Flexible Parametric Alternatives to the Cox Model

Paul C Lambert[1,2]

Work in collaboration with Paul Dickman, Chris Nelson, Patrick Royston

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

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

Seminar 19/11/2009
University of Birmingham

The Cox Model I

- Web of Science: over 23,300 citations (October 2008).
- Has an h-index of 13 from repeat mis-citations[1].

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

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

[1] http://network.nature.com/people/boboh/blog/2008/06/24/outdone-by-mis-prints

The Cox Model II

- The crucial assumption of the Cox model is that the estimated parameters are not associated with time, i.e. we assume proportional hazards.
- If you are only interested in the relative effect of a covariate on the hazard rate and the assumption of proportional hazards is reasonable, then the Cox model is probably the most appropriate model. In other situations alternative models may be more appropriate.
- However, whenever we estimate a relative effect we should ask "relative to what?"

Quote from Sir David Cox (Reid 1994 [10])

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."

Flexible Parametric Survival Models

- Parametric estimate of the survival and hazard functions.
- Useful for 'standard' and relative survival models.
- First introduced by Royston and Parmer[11].
- Parametric Models have advantages for
  - Understanding.
  - Prediction.
  - Extrapolation.
  - Quantification (e.g. absolute and relative measures of risk).
  - Modelling time-dependent effects.

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.

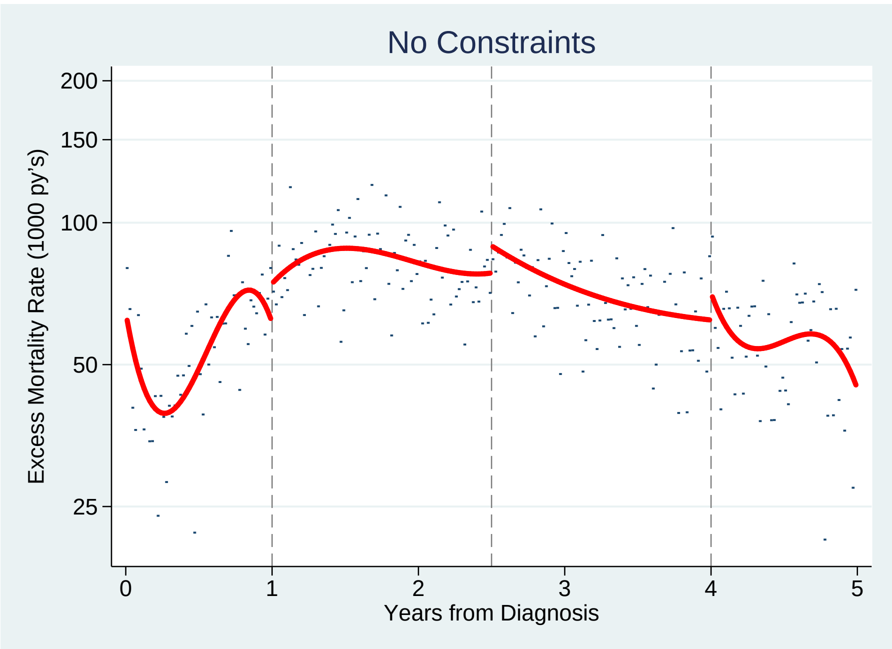
$$\ln\left[H(t)\right] = \ln[-\ln(S(t))]$$
$$\ln\left[H(t)\right] = \ln(\lambda) + \gamma \ln(t)$$

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

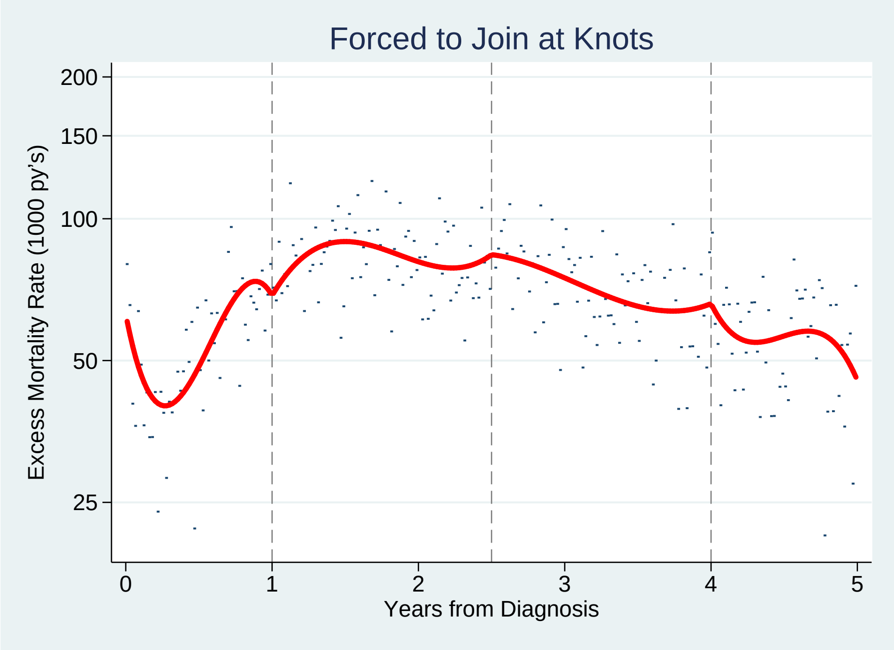
$$\ln\left[H(t|\mathbf{x}_i)\right] = \ln(\lambda) + \gamma \ln(t) + \mathbf{x}_i\boldsymbol{\beta}$$

