Up-to-date survival estimates from prognostic models using temporal recalibration

Sarah Booth[1] Mark J. Rutherford[1] Paul C. Lambert [1,2]

[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK

[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

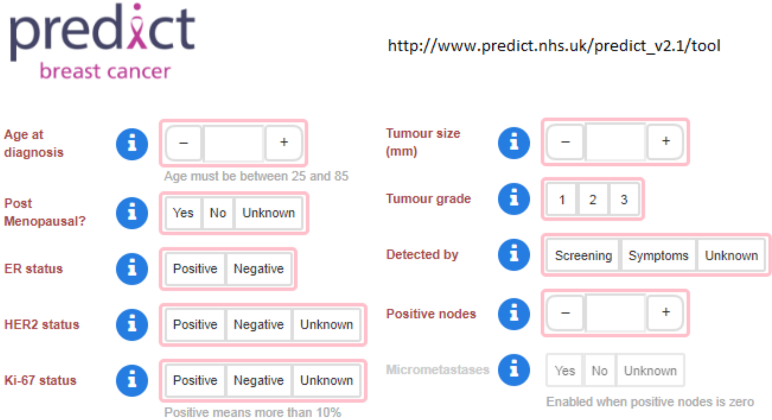
12th September 2018

Nordic and Baltic Stata Users Group Meeting

Overview

- Prognostic models for cancer
- Flexible parametric survival models (`stpm2`)
- Period analysis (`stset`)
- Method of temporal recalibration
- Comparison of cohort, recalibrated and period analysis models
- Importance of updating prognostic models

http://www.predict.nhs.uk/predict_v2.1/tool

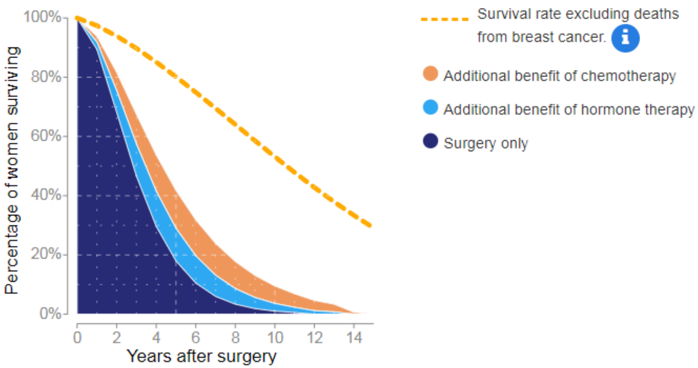


dos Reis, F. J. C., Wishart, G. C., Dicks, E. M. et al. (2017), 'An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation', Breast Cancer Research 19(1). PREDICT Version 2.1 tool available from: `http://www.predict.nhs.uk/predict_v2.1/`

PREDICT: Prognostic Model for Breast Cancer

Table | Curves | Chart | Texts | Icons

These results are for women who have already had surgery. This graph shows the percentage of women surviving up to 15 years. These results are based on the inputs and treatments you selected.



dos Reis, F. J. C., Wishart, G. C., Dicks, E. M. et al. (2017), 'An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation', Breast Cancer Research 19(1). PREDICT Version 2.1 tool available from: `http://www.predict.nhs.uk/predict_v2.1/`

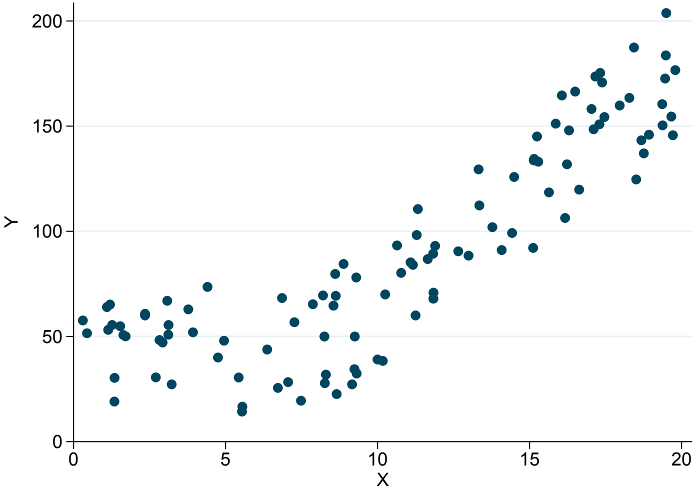
Flexible Parametric Survival Models

- Unlike the Cox model, parametric models specify the baseline hazard
- The Weibull model requires linearity on the log cumulative hazard scale

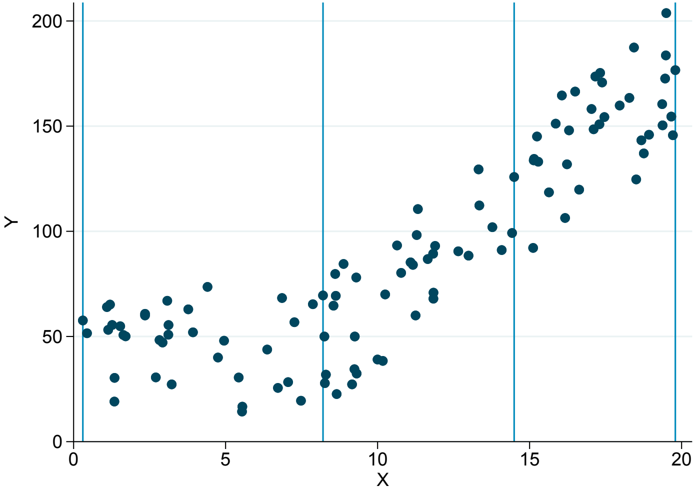
$$\ln[H(t|x_i)] = \ln(\lambda) + \gamma \ln(t) + x_i\beta$$

