Producing up-to-date survival predictions from prognostic models

Sarah Booth[1]   Richard D. Riley[2]   Joie Ensor[2]
Paul C. Lambert[1,3]   Mark J. Rutherford[1]

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

[2]Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University, Keele, UK

[3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
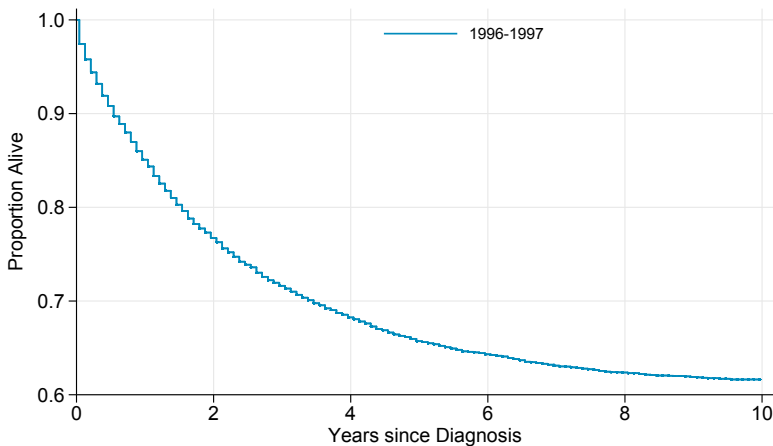
7th October 2019

PhD Project

Prognostic Models

- Prognostic models can be used to inform patients and aid treatment decisions
- Often built using data collected over a long time period
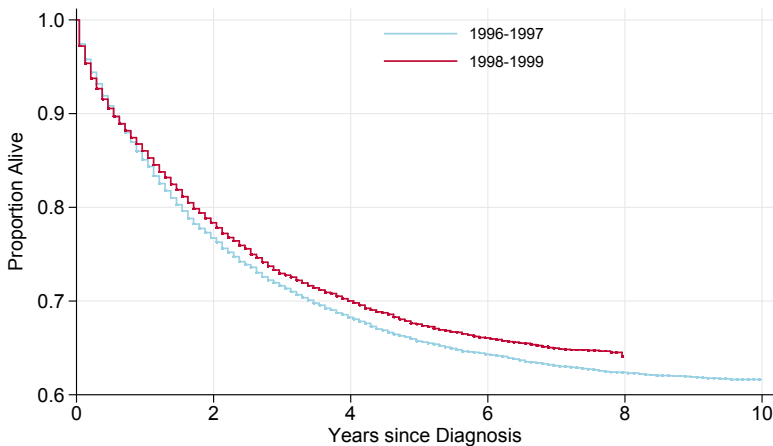- Improvements in survival may lead to out-dated survival predictions

