Flexible Parametric Alternatives to the Cox Model

Paul C Lambert[1,2] and 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

UK Stata User Group 2009
London, 11th September 2009

University *of* Leicester

Karolinska Institutet

Outline

1. The `stpm2` command

2. Why We Need Flexible Models

3. Time-Dependent Effects

4. Quantifying Differences

5. Average Survival Curve

6. Attained Age as the Time-Scale

7. Relative Survival

8. Crude Mortality

stpm2: A brief history

- Patrick Royston wrote `stpm` in 2001(Royston, 2001).
- Chris Nelson extended the methodology in `stpm` to relative survival(Nelson et al., 2007) in `strsrcs`.
- Time-dependent effects could be incorporated, but they tended to be over parameterised.
- I wrote `stpm2` (Lambert and Royston, 2009) to
  - Improve the modelling of time-dependent effects.
  - Combine the methods for standard and relative survival.
  - Make it easier to obtain useful predictions.
- `stpm2` is much faster than `stpm`, especially with large datasets.

How `stpm2` works

- In Royston-Parmar models the linear predictor is

Linear Predictor

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

- For models on the log cumulative hazard scale.

Survival and hazard functions

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

- Feed these into the likelihood.

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

How `stpm2` works

- A simplified version of the `ml` program is as follow,

stpm2_ml_hazard.ado

```
program stpm2_ml_hazard
  version 10.0
  args todo b lnf g negH g1 g2

  tempvar xb dxb
  mleval `xb' = `b', eq(1)
  mleval `dxb' = `b', eq(2)

  local st exp(-exp(`xb'))
  local ht `dxb'*exp(`xb')

  mlsum `lnf' = _d*ln(`ht') + ln(`st')
  /** then deal with late entry, first and ***
  *** second derivatives etc              **/
end
```

Run Rotterdam Example

England & Wales Breast Cancer Data

- Women diagnosed with breast cancer in England and Wales 1986-1990 with follow-up to 1995(Coleman et al., 1999).
- As an example I will investigate the effect of deprivation (in five groups) on all-cause mortality in women who were diagnosed under the age of 50 years.
- Follow-up will be restricted to 5 years.
- Due to their age, most of the women who die within 5 years will die due to their cancer.

Kaplan-Meier Graphs for Breast Cancer Data
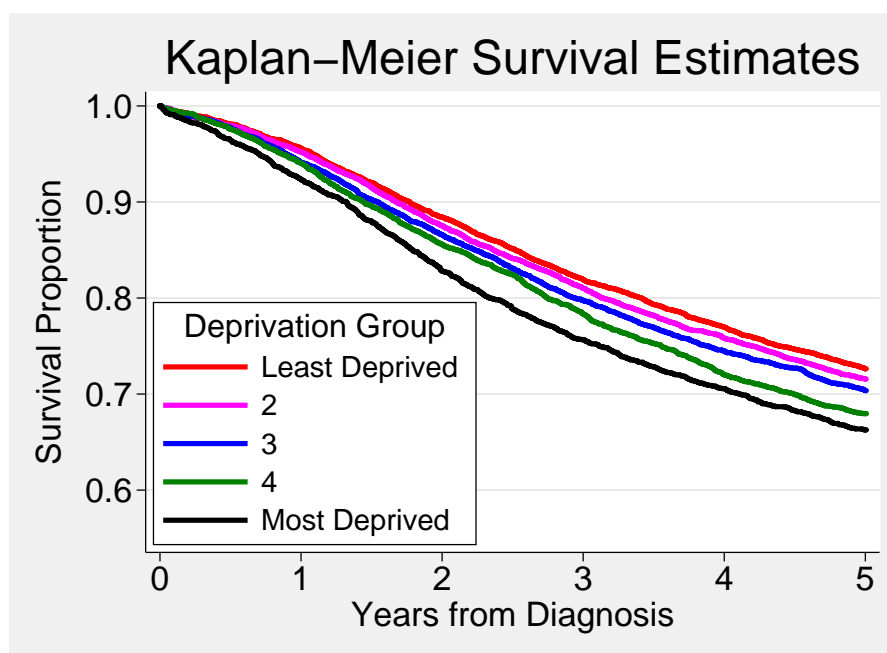
Why We Need Flexible Models
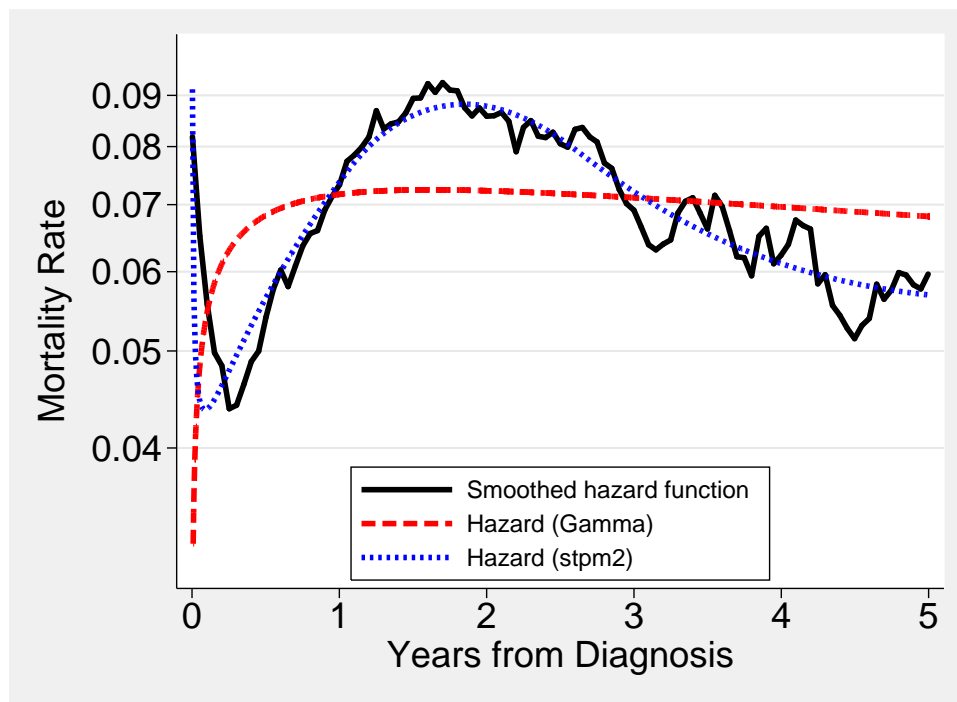
Time-Dependent Effects

- 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

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

- With D time-dependent effects we write,

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

- 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, $\mathbf{k}_j$

stpm2 and Time-Dependent Effects

- Non-proportional effects can be fitted by use of the `tvc()` and `dftvc()` options.

Non-proportional hazards models

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

- There is no need to split the time-scale when fitting time-dependent effects.
- When time-dependence is a linear function of $\ln(t)$ and $N = 50,000$, 50% censored and no ties.
  - `stcox` using `tvc()` - 28 minutes, 24 seconds.
 - `stpm2` using `dftvc(1)` - 0 minutes, 2.5 seconds.

