difference in two rates.
- It is the observed mortality that drives the variance.
- Can use Poisson regression (Dickman *et al.* 2004) [9].
- Even better: flexible parametric models (Royston and Parmar 2002 [10], Nelson *et al.* [11]).



## Flexible Parametric Survival Models [10, 13, 14]

- First introduced by Royston and Parmar (2002) [10].
- 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) [12].
- 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).



## This paper has been cited over 27,000 times [15]

**Regression Models and Life-Tables**

By D. R. Cox

*Imperial College, London*

[Read before the Royal Statistical Society, at a meeting organized by the Research Section, on Wednesday, March 8th, 1972, Mr M. J. R. Healy in the Chair]

Summary

The analysis of censored failure times is considered. It is assumed that on each individual are available values of one or more explanatory variables. The hazard function (age-specific failure rate) is taken to be a function of the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time. A conditional likelihood is obtained, leading to inferences about the unknown regression coefficients. Some generalizations are outlined.

## The Cox model[15]

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

- 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.



## Quote from Sir David Cox (Reid 1994 [16])

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."

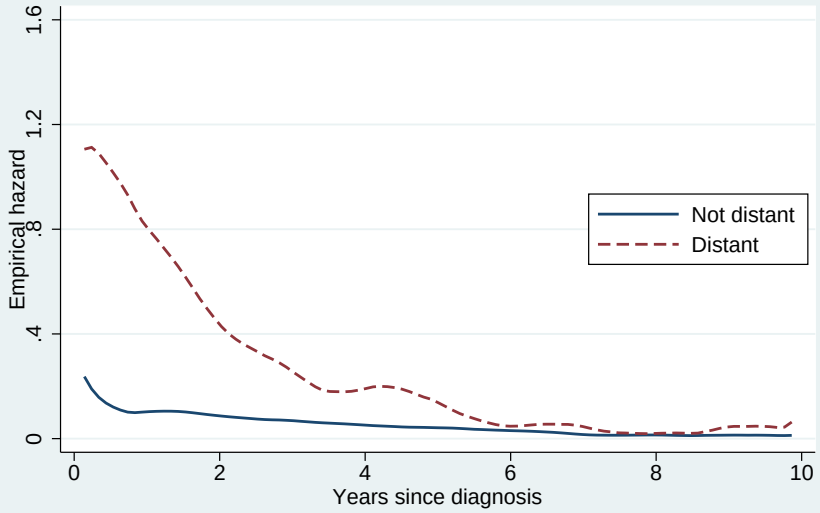


## Example: survival of patients diagnosed with colon carcinoma in Finland

- Patients diagnosed with colon carcinoma in Finland 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?

## Smoothed empirical hazards (cancer-specific mortality rates)

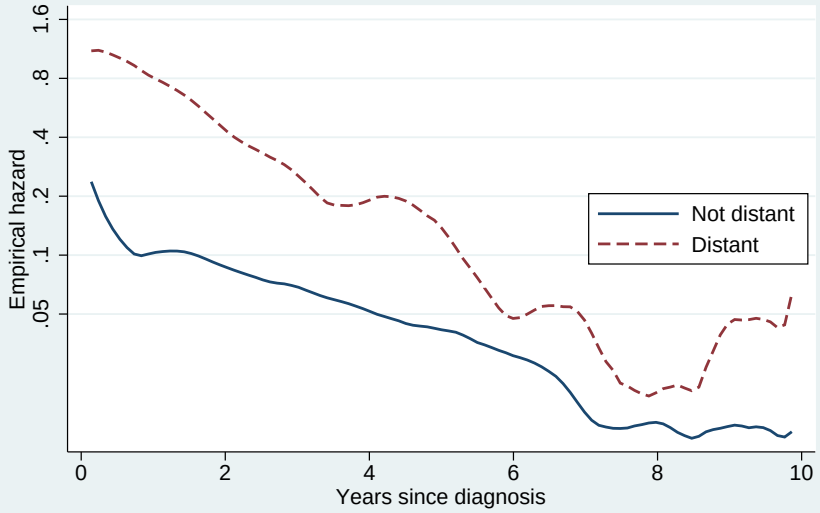
sts graph, by(distant) hazard kernel(epan2)