Methods

- Developed temporal recalibration which combines period analysis with recalibration techniques
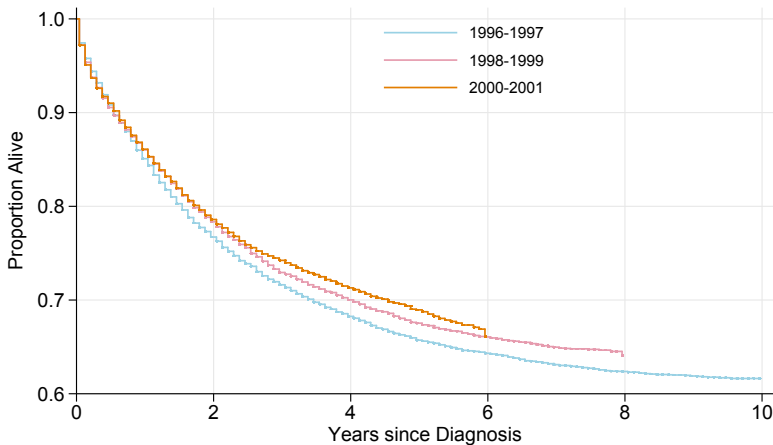- Alternative approach involving modelling calendar time
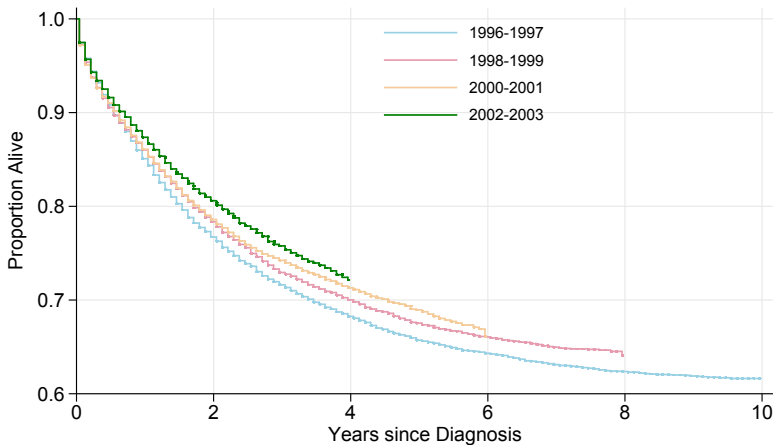
Assessing Improvements in Survival

