An overview and some recent advances in statistical methods for population-based cancer survival analysis: relative survival, cure models, and flexible parametric models

Paul W Dickman[1] Paul C Lambert[1,2]
Sandra Eloranta[1] Therese Andersson[1]

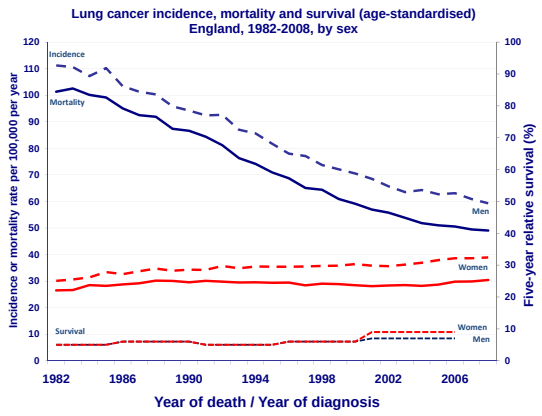
[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[2]Department of Health Sciences, University of Leicester, UK

FMS Jubilee meeting, Utö, October 2012

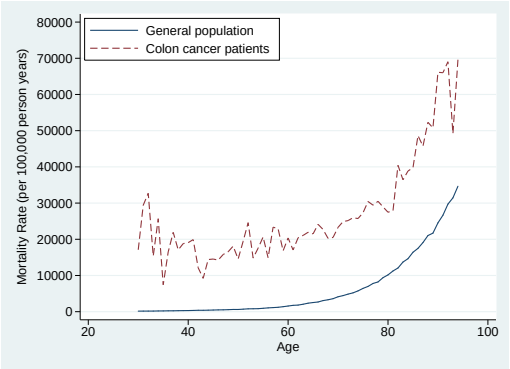
University of Leicester Karolinska Institutet

Overview

- 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.
- The concept of statistical cure; cure models.
- Estimating crude and net probabilities of death.
- Partitioning excess mortality; estimating treatment related CVD mortality.
- Cool stuff that I definitely won't have time to talk about.
  - Estimating the number of avoidable premature deaths.
  - Loss in expectation of life.



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



Relative survival

- 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.

Modelling excess mortality

Relative Survival Models

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

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

- 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) [1].
- Even better: flexible parametric models (Royston and Parmar 2002 [2], Nelson *et al.* [3]).

Flexible Parametric Survival Models

- First introduced by Royston and Parmar (2002) [2].
- 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) [4].
- 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).

Quote from Sir David Cox (Reid 1994 [5])

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."

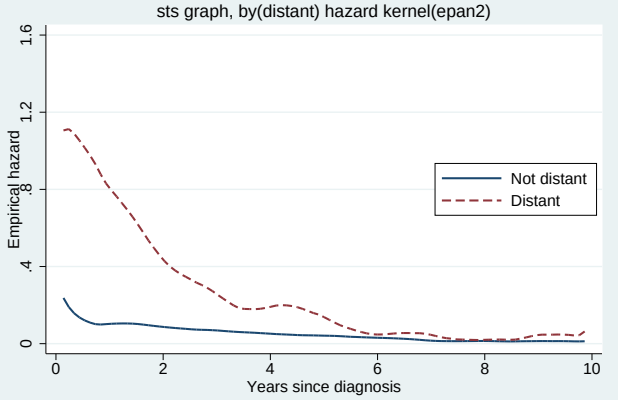
Cox "That's right, but since then various people have shown that the answers are very insensitive to the parametric formulation of the underlying distribution. And if you want to do things like predict the outcome for a particular patient, it's much more convenient to do that parametrically."

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?