## Smoothed empirical hazards on log scale

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

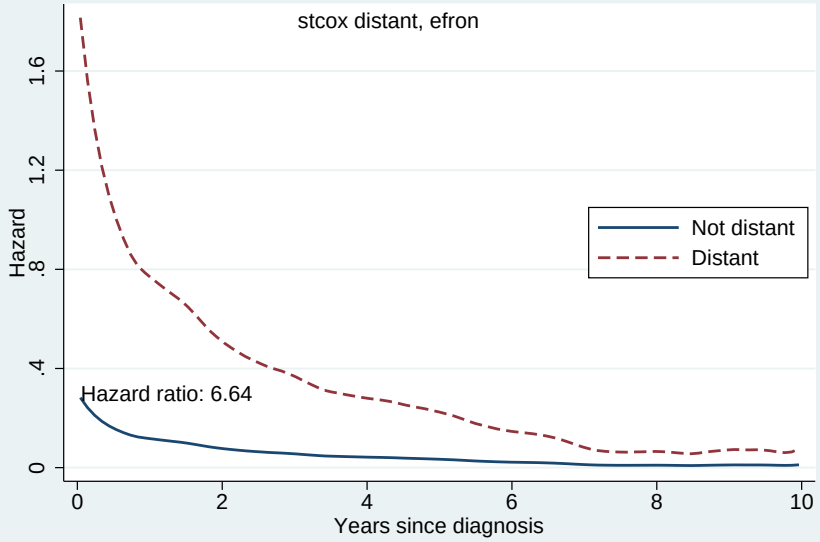





## Fit a Cox model to estimate the mortality rate ratio

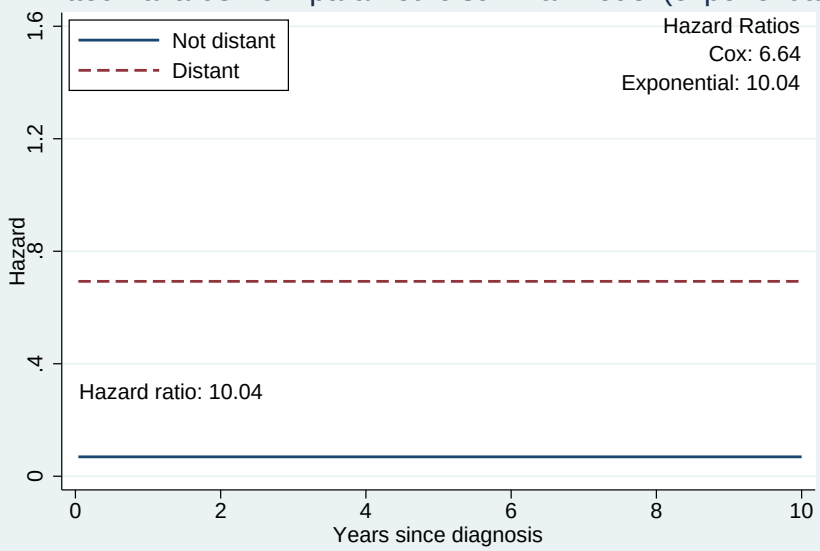
```
. stcox distant

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 |    6.64        .1689   73.00    0.000  6.24  6.90
------------------------------------------------------------------
```

