Modelling Relative Survival: Flexible Parametric Models and the Estimation of Net and Crude Mortality.

Paul C Lambert[1,2],
Paul W Dickman[2], Christopher P Nelson[1], Patrick Royston[3]

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

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

[3]MRC Clinical Trials Unit, London, UK

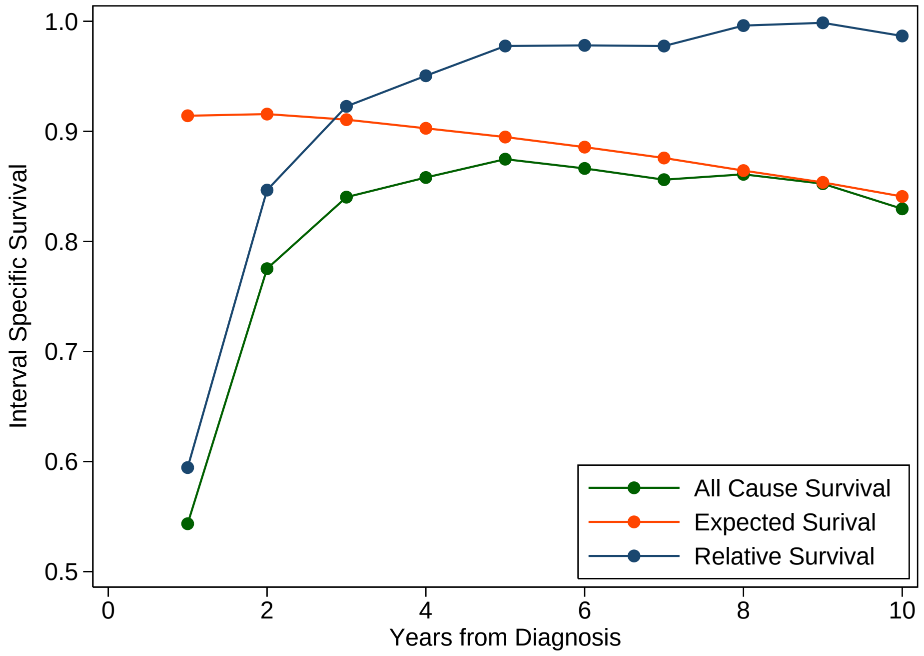
Seminar 18/3/2009
London School of Hygiene and Tropical Medicine

What is Relative Survival

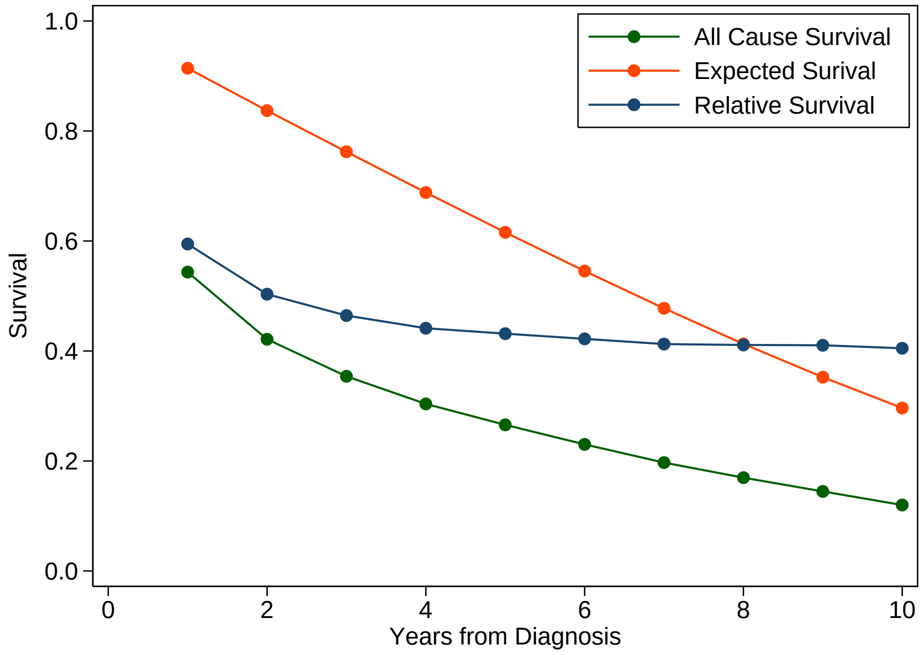
$$\text{Relative Survival} = \frac{\text{Observed Survival}}{\text{Expected Survival}} \qquad R(t) = S(t)/S^*(t)$$

- Expected survival obtained from national population life tables stratified by age, sex, calendar year, other covariates.
- Estimate of mortality associated with a disease without requiring information on cause of death.
- Traditionally estimated in life tables.

Interval Specific Survival - Colon Cancer - Finland - 75+



Cumulative Survival - Colon Cancer - Finland - 75+



Excess Mortality

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

- Transforming to the hazard scale gives

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

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

- We are usually interested in modelling on the log excess hazard scale.
- Assume that competing risks of other causes are acting independently.

Why do we use Relative Survival

- Estimate of mortality associated with a diagnosis of a particular cancer without the need for cause of death information.
- If we had perfect cause-of-death information then treat those that die from another cause as censored at their time of death.
- The quality of cause-of-death information varies over time, between types of cancer and between regions/countries.
- Many cancer registries do not record cause of death.
- Cause of death is rarely a simple dichotomy.

Net Survival

- Relative survival and cause-specific survival are both estimates of net survival.
- Net survival is the probability of survival in the hypothetical situation where the cancer of interest is the only possible cause of death, i.e in the absence of other causes.
- This is useful for national and international comparisons, changes in survival over calendar time etc.
- However, a patient and the treating clinician are also interested in the probability of death in the presence of other causes.
- I will return to this later.
- Assume that the estimated expected survival is appropriate for the group in question and that non-cancer mortality is independent of cancer mortality.

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[10].
- Smooth estimates of the hazard and survival functions.
- The hazard function is of particular interest.
- Also useful for modelling complex time-dependent functions.
- In their simple form they assume proportional hazards.

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 $\ln(t)$.

Why model on the log cumulative hazard scale? I

- We are used to modeling on the log hazard scale, so why model on the log cumulative hazard scale?
- Under the proportional hazards assumption covariate effects can still be interpreted as hazard ratios.

$$h_i(t|\mathbf{x}_i) = h_0(t)\exp(\mathbf{x}_i\beta) \qquad H_i(t|\mathbf{x}_i) = H_0(t)\exp(\mathbf{x}_i\beta)$$

- It is easy to transform to the survival and hazard functions.

$$S(t) = \exp[-H(t)] \qquad h(t) = \frac{d}{dt}H(t)$$

- The log cumulative hazard as a function of log time is a generally stable function, e.g. in all Weibull models it is a straight line. It is easier to capture the shape of simple functions.

Cubic Splines

- Flexible mathematical functions defined by piecewise polynomials.
- The points at which the polynomials join are called knots.
- Constraints ensure the function is smooth.
- The most common splines used in practice are cubic splines.
- Function is forced to have continuous 0^{th}, 1^{st} and 2^{nd} derivatives.
- Regression splines can be incorporated into any regression model with a linear predictor.
- Restricted cubic splines are forced to be linear beyond the first and last knots[4].
- Restricted cubic splines are used in the models described here.

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.

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

- For example, with 4 knots we can write

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

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

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.

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

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

Likelihood

- The general log-likelihood for a survival model can be written

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

- Thus

$$\ln L_i = d_i \left(\ln\left[s'(\ln(t)|\gamma, \mathbf{k}_0)\right] + \eta_i\right) - \exp(\eta_i)$$

- The likelihood can be maximized (using a few tricks) using Stata's optimizer, `ml`.
- This is implemented in `stpm2`. A description of this command will shortly appear in *The Stata Journal* [8].