Assessing Improvements in Survival

Assessing Improvements in Survival

Assessing Improvements in Survival

Assessing Improvements in Survival

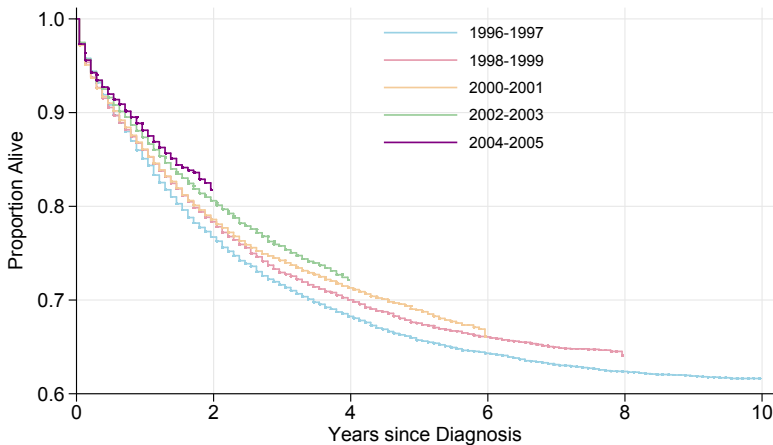## Assessing Improvements in Survival

Assessing Improvements in Survival

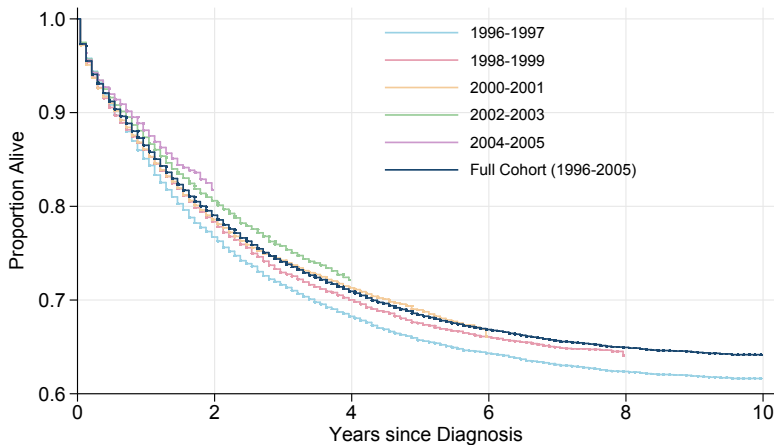## Assessing Improvements in Survival

Assessing Improvements in Survival

Assessing Improvements in Survival

Assessing Improvements in Survival

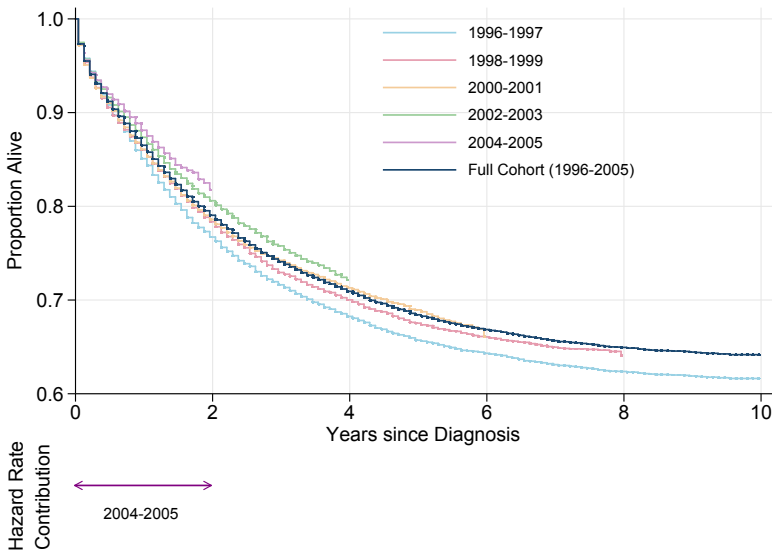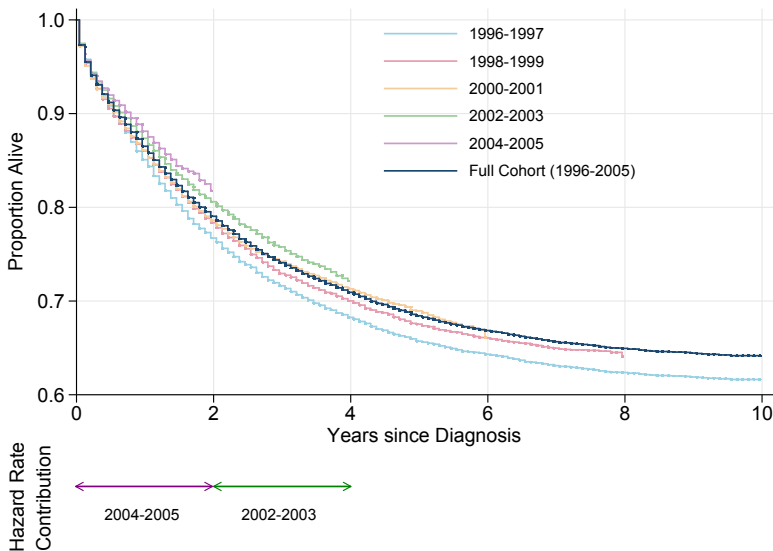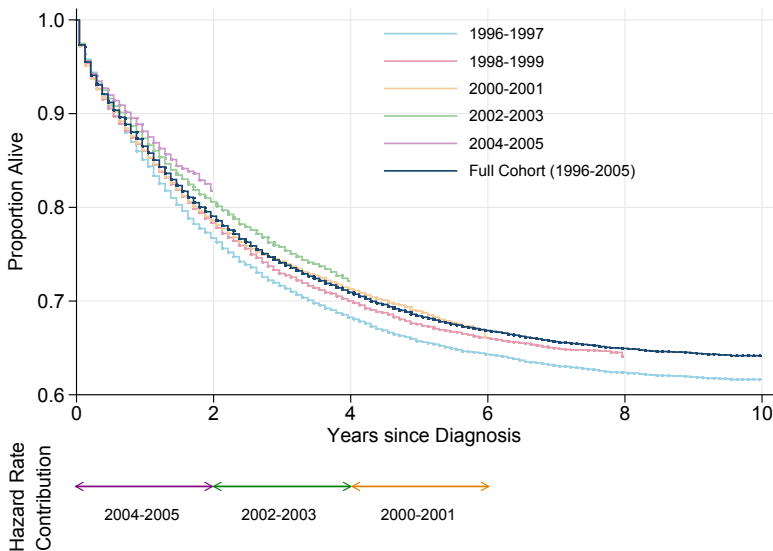## Assessing Improvements in Survival