## Fitted hazards from Cox model with Efron method for ties



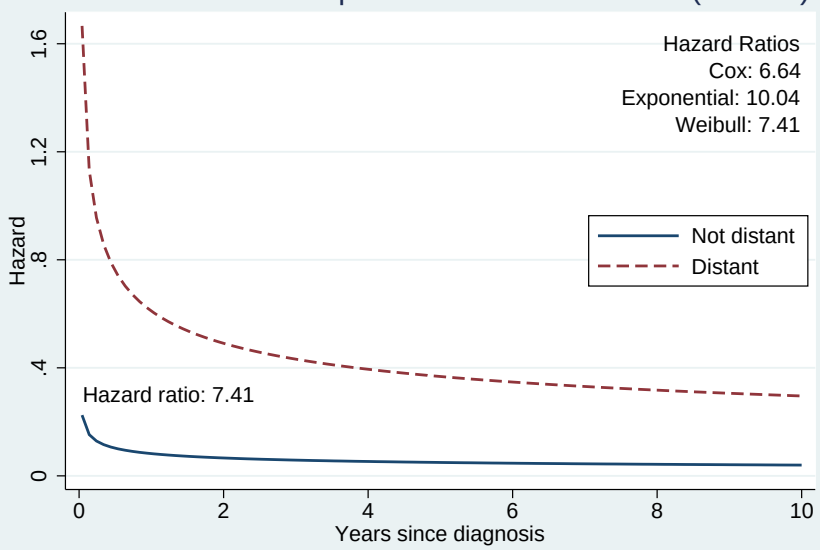


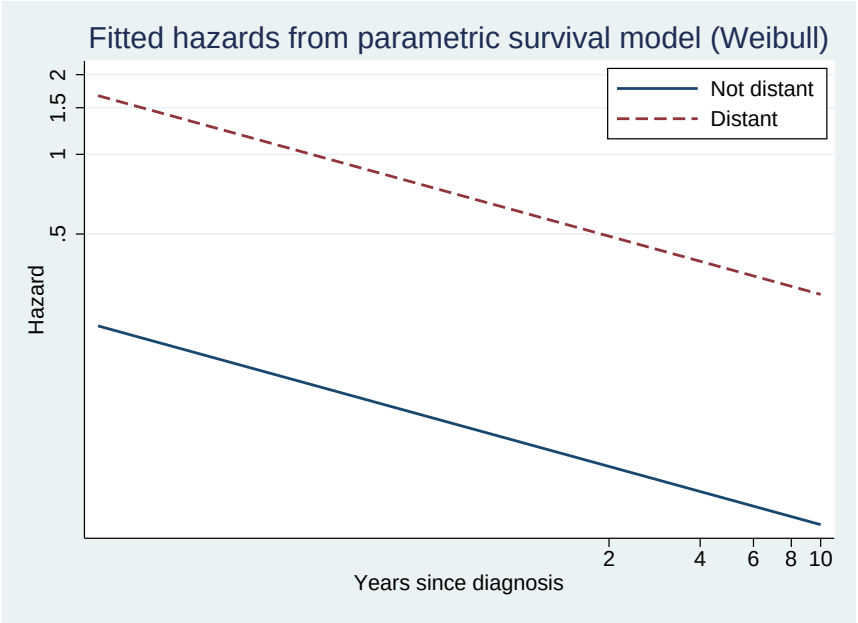

## Fitted hazards from parametric survival model (exponential)



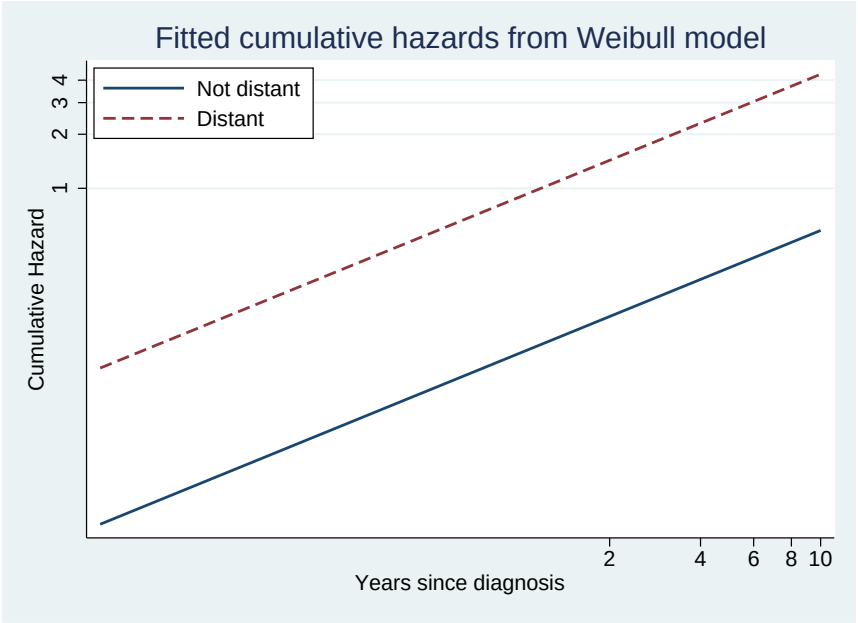



## Fitted hazards from parametric survival model (Weibull)










# Demography and epidemiology: Practical use of the Lexis diagram in the computer age.

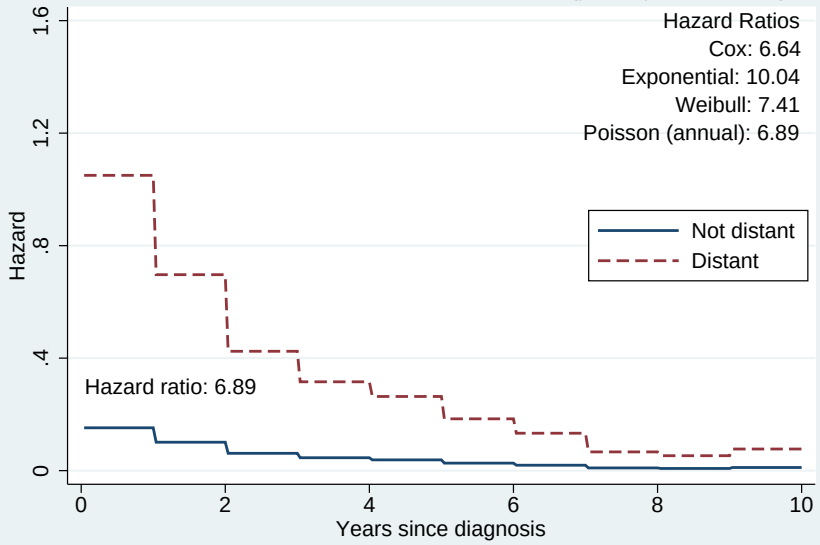
## or:

## Who needs the Cox-model anyway?

Annual meeting of Finnish Statistical Society
23–24 May 2005
Revised December 2005.

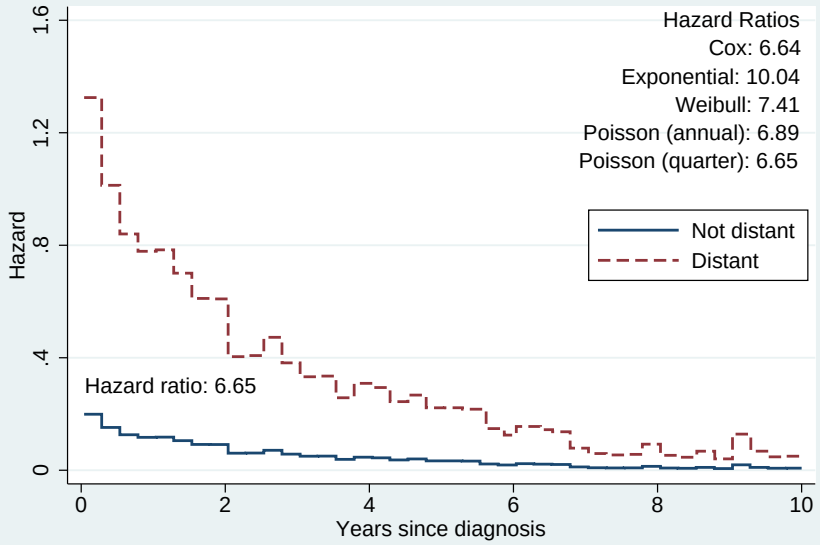
Bendix Carstensen
Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
www.biostat.ku.dk/~bxc