Breast Cancer Example

- Data were obtained from the public-use data set of all England and Wales cancer registrations between 1 January 1971 and 31 December 1990 with follow-up to 31 December 1995[1].
- As an example I will investigate the effect of deprivation (based Carstairs score[1]) on all-cause mortality in women who were diagnosed with breast cancer under the age of 50 years.
- There are five deprivation groups ranging from the least deprived (affluent) to the most deprived quintile in the population.

Fitting a Proportional Hazards Model I

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.

Fitting a Proportional Hazards Model II

Proportional hazards models

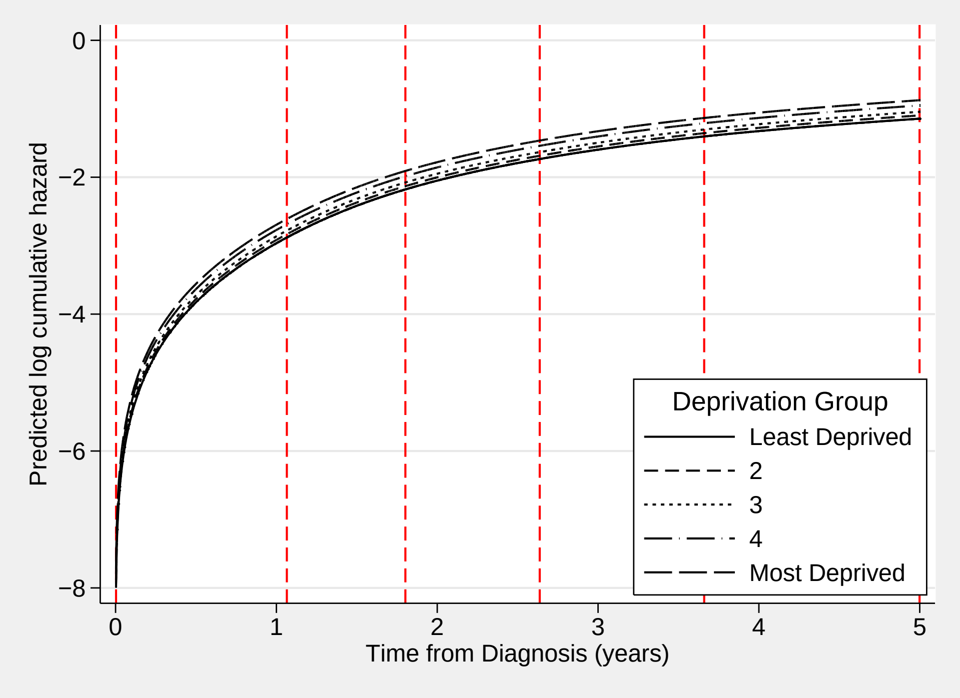
```
.

. stcox dep2-dep5, noshow nolog
No. of subjects =        24889                  Number of obs   =      24889
                                                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
             |
------------------------------------------------------------------------------
. stpm2 dep2-dep5, df(5) scale(hazard) eform nolog
                                               Number of obs   =      24889
                                               Wald chi2(4)    =      63.32
Log likelihood = -22502.633                    Prob > chi2     =     0.0000
------------------------------------------------------------------------------
             |     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
------------------------------------------------------------------------------
```

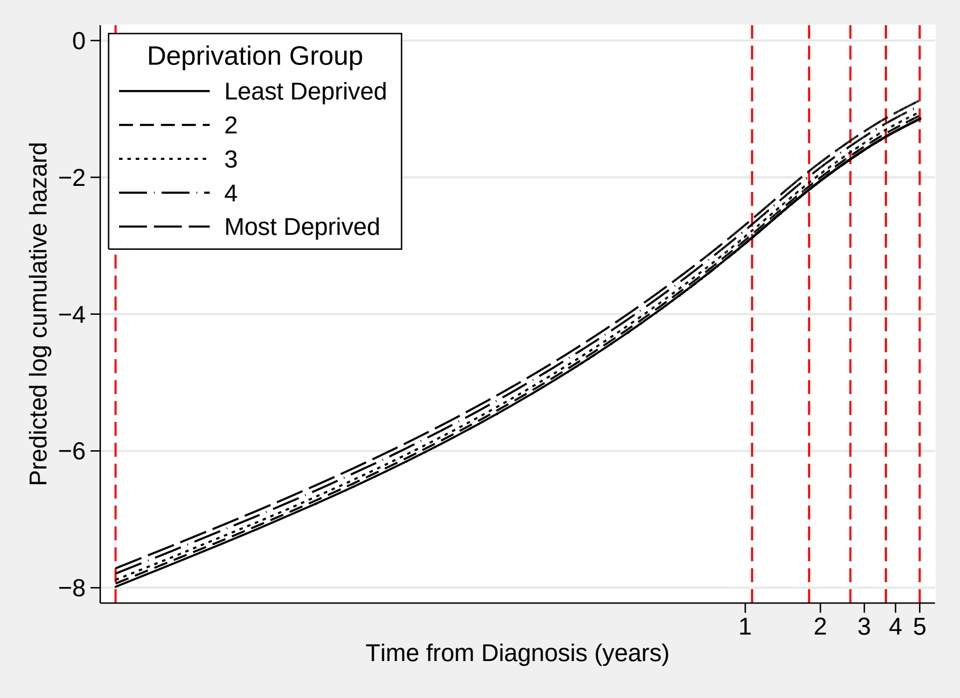
Proportional hazards models

- The estimated hazard ratios and their 95% confidence intervals are very similar.
- I have yet to find an example of a proportional hazards model, where there is a large difference in the estimated hazard ratios.
- If you are just interested in hazard ratios in a proportional hazards model, then you can get away with poor modelling of the baseline hazard.
- One important exception is when the follow-up time differs between groups.
- It is of course better to model the baseline hazard well!