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

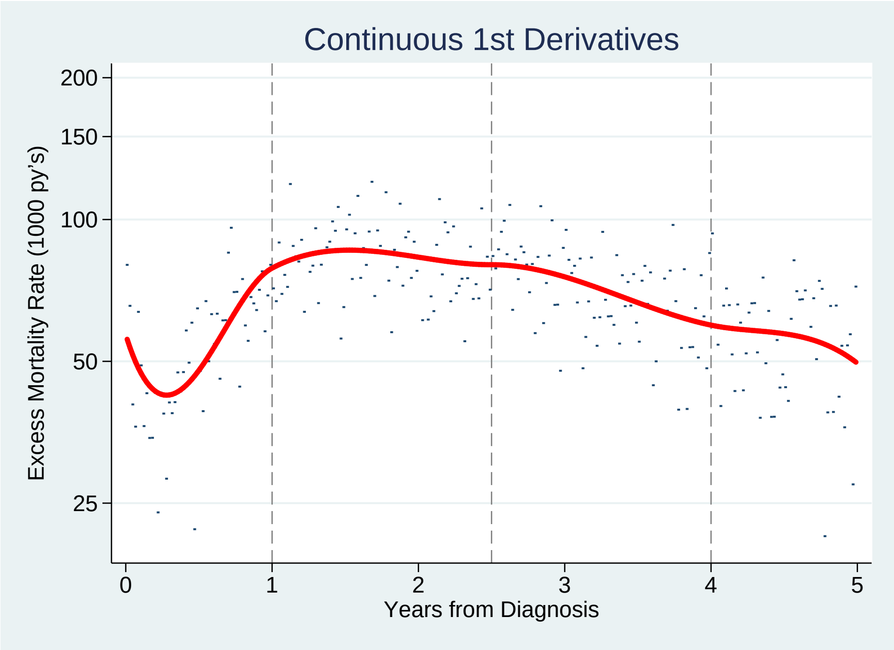
No Continuity Corrections



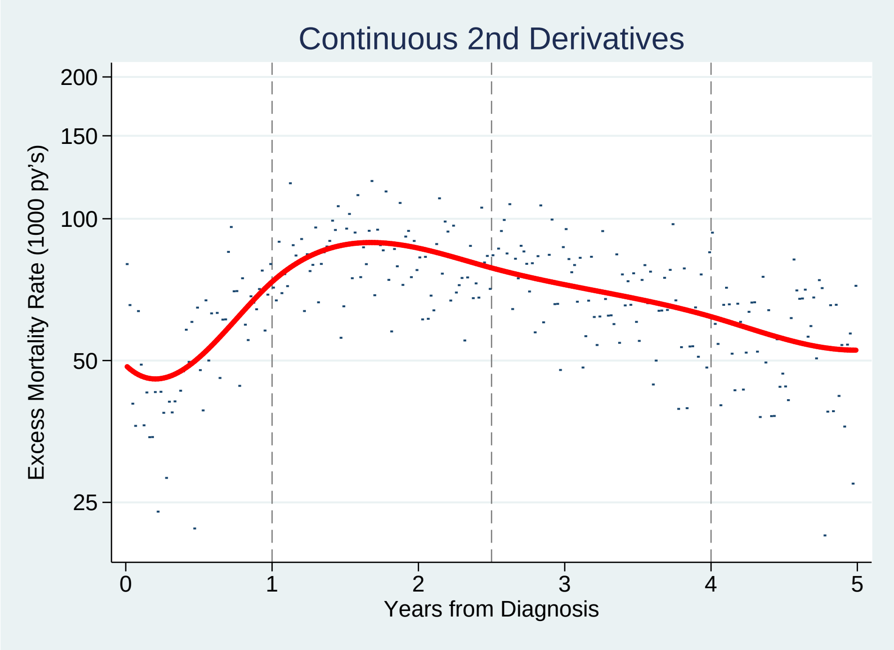
Function forced to join at knots



Continuous first derivative



Continuous second derivative



Incorporating Restricted Cubic Splines

- The linear predictor is

Linear Predictor

$$\eta_i = s\left(\ln(t)|\gamma, \mathbf{k_0}\right) + \mathbf{x}\beta$$

- For models on the log cumulative hazard scale.

Survival and hazard functions

$$S(t) = \exp\left(-\exp\left(\eta_i\right)\right) \qquad h(t) = \frac{ds(\ln(t)|\gamma, \mathbf{k_0})}{dt}\exp\left(\eta_i\right)$$

- Feed these into the likelihood.

$$\ln L_i = d_i \ln\left[h(t_i)\right] + \ln\left[S(t_i)\right]$$

Fitting a Proportional Hazards Model I

- Example: 24,889 women aged under 50 diagnosed with breast cancer 1986-1990.
- Compare five deprivation groups from most affluent to most deprived.

Proportional hazards models

```
. stcox dep2-dep5,
. stpm2 dep2-dep5, df(5) scale(hazard) eform
```

- The `df(5)` option implies using 4 internal knots and 2 boundary knots at their default locations.
- The `scale(hazard)` requests the model to be fitted on the log cumulative hazard scale.

Cox Model

Cox Proportional Hazards Model

```
.

. stcox dep2-dep5,
         failure _d:  dead == 1
   analysis time _t:  survtime
  exit on or before:  time 5
Iteration 0:   log likelihood = -73334.091
Iteration 1:   log likelihood = -73303.081
Iteration 2:   log likelihood = -73302.997
Iteration 3:   log likelihood = -73302.997
Refining estimates:
Iteration 0:   log likelihood = -73302.997
Cox regression -- Breslow method for ties
No. of subjects =        24889                     Number of obs    =     24889
No. of failures =         7366
Time at risk    =   104638.953
                                                   LR chi2(4)       =     62.19
Log likelihood  =   -73302.997                     Prob > chi2      =    0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dep2	1.048716	.0353999	1.41	0.159	.9815786	1.120445
dep3	1.10618	.0383344	2.91	0.004	1.03354	1.183924
dep4	1.212892	.0437501	5.35	0.000	1.130104	1.301744
dep5	1.309478	.0513313	6.88	0.000	1.212638	1.414051

Flexible Parametric Proportional Hazards Model

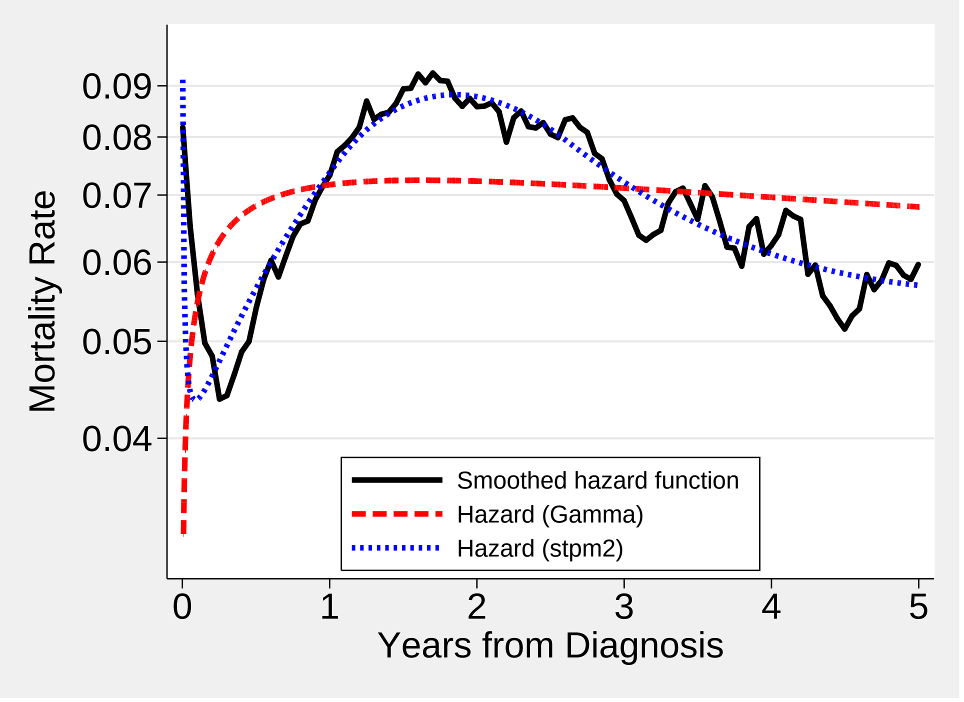
Flexible Parametric Proportional Hazards Model

.

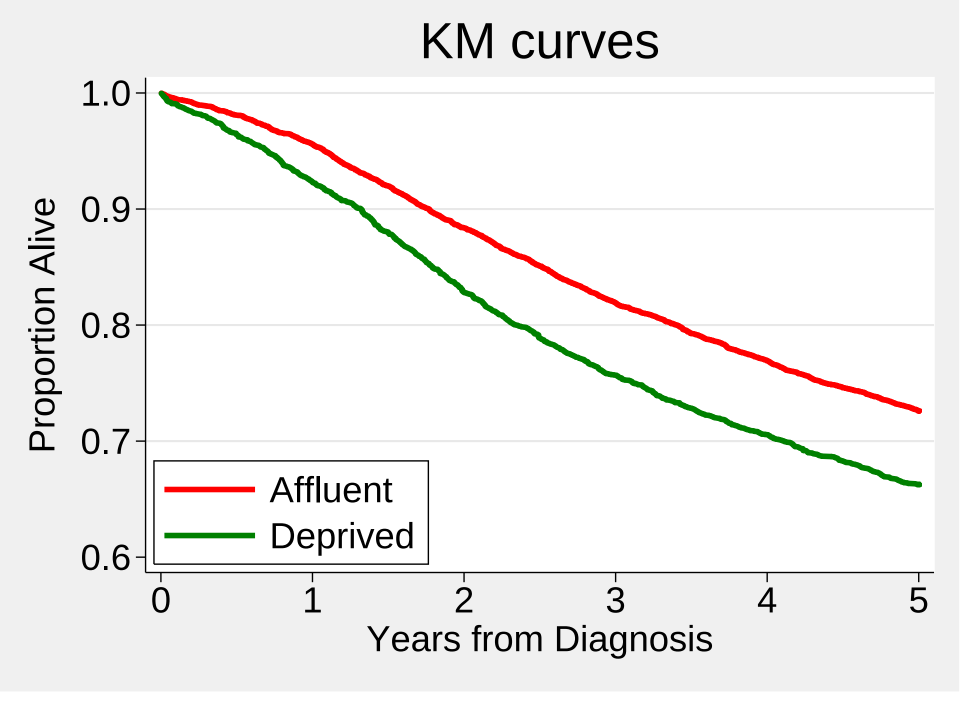
```
. stpm2 dep2-dep5, df(5) scale(hazard) eform
Iteration 0:   log likelihood = -22507.096
Iteration 1:   log likelihood = -22502.639
Iteration 2:   log likelihood = -22502.633
Iteration 3:   log likelihood = -22502.633
Log likelihood = -22502.633                        Number of obs   =      24889
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
dep2	1.048752	.0354011	1.41	0.158	.9816125	1.120483
dep3	1.10615	.0383334	2.91	0.004	1.033513	1.183893
dep4	1.212872	.0437493	5.35	0.000	1.130085	1.301722
dep5	1.309479	.0513313	6.88	0.000	1.212639	1.414052
_rcs1	2.126897	.0203615	78.83	0.000	2.087361	2.167182
_rcs2	.9812977	.0074041	-2.50	0.012	.9668927	.9959173
_rcs3	1.057255	.0043746	13.46	0.000	1.048715	1.065863
_rcs4	1.005372	.0020877	2.58	0.010	1.001288	1.009472
_rcs5	1.002216	.0010203	2.17	0.030	1.000218	1.004218