## Fitted hazards from Poisson model (yearly intervals)



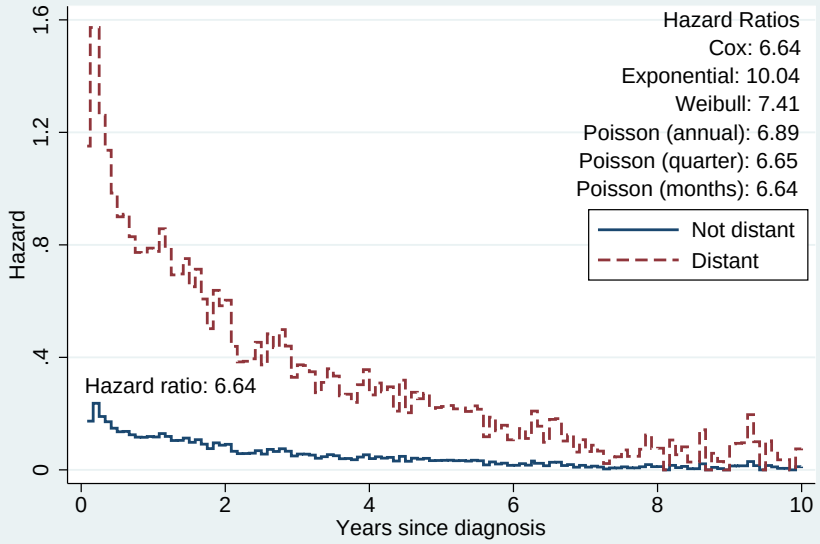


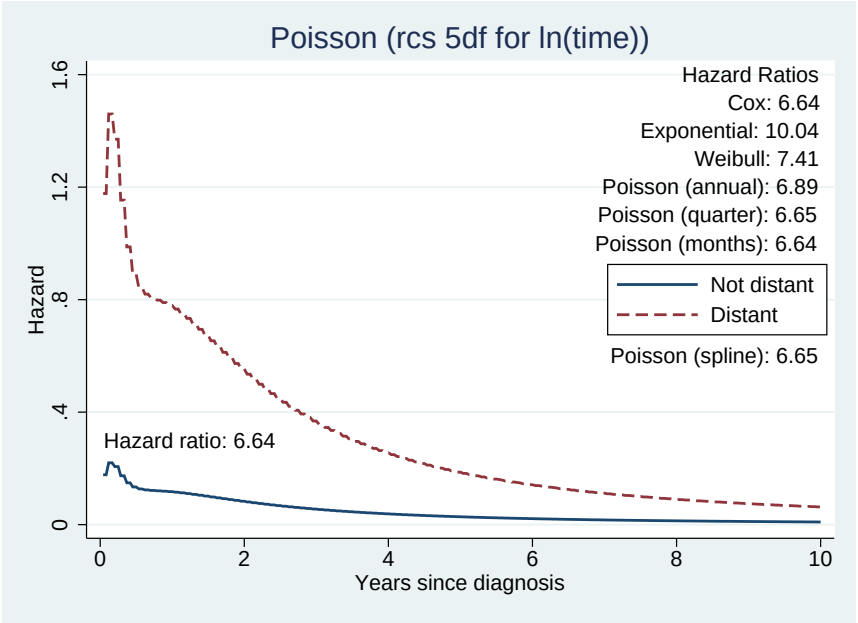

## Fitted hazards from Poisson model (3-months)



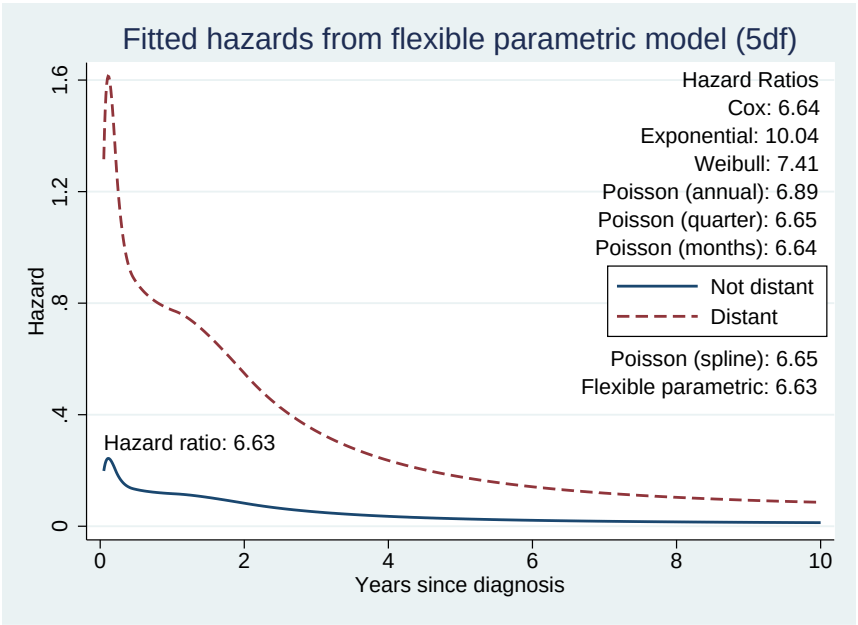



## Fitted hazards from Poisson model (months)










## Flexible Parametric Models: Basic Idea

- Consider a Weibull survival curve.

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

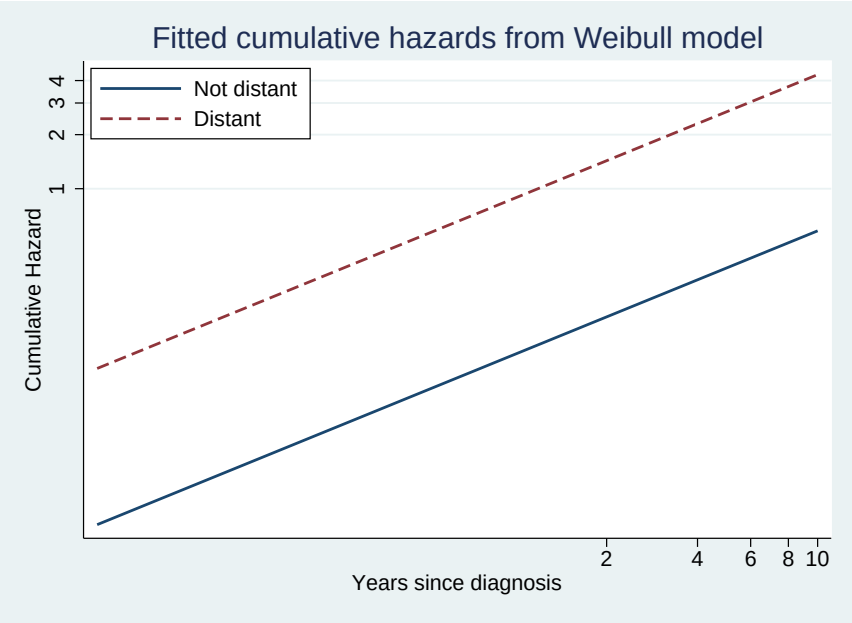
- If we transform to the log cumulative hazard scale.