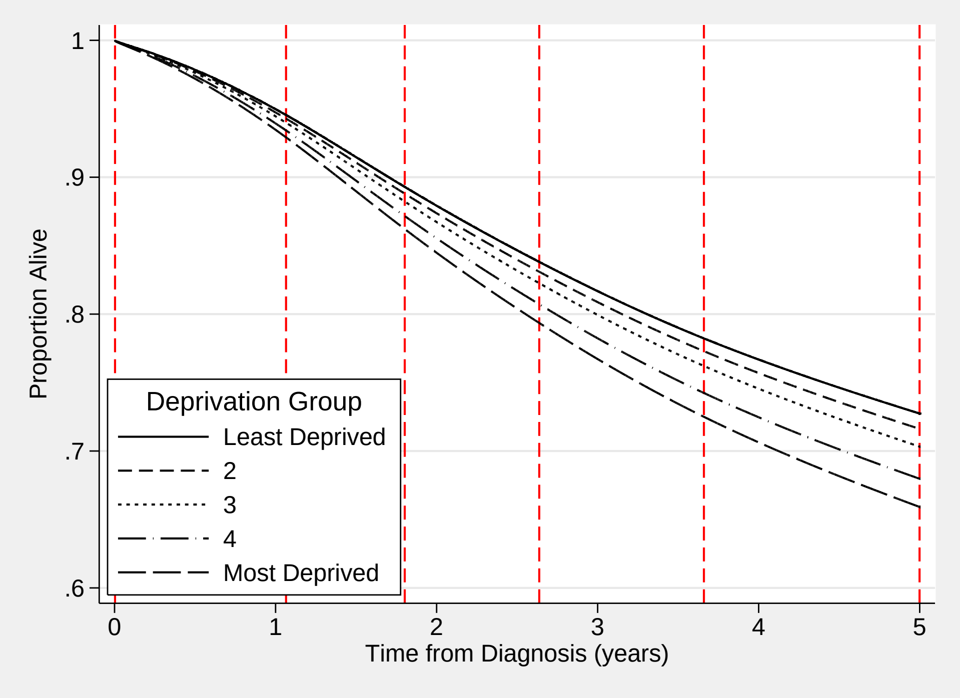
Log Cumulative Hazard



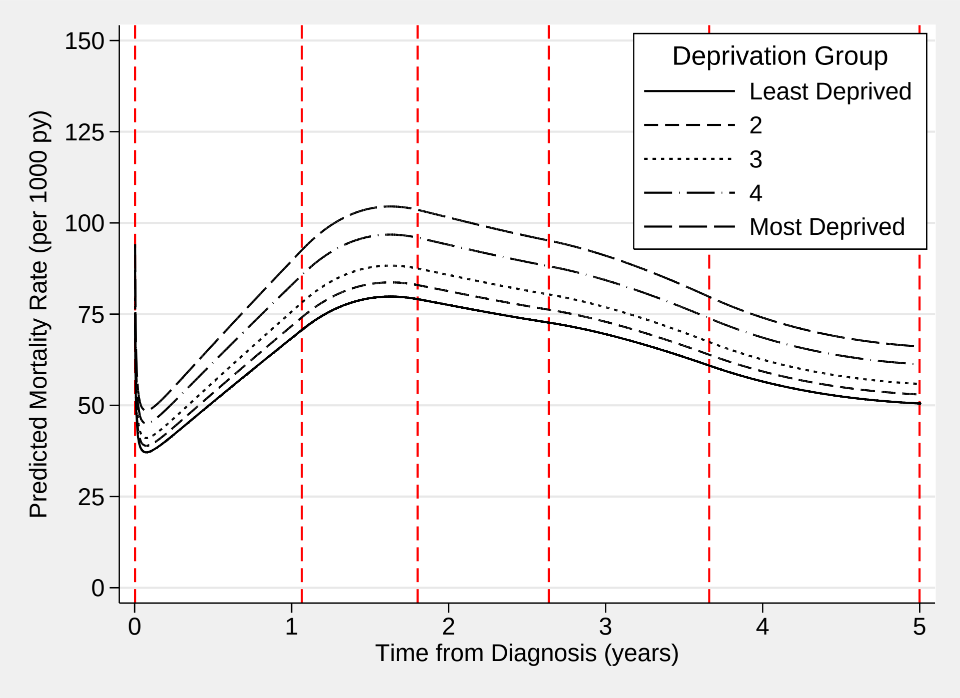
Log Cumulative Hazard vs log(time)



Survival Function



Hazard Function ×1000



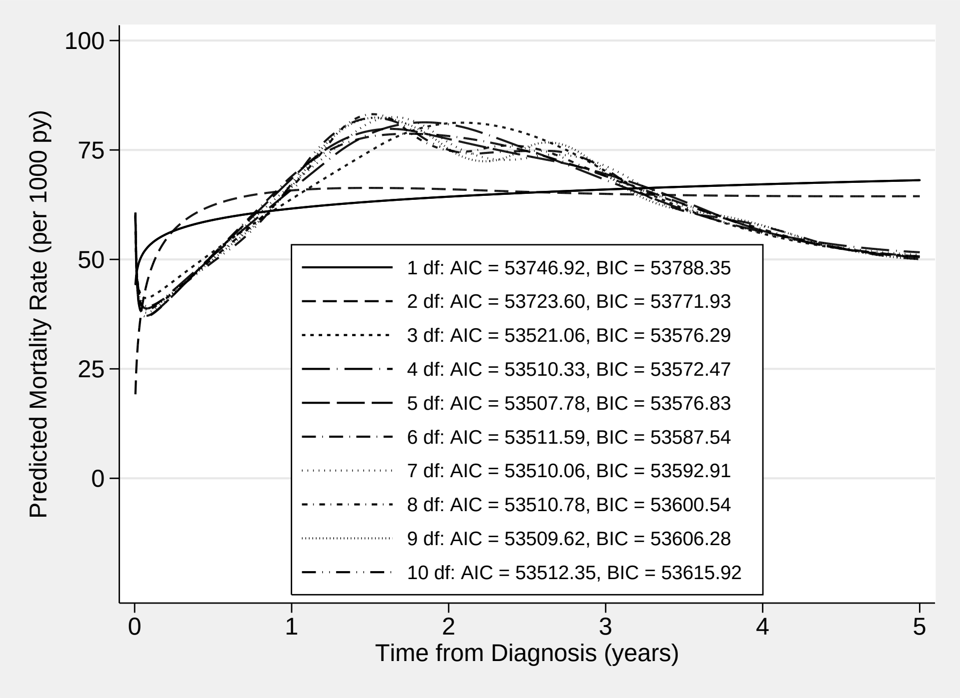
Location of the Knots

- The `df()` option specifies that a certain number of knots are to be used using the default locations.
- If `df(1)` is specified then the log cumulative hazard function is assumed to be a linear function of $\ln(t)$, i.e. a Weibull model.

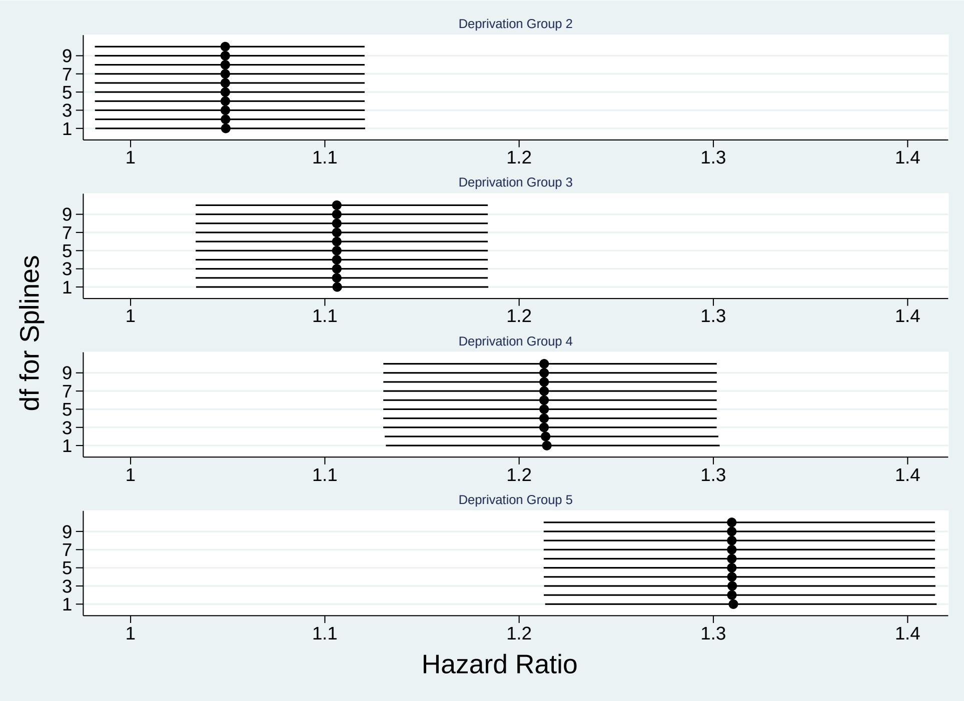
Knots	df	Centiles
1	2	50
2	3	33, 67
3	4	25, 50, 75
4	5	20, 40, 60, 80
5	6	17, 33, 50, 67, 83
6	7	14, 29, 43, 57, 71, 86
7	8	12.5, 25, 37.5, 50, 62.5, 75, 87.5
8	9	11.1, 22.2, 33.3, 44.4, 55.6, 66.7, 77.8, 88.9
9	10	10, 20, 30, 40, 50, 60, 70, 80, 90

Default positions of internal knots for modelling the baseline distribution function and time-dependent effects in Royston-Parmar models. Knots are positions on the distribution of uncensored log event-times