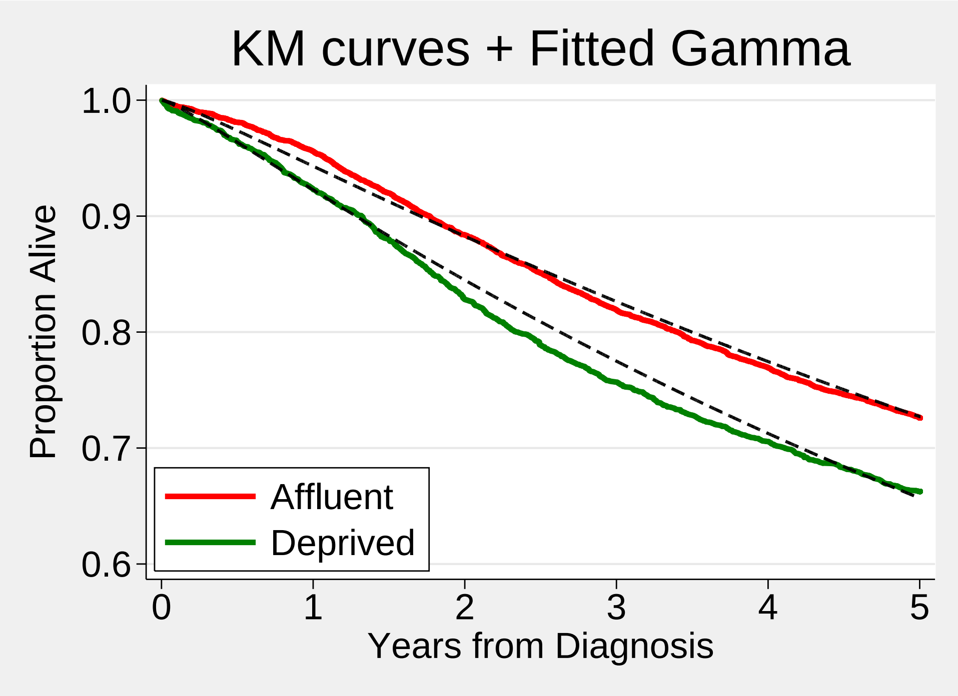
Why We Need Flexible Models



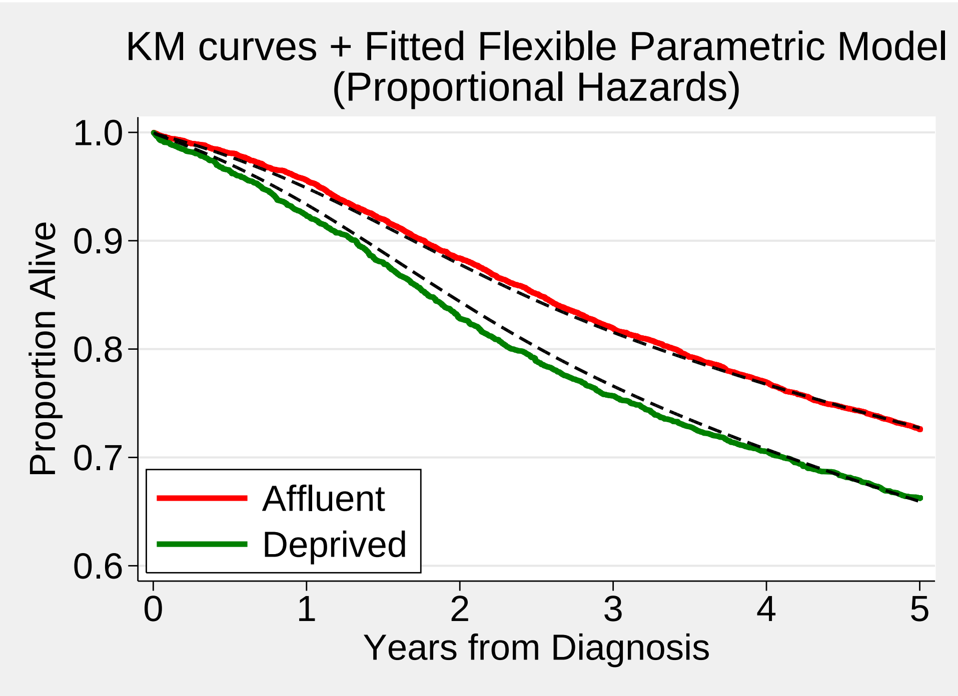
KM Curves



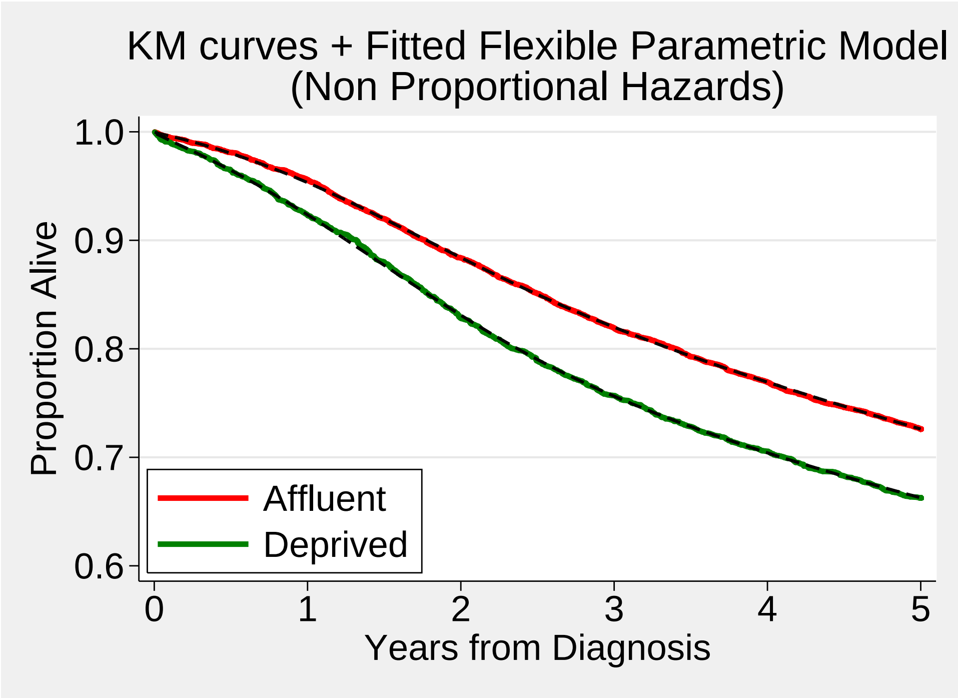
KM Curves + Gamma Model



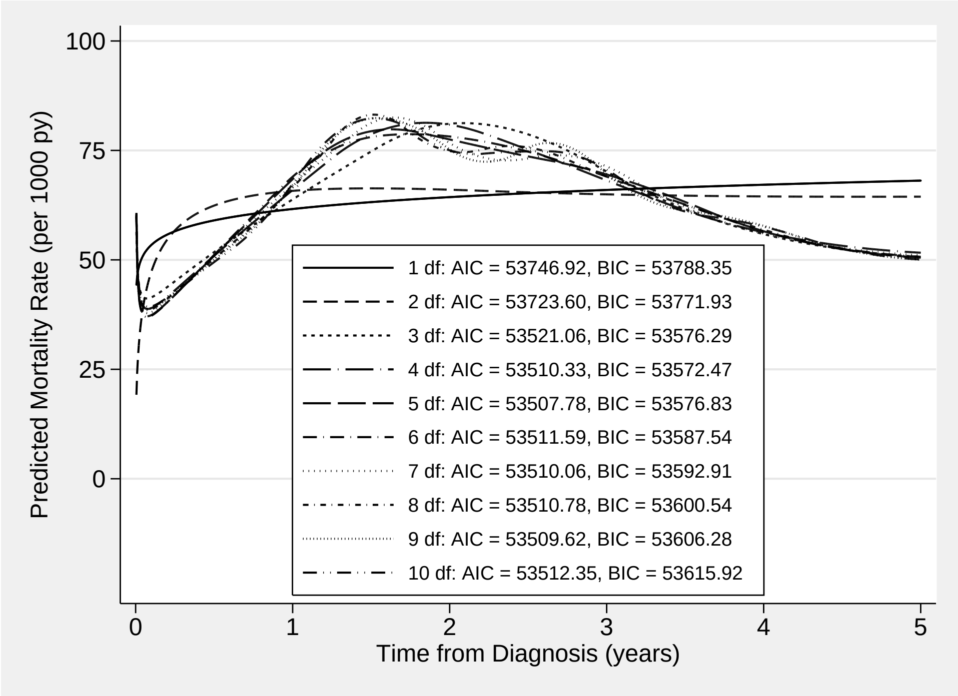
KM Curves + Flexible Parametric Model (PH)