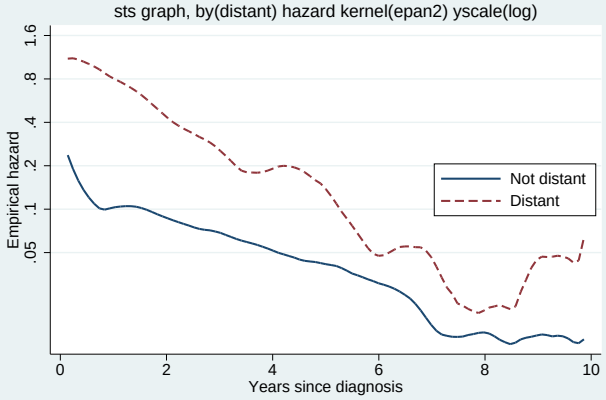
Smoothed empirical hazards (cancer-specific mortality rates)

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



Smoothed empirical hazards on log scale

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

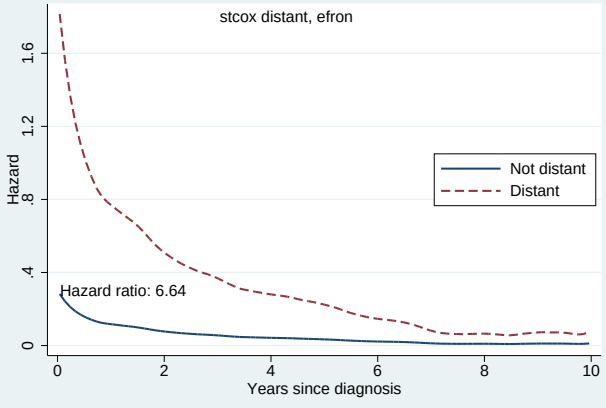


Fit a Cox model to estimate the mortality rate ratio

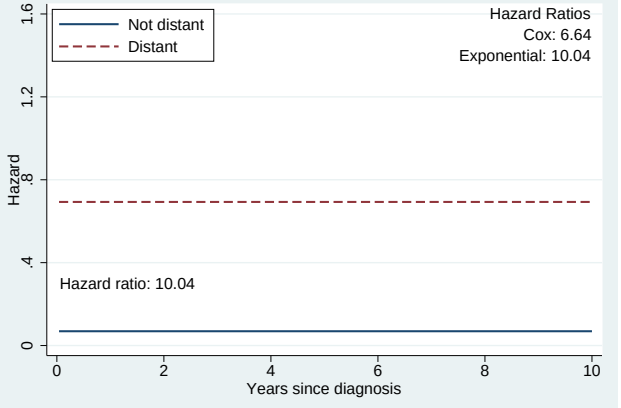
```
. stcox distant

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

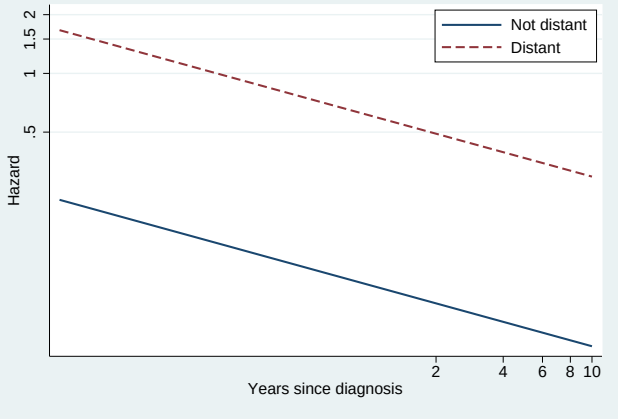


Fitted hazards from parametric survival model (exponential)

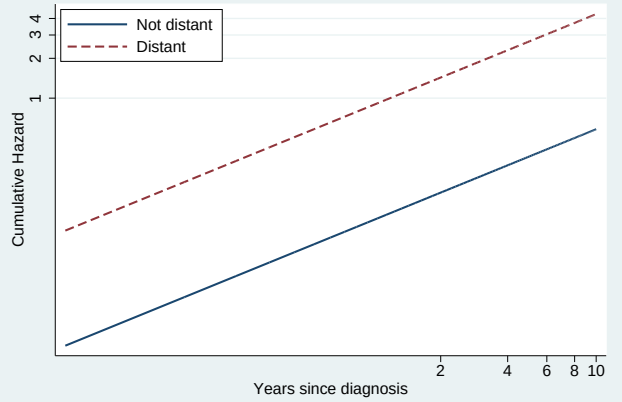


Fitted hazards from parametric survival model (Weibull)

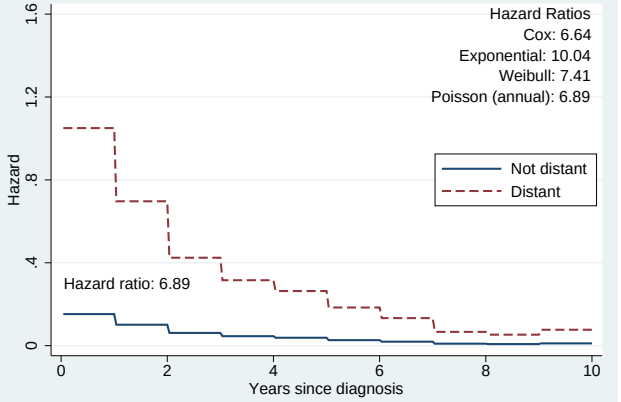
Fitted hazards from parametric survival model (Weibull)



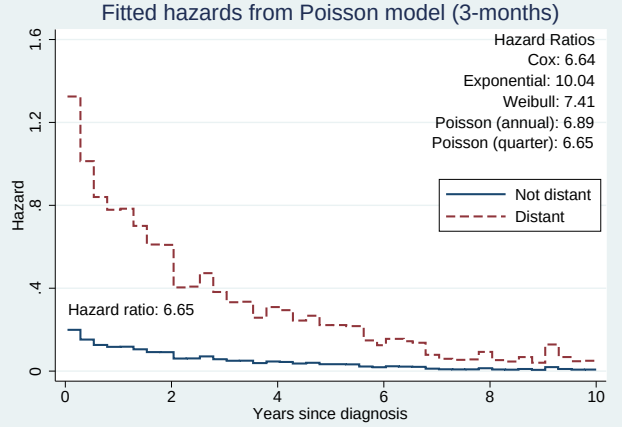
Fitted cumulative hazards from Weibull model



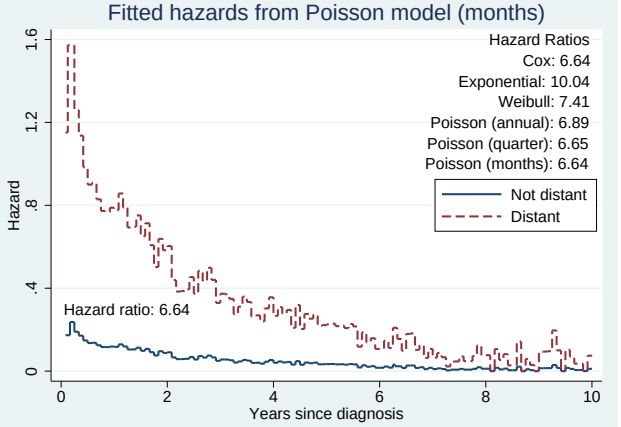
Fitted hazards from Poisson model (yearly intervals)



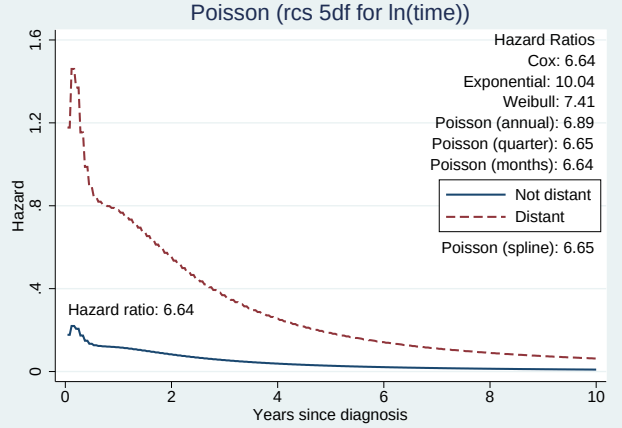
Fitted hazards from Poisson model (3-months)



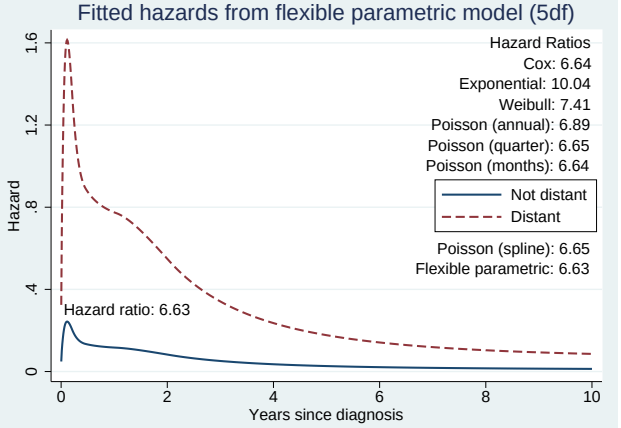
Fitted hazards from Poisson model (months)



Poisson (rcs 5df for ln(time))