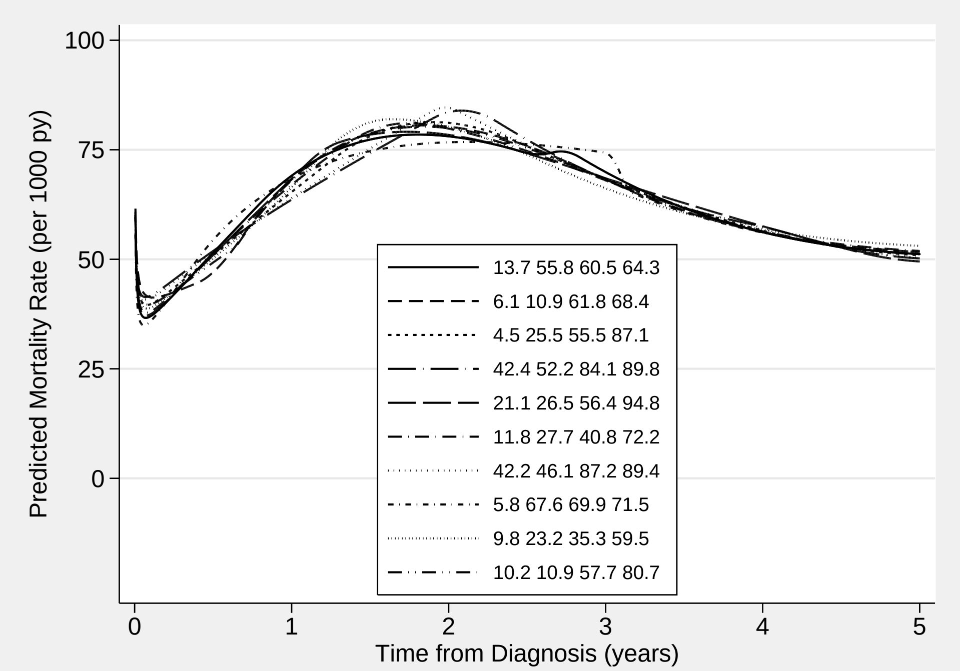
Example of different knots for baseline hazard



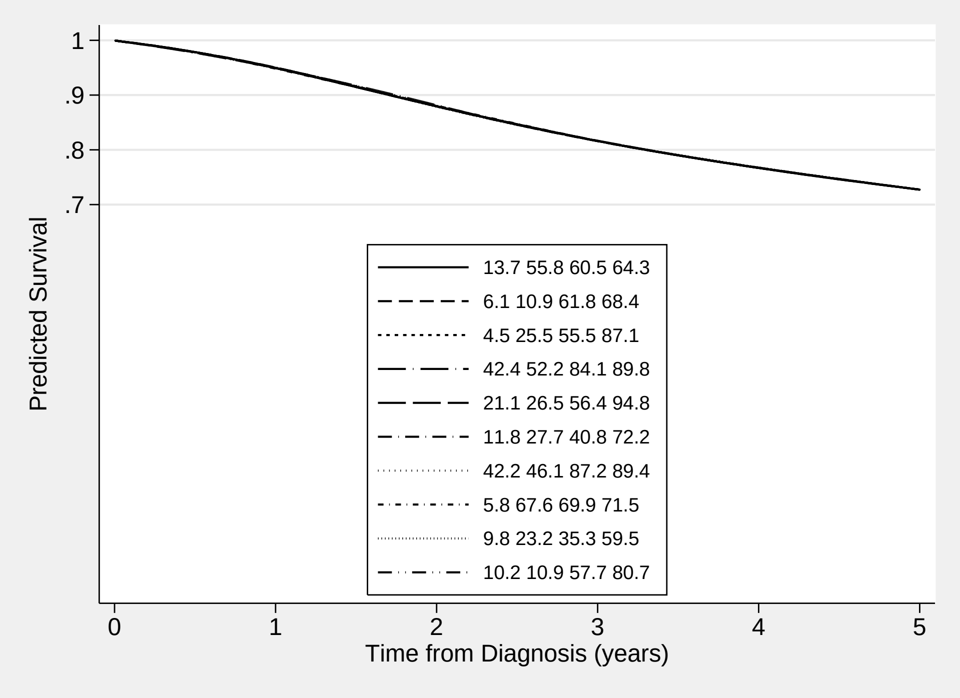
Effect of number of knots on hazard ratios



Effect of location of knots on baseline hazard - random knots



Effect of number of knots on baseline survival



Time-Dependent Effects I

- A proportional cumulative hazards model can be written

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

- There is a new set of spline variables for each time-dependent effect.
- If there are D time-dependent effects then

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

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

Example

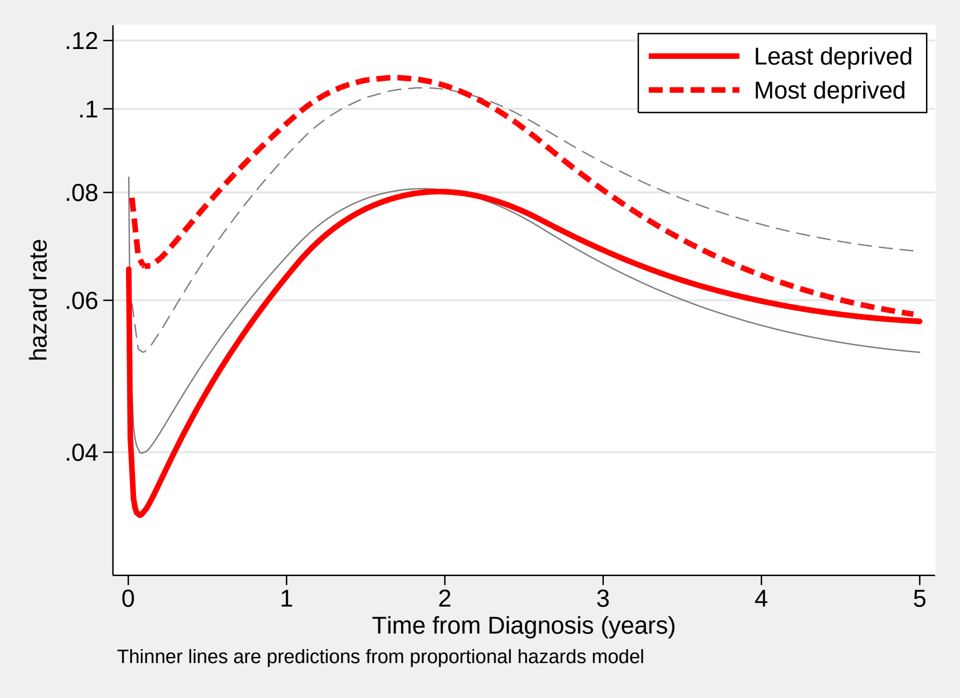
- I will look at the time-dependent effect of deprivation. For simplicity I will initially consider a model comparing the most deprived with the least deprived group.
- Time-dependent effects are fitted using the `tvc()` and `dftvc()` options.
- The `dftvc()` option controls the number of knots in the same wayas for the baseline hazard. Note that `dftvc(1)` means that the time-dependent effect is modelled as a function of log time.