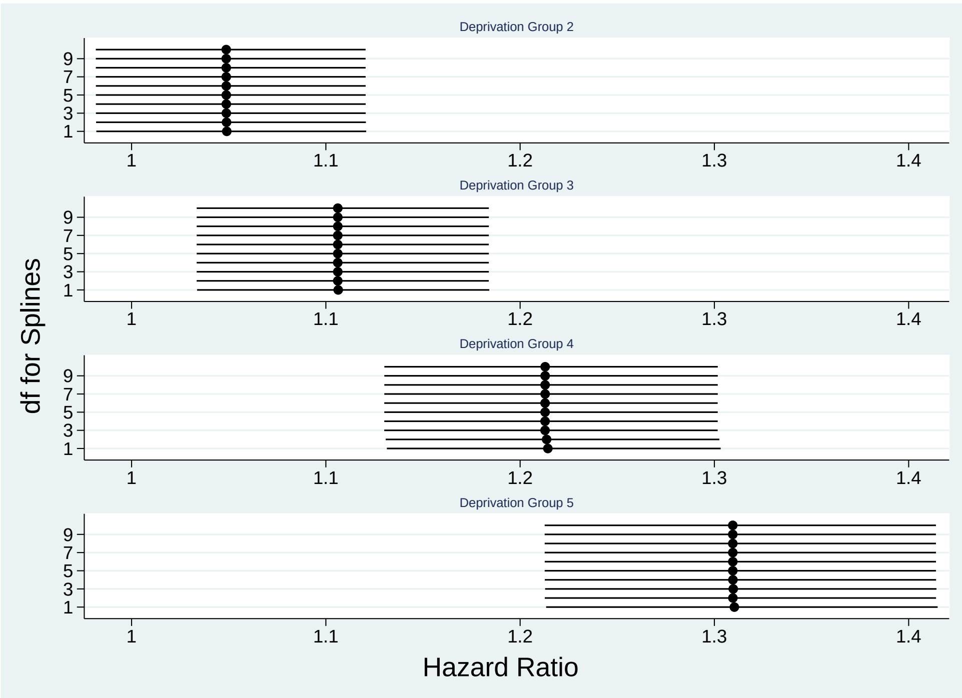
KM Curves + Flexible Parametric Model (Non-PH)



Example of different knots for baseline hazard



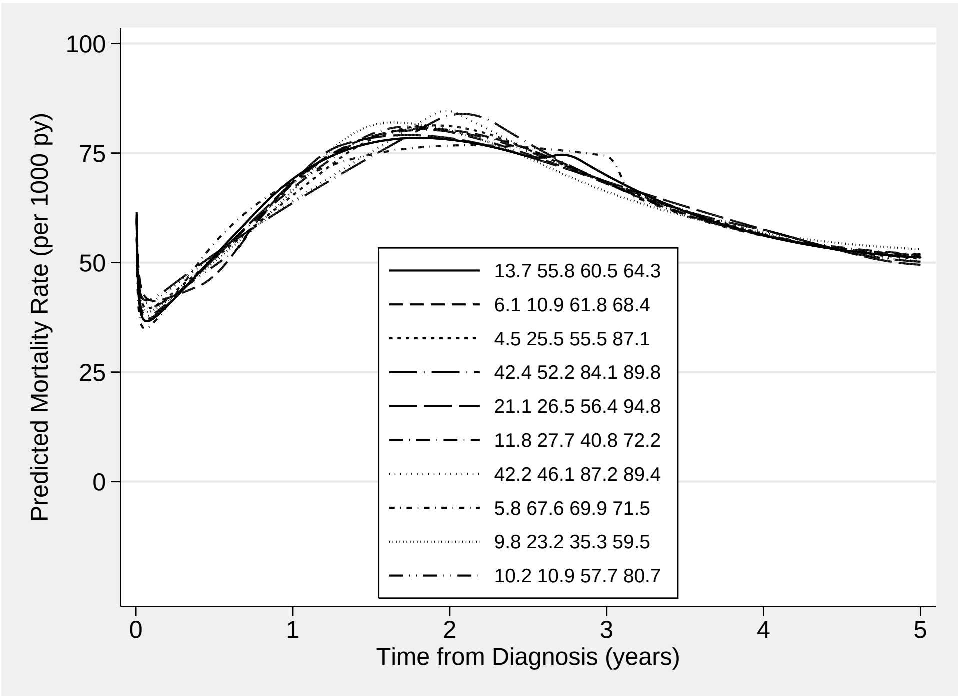
Effect of number of knots on hazard ratios



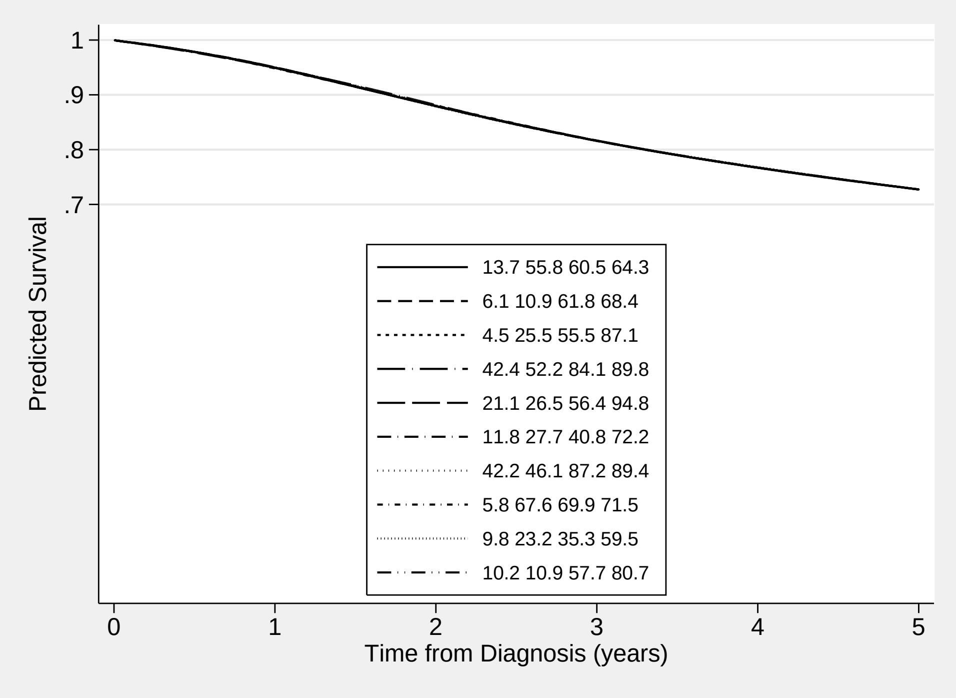
Where to place the knots?