- Flexible parametric survival models use restricted cubic splines which allow more complex shapes to be captured

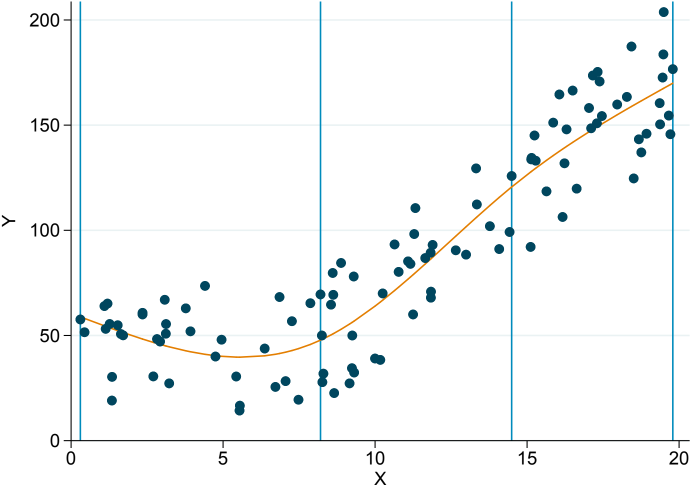
Restricted Cubic Splines



Restricted Cubic Splines



Restricted Cubic Splines



Flexible Parametric Survival Models

$$\ln[H(t|x_i)] = \gamma_0 + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \gamma_3 z_{3i} + ... + x_i\beta$$

- z_i = derived variables for the restricted cubic splines
- $x_i\beta$ = linear predictor = prognostic index
- `stpm2` command in Stata

Cohort vs Period Analysis

Participant	Follow-Up											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A	1	2	3	4	5	6	7	8	9	10	11	
B		1	2	3	4	5						
C						1	2	3	4	5	6	7
D									1	2	3	

Cohort Analysis

- All 4 participants would be included in cohort analysis
- Referred to as "complete analysis" by Brenner et al. (2009)

Cohort vs Period Analysis

Participant	Follow-Up											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A	1	2	3	4	5	6	7	8	9	10	11	
B		1	2	3	4	5			-	-		
C						1	2	3	4	5	6	7
D									1	2	3	

Cohort vs Period Analysis

Participant	Follow-Up											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A	1	2	3	4	5	6	7	8	9	10	11	
B		1	2	3	4	5			-	-		
C						1	2	3	4	5	6	7
D									1	2	3	

Advantages of Period Analysis

- Creates more up-to-date survival estimates because people diagnosed many years ago only contribute to long-term survival estimates

Cohort vs Period Analysis

Participant	Follow-Up											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A	1	2	3	4	5	6	7	8	9	10	11	
B		1	2	3	4	5			-	-		
C						1	2	3	4	5	6	7
D									1	2	3	

Advantages of Period Analysis

- Creates more up-to-date survival estimates because people diagnosed many years ago only contribute to long-term survival estimates

Disadvantages of Period Analysis

- Reduces sample size

Temporal Recalibration

Participant	Follow-Up											
	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
A	1	2	3	4	5	6	7	8	9	10	11	
B		1	2	3	4	5			-	-		
C						1	2	3	4	5	6	7
D									1	2	3	

Method

- Fit a cohort model
- Use a period analysis sample to recalibrate the model
- The covariate effects are constrained to be the same
- The baseline hazard function is allowed to vary which can capture any improvements in survival

Data

- Colon cancer data from Surveillance, Epidemiology, and End Results Program (SEER) database
- National Cancer Institute: Data collected from the United States
- Variables used in this analysis are: age at diagnosis, sex, ethnicity
- Survival times measured in months but for period analysis dates are required

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission

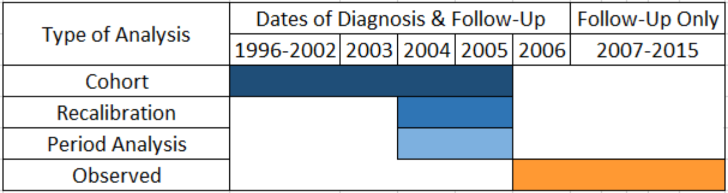
Data

- Colon cancer data from Surveillance, Epidemiology, and End Results Program (SEER) database
- National Cancer Institute: Data collected from the United States
- Variables used in this analysis are: age at diagnosis, sex, ethnicity
- Survival times measured in months but for period analysis dates are required

```
. gen dx = mdy(mmdx,1,yydx)
. format dx %td
. gen exit = dx+survmm*30.5
. format exit %td
```