Non-proportional hazards models

```
.
stpm2 dep5, scale(hazard) df(5) tvc(dep5) dftvc(3)
```

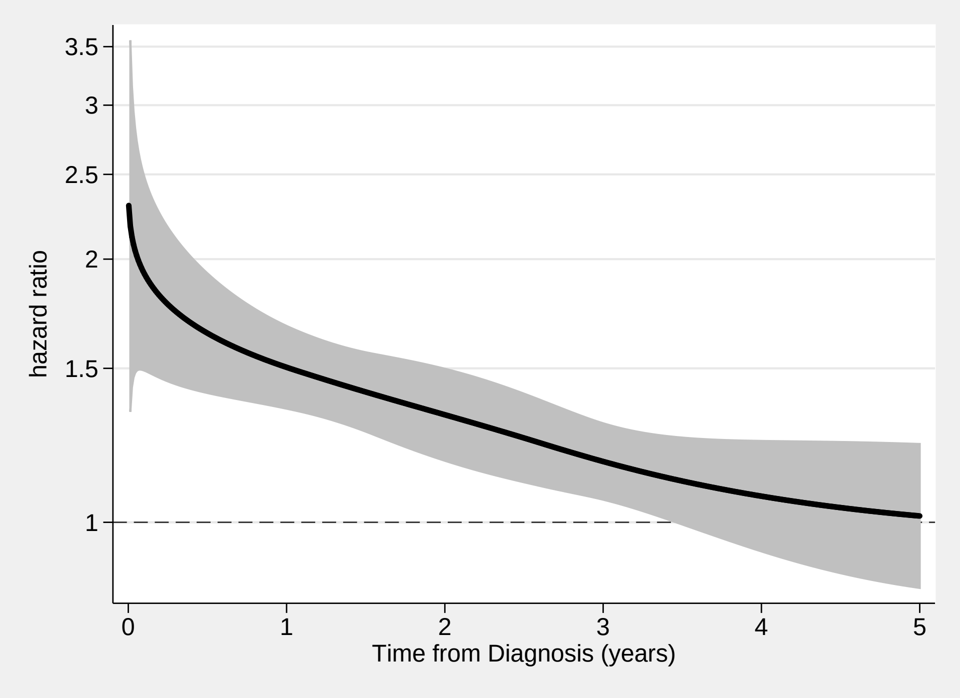
- Differences quantified by (time-dependent) hazard ratios, hazard differences and survival differences.