Period Analysis

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

Advantages

- More up-to-date survival estimates, people diagnosed many years ago only contribute to long-term hazard rates

---

H. Brenner and O. Gefeller, "An alternative approach to monitoring cancer patient survival," Cancer, vol. 78, no. 9, pp. 2004–2010, 1996.

Period Analysis

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

Advantages

- More up-to-date survival estimates, people diagnosed many years ago only contribute to long-term hazard rates

Disadvantages

- Reduces sample size and number of events
- Larger standard errors

H. Brenner and O. Gefeller, "An alternative approach to monitoring cancer patient survival," Cancer, vol. 78, no. 9, pp. 2004–2010, 1996.

Temporal Recalibration

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

Method

- Fit a full cohort model

Temporal Recalibration

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

Method

- Fit a full cohort model
- Use a period analysis sample to recalibrate the model

Temporal Recalibration

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

Method

- Fit a full cohort model
- Use a period analysis sample to recalibrate the model
- The predictor effects are constrained to be the same (i.e hazard ratios for age, sex, stage are the same)
- The baseline hazard function is allowed to vary which can capture any improvements in survival

Summary of Methods

Type of Analysis	Predictor Effects	Baseline
Full Cohort	All	All
Period Analysis	Recent	Recent
Temporal Recalibration	All	Recent

Models

Type of Analysis	Year of Diagnosis & Follow-Up					Follow-Up Only
	1996-2002	2003	2004	2005	2006	2007-2015
Full Cohort						
Temporal Recalibration						
Period Analysis						
Validation						

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

Background
○○

Methods
○○○

Models
●

Results
○○

Conclusion
○

Discussion
○

Models

Type of Analysis	Year of Diagnosis & Follow-Up					Follow-Up Only
	1996-2002	2003	2004	2005	2006	2007-2015
Full Cohort						
Temporal Recalibration						
Period Analysis						
Validation						

- US colon cancer registry data from the Surveillance, Epidemiology, and End Results (SEER) Program

---

Models

| Type of Analysis	Year of Diagnosis & Follow-Up					Follow-Up Only
	1996-2002	2003	2004	2005	2006	2007-2015
Full Cohort						
Temporal Recalibration						
Period Analysis						
Validation						

- US colon cancer registry data from the Surveillance, Epidemiology, and End Results (SEER) Program
- Cause-specific flexible parametric survival models

Models

Type of Analysis	Year of Diagnosis & Follow-Up					Follow-Up Only
	1996-2002	2003	2004	2005	2006	2007-2015
Full Cohort						
Temporal Recalibration						
Period Analysis						
Validation						

- US colon cancer registry data from the Surveillance, Epidemiology, and End Results (SEER) Program
- Cause-specific flexible parametric survival models
- No variable selection: Age (modelled with splines), sex, race, stage and grade
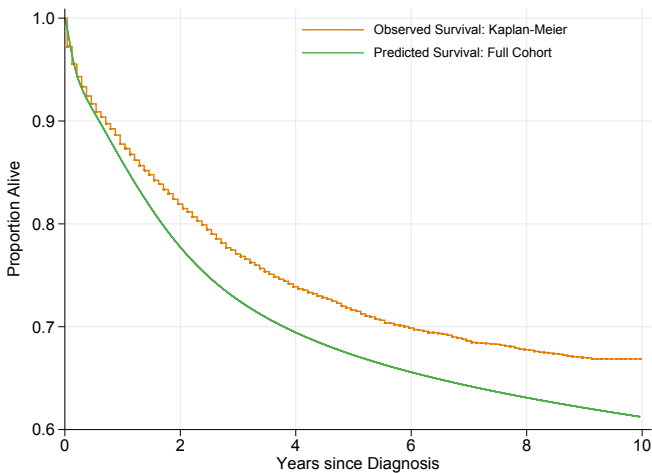

Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2015), National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission
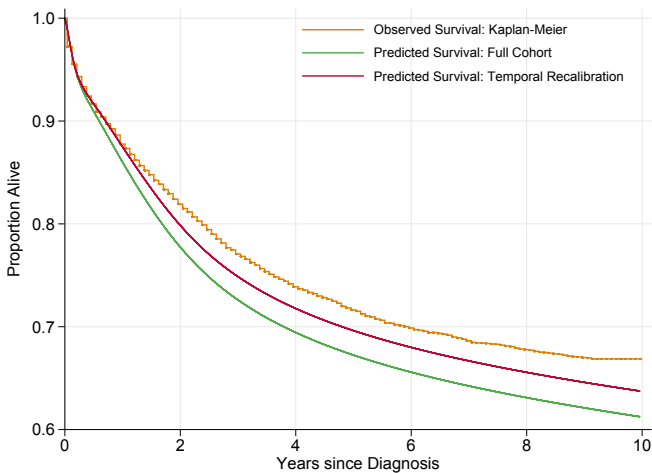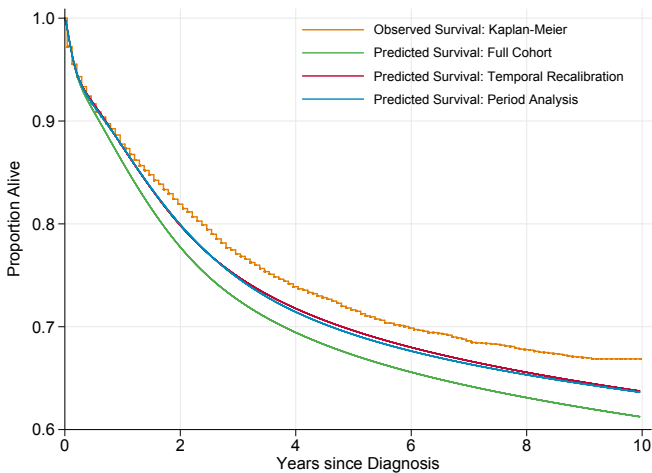
Model Validation: Individuals diagnosed in 2006
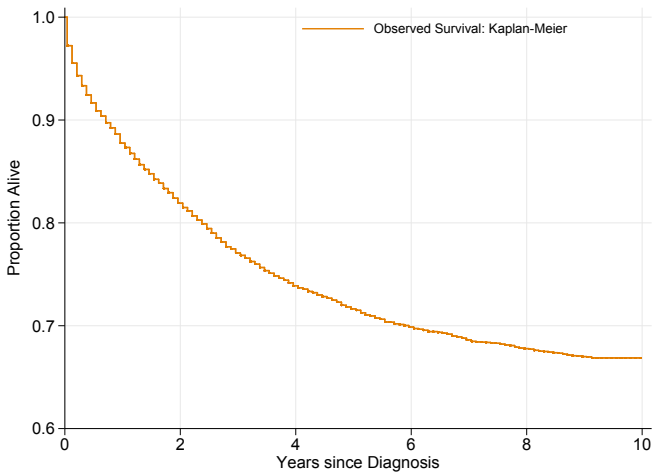
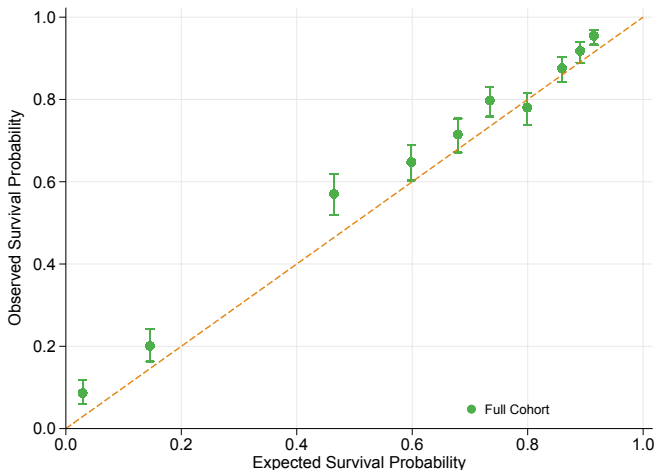# Model Validation: Individuals diagnosed in 2006