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)
```
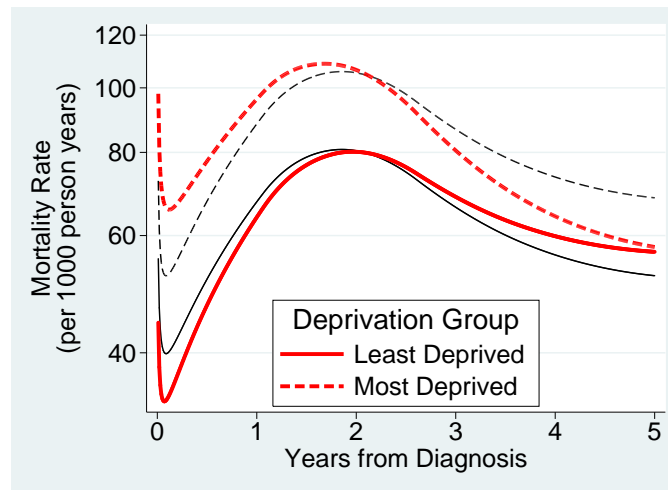
Predicting Hazard Ratios

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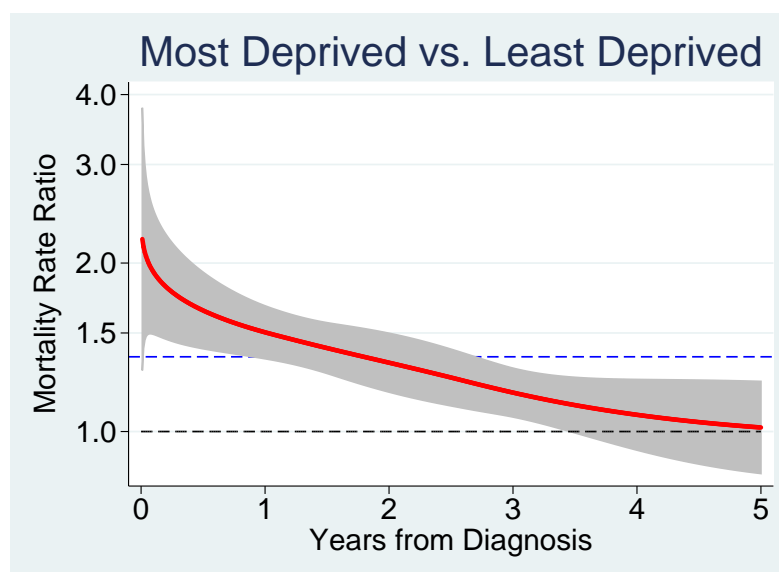
Quantifying Differences

- 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.
- The `predict` command of `stpm2` makes the predictions easy.
- They work in a similar way as the `hrnumerator()` and `hrdenominator()` commands.
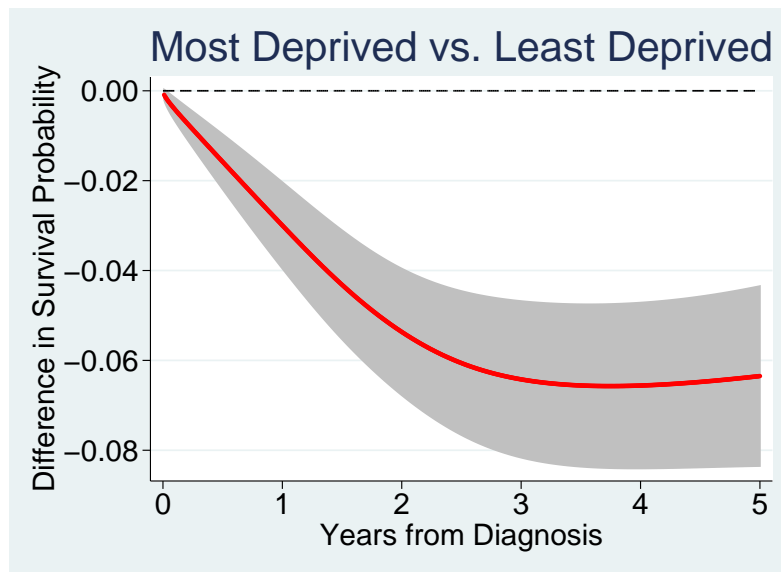
Difference in Hazard Rates

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


Most Deprived vs. Least Deprived
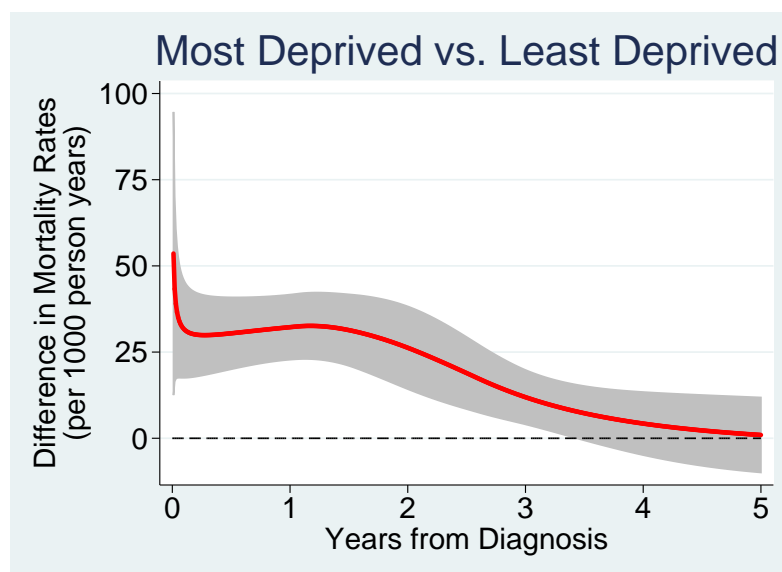
Difference in Survival Proportions

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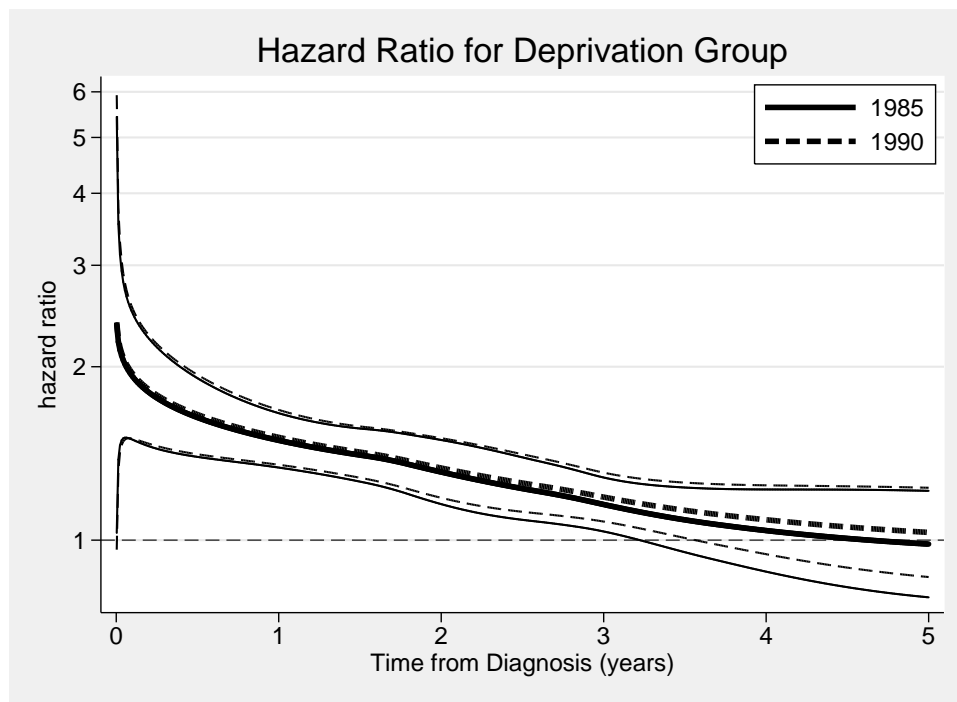

Most Deprived vs. Least Deprived

More than one time-dependent effect

- As we are modelling on the log cumulative hazard scale, we are essentially modelling non-proportional cumulative hazards.
- So far we have just considered one time-dependent factor.
- If we have two time-dependent effects (e.g. deprivation group and year of diagnosis) then the time-dependent hazard ratio for deprivation group may be different at different levels of year of diagnosis.
- Modelling on the log hazard scale would not have this problem.

Two time-dependent effects