Estimated Hazard Functions

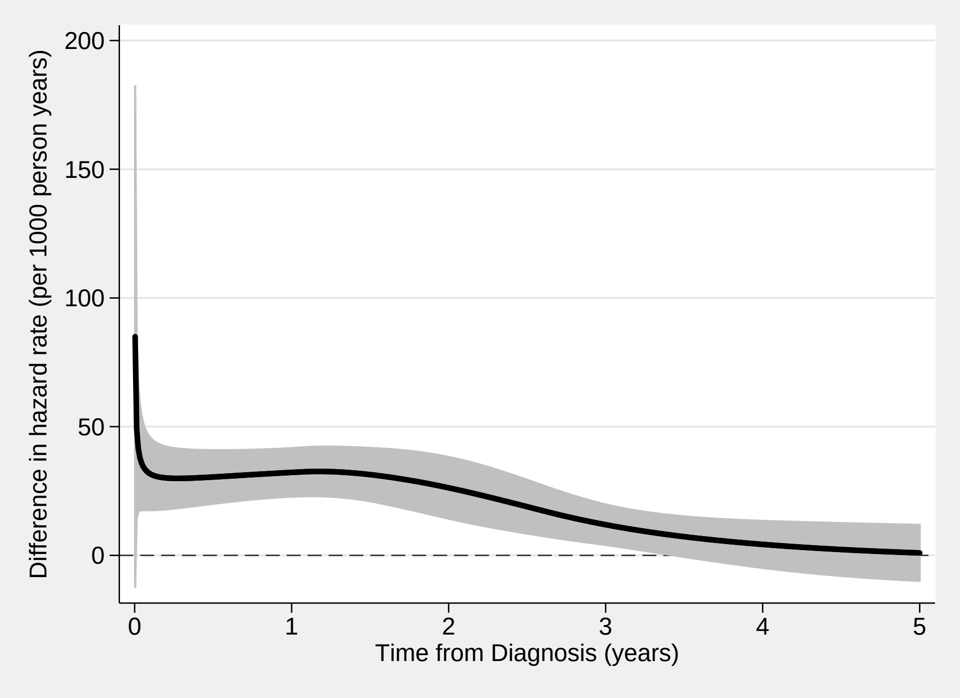


Thinner lines are predictions from proportional hazards model

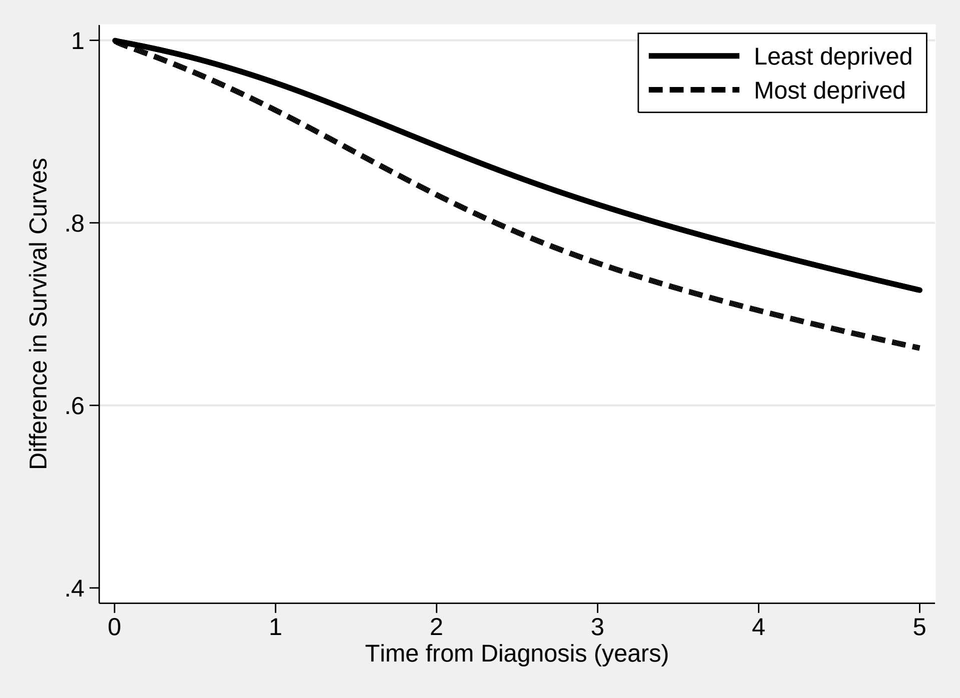
Estimated Hazard Ratio



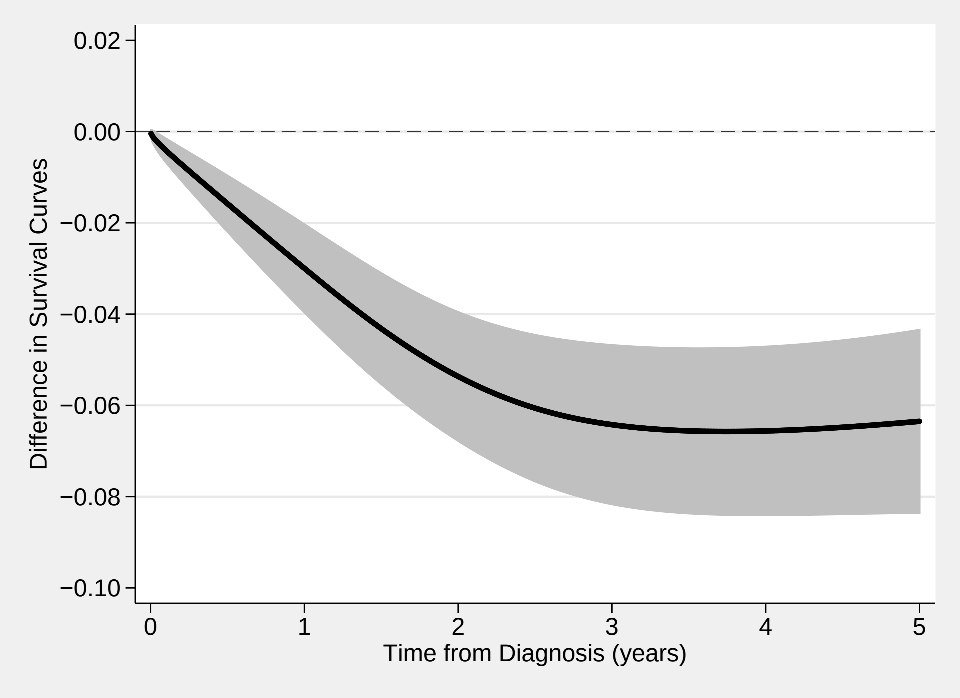
Estimated Hazard Difference



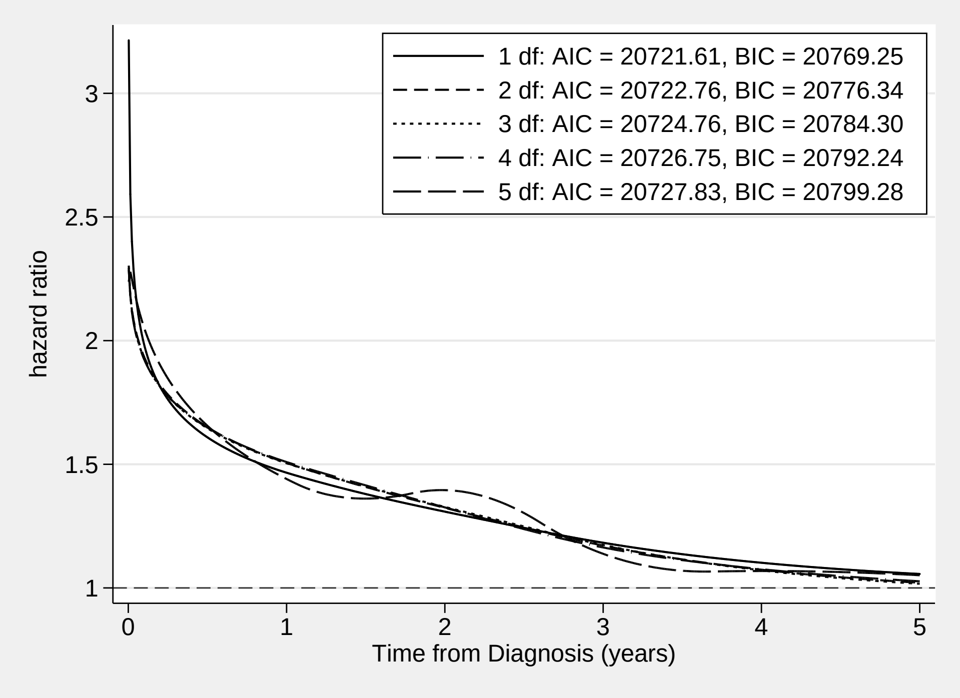
Estimated Survival curves



Estimated Survival Difference



Knots for Time-dependent effects



Flexible Parametric Models: Extensions

- Modelling on other scales (probit, cumulative odds).
- Age as the time-scale.
- Multiple time-scales.
- Multiple events.
- Time varying covariates (with time-dependent effects).
- Adjusted survival curves.
- Relative Survival
- Estimating crude and net mortality.

Relative Survival Models

- Usually model on the log excess hazard (mortality) scale[3].

$$h(t) = h^*(t) + \exp(\mathbf{x}\beta)$$

- Parameters are log excess hazard ratios.
- Models have proportional excess hazards as a special case, but often non-proportional excess hazards are observed.
- Non-proportionality modelled piecewise[3], using fractional polynomials[6], or splines[5].

Modelling on the Log Cumulative Excess Hazard Scale

- Nelson *et al.*[9] extended the flexible parametric modelling approach to model on the log cumulative excess hazard scale.

$$H_i(t) = H_i^*(t) + \Lambda_i(t)$$

$$\ln\left(-\ln R(t|\mathbf{x}_i)\right) = \ln\left(\Lambda(t)\right) = \ln\left(\Lambda_0(t)\right) + \mathbf{x}\boldsymbol{\beta}$$

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.

1. Relative Survival
2. Flexible Parametric Survival Models
3. Modelling Relative Survival
4. **Estimating Net and Crude Mortality**
5. Conclusion

Mortality in the absence/presence of other causes

- Relative Survival is a measure of survival in the absence of other causes, i.e. net probability.
- 1 - $R(t)$ is an estimate of the net probability of death due to cancer.

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

Estimation of Crude Mortality

- Cronin and Feuer[2] showed how this can be calculated from life tables.
- Calculated separately in age groups.
- Mortality may increase dramatically between the lower and upper boundaries of these age groups.
- Time-scale split into large (yearly) time intervals.
- Flexible parametric approach allows individual level covariates to be modelled

---

Brief Mathematical Details

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

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

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

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

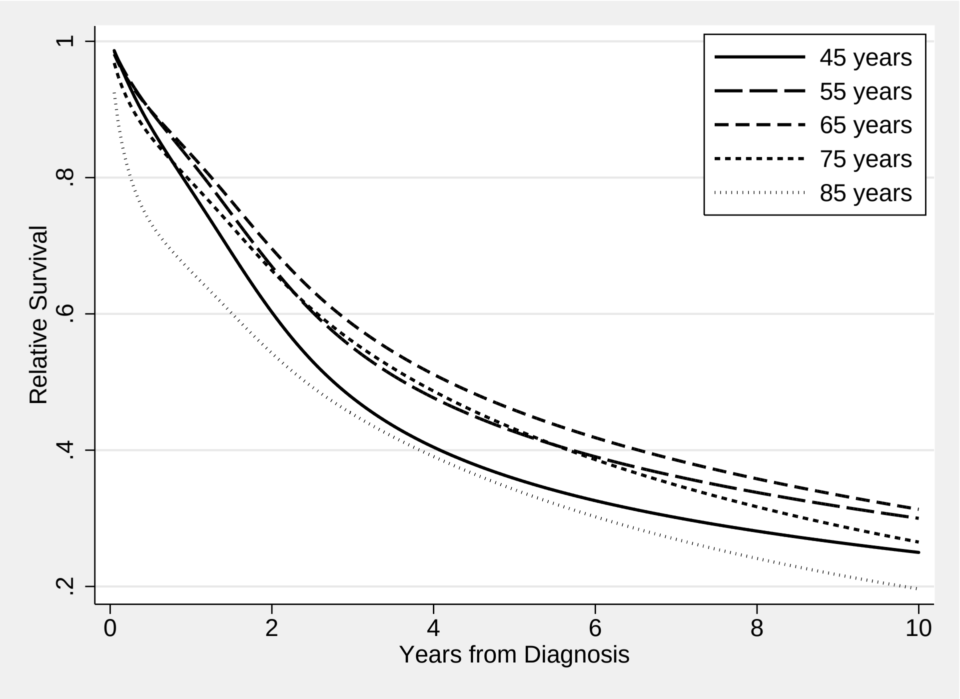
$$\text{Crude Prob of Death (other causes)} = \int_0^t S^*(t)R(t)h^*(t)$$

- Integration performed numerically.
- Delta-method to obtain variance.