- **mmdx:** month of diagnosis
- **yydx:** year of diagnosis
- **survmm:** survival time in months
- **dx:** date of diagnosis
- **exit:** date of death or censoring

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission



- **Cause-specific survival:** deaths due to colon cancer
- **Proportional hazards models:** for simplicity but also possible with time-dependent effects
- **Cohort:** 63,223 participants, 22,119 deaths
- **Period Analysis:** 39,743 participants, 4,889 deaths
- **Observed:** 6,300 participants, 2,474 deaths

stset: Cohort

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> exit(time min(dx+10*365.25,mdy(12,31,2005)))

                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx

------------------------------------------------------------------------------
    124,579  total observations
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     63,223  observations remaining, representing
     63,223  subjects
     22,119  failures in single-failure-per-subject data
 184,050.03  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  9.998905
```

exit: date of death or censoring

stset: Cohort

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> exit(time min(dx+10*365.25,mdy(12,31,2005)))

                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx
------------------------------------------------------------------------------
    124,579  total observations
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     63,223  observations remaining, representing
     63,223  subjects
     22,119  failures in single-failure-per-subject data
 184,050.03  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  9.998905
```

origin: when people become at risk, **dx** date of diagnosis

stset: Cohort

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> exit(time min(dx+10*365.25,mdy(12,31,2005)))
                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx
------------------------------------------------------------------------------
    124,579  total observations
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     63,223  observations remaining, representing
     63,223  subjects
     22,119  failures in single-failure-per-subject data
 184,050.03  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  9.998905
```

scale(365.24): convert to survival time in years

stset: Cohort

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> exit(time min(dx+10*365.25,mdy(12,31,2005)))

                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx
------------------------------------------------------------------------------
    124,579  total observations
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     63,223  observations remaining, representing
     63,223  subjects
     22,119  failures in single-failure-per-subject data
 184,050.03  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  9.998905
```

fail: event indicator, **cancer==1:** death due to colon cancer

stset: Cohort

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> exit(time min(dx+10*365.25,mdy(12,31,2005)))
                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx
------------------------------------------------------------------------------
    124,579  total observations
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     63,223  observations remaining, representing
     63,223  subjects
     22,119  failures in single-failure-per-subject data
 184,050.03  total analysis time at risk and under observation
                                                 at risk from t =         0
                                      earliest observed entry t =         0
                                           last observed exit t =  9.998905
```

exit(): follow-up until end of 2005 or for a maximum of 10 years

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                    Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

agercs* female black: covariates in the model

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                        Number of obs    =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

scale(hazard): scale used e.g. hazards, odds

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                     Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

df(5): degrees of freedom for modelling the baseline

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                    Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

noorthog: splines are not orthogonalised (simplifies recalibration)

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                    Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

eform: display the hazard ratios instead of log hazard ratios

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                    Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                    Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

Model: Cohort

```
. stpm2 agercs* female black, scale(hazard) df(5) noorthog eform
Log likelihood = -73439.283                   Number of obs     =     63,223
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	.0025474	4.96	0.000	1.007577	1.017563
agercs2	1.00005	7.46e-06	6.68	0.000	1.000035	1.000064
agercs3	.9999177	8.99e-06	-9.15	0.000	.9999001	.9999353
female	.9098671	.0125303	-6.86	0.000	.8856366	.9347606
black	1.403117	.0286116	16.61	0.000	1.348145	1.46033
_rcs1	12.69938	.6035658	53.48	0.000	11.56984	13.93919
_rcs2	1.150777	.0046616	34.67	0.000	1.141677	1.15995
_rcs3	.8279092	.0097947	-15.96	0.000	.8089329	.8473307
_rcs4	1.009746	.0174485	0.56	0.575	.9761203	1.04453
_rcs5	1.113578	.0115556	10.37	0.000	1.091159	1.136459
_cons	308.5041	53.719	32.92	0.000	219.3025	433.9887

```
. estimates store cohort
. range timevar10 0 10 1000
. predict cohort2006 if yydx==2006, timevar(timevar10) meansurv
```

stset: Temporal Recalibration & Period Analysis

```
. stset exit, origin(dx) fail(cancer==1) scale(365.24) ///
> entry(time mdy(1,1,2004)) exit(time min(dx+10*365.25,mdy(12,31,2005)))
                id:  id
     failure event:  cancer == 1
obs. time interval:  (exit[_n-1], exit]
 enter on or after:  time mdy(1,1,2004)
 exit on or before:  time min(dx+10*365.25,mdy(12,31,2005))
    t for analysis:  (time-origin)/365.24
            origin:  time dx
------------------------------------------------------------------------------
    124,579  total observations
     23,480  observations end on or before enter()
     61,356  observations begin on or after exit
------------------------------------------------------------------------------
     39,743  observations remaining, representing
     39,743  subjects
      4,889  failures in single-failure-per-subject data
 59,904.493  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =  9.998905
```