```
. stpm2 dep5 yeardiag, scale(hazard) df(5) tvc(dep5 yeardiag) dftvc(3)
. predict hr_early, hrnum(dep5 1 yeardiag 1985) ///
                    hrdenom(dep5 0 yeardiag 1985) ///
                    timevar(timevar) ci
. predict hr_late, hrnum(dep5 1 yeardiag 1990) ///
                   hrdenom(dep5 0 yeardiag 1990) ///
                   timevar(timevar) ci
```

Hazard Ratio for Deprivation Group

Average Survival Curves

- It can be useful to summarise the average survival curve.
- The "easy" method is at the mean of the covariates.

$$\hat{S}_{ind}(t) = \exp\left(-\hat{H}_0(t)\exp\left(\bar{\mathbf{x}}\hat{\boldsymbol{\beta}}\right)\right)$$

  - Prediction for an individual who happens to have the mean values of each covariate.
  - Problem with binary covariates, e.g. a person of average sex.
- This is what `stcurve` does.
- A different concept is the mean survival for a population with a particular covariate distribution.

$$\hat{S}_{pop}(t) = \sum_{i=1}^{N}\exp\left(-\hat{H}_0(t)\exp\left(\mathbf{x}\hat{\boldsymbol{\beta}}\right)\right)$$

- These are not equivalent.

Adjusted/Standardised Survival Curve

- We can extend the ideas of average survival curves to obtain adjusted survival curves.
- The key is to obtain the predicted mean population survival curves for two or more groups, while allowing the distribution of other covariates (e.g. age) to be the same for the two groups.
- The most common method is to use the covariate distribution in the study population as a whole, but other covariate distributions can also be used.
- The basic idea is similar to the "correct group prognostic method" (Nieto and Coresh, 1996).
- Using flexible parametric survival models we can allow for time-dependent covariates, continuous covariates etc.

Kaplan-Meier Curves - Renal Replacement Therapy
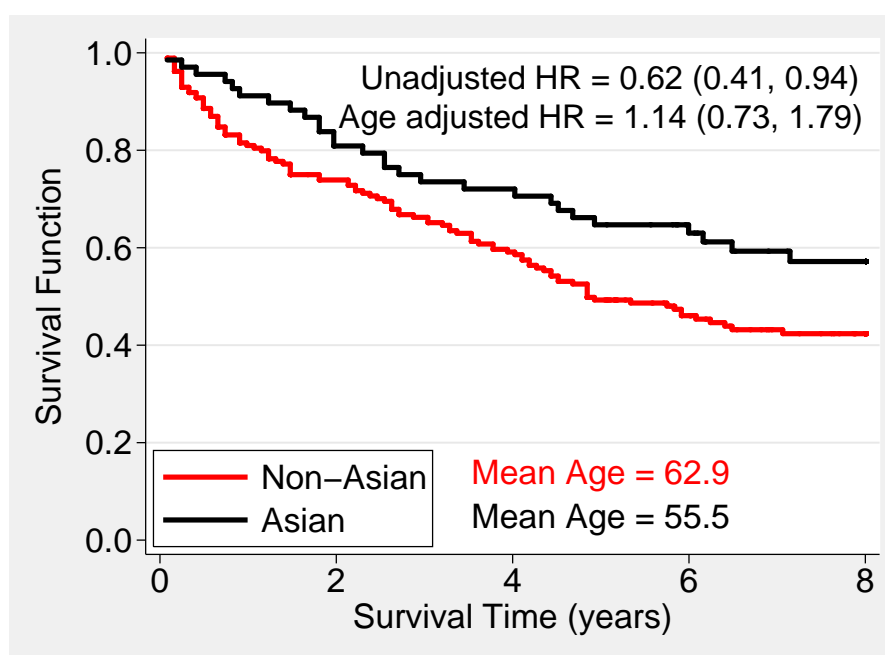
# Predictions for Adjusted Survival Curves

The `meansurv` option

```
stpm2 asian age, df(3) scale(hazard)
/* Age distribution for study population as a whole */
predict meansurv_pop0, meansurv at(asian 0)
predict meansurv_pop1, meansurv at(asian 1)

/* Age distribution for non-asians */
predict meansurv_pop0b if asian == 0, meansurv at(asian 0)
predict meansurv_pop1b if asian == 0, meansurv at(asian 1)