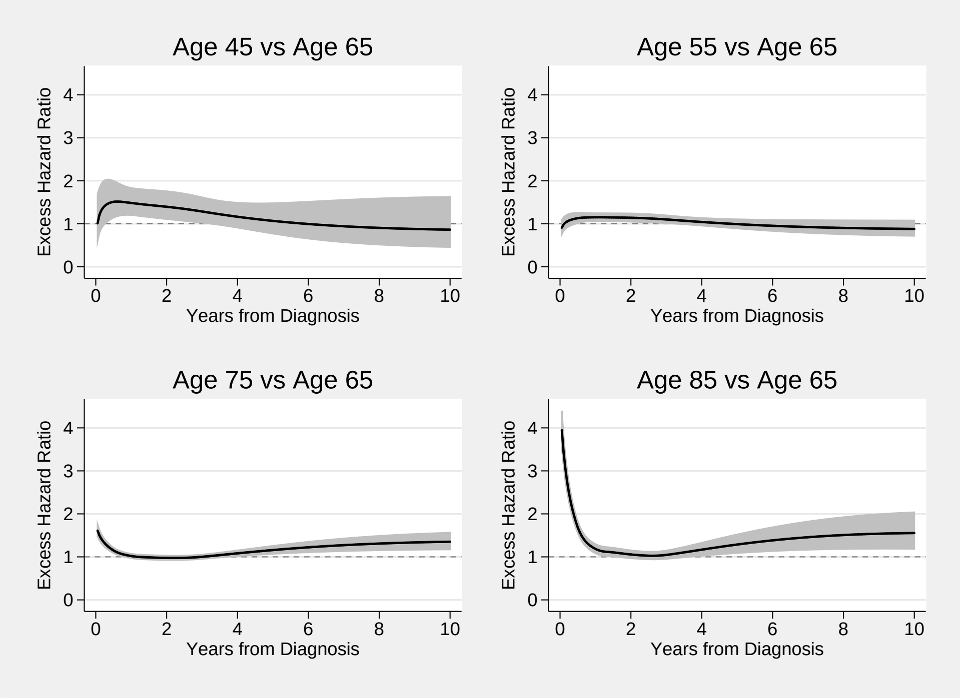
Example

- 28,943 men diagnosed with prostate cancer aged 40-90 in England and Wales between 1986-1988 inclusive and followed up to 1995[1].
- A model is fitted on the log cumulative hazard scale using restricted cubic splines are used to model the baseline excess hazard (6 knots).
- Restricted cubic splines are also used to model the effect of age (4 knots).
- The effects of age is also allowed to vary over time by incorporating interactions between the restricted cubic spline terms for age at diagnosis and a further set of restricted cubic splines for $\ln(t)$ (4 knots).
- Background mortality is incorporated and so this is a relative survival (excess mortality) model.

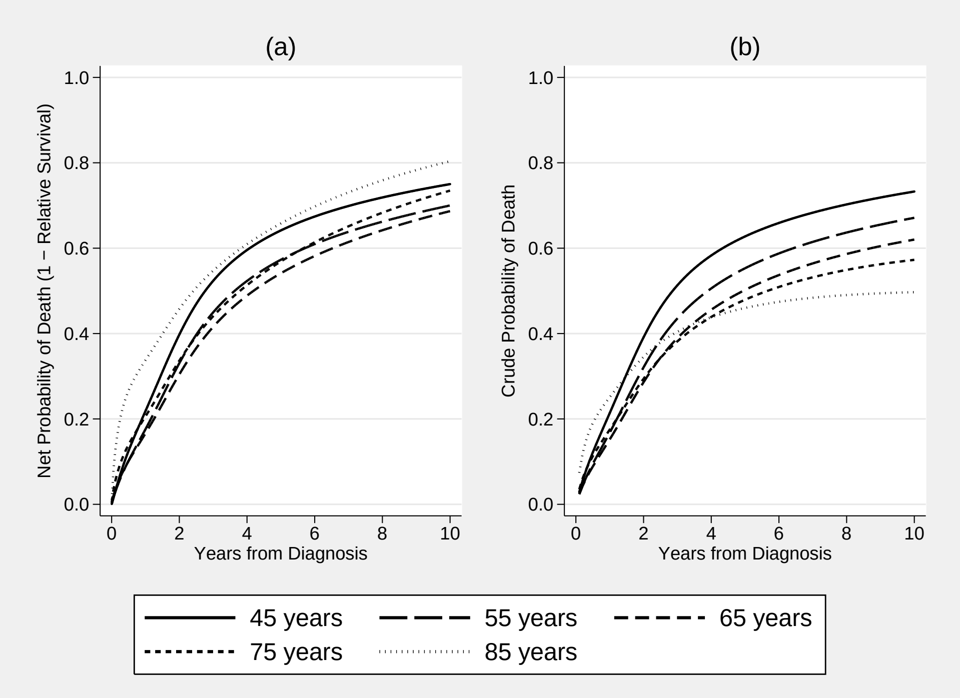
Estimated Relative Survival



Excess Hazard Ratios

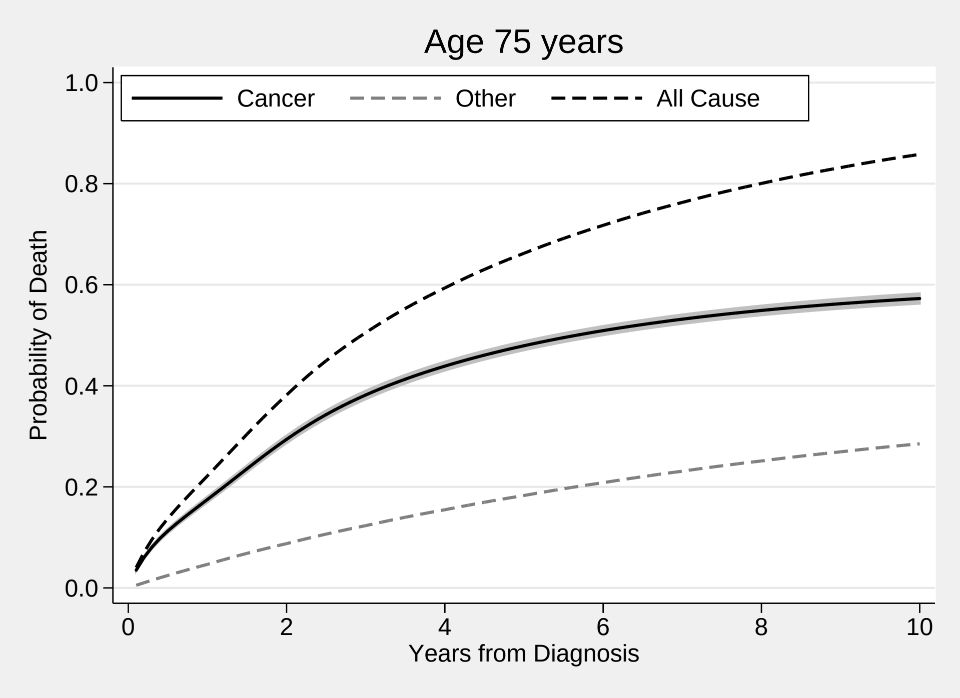


Net and Crude Probability of Death due to Cancer

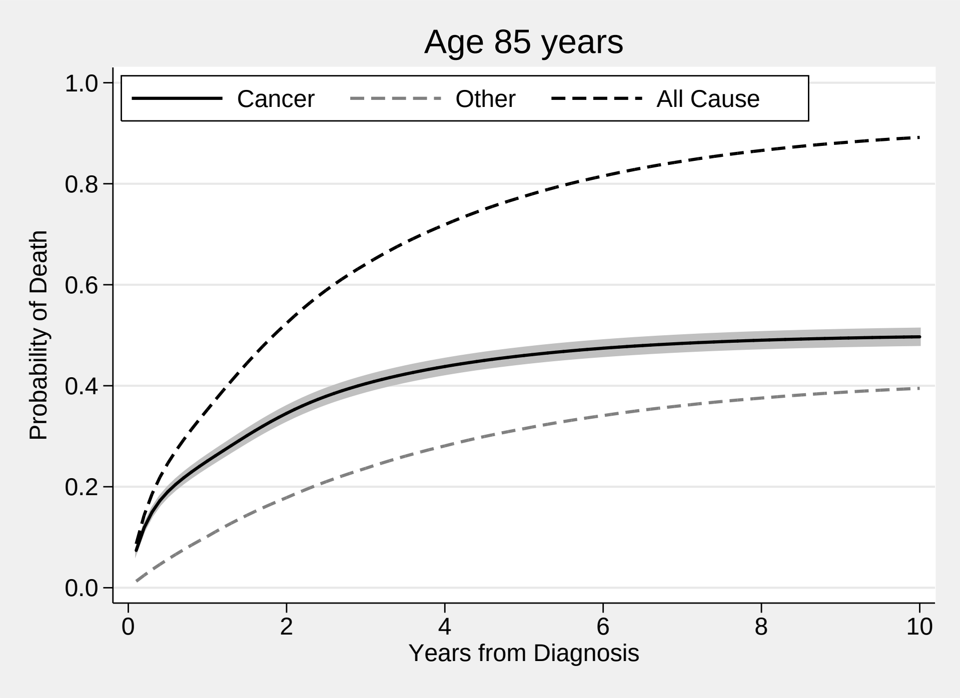


Crude Probability of Death - Age 45

Crude Probability of Death - Age 75



Crude Probability of Death - Age 85



Natural Frequencies

55 year old man

- Out of 100 people like you - by 5 years
  - 55 will die of cancer
  - 2 will die of other causes
  - 43 will be alive

