Application of `stpm2` to estimate relative survival for cancer patients in the Nordic countries

Frida Lundberg

Paul Lambert, Anna Johansson, Therese Andersson, Mats Lambe, David Pettersson, Gerda Engholm, Lina Mørch, Tom Johannesen, Anni Virtanen, Helgi Birgisson, Elínborg Ólafsdóttir

The NORDCAN Survival Project

- Comparisons of cancer survival across the Nordic countries
  - Denmark, Finland, Iceland, Norway and Sweden

- Investigate possible differences in survival for nine cancer sites
  - Colon, rectum, lung, skin melanoma, kidney
  - Breast, uterus, ovary, prostate

- Highlight differences using novel measures of survival

Measures of cancer-specific survival

- Crude probability of death due to cancer
  - In the presence of competing risks (other causes of death)
  - Patient prognostic measure

- Net probability of death due to cancer
  - Hypothetical world where you cannot die of other causes
  - Competing risks are assumed to be eliminated
  - Independent of background mortality
  - Comparable across age, calendar time or country

Estimation of cancer survival

- Cause-specific framework: using cause of death information

$$\text{Cause specific mortality} = \frac{\text{number of deaths due to cancer}}{\text{person time at risk}}$$

- Relative survival framework: using expected mortality tables

$$\text{Excess mortality} = \text{all cause mortality} - \text{expected mortality}$$

$$\text{Relative survival ratio} = \frac{\text{all cause survival proportion}}{\text{expected survival proportion}}$$

Material

- Data from the NORDCAN database from five countries
- Individual level data on patients diagnosed 1990-2016
  - Year and month of diagnosis
  - Follow-up time in days
  - Status at end of follow-up (alive, dead, emigrated)
  - Patients' sex, age at diagnosis and country
  - Cancer site
- Population based mortality rates from national statistics offices (expected mortality)
  - By country, age, year and sex

Flexible parametric RS models

- Time since diagnosis as primary time-scale
- Log cumulative baseline excess hazard modelled continuously
  - Restricted cubic splines (rcs) with 5 degrees of freedom
- Age and calendar year included as continuous variables
- Sex included as binary variable where appropriate
- Two-way interactions between age, year and sex
- Two- and three-way interactions with time since diagnosis
  - Relaxing proportional excess hazard assumption
- Separate models for each country and cancer site

Main model – `stpm2`

```
stset followup_days, failure(status==2)                    ///
   exit(time 10*365.24) scale(365.24) id(id)


stpm2 rcs3age* rcs3year* sex                               ///
   rcs2age2year* rcs2agesex* rcs2yearsex* ,                ///
   tvc(rcs3age* rcs3year* sex                              ///
   rcs2age2year* rcs2agesex* rcs2yearsex* )                ///
   scale(hazard) bhazard(rate)                             ///
   df(5) dftvc(2)
```

- Models the log cumulative excess hazard over time since diagnosis

Model stability

- Started with an ’ideal model’
- Defined an algorithm to simplify the model if convergence failed
  - Winsorizing for tails of age (at different percentiles)
  - Fewer degrees of freedom for interaction terms
  - Fewer degrees of freedom for time-varying effects
  - Dropped three-way interactions with time (for age, year and sex)
  - Non-parametric Pohar Perme approach using **`strs`**
- Tested models for all nine sites in each of the five countries

Winsorizing

- 96% of age distribution modelled continuously, individuals outside the 2nd and 98th percentiles of age reassigned to percentile limits
  - Assumed to have the same relative survival

```
_pctile age, per(2)
global age_lo `r(r1)'
gen ageadj = cond(age < $age_lo , $age_lo , age)

_pctile age, per(98)
global age_hi `r(r1)'
replace ageadj = cond(ageadj > $age_hi , $age_hi , ageadj)
```

Outcome measures

- Post-estimation to obtain additional measures using **`standsurv`**
- 1- and 5-year relative survival
  - Age-standardised and age-specific estimates
  - By cancer site, country, sex and calendar year of diagnosis
  - Percentage points change since 1990
- 5-y relative survival conditional on surviving one year
- Period approach for 5-y RS in the most recent period
- Crude probability of death and average number of life-years lost

Age-standardization

- Adapted versions of the International Cancer Survival Standard (ICCS) age-standard weights by 10-year age groups
- Makes estimates comparable across countries with different age distributions among cancer cases
- Regression standardization stratified by calendar year and sex

```
recode age (min/49=1) (50/59=2) (60/69=3) (70/79=4) ///
        (80/max=5), gen(agegrp)

recode agegrp (1=0.11906) (2=0.16735) (3=0.27593)    ///
              (4=0.28897) (5=0.14869) , gen(agewt)
```

5-year relative survival – `standsurv`

```
keep if female == `sex' & yydx == `year'
local totalobs = _N
bysort agegrp: gen standwt = _N/`totalobs'
gen indwt = agewt/standwt

gen tflag = 1 in 1
gen t5 = 5 if tflag == 1

standsurv, at1(.)                                ///
   atvar(surv5_`year'_`sex')                     ///
   timevar(t5) ci indweights(indwt)
```