/* Age distribution for asians */
predict meansurv_pop0c if asian == 1, meansurv at(asian 0)
predict meansurv_pop1c if asian == 1, meansurv at(asian 1)
```
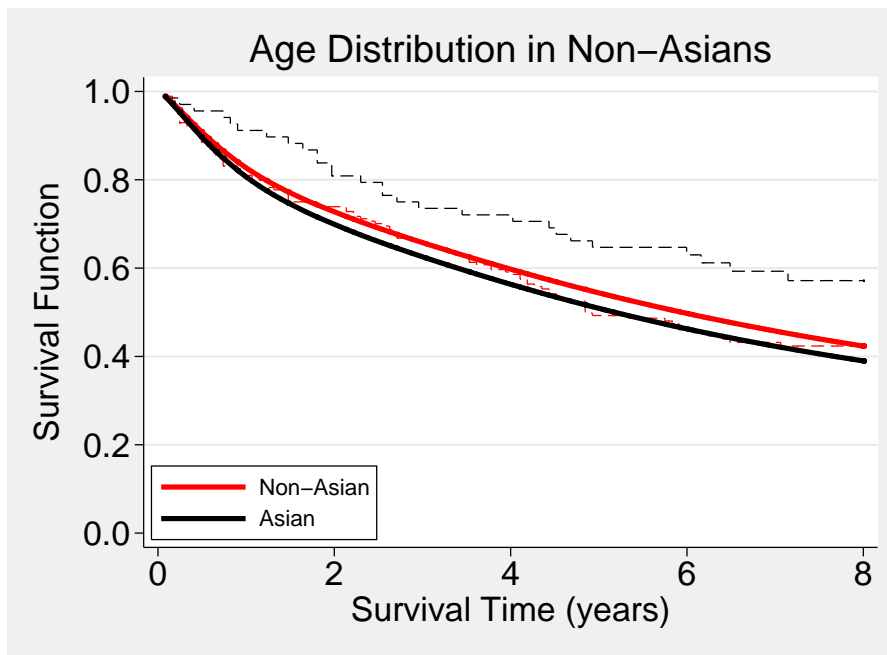
- Survival curve calculated for each subject in the study population and then averaged.
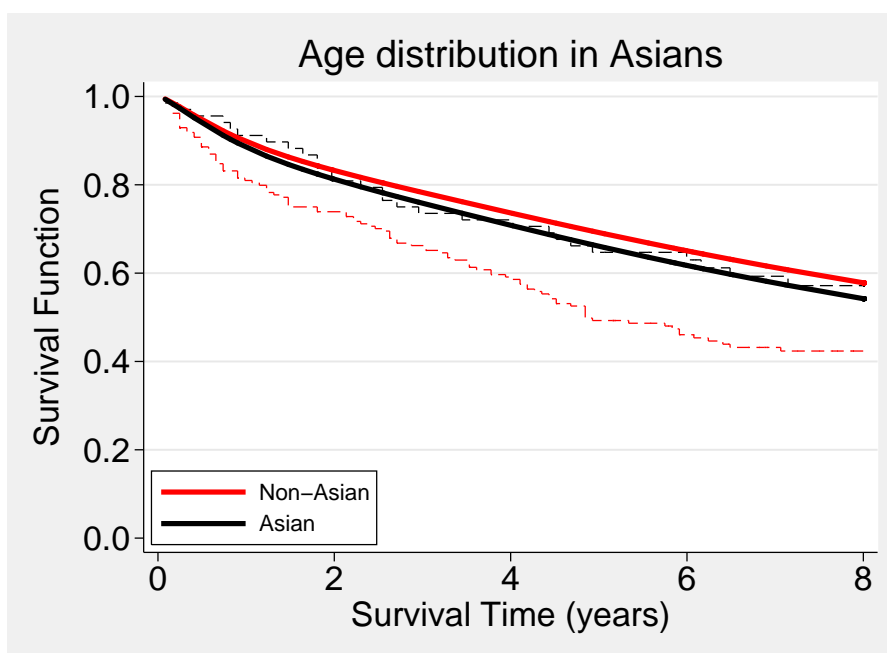- In large studies use the `timevar()` option.

Adjusted Survival Curve 1

Adjusted Survival Curve 2

Age Distribution in Non–Asians

Adjusted Survival Curve 3

Age distribution in Asians
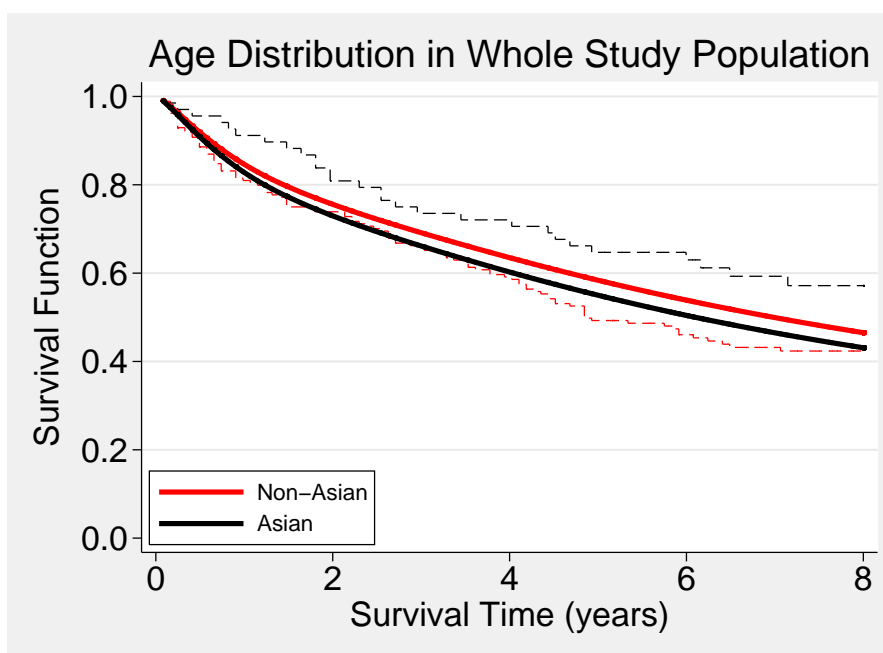
Example of Attained Age as the Time-scale

- Study from Sweden(Dickman et al., 2004) 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)
```
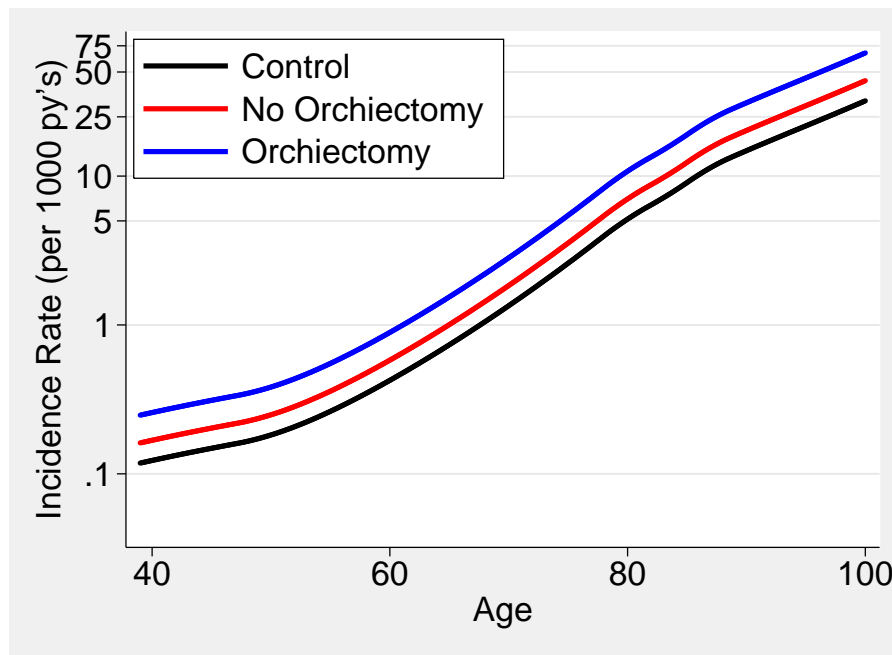
```
.   stcox noorc orc
```