- The knots are evenly spaced at the centiles of the uncensored survival times.
- The default knots positions tend to work fairly well.
- Unless the knots are in stupid places then there is usually very little difference in the fitted values.
- The graphs on the following page shows for 5 df (4 internal knots) the fitted hazard and survival functions with the internal knot locations randomly selected.

Baseline hazard - random knots



Baseline survival - random knots



Time-Dependent Effects

- The difference between two hazard rates may not be proportional.
- We can choose to,
  1. Ignore.
  2. Model on a different scale.
  3. Fit an interaction between the covariate and time.

Time-Dependent Effects

- A proportional hazards model can be written

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

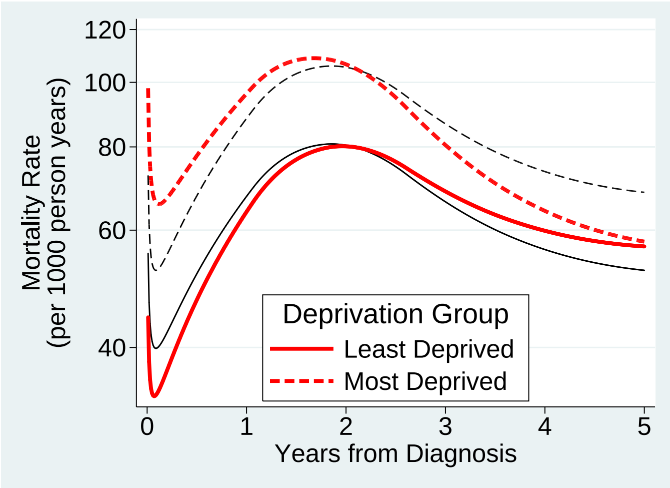
- With D time-dependent effects we write,

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

- There is a set of spline variables for each time-dependent effect.
- For any time-dependent effect there is an interaction between the covariate and the spline variables.
- The number of spline variables for a particular time-dependent effect will depend on the number of knots, $\mathbf{k}_j$

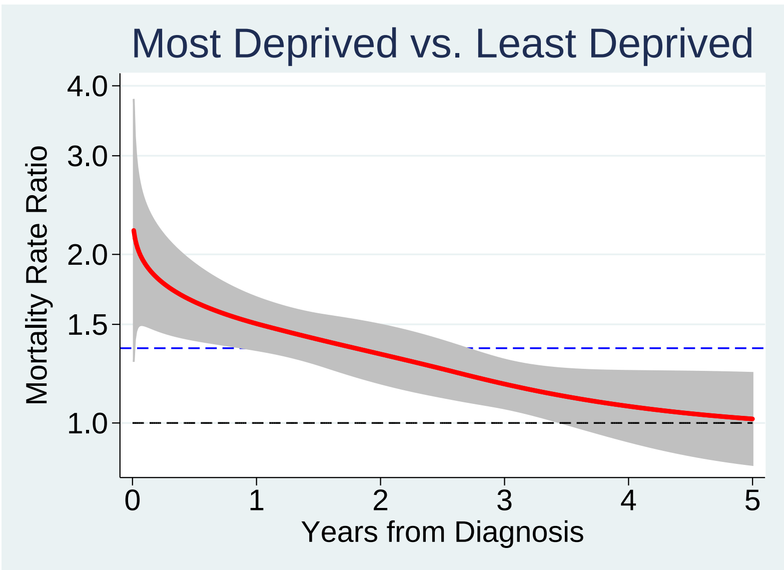
Predicted Hazard Rates

```
. stpm2 dep5, scale(hazard) df(5) tvc(dep5) dftvc(3)
. range temptime 0 5 200
. predict h1, hazard timevar(temptime) at(dep5 0) per(1000)
. predict h5, hazard timevar(temptime) at(dep5 1) per(1000)
```



Predicting Hazard Ratios

```
. stpm2 dep5, scale(hazard) df(5) tvc(dep5) dftvc(3)
. predict hr_tvc, hrnumerator(dep5 1) hrdenominator(dep5 0) ci
```

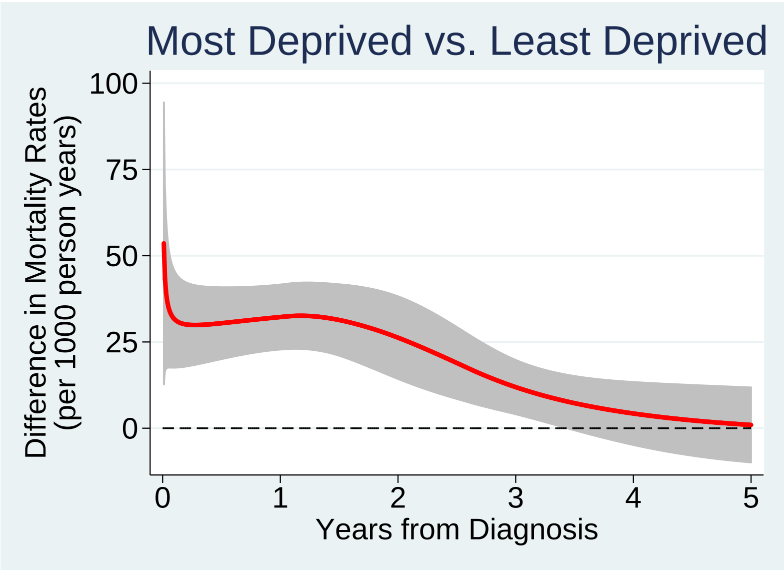


Quantifying Differences

- A key advantage of using a parametric model over the Cox model is that we can transform the model parameters to express differences between groups in different ways.
- The hazard ratio is a relative measure and a greater understanding of the impact of an exposure can be obtained by also looking at absolute differences.
- Use the delta-method to calculate confidence interavls.

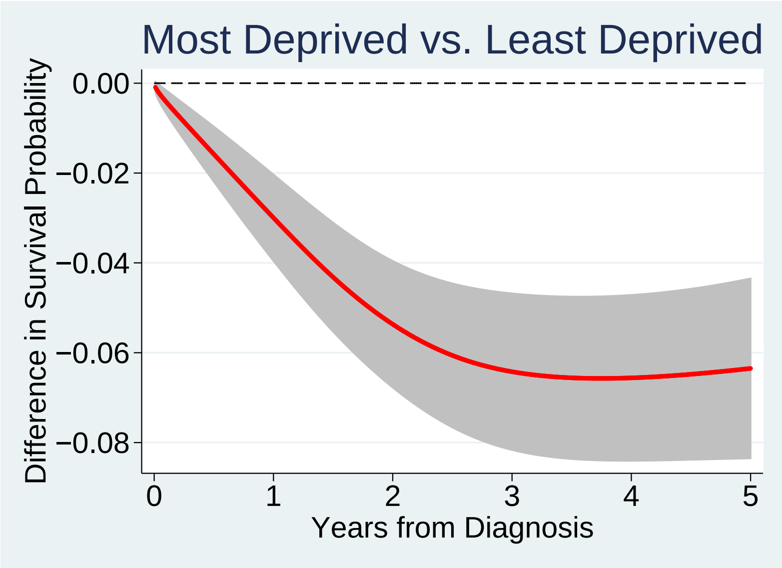
Difference in Hazard Rates

```
. predict hdiff, hdiff1(dep5 1) hdiff2(dep5 0) ci
```



Difference in Survival Proportions

```
. predict sdiff, sdiff1(dep5 1) sdiff2(dep5 0) ci
```



Example of Attained Age as the Time-scale

- Study from Sweden[5] 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.
- Outcome is for femoral neck fractures.
- Risk of fracture varies by age.
- Age is used as the main time-scale.
- Alternative way of "adjusting" for age.
- Gives the age specific incidence rates.