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

- 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.
- 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

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



## Survival and Hazard Functions

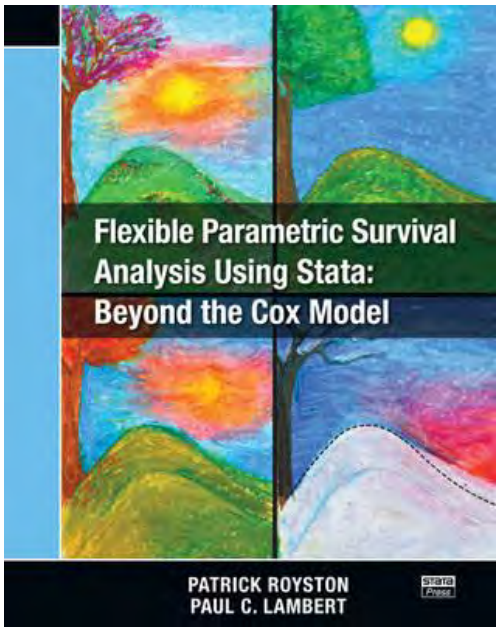
- We can transform to the survival scale

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

- The hazard function is a bit more complex.

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

- This involves the derivatives of the restricted cubic splines functions, although these are relatively easy to calculate.


Flexible Parametric Survival
Analysis Using Stata:
Beyond the Cox Model
PATRICK ROYSTON
PAUL C. LAMBERT
stata
Press




www.cansurv.net

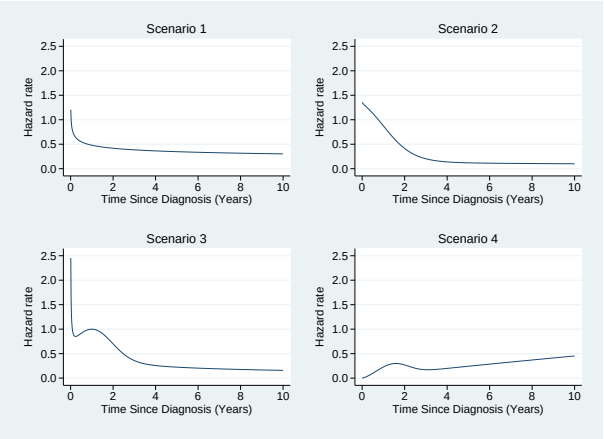
## Sensitivity to choice of knots; Simulation study by Rutherford et al. [17]

- '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.'



## Simulation Study (Rutherford et al.) [17]

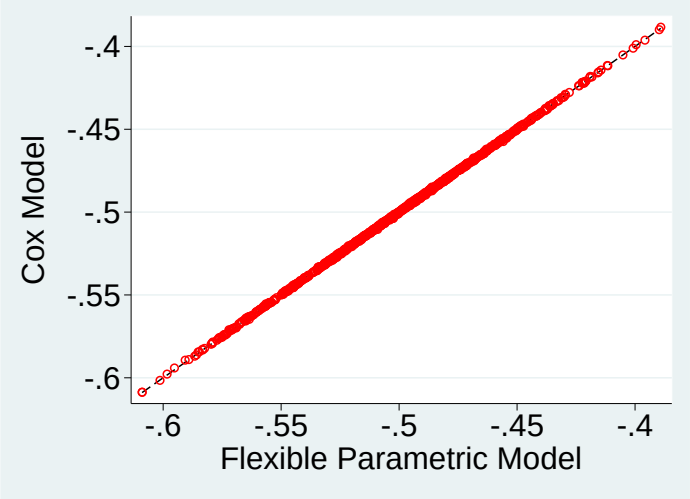
- Generate data assuming a mixture Weibull distribution.



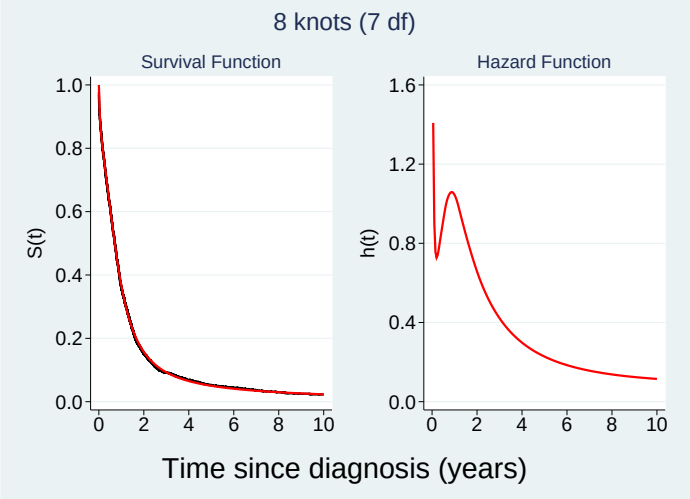

- Fit models using restricted cubic splines.



## Scenario 3 comparison of Log Hazard Ratios



## 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 inference based on the 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.



## Implementation in Stata [13]

**stpm2 available from SSC**

```
. ssc install stpm2
```

**All cause survival**

```
. stpm2 eng, scale(hazard) df(5)
```

**Relative survival**

```
. stpm2 eng, scale(hazard) df(5) hazard(rate)
```

**Time-dependent effects**

```
. stpm2 eng, scale(hazard) df(5) hazard(rate) tvc(eng) dftvc(3)
```

**Cure model**

```
. stpm2 eng, scale(hazard) df(5) hazard(rate) tvc(eng) dftvc(3) cure
```

## Example using attained age as the time-scale

- Study from Sweden [9] 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 (hip) fracture.
- Risk of fracture varies by age.
- Attained age is used as the primary time-scale.
- Actually, two timescales, but will initially ignore time from diagnosis.