Constraints: Temporal Recalibration

```
. estimates restore cohort
(results cohort are active now)
. local agercs1 = _b[agercs1]
. local agercs2 = _b[agercs2]
. local agercs3 = _b[agercs3]
. local female = _b[female]
. local black = _b[black]
. constraint 1 _b[agercs1] = `agercs1´
. constraint 2 _b[agercs2] = `agercs2´
. constraint 3 _b[agercs3] = `agercs3´
. constraint 4 _b[female] = `female´
. constraint 5 _b[black] = `black´
. local knots = e(bhknots)
. local bknots = e(boundary_knots)
```

Constraints: Temporal Recalibration

```
. estimates restore cohort
(results cohort are active now)
. local agercs1 = _b[agercs1]
. local agercs2 = _b[agercs2]
. local agercs3 = _b[agercs3]
. local female = _b[female]
. local black = _b[black]
. constraint 1 _b[agercs1] = `agercs1´
. constraint 2 _b[agercs2] = `agercs2´
. constraint 3 _b[agercs3] = `agercs3´
. constraint 4 _b[female] = `female´
. constraint 5 _b[black] = `black´
. local knots = e(bhknots)
. local bknots = e(boundary_knots)
```

Constraints: Temporal Recalibration

```
. estimates restore cohort
(results cohort are active now)
. local agercs1 = _b[agercs1]
. local agercs2 = _b[agercs2]
. local agercs3 = _b[agercs3]
. local female = _b[female]
. local black = _b[black]
. constraint 1 _b[agercs1] = `agercs1´
. constraint 2 _b[agercs2] = `agercs2´
. constraint 3 _b[agercs3] = `agercs3´
. constraint 4 _b[female] = `female´
. constraint 5 _b[black] = `black´
. local knots = e(bhknots)
. local bknots = e(boundary_knots)
```

Model: Temporal Recalibration

```
. stpm2 agercs* female black, scale(hazard) noorthog constraints(1 2 3 4 5) ///
> bknots(`bknots´) knots(`knots´) eform
note: delayed entry models are being fitted

Log likelihood = -16015.094                       Number of obs     =     39,743
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
agercs1	1.012557	(constrained)				
agercs2	1.00005	(constrained)				
agercs3	.9999177	(constrained)				
female	.9098671	(constrained)				
black	1.403117	(constrained)				
_rcs1	23.11036	2.852501	25.44	0.000	18.14443	29.43541
_rcs2	1.201228	.0117535	18.74	0.000	1.178411	1.224486
_rcs3	.7933542	.0212882	-8.63	0.000	.7527083	.8361949
_rcs4	.9970216	.0372468	-0.08	0.936	.9266278	1.072763
_rcs5	1.144055	.02421	6.36	0.000	1.097575	1.192504
_cons	2544.921	1128.816	17.68	0.000	1066.887	6070.581

```
. predict recalibration2006 if yydx==2006, timevar(timevar10) meansurv
```

Model: Temporal Recalibration

```
. stpm2 agercs* female black, scale(hazard) noorthog constraints(1 2 3 4 5) ///
> bknots(`bknots´) knots(`knots´) eform
note: delayed entry models are being fitted

Log likelihood = -16015.094                     Number of obs     =     39,743

------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
    agercs1  |   1.012557  (constrained)
    agercs2  |    1.00005  (constrained)
    agercs3  |   .9999177  (constrained)
     female  |   .9098671  (constrained)
      black  |   1.403117  (constrained)
      _rcs1  |   23.11036   2.852501    25.44   0.000     18.14443    29.43541
      _rcs2  |   1.201228   .0117535    18.74   0.000     1.178411    1.224486
      _rcs3  |   .7933542   .0212882    -8.63   0.000     .7527083    .8361949
      _rcs4  |   .9970216   .0372468    -0.08   0.936     .9266278    1.072763
      _rcs5  |   1.144055     .02421     6.36   0.000     1.097575    1.192504
      _cons  |   2544.921   1128.816    17.68   0.000     1066.887    6070.581
------------------------------------------------------------------------------

. predict recalibration2006 if yydx==2006, timevar(timevar10) meansurv
```

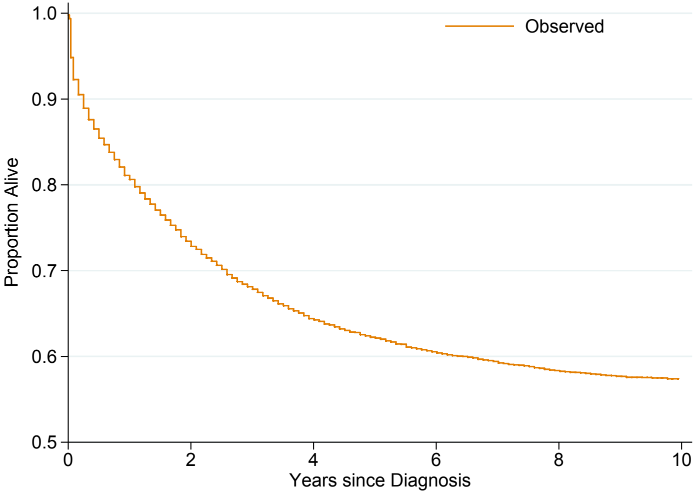
Model: Temporal Recalibration

```
. stpm2 agercs* female black, scale(hazard) noorthog constraints(1 2 3 4 5) ///
> bknots(`bknots´) knots(`knots´) eform
note: delayed entry models are being fitted