Estimates from a PH 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)
```

```
.  stcox noorc orc
```

Cox Model

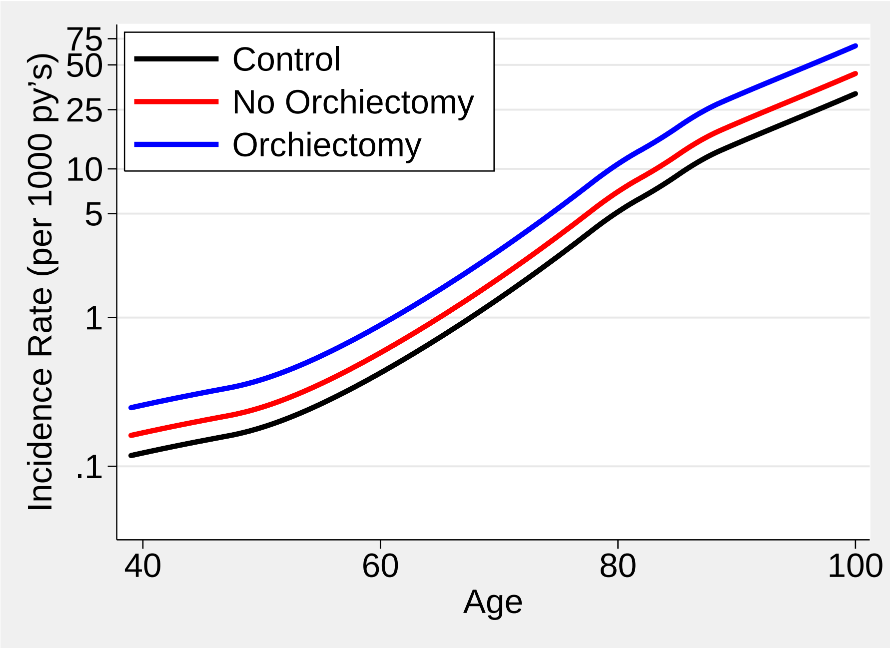
Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.10 (1.93 to 2.28)

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

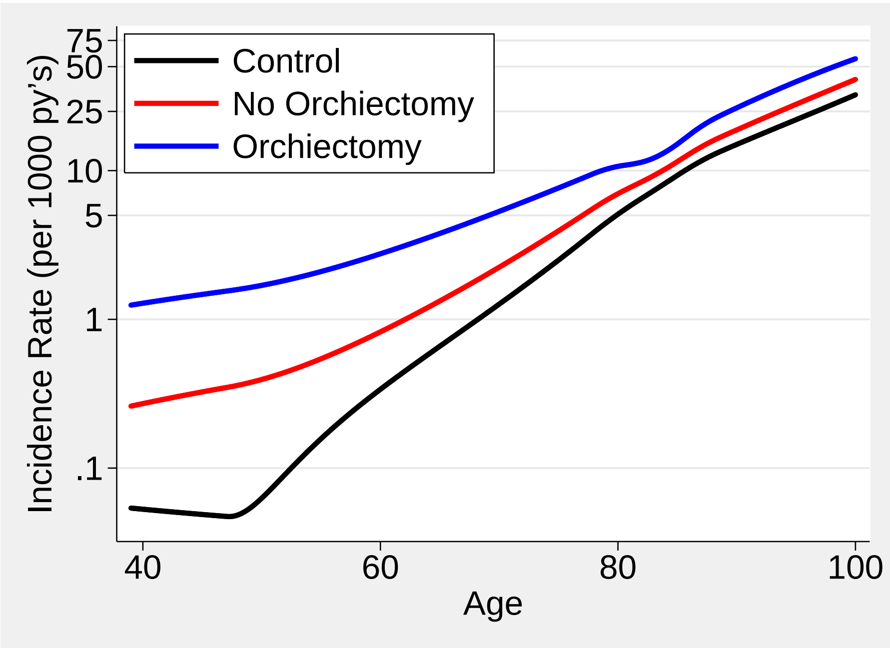
Flexible Parametric Model

Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.10 (1.93 to 2.28)

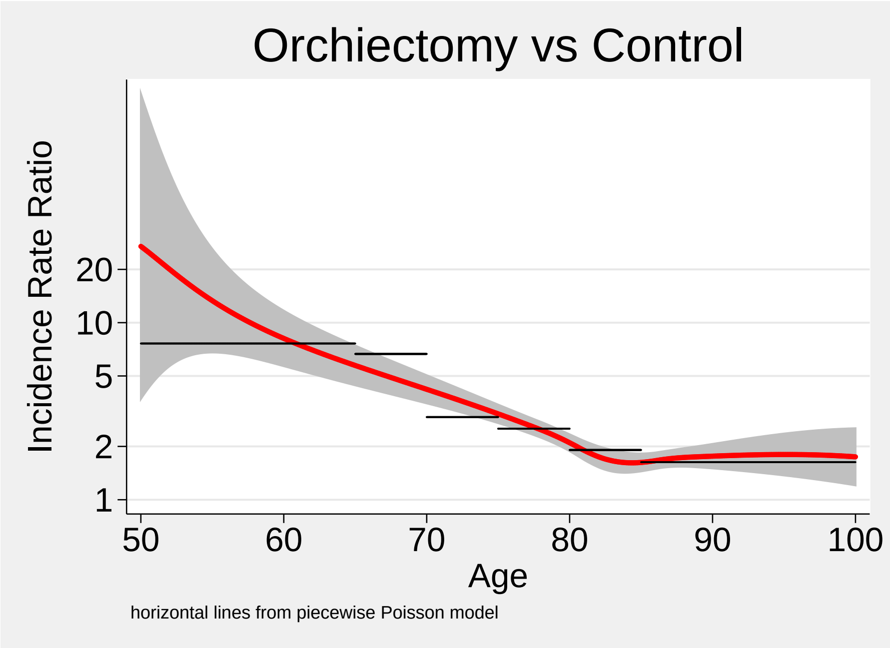
Proportional Hazards



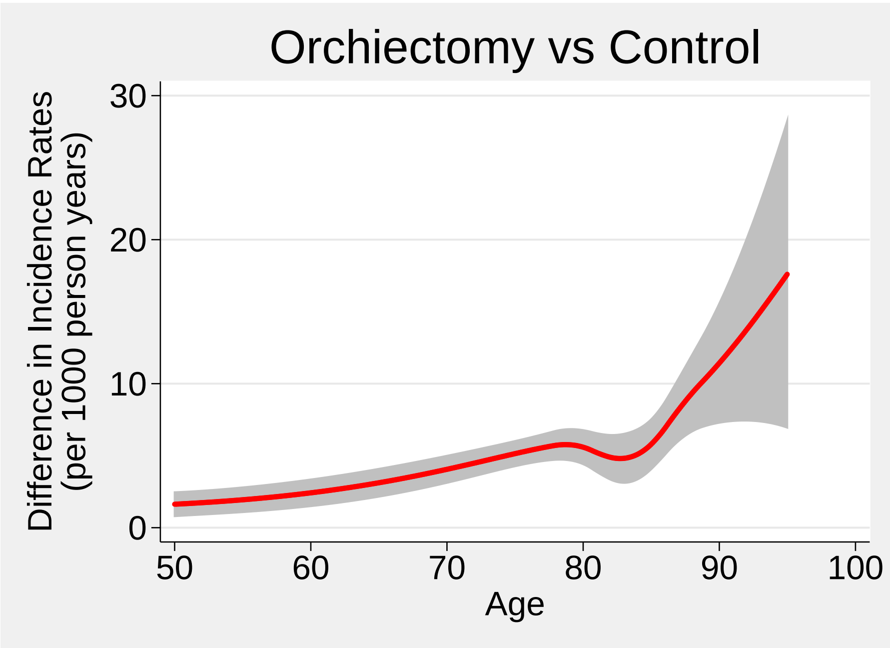
Non Proportional Hazards



Incidence Rate Ratio



Incidence Rate Difference



Relative Survival/Excess Mortality 1

- Relative Survival is used in population-based cancer studies.
- Growing interest in other disease areas: HIV [1], CHD [9].
- Relative Survival is used to measure mortality associated with a particular disease.
- Avoids needing information on cause of death.
- Important as cause of death may not be recorded or may be inaccurately recorded.
- We use expected mortality (from routine data sources).
- Flexible parametric models are easily extended to relative survival [8].

Relative Survival/Excess Mortality 1

- The total mortality (hazard) rate is the sum of two components.

Observed Mortality Rate	=	Expected Mortality Rate	+	*Excess Mortality Rate*
$h(t)$	=	$h^*(t)$	+	$\lambda(t)$

- If we transform to the survival scale,

$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \qquad R(t) = \frac{S(t)}{S^*(t)}$$

Likelihood for Relative Survival Models

Relative Survival Models

$$\ln L_i = d_i \ln(h^*(t_i) + \lambda(t_i)) + \ln(S^*(t_i)) + \ln(R(t_i))$$

- $S^*(t_i)$ does not depend on the model parameters and can be excluded from the likelihood.
- Merge in expected mortality rate at time of death, $h^*(t_i)$.
- This is important as many of other models for relative survival involve fine splitting of the time-scale and/or numerical integration. With large datasets this can be computationally intensive.
- Relative survival models can be fitted in `stpm2` by specifying the `bhazard()` option that gives the expected mortality rate at death.

Fitting Relative Survival Models using stpm2

- Analyse all 115,331 women diagnosed with breast cancer.
- Compare 5 age groups.

All Cause Survival