## Estimates from a proportional hazards model

### stset using age as the time-scale

```
. stset dateexit,fail(frac = 1) enter(datecancer) origin(datebirth) ///
                id(id) scale(365.25) exit(time datebirth + 100*365.25)
```

**Cox Model**

```
.  stcox noorc orc
```

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

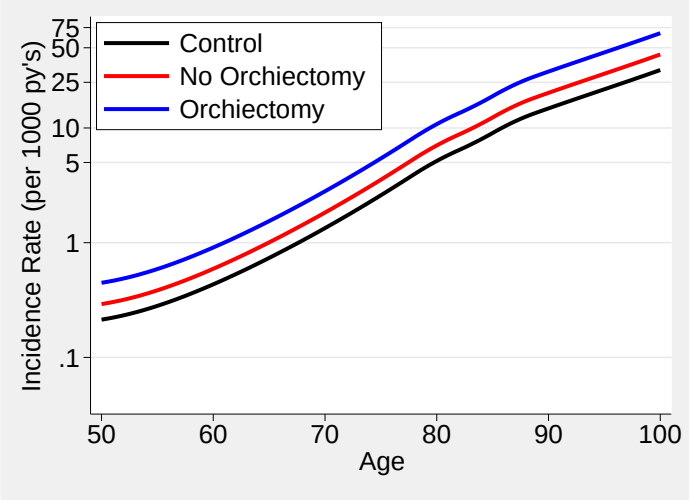
**Flexible Parametric Model**

```
.  stpm2 noorc orc, df(5) scale(hazard)
```

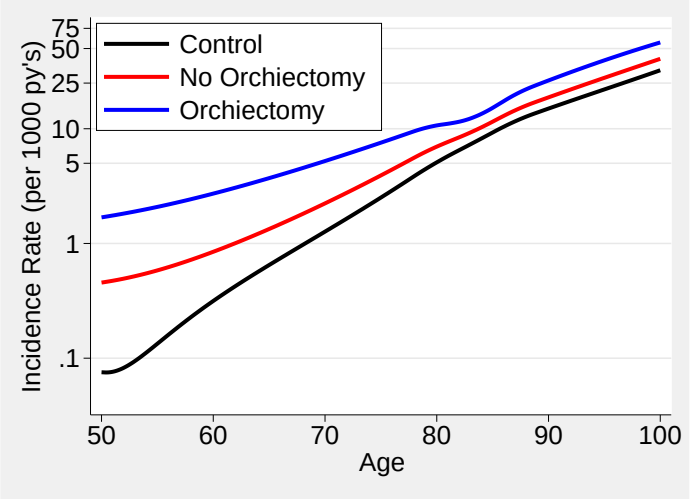
| | | |
|---|---|---|
| Incidence rate ratio (no orchiectomy) | = | 1.37 (1.28 to 1.46) |
| Incidence rate ratio (orchiectomy) | = | 2.09 (1.93 to 2.27) |



## Proportional Hazards

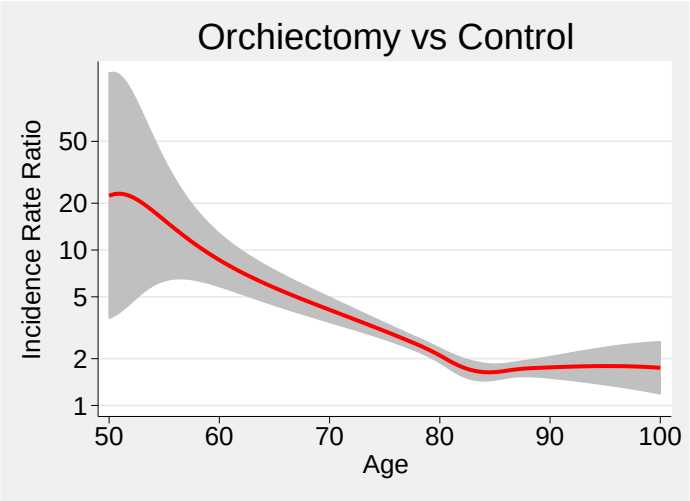


## Non Proportional Hazards



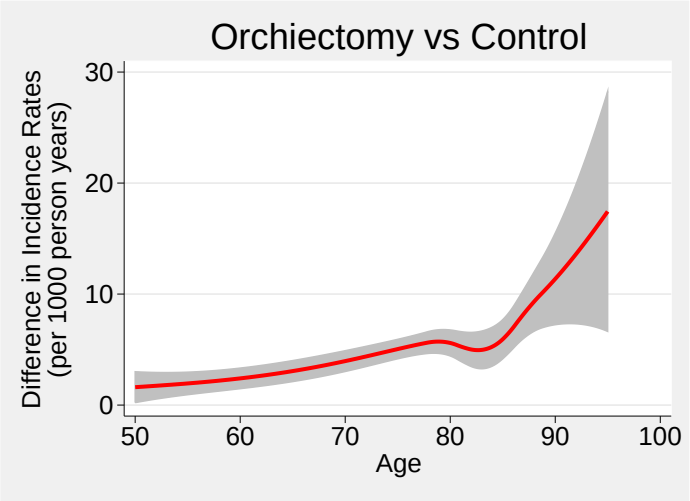



## Incidence Rate Ratio






## Incidence Rate Difference



## stpm2 postestimation commands

### Fit model (non-proportional hazards)

```
stpm2 noorc orc, df(5) scale(h) tvc(noorc orc) dftvc(3)
```

### Fitted hazards

```
predict h1, h zeros
predict h2, h at(noorc 1 orc 0)
predict h3, h at(noorc 0 orc 1)
```

### Hazard ratios

```
predict hr2, hrnum(noorc 1) ci
predict hr3, hrnum(orc 1) ci
```

### Hazard differences

```
predict hdiff2, hdiff1(noorc 1) ci
predict hdiff3, hdiff1(orc 1) ci
```



## Net Survival (revisited)

- Relative Survival aims to estimate net survival.
- This is the probability of not dying of cancer in the hypothetical world where it is impossible to die of other causes.