Log likelihood = -16015.094                       Number of obs     =     39,743

------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
     agercs1 |   1.012557  (constrained)
     agercs2 |    1.00005  (constrained)
     agercs3 |   .9999177  (constrained)
      female |   .9098671  (constrained)
       black |   1.403117  (constrained)
       _rcs1 |   23.11036   2.852501    25.44   0.000     18.14443    29.43541
       _rcs2 |   1.201228   .0117535    18.74   0.000     1.178411    1.224486
       _rcs3 |   .7933542   .0212882    -8.63   0.000     .7527083    .8361949
       _rcs4 |   .9970216   .0372468    -0.08   0.936     .9266278    1.072763
       _rcs5 |   1.144055     .02421     6.36   0.000     1.097575    1.192504
       _cons |   2544.921   1128.816    17.68   0.000     1066.887    6070.581
------------------------------------------------------------------------------

. predict recalibration2006 if yydx==2006, timevar(timevar10) meansurv
```

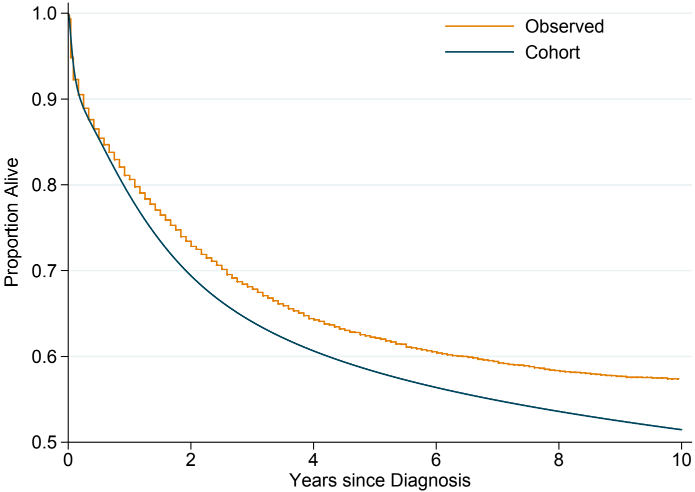
Model: Period Analysis

```
. stpm2 agercs* female black, scale(hazard) df(5) eform
note: delayed entry models are being fitted
Log likelihood =  -16080.35                    Number of obs     =     39,743

------------------------------------------------------------------------------
             |     exp(b)   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
xb           |
     agercs1 |   1.004674   .0051795     0.90   0.366     .9945736    1.014877
     agercs2 |   1.000028   .0000157     1.80   0.072     .9999974    1.000059
     agercs3 |   .9999383    .000019    -3.24   0.001      .999901    .9999756
      female |   .9084784   .0266046    -3.28   0.001     .8578025     .962148
       black |   1.441617   .0606779     8.69   0.000     1.327464    1.565587
       _rcs1 |   2.014562   .0187427    75.28   0.000      1.97816    2.051634
       _rcs2 |   1.124344   .0079382    16.60   0.000     1.108892     1.14001
       _rcs3 |   .9535394   .0044961   -10.09   0.000     .9447678    .9623925
       _rcs4 |   1.069052    .003847    18.56   0.000     1.061538    1.076618
       _rcs5 |   1.008206   .0025619     3.22   0.001     1.003198     1.01324
       _cons |   .3234849   .0094079   -38.81   0.000     .3055615    .3424596
------------------------------------------------------------------------------