Fitted hazards from flexible parametric model (5df)



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.

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

- This is a proportional hazards model.
- Restricted cubic splines with knots, $\mathbf{k_0}$, are used to model the log baseline cumulative hazard.

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

- For example, with 4 knots we can write

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

Survival and Hazard Functions

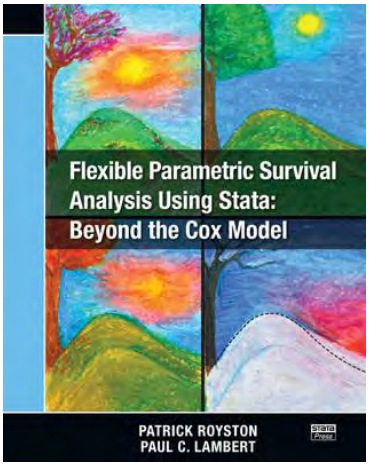
- We can transform to the survival scale

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

- The hazard function is a bit more complex.

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

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



Net Survival

- Relative Survival aims to estimate of 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 cumulative incidence function.

Brief Mathematical Details [6]

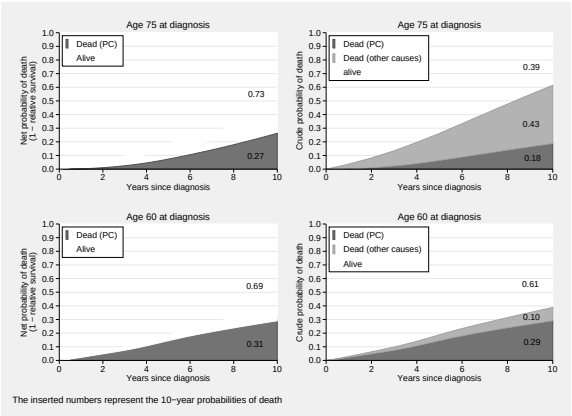
$h(t) = h^*(t) + \lambda(t)$	-	all-cause mortality rate
$h^*(t)$	-	expected mortality rate
$\lambda(t)$	-	excess mortality rate
$S^*(t)$	-	Expected Survival
$R(t)$	-	Relative Survival