### Key Assumptions

Independence between mortality due to cancer and mortality due to other causes & an appropriate estimate of expected survival.

- Same interpretation/assumption for cause-specific survival.
- We also assume that we have modelled covariates appropriately .



## Crude and Net Probabilities

| | | |
|---|---|---|
| Net Probability of Death Due to Cancer | = | Probability of death due to cancer 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 due to cancer 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 the cause-specific cumulative incidence function (Geskus) or the cumulative incidence function.

## Brief Mathematical Details [18]

| | | |
|---|---|---|
| $h(t) = h^*(t) + \lambda(t)$ | - | all-cause mortality rate |
| $h^*(t)$ | - | expected mortality rate |
| $\lambda(t)$ | - | excess mortality rate |
| $S^*(t)$ | - | Expected Survival |
| $R(t)$ | - | Relative Survival |

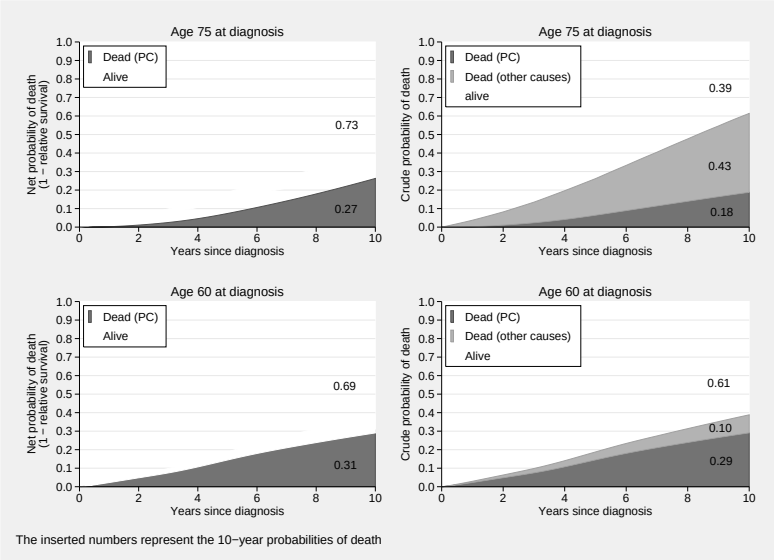
$$\text{Net Prob of Death} = 1 - R(t) = 1 - \exp\left(-\int_0^t \lambda(t)\right)$$

$$\text{Crude Prob of Death (cancer)} = \int_0^t S^*(t)R(t)\lambda(t)$$

$$\text{Crude Prob of Death (other causes)} = \int_0^t S^*(t)R(t)h^*(t)$$



## Probabilities of death due to prostate cancer [19]



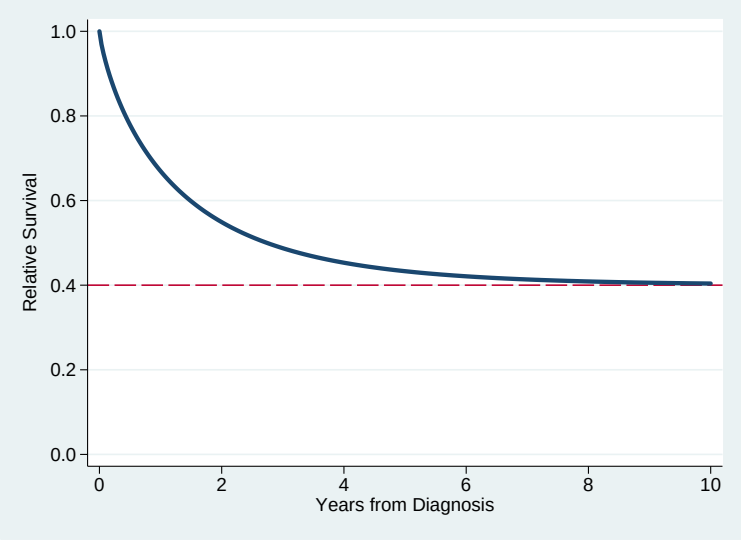

The inserted numbers represent the 10–year probabilities of death



## What is cure?

- Medical cure occurs when all signs of cancer have been removed in a patient; this is an individual-level definition of cure.
- It is difficult to prove that a patient is medically cured.
- Population or statistical cure occurs when mortality among patients with the disease returns to the same level as that expected for the general population.
- Equivalently the *excess* mortality rate approaches zero.
- This is a population-level definition of cure.
- When the excess mortality reaches (and stays) at zero, the relative survival curve is seen to reach a plateau.

## Plateau for relative survival






## Mixture cure model

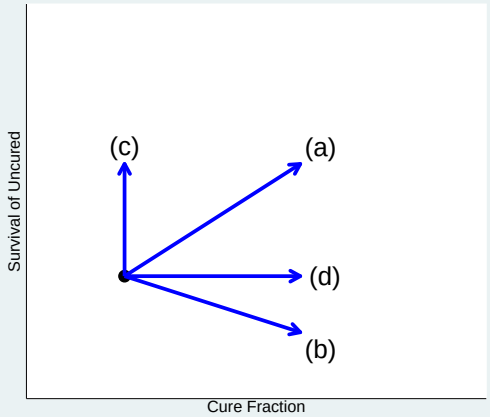
### Mixture cure model

$$S(t) = S^*(t)(\pi + (1-\pi)S_u(t)); \qquad \lambda(t) = h^*(t) + \frac{(1-\pi)f_u(t)}{\pi+(1-\pi)S_u(t)}$$

- $S^*(t)$ is the expected survival.
- $\pi$ is the proportion cured (the cure fraction).
- $(1-\pi)$ is the proportion 'uncured' (those 'bound to die').
- $S_u(t)$ is the net survival for the 'uncured' group.
- The excess mortality rate has an asymptote at zero.
- See De Angelis *et al.* [20], Verdecchia *et al.* [21] and Lambert *et al.*[22] for details.