85 year old man

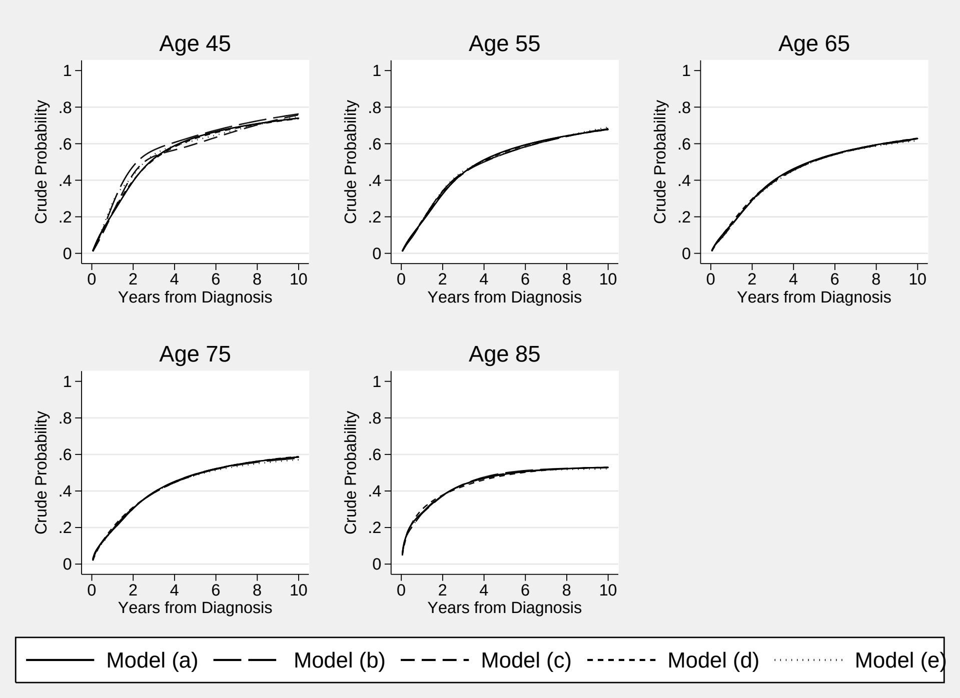
- Out of 100 people like you - by 5 years
  - 46 will die of cancer
  - 32 will die of other causes
  - 22 will be alive

Sensitivity to the number of knots

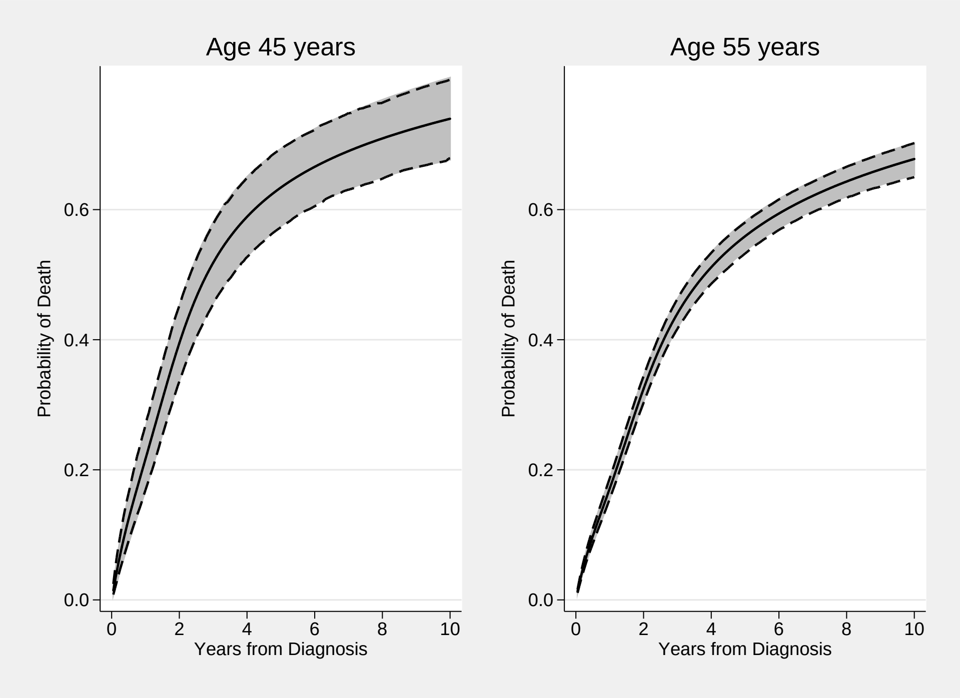
- A potential criticism of these models is the subjectivity in the number and the location of the knots.
- A small sensitivity analysis was carried out where the following models were fitted.

Model	Baseline	Time-dependent	age	No. of	AIC	BIC
	df_b	df_t	df_a	Parameters		
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



Variance Estimates



Conclusions

- Measuring the cumulative cause-specific mortality in the presence of other causes is a useful measure.
- It is a complement to relative survival.
- At present this is estimated separately for subgroups of interest in life-tables.
- Modelling has the advantage of estimating in continuous time, smaller standard errors, and making predictions for individual patients.
- The flexible parametric approach is a useful framework to estimate these quanities.
- It is also possible to obtain these estimates from other relative survival models, for example, cure models[7].
- Aim now is to move to more recent data to obtain individual level predictions using stage, risk factors, biomarkers etc.

References I

[1] M.P. Coleman, P. Babb, P. Damiecki, P. Grosclaude, S. Honjo, J. Jones, G. Knerer, A. Pitard, Quinn.M.J., A. Sloggett, and B. De Stavola. *Cancer survival trends in England and Wales, 1971-1995: deprivation and NHS Region*. Office for National Statistics, London, 1999.

[2] 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.

[3] Paul W Dickman, Andy Sloggett, Michael Hills, and Timo Hakulinen. Regression models for relative survival. *Statistics in Medicine*, 23(1):51–64, Jan 2004.

[4] S. Durrleman and R. Simon. Flexible regression models with cubic splines. *Statistics in Medicine*, 8(5):551–561, May 1989.

[5] Roch Giorgi, Michal Abrahamowicz, Catherine Quantin, Philippe Bolard, Jacques Esteve, Joanny Gouvernet, and Jean Faivre. A relative survival regression model using b-spline functions to model non-proportional hazards. *Statistics in Medicine*, 22(17):2767–2784, Sep 2003.

References II

[6] Paul C Lambert, Lucy K Smith, David R Jones, and Johannes L Botha. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Statistics in Medicine*, 24(24):3871–3885, Dec 2005.

[7] Paul C Lambert, John R Thompson, Claire L Weston, and Paul W Dickman. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3):576–594, Jul 2007.

[8] P.C. Lambert and P. Royston. Further development of flexible parametric models for survival analysis. *The Stata Journa (in press)*, 2009.

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

[10] Patrick Royston and Mahesh 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.