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

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

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

Probabilities of death due to prostate cancer

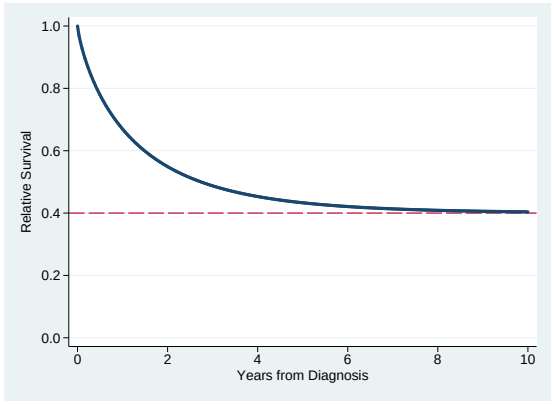


The inserted numbers represent the 10–year probabilities of death

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



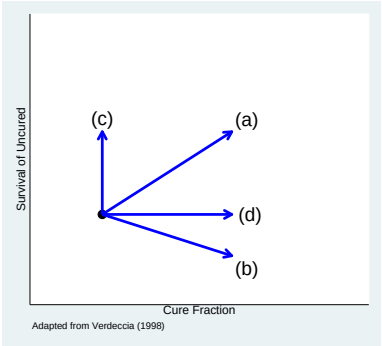
Mixture cure model

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.
- π 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.* [7], Verdecchia *et al.* [8] and Lambert *et al.*[9] for more details.

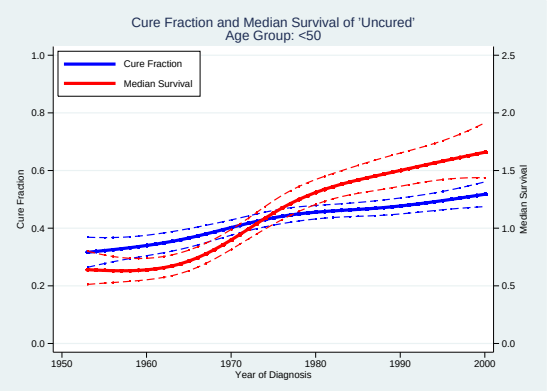
Cure models: Interpreting changes over time



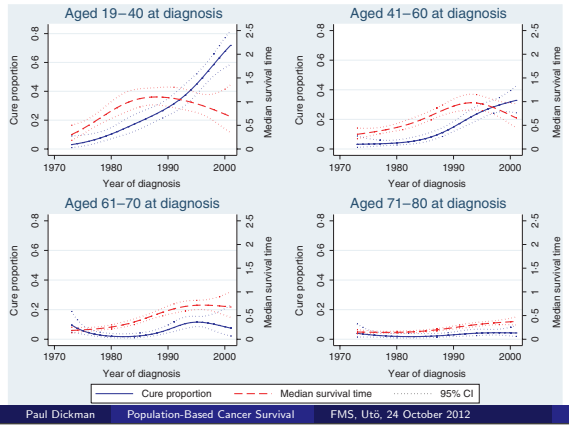
Adapted from Verdeccia (1998)

(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



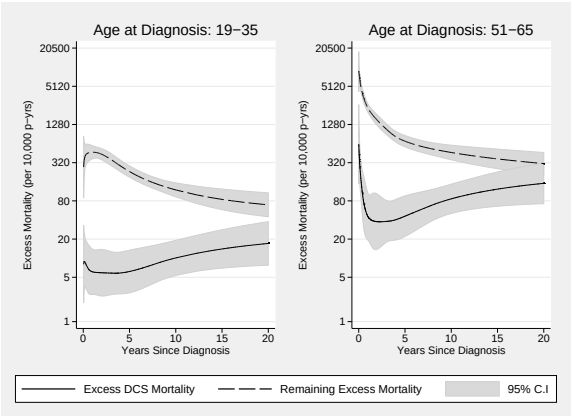
Andersson 2010 [10]: trends for AML



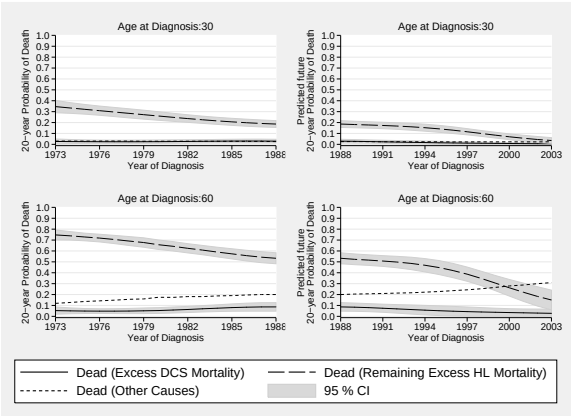
Partitioning Excess Mortality (Eloranta [11])

- We have extended flexible parametric models for relative survival to simultaneously estimate the excess mortality due to diseases of the circulatory system (DCS) and the remaining excess mortality among patients diagnosed with Hodgkin lymphoma.
- Results are presented both as excess mortality rates and crude probabilities of death.
- The outcomes (DCS and non-DCS mortality) can be regarded as competing events and the total excess mortality is partitioned using ideas from classical competing risks theory.
- The model requires population mortality files stratified on cause of death (i.e., DCS and other deaths) to identify those deaths in excess of expected.

Excess mortality for males with Hodgkin lymphoma



Temporal trends in 20-year probability of death



References

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

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

References 2

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

[9] 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.

[10] 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. *British Journal of Haematology* 2010;**148**:918–924.

[11] Eloranta S, Lambert PC, Andersson TML, Czene K, Hall P, Björkholm M, Dickman PW. Partitioning of excess mortality in population-based cancer patient survival studies using flexible parametric survival models. *BMC Med Res Methodol* 2012;**12**:86.