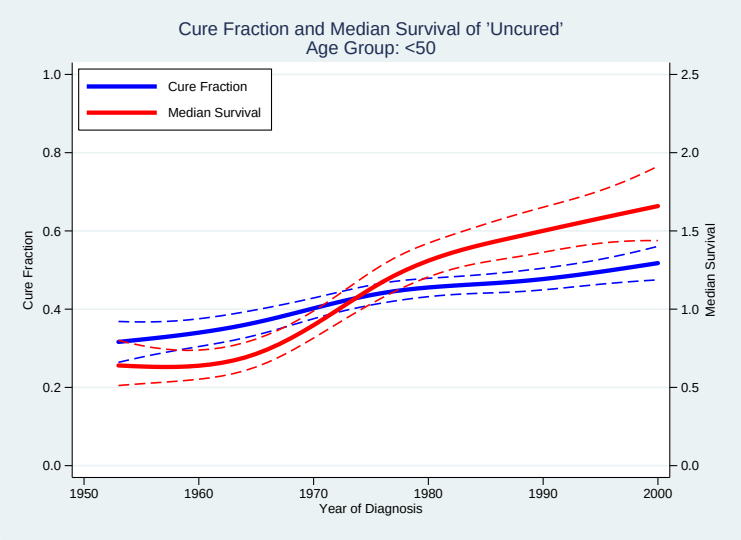
## Cure models: Interpreting changes over time




Adapted from Verdeccia (1998)

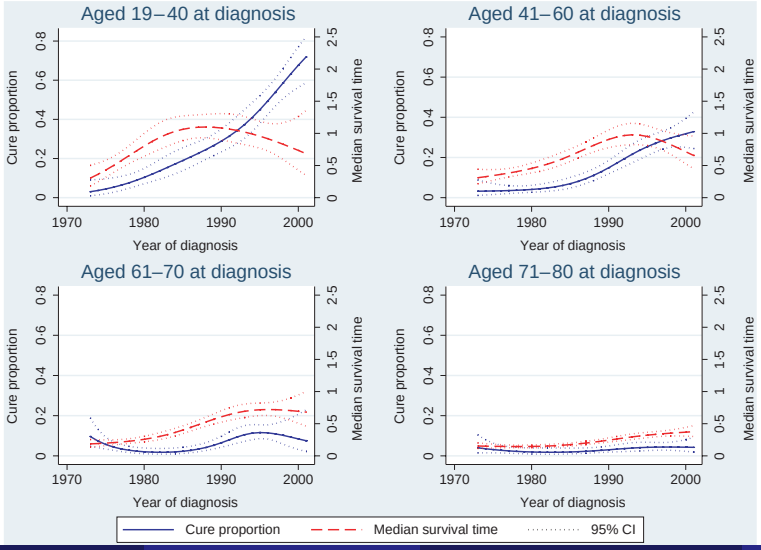
(a) General Improvement
(b) Selective Improvment
(c) Improved palliative care or lead time
(d) Inclusion of subjects with no excess risk

# Time trends for cancer of the colon age <50 [23]






# Andersson 2010 [24]: trends for AML






# References


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

[2] Beral V, Peto R. UK cancer survival statistics. *BMJ* 2010;**341**:c4112.

[3] Rutherford MJ, Møller H, Lambert PC. A comprehensive assessment of the impact of errors in the cancer registration process on 1- and 5-year relative survival estimates. *Br J Cancer* 2013;**108**:691–698.

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

[5] Cancer Registry of Norway. *Cancer in Norway 2012 - Cancer incidence, mortality, survival and prevalence in Norway*. Oslo: Cancer Registry of Norway, 2014.

[6] Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. *Stat Methods Med Res* 2002;**11**:203–215.

[7] 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.

[8] Pohar Perme M, Stare J, Estève J. On estimation in relative survival. *Biometrics* 2012; **68**:113–120.

# References 2


[9] Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Stat Med* 2004;**23**:51–64.

[10] 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.

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

[12] 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 Med Res Methodol* 2011;**11**:96.

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

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

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

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




# References 3


[17] Rutherford MJ, Hinchliffe SR, Abel GA, Lyratzopoulos G, Lambert PC, Greenberg DC. How much of the deprivation gap in cancer survival can be explained by variation in stage at diagnosis: An example from breast cancer in the east of england. *International Journal of Cancer* 2013;.

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

[19] Eloranta S, Adolfsson J, Lambert PC, Stattin Pär and Akre O, Andersson TML, Dickman PW. How can we make cancer survival statistics more useful for patients and clinicians: An illustration using localized prostate cancer in sweden. *Cancer Causes Control* 2013; **24**:505–515.

[20] Angelis RD, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Statistics in Medicine* 1999;**18**:441–454.

[21] Verdecchia A, Angelis RD, Capocaccia R, Sant M, Micheli A, Gatta G, Berrino F. The cure for colon cancer: results from the EUROCARE study. *Int J Cancer* 1998;**77**:322–329.

[22] Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 2007;**8**:576–594.




# References 4


[23] Lambert PC, Dickman PW, Österlund P, Andersson T, Sankila R, Glimelius B. Temporal trends in the proportion cured for cancer of the colon and rectum: a population-based study using data from the Finnish cancer registry. *International Journal of Cancer* 2007; **121**:2052–2059.

[24] Andersson TML, Lambert PC, Derolf AR, Kristinsson SY, Eloranta S, Landgren O, *et al.*. Temporal trends in the proportion cured among adults diagnosed with acute myeloid leukaemia in Sweden 1973-2001, a population-based study. *Br J Haematol* 2010; **148**:918–924.