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)
```

Royston-Parmar Model

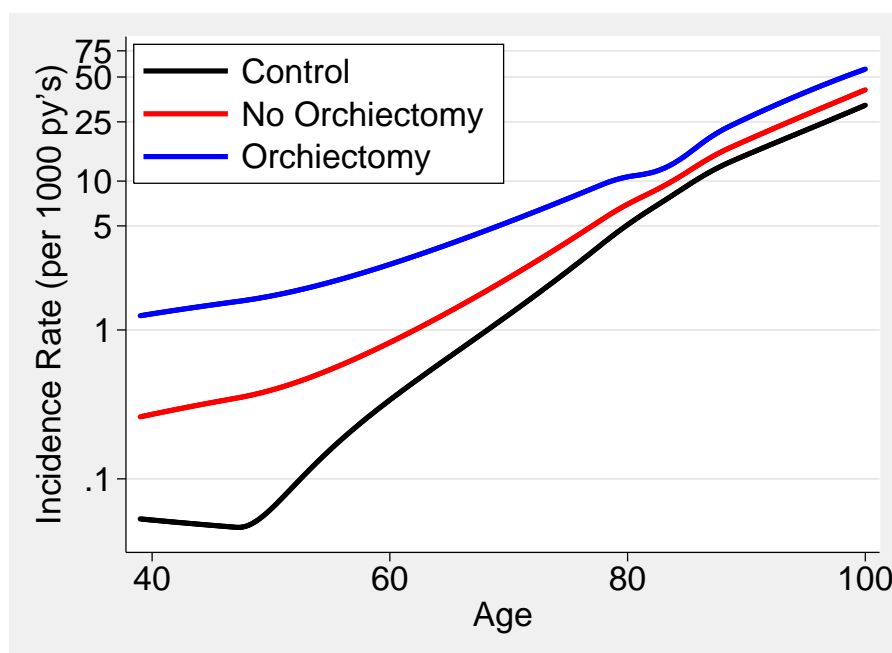|		
Incidence rate ratio (no orchiectomy)	=	1.37 (1.28 to 1.46)
Incidence rate ratio (orchiectomy)	=	2.10 (1.93 to 2.28)

Proportional Hazards
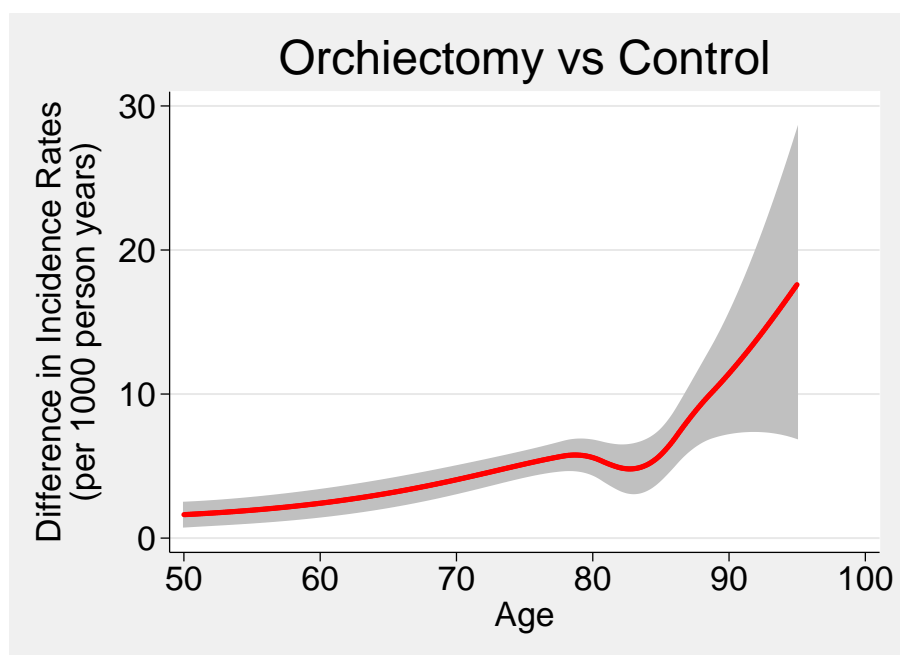
Non Proportional Hazards

Incidence Rate Ratio

Orchiectomy vs Control

horizontal lines from piecewise Poisson model
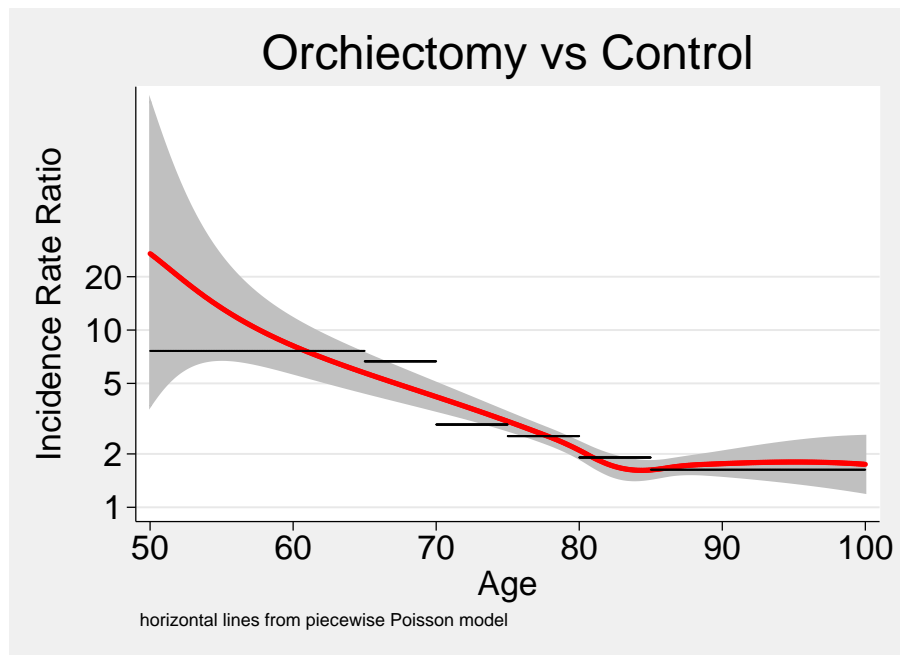
Incidence Rate Difference

Orchiectomy vs Control

Relative Survival/Excess Mortality 1

- Relative Survival is used in population-based cancer studies.
- Growing interest in other disease areas: HIV (Bhaskaran et al., 2008), CHD (Nelson et al., 2008).
- 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).

Relative Survival/Excess Mortality 1

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

$$
\begin{array}{ccccc}
\text{Observed} & = & \text{Expected} & + & \textit{Excess} \\
\text{Mortality Rate} & & \text{Mortality Rate} & & \textit{Mortality Rate} \\
h(t) & = & h^*(t) & + & \lambda(t)
\end{array}
$$

- 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 is the most important single measure of cancer patient care (the diagnosis and treatment of cancer) and is of considerable interest to clinicians, patients, researchers, politicians, health administrators, and public health professionals (Dickman and Adami, 2006).
- 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(Cronin and Feuer, 2000) 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 et al. (2009) 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(Carstensen, 2006)

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)
```