```
. stpm2 agegrp2-agegrp5, df(5) scale(hazard)
```

- For relative survival models, just add the `bhazard()` option.

Relative Survival

```
. stpm2 agegrp2-agegrp5, df(5) scale(hazard) bhazard(rate)
```


Hazard Ratios vs Excess Hazard Ratios

	All Cause Survival (Hazard Ratio)	**Relative Survival** (Excess Hazard Ratio)
< 50	-	-
50-59	**1.12** (1.08 to 1.15)	**1.05** (1.02 to 1.09)
60-69	**1.28** (1.25 to 1.32)	**1.07** (1.04 to 1.11)
70-79	**1.98** (1.92 to 2.04)	**1.41** (1.36 to 1.46)
80+	**4.15** (4.02 to 4.28)	**2.65** (2.55 to 2.75)

- The excess hazard ratios come from a poor fitting model.
- The effect of age is nearly always time-dependent.
- The inclusion of time-dependent effects is the same as for standard survival models.
- Relative and standard survival are now analysed within the same framework.

Crude Mortality

- Patient survival of interest to clinicians, patients, researchers, politicians, health administrators, and public health professionals [4].
- Little attention has been paid to the fact that each of these consumers of survival statistics have quite different needs.
- The standard approach of estimating net survival (relative survival or cause-specific survival) is useful for comparing populations but not necessarily relevant to individual patients.

Interpreting Relative Survival

- The cumulative relative survival ratio can be interpreted as the proportion of patients alive after t years of follow-up in the hypothetical situation where the cancer in question is the only possible cause of death.
- Same interpretation for cause-specific survival.
- None of us live in this hypothetical world.
- An individual should understand their personal risk, which includes their risk of dying of other causes.
- To calculate "real world" probabilities we need to borrow ideas from competing risks theory.

Net and Crude Mortality

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

- Some people refer to the crude probability as cumulative incidence.

Life table calculation of crude mortality

- Cronin and Feuer[3] showed how crude mortality due to cancer and due to other causes can be calculated from life tables.
- Available in Paul Dickman's `strs` command.
- Calculated separately in age groups.
- Time-scale split into large (yearly) time intervals.
- No individual level prediction using continuous covariate.

Crude Mortality in Relative Survival Models

- Crude mortality can be estimated after fitting a relative survival model.
- The fitting of the relative survival model is not any different, but we do some tricky calculations postestimation.
- The flexible parametric models allow individual level covariates to be modelled.
- See Lambert it et al.[6] for details.

Brief Mathematical Details

$h^*(t)$	-	Expected mortality rate
$\lambda(t)$	-	Excess mortality rate
$h(t) = h^*(t) + \lambda(t)$	-	All-cause mortality rate
$S^*(t)$	-	Expected Survival
$R(t)$	-	Relative Survival
$S(t) = S^*(t)\lambda(t)$	-	Overall Survival

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

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

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

Integrating

- The integration is performed numerically by splitting the time-scale into a large number, n, of small intervals (e.g. 1000).
- The predicted value of the integrand at each of the n values of t is obtained.
- The crude probability of death is the sum of the these predicted values.
- The variance is a bit trickier, as the observation-specific derivatives need to be obtained. These are calculated numerically (Stata's `predictnl` command).
- The approach is similar to that used by Carstensen when calculating survival functions from Poisson based survival models[2]

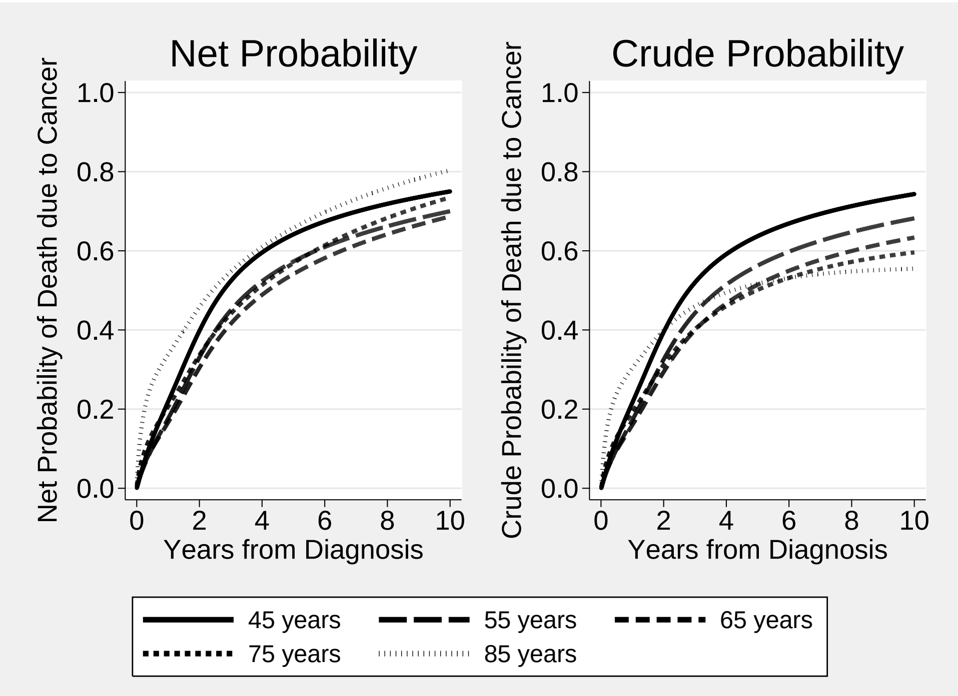
Example

- 28,943 men diagnosed with prostate cancer aged 40-90 in England and Wales between 1986-1988 inclusive and followed up to 1995.
- Restricted cubic splines are used to
  - Model the baseline excess hazard (6 knots).
  - Model the main effect of age (4 knots).
  - Model time-dependence of age (4 knots).

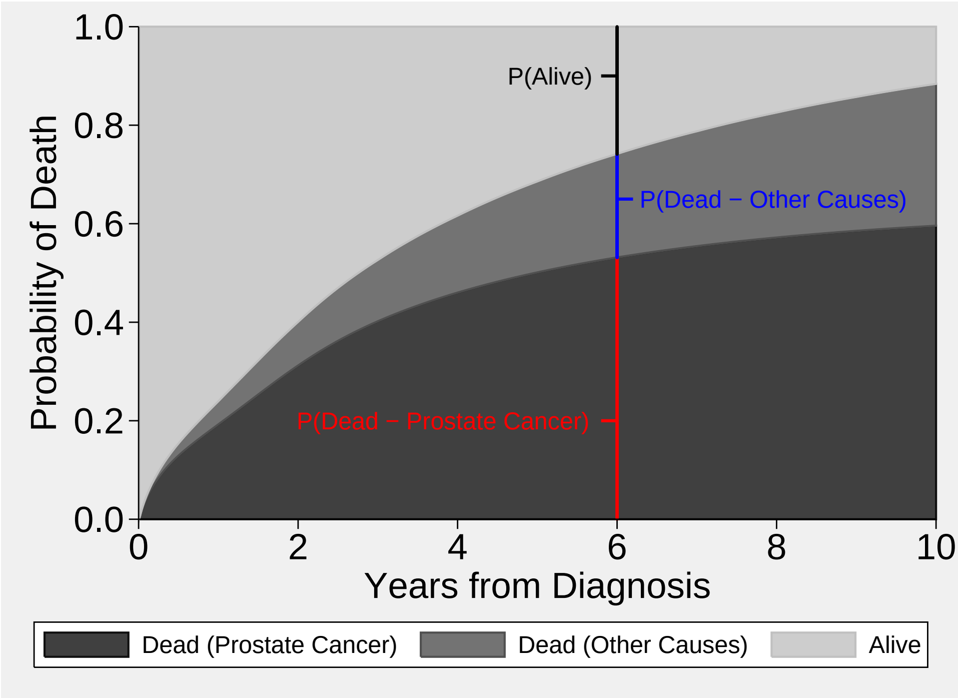
Splines, Splines, Splines