. predict period2006 if yydx==2006, timevar(timevar10) meansurv
```

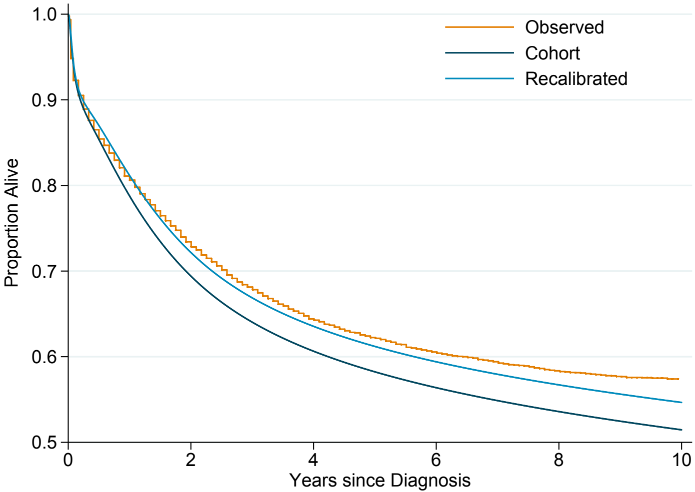
10 Year Marginal Survival



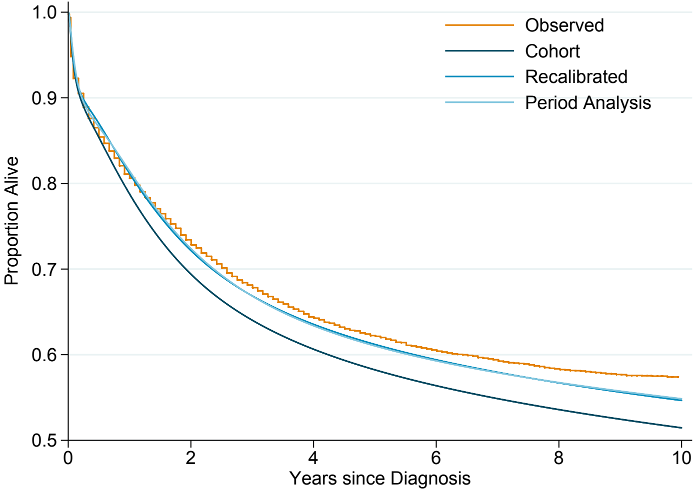
10 Year Marginal Survival



10 Year Marginal Survival

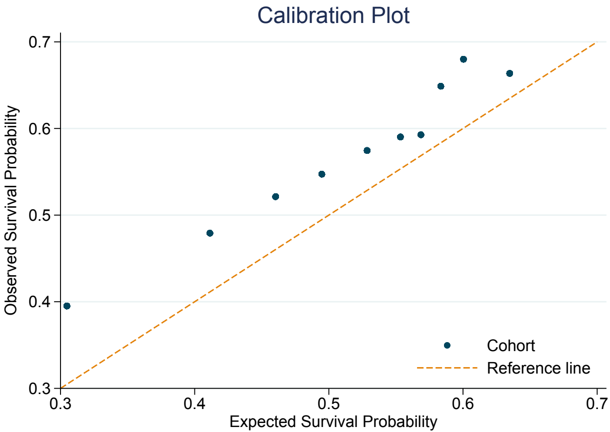


10 Year Marginal Survival



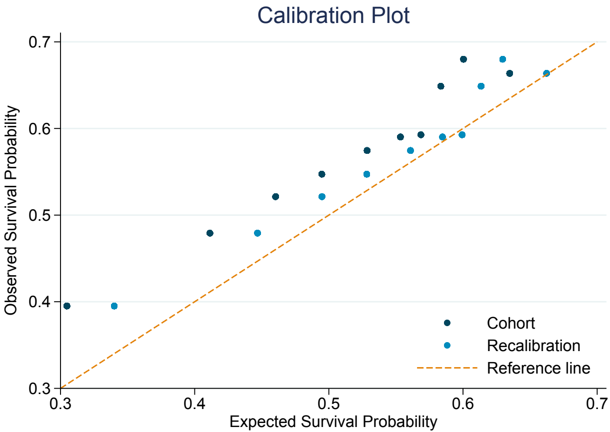
Calibration of Models

```
. predict prognosticindex, xbnobaseline
. xtile calibrationgroup = prognosticindex, n(10)
```



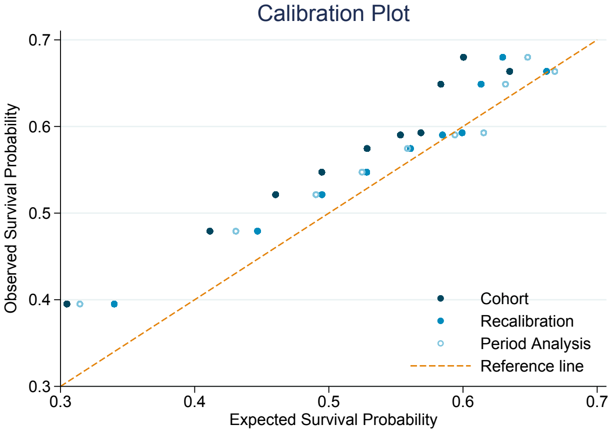
Calibration of Models

```
. predict prognosticindex, xbnobaseline
. xtile calibrationgroup = prognosticindex, n(10)
```

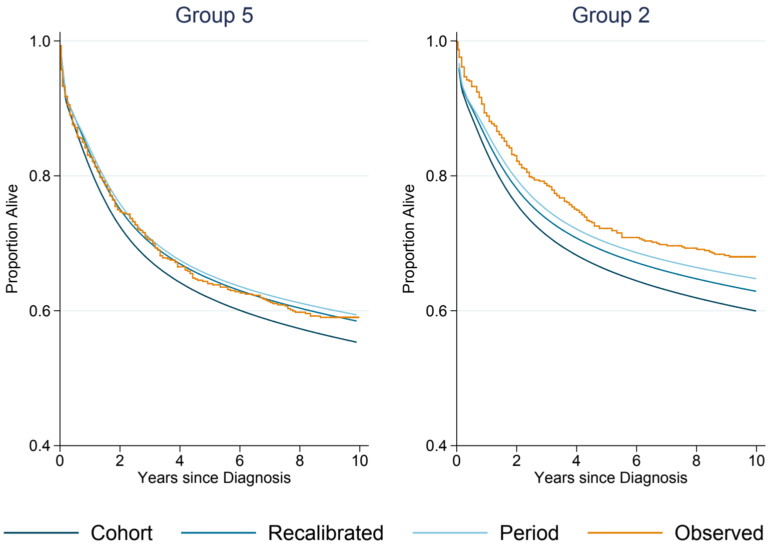


Calibration of Models