The `stpm2cm` Command

- stpm2cm is a post estimation command.
- It will calculate the crude probability of death due to cancer and other causes with associated confidence intervals.
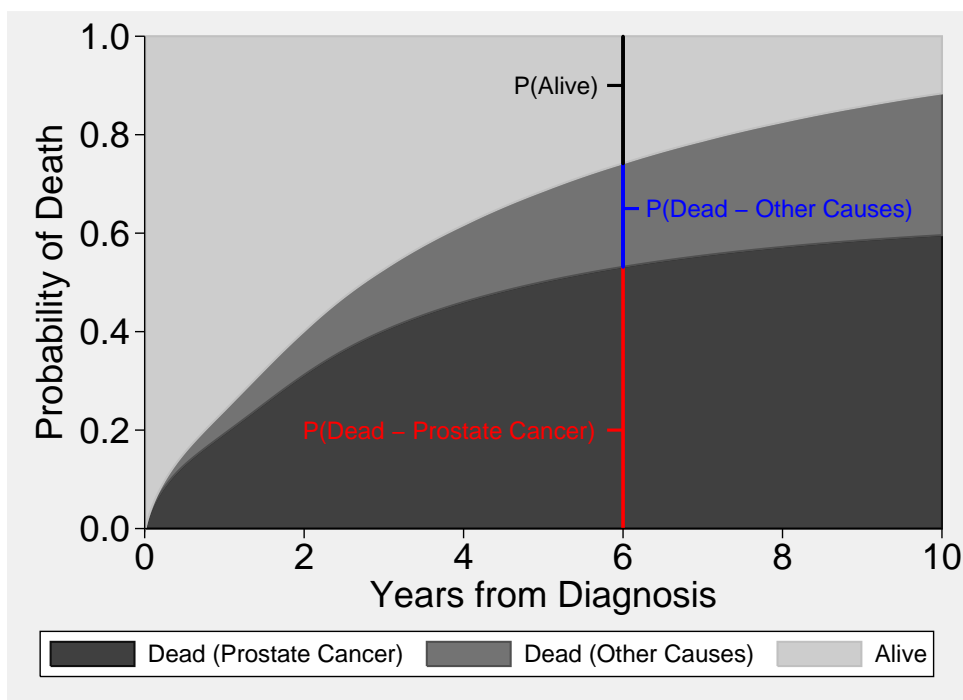
stpm2cm

```
.   stpm2cm using uk_popmort, ///
        mergeby(_year sex region caquint _age) maxt(10) ///
        diagage('agediag') diagyear(1986) attyear(_year) ///
        attage(_age) diagsex(1) ///
        at(agercs1 'a1' agercs2 'a2' agercs3 'a3') ///
        stub(ci'agediag') tgen(ci_t'agediag') ///
        mergegen(region 1 caquint 1) nobs(1000) ci
```

Net and Crude Probability of Death
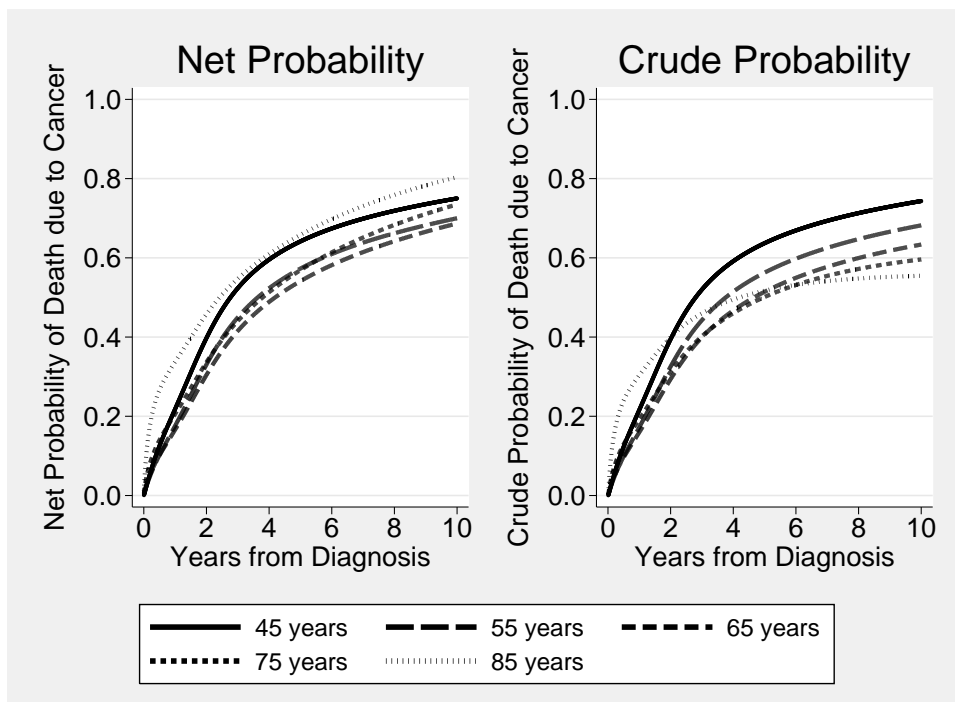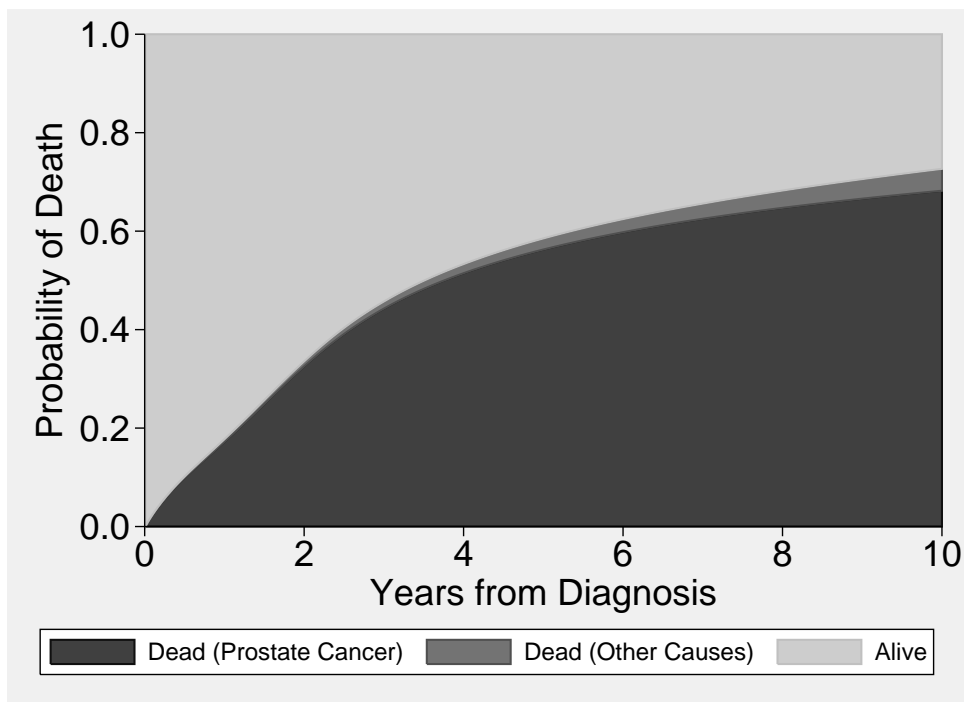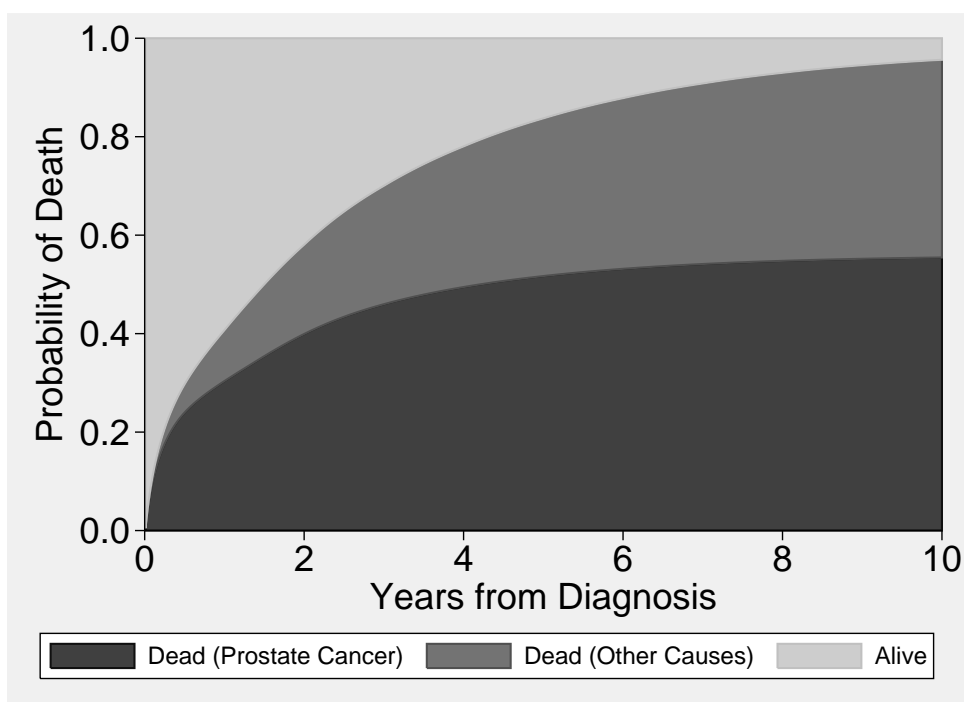
Predictions for a 75 year old man