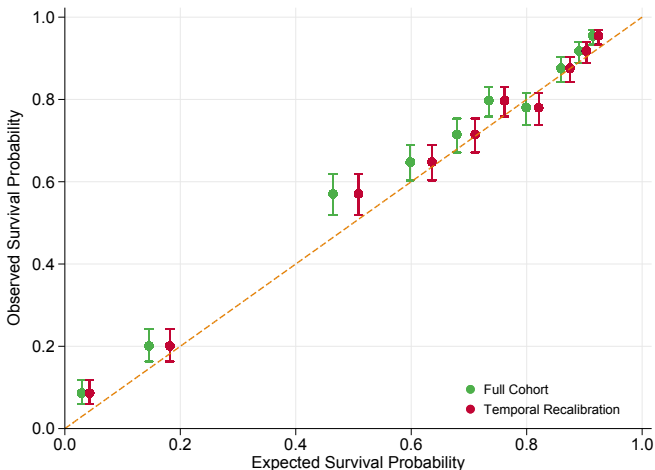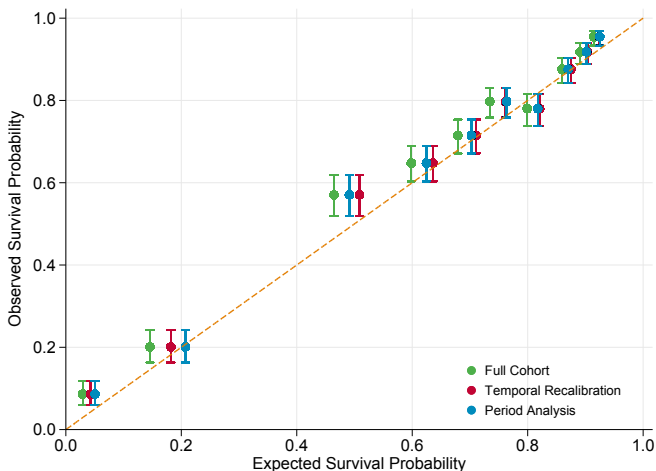# Model Validation: Individuals diagnosed in 2006

Model Validation: Individuals diagnosed in 2006

Calibration of Models

Calibration of Models

Background
○○

Methods
○○○

Models
○

Results
○●

Conclusion
○

Discussion
○

Calibration of Models

Summary

- Full cohort models often underestimate survival if there have been recent improvements in survival

- Period analysis uses a subset of data to create more up-to-date survival predictions

- Temporal recalibration also produces more up-to-date survival predictions but all the data is used to estimate the predictor effects

- Temporal recalibration can also be used to easily update existing prognostic models

Modelling calendar time (year of diagnosis)

Functional form

- Linear, categorical, restricted cubic splines
- Incorporate month of diagnosis for a smoother function
- Time dependent effects
- Interactions between predictor effects and year of diagnosis

Survival predictions e.g. new patient diagnosed in 2007 and the model is fitted using data from 1996-2005

- Use the most recent year (2005) included in the model
- Extrapolate the trend to 2007
- Update the prognostic model every year