```
. predict prognosticindex, xbnobaseline
. xtile calibrationgroup = prognosticindex, n(10)
```

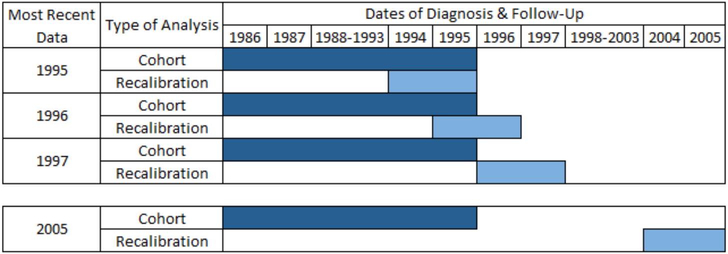


Risk Groups



Group 5
Group 2
Proportion Alive
1.0
0.8
0.6
0.4
0
2
4
6
8
10
Years since Diagnosis
Cohort
Recalibrated
Period
Observed

Use of New Data

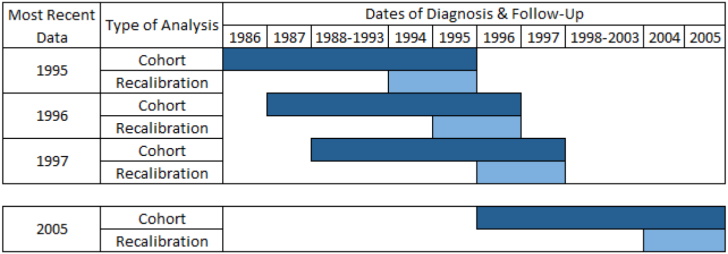


- Keep the original model from 1986-1995
- Recalibrate using a period window for the 2 most recent years

  ⋮

- Continue until recalibrating with a period window of 2004-2005

Use of New Data

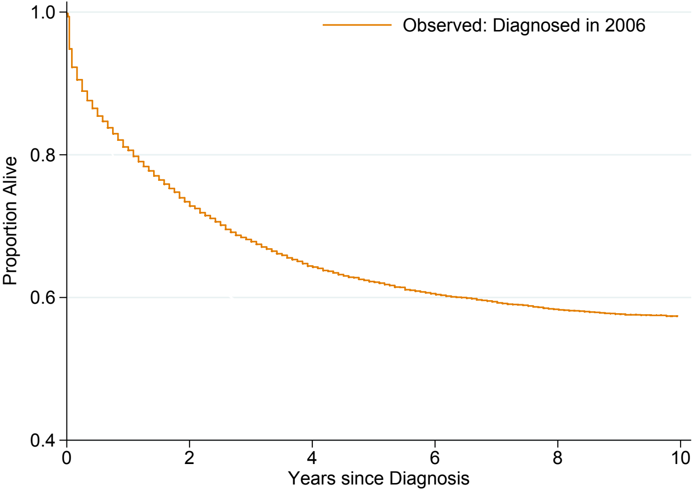


- Refit the model each year using the most recent 10 years of data
- Recalibrate using a period window for the 2 most recent years

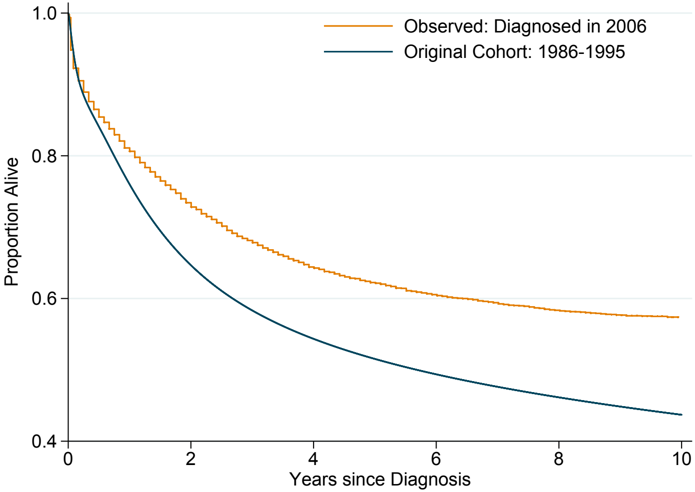
⋮

- Continue until recalibrating with a period window of 2004-2005

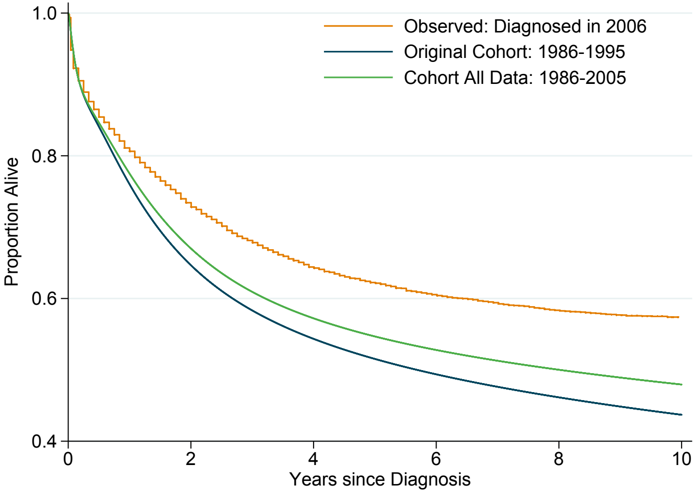
10 Year Marginal Survival Predictions



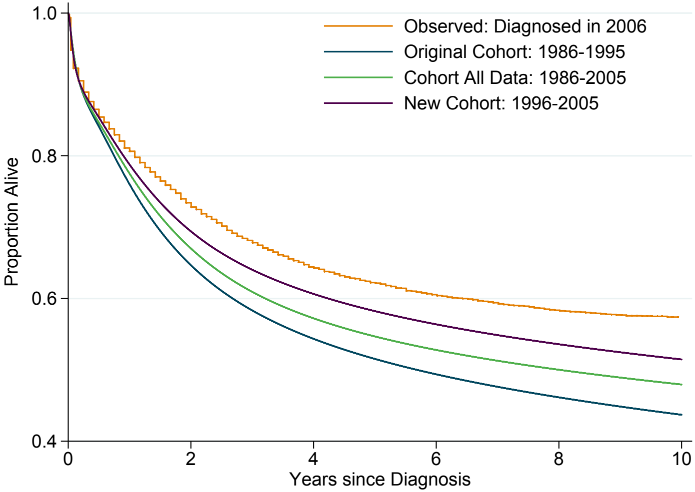
10 Year Marginal Survival Predictions



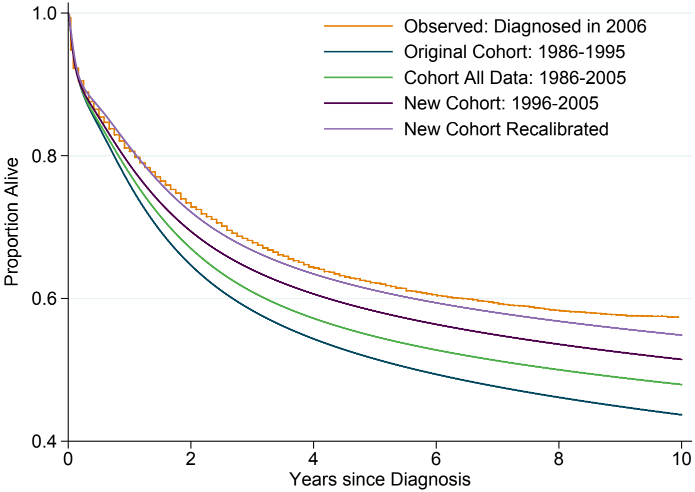
10 Year Marginal Survival Predictions



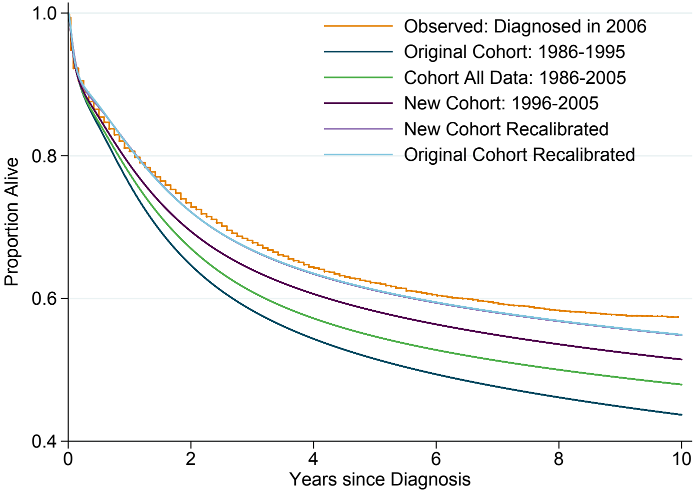
10 Year Marginal Survival Predictions



10 Year Marginal Survival Predictions



10 Year Marginal Survival Predictions



Summary

- Cohort models often underestimate survival
- Period analysis uses a subset of data to create more up-to-date survival predictions
- Very similar predictions are produced using temporal recalibration but all the data is used
- Simple to fit these types of models in Stata using `stset` to define the sample, `stpm2` and `constraints` to fit the models
- Importance of regularly updating models when new data becomes available
- These methods can also be used for non-proportional hazard models

Selected References

dos Reis, F. J. C., Wishart, G. C., Dicks, E. M. et al. (2017)
An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation, *Breast Cancer Research*, 19. PREDICT Version 2.1 tool available from: `http://www.predict.nhs.uk/predict_v2.1/`

Surveillance, Epidemiology, and End Results (SEER) Program (`www.seer.cancer.gov`)
Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.

Royston, P. & Lambert, P. C. (2011)
Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model, *Stata Press*.

Hinchliffe, S. R. & Lambert, P. C. (2013)
Flexible parametric modelling of cause-specific hazards to estimate cumulative incidence functions, *BMC Medical Research Methodology, Springer Nature*, 13.

Brenner, H. & Gefeller, O (1996)
An alternative approach to monitoring cancer patient survival, *Cancer*, 78, 2004-2010.

Brenner, H. & Hakulinen, T. (2009)
Up-to-date cancer survival: Period analysis and beyond, *International Journal of Cancer*, 124, 1384-1390.