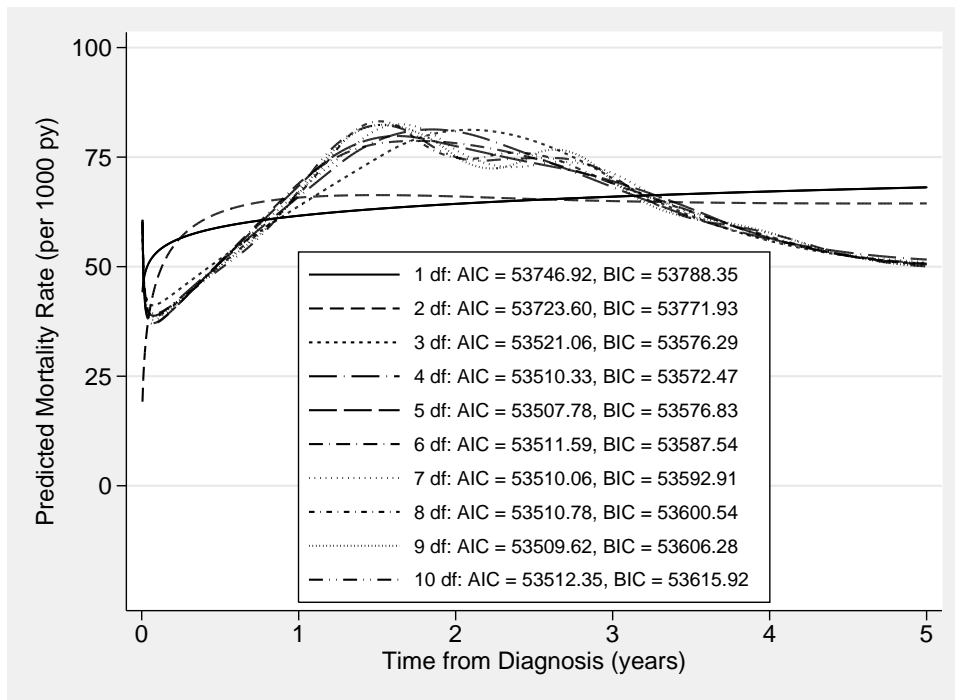# Predictions for a 55 year old man

Predictions for a 85 year old man

Further Extensions

- Update to Stata 11.
- Univariate and shared frailty models.
- Multiple Events.
- Competing Risks.
- Survey options?
- Cure models.
- Estimation of loss in expectation of life.
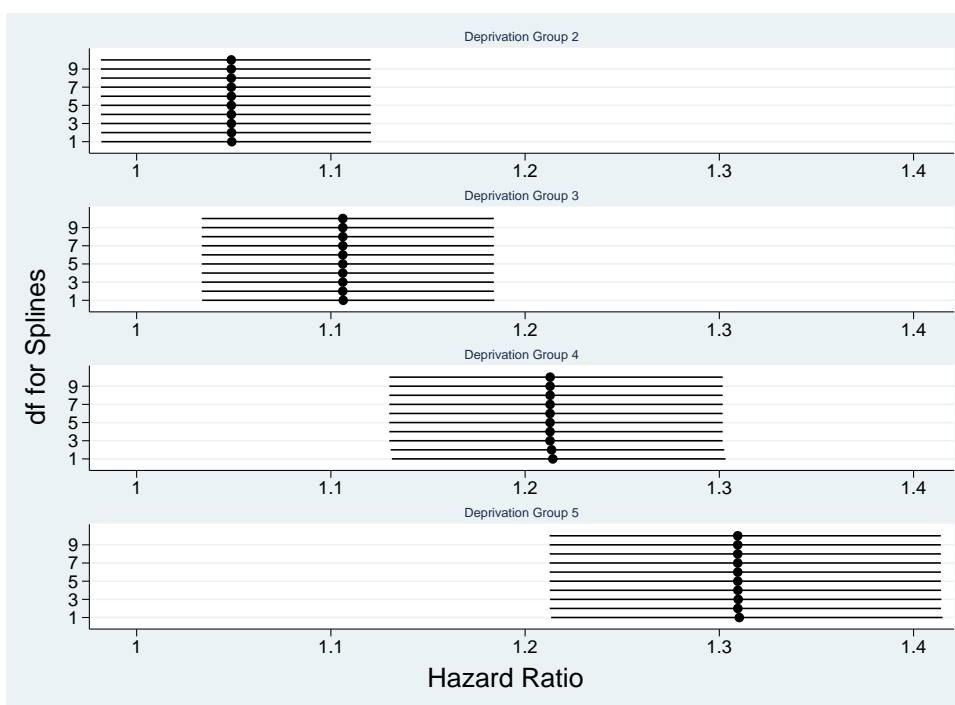- Enhance ability to model multiple time-scale.