```
. rcsgen agediag, gen(agercs) df(3) orthog
. stpm2 agercs1-agercs3, scale(h) df(5) bhazard(rate) ///
        tvc(agercs1-agercs3) dftvc(3)
```

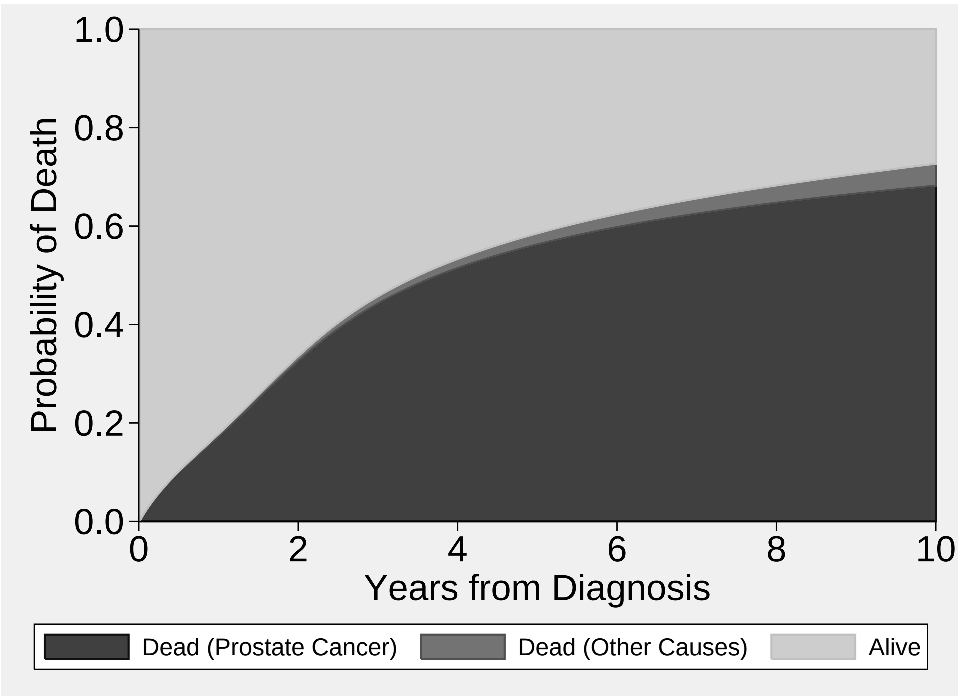

Net and Crude Probability of Death



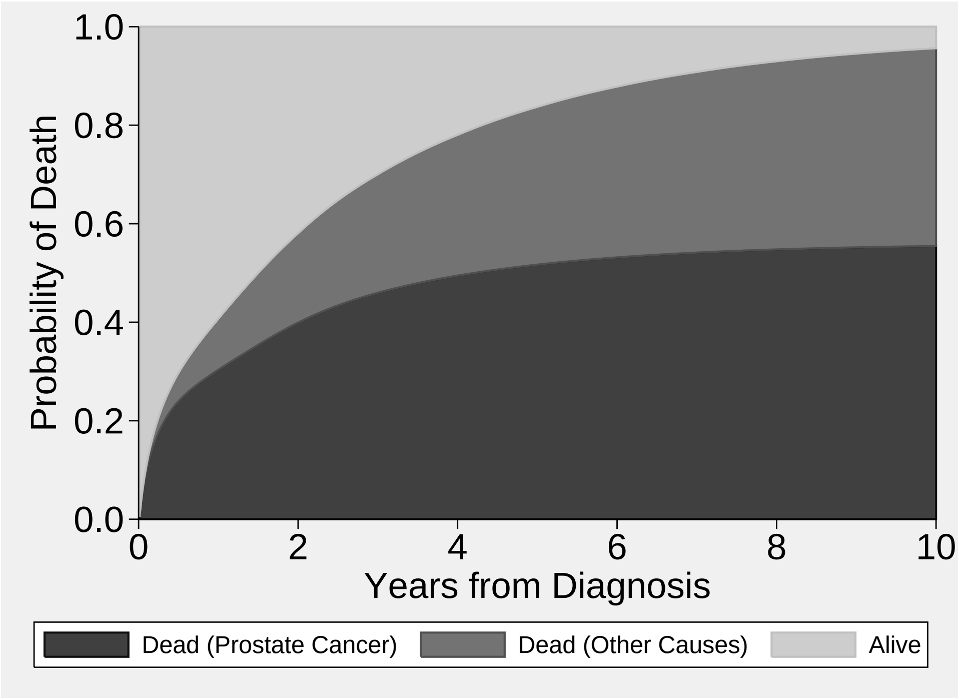
Predictions for a 75 year old man



Predictions for a 55 year old man



Predictions for a 85 year old man



Conclusion

- Flexible parametric PH models give essentially the same estimates as a Cox model.
- Predictions are much easier from a parametric model.
- Simple extension to time-dependent effect and relative survival.
- User friendly software available, Lambert *et al.* 2009 [7].

Further Extensions

- Univariate and shared frailty models.
- Multiple Events.
- Competing Risks.
- Cure models.
- Estimate loss in expectation in life.
- Enhance ability to model multiple time-scales.

References I

[1] K. Bhaskaran, O. Hamouda, M. Sannes, F. Boufassa, A. M. Johnson, P. C. Lambert, K. Porter, and C. A. S. C. A. D. E. Collaboration. Changes in the risk of death after hiv seroconversion compared with mortality in the general population. *JAMA*, 300(1):51–59, Jul 2008.

[2] B. Carstensen. Demography and epidemiology: Practical use of the lexis diagram in the computer age or: Who needs the cox-model anyway? Technical report, Department of Biostatistics, University of Copenhagen, 2006.

[3] K. A. Cronin and E. J. Feuer. Cumulative cause-specific mortality for cancer patients in the presence of other causes: a crude analogue of relative survival. *Statistics in Medicine*, 19(13):1729–1740, Jul 2000.

[4] P. W. Dickman and H.-O. Adami. Interpreting trends in cancer patient survival. *J Intern Med*, 260(2):103–117, Aug 2006.

[5] P. W. Dickman, J. Adolfsson, K. Astrm, and G. Steineck. Hip fractures in men with prostate cancer treated with orchiectomy. *Journal of Urology*, 172(6 Pt 1):2208–2212, Dec 2004.

References II

[6] P. C. Lambert, P. W. Dickman, C. P. Nelson, and P. Royston. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Statistics in Medicine*, (in press), 2009.

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

[8] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Flexible parametric models for relative survival, with application in coronary heart disease. *Statistics in Medicine*, 26(30):5486–5498, Dec 2007.

[9] C. P. Nelson, P. C. Lambert, I. B. Squire, and D. R. Jones. Relative survival: what can cardiovascular disease learn from cancer? *European Heart Journal*, 29(7):941–947, Apr 2008.

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

[11] P. Royston and M. K. B. Parmar. 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*, 21(15):2175–2197, Aug 2002.

Sensitivity to the number of knots

- A potential criticism of these models is the subjectivity in the number and the location of the knots.
- A small sensitivity analysis was carried out where the following models were fitted.

Model	Baseline df_b	Time-dependent df_t	age df_a	No. of Parameters	AIC	BIC
Model (a)	5	3	3	18	97250.11	97399.02
Model (b)	8	5	5	39	97059.30	97381.95
Model (c)	5	5	3	24	97235.68	97434.23
Model (d)	3	3	3	16	97447.35	97579.72
Model (e)	8	8	8	81	97105.8	97775.92

Knot sensitivity analysis