References I

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

Carstensen, B. (2006). 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.

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

Cronin, K. A. and Feuer, E. J. (2000). 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.

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

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

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

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

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

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

Nieto, F. J. and Coresh, J. (1996). Adjusting survival curves for confounders: a review and a new method. *Am J Epidemiol*, 143(10):1059–1068.

Royston, P. (2001). Flexible parametric alternatives to the Cox model, and more. *The Stata Journal*, 1:1–28.

Example of different knots for baseline hazard

Legend:
- 1 df: AIC = 53746.92, BIC = 53788.35
- 2 df: AIC = 53723.60, BIC = 53771.93
- 3 df: AIC = 53521.06, BIC = 53576.29
- 4 df: AIC = 53510.33, BIC = 53572.47
- 5 df: AIC = 53507.78, BIC = 53576.83
- 6 df: AIC = 53511.59, BIC = 53587.54
- 7 df: AIC = 53510.06, BIC = 53592.91
- 8 df: AIC = 53510.78, BIC = 53600.54
- 9 df: AIC = 53509.62, BIC = 53606.28
- 10 df: AIC = 53512.35, BIC = 53615.92

# Effect of number of knots on hazard ratios

Where to place the knots?

- 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
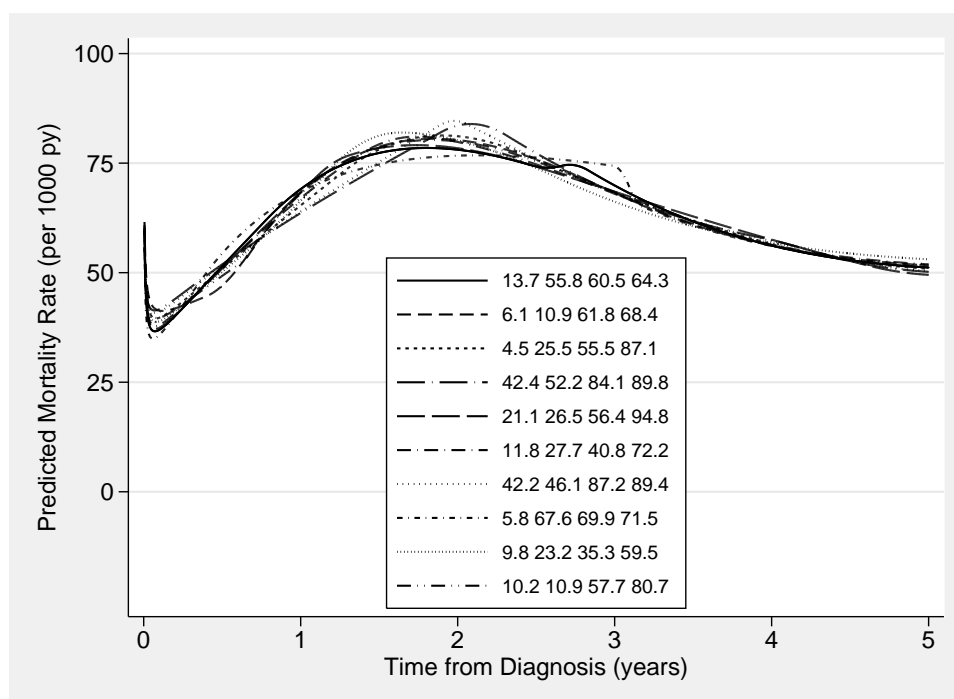
Baseline survival - random knots
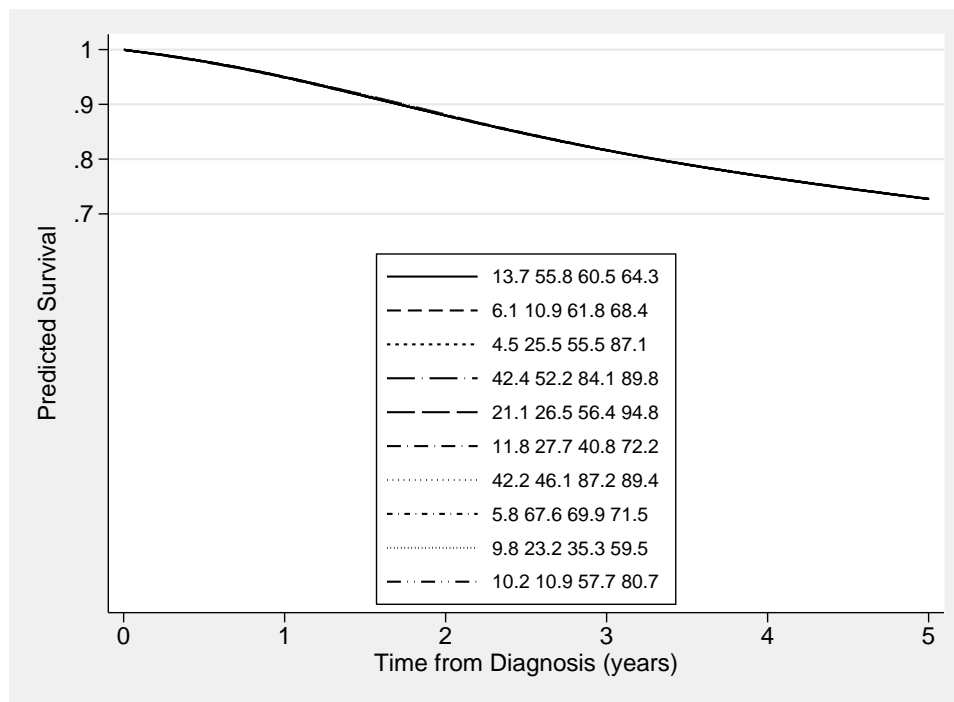
Sensitivity to the number of knots

- 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