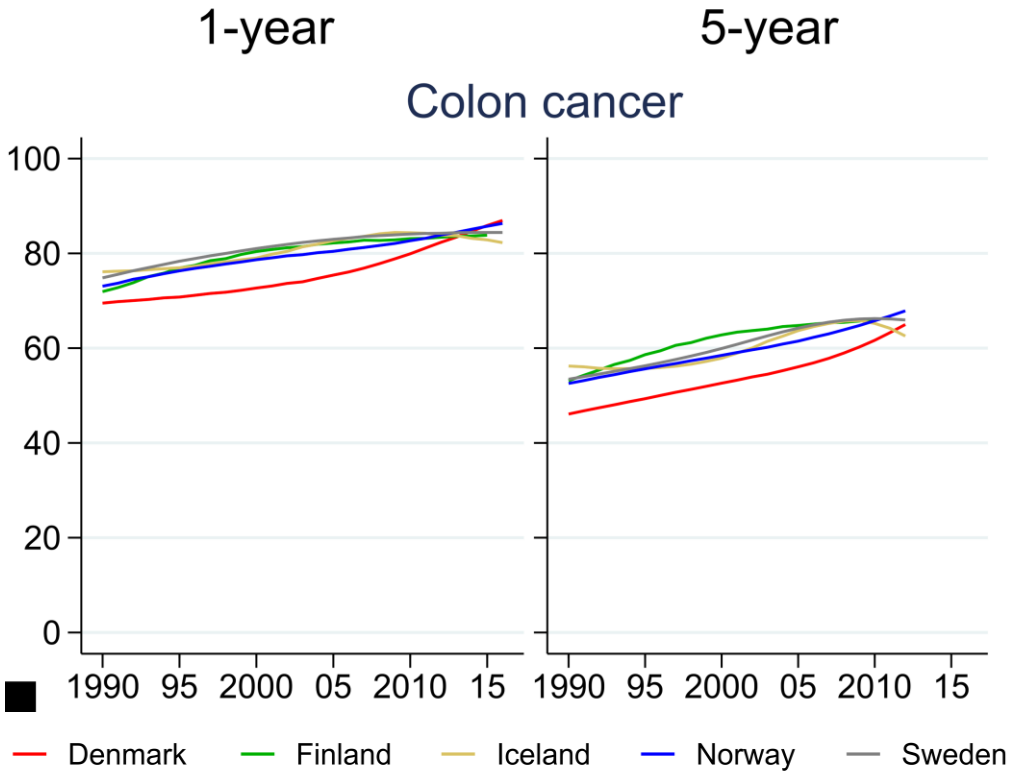
Conditional relative survival – `standsurv`

```
gen t1 = 1 if tflag == 1


standsurv,                                          ///
   at1(., attimevar(t1))                            ///
   at2(., attimevar(t5))                            ///
   atvar(a5_t1_`year'_`sex'                         ///
   b5_t1_`year'_`sex')                              ///
   contrast(ratio)                                  ///
   contrastvar(surv5_c1_`year'_`sex')               ///
   ci indweights(indwt)
```

Loops and output

```
foreach country in se dk no fi is {
   foreach site in $sitelist {
      //load data, stset, merge with population mortality rates,
      //winsorize, create spline variables and interaction terms
      //run stpm2 model
      forvalues year = 1990/2016 {
         foreach sex in 0 1 {
            foreach endtime in 1 5 {
               foreach starttime in 0 1 {
                  //standsurv
                  //save estimates to dataset
               }
            }
         }
      }
   }
}
//create tables and graphs
```

1- and 5-y RS, women with colon cancer



Supplementary table 4. Trends in 1-year relative survival 1990-intervals, the NORDCAN survival studies

Sex	Country	Site	1990	1995	2000
Women	Denmark	Colon	70 (68-71)	71 (70-72)	73 (72-73)
Women	Finland	Colon	72 (70-74)	77 (76-78)	80 (80-81)
Women	Iceland	Colon	76 (69-84)	77 (73-81)	79 (75-83)
Women	Norway	Colon	73 (72-75)	76 (76-77)	79 (78-79)
Women	Sweden	Colon	75 (74-76)	78 (78-79)	81 (80-82)

Supplementary table 5. Trends in 5-year relative survival 1990-intervals, the NORDCAN survival studies

Sex	Country	Site	1990	1995	2000
Women	Denmark	Colon	46 (44-48)	49 (48-50)	53 (52-54)
Women	Finland	Colon	53 (51-56)	59 (57-60)	63 (61-64)
Women	Iceland	Colon	56 (47-68)	56 (50-62)	58 (53-64)
Women	Norway	Colon	53 (50-55)	56 (55-57)	58 (57-60)
Women	Sweden	Colon	53 (52-55)	56 (55-57)	60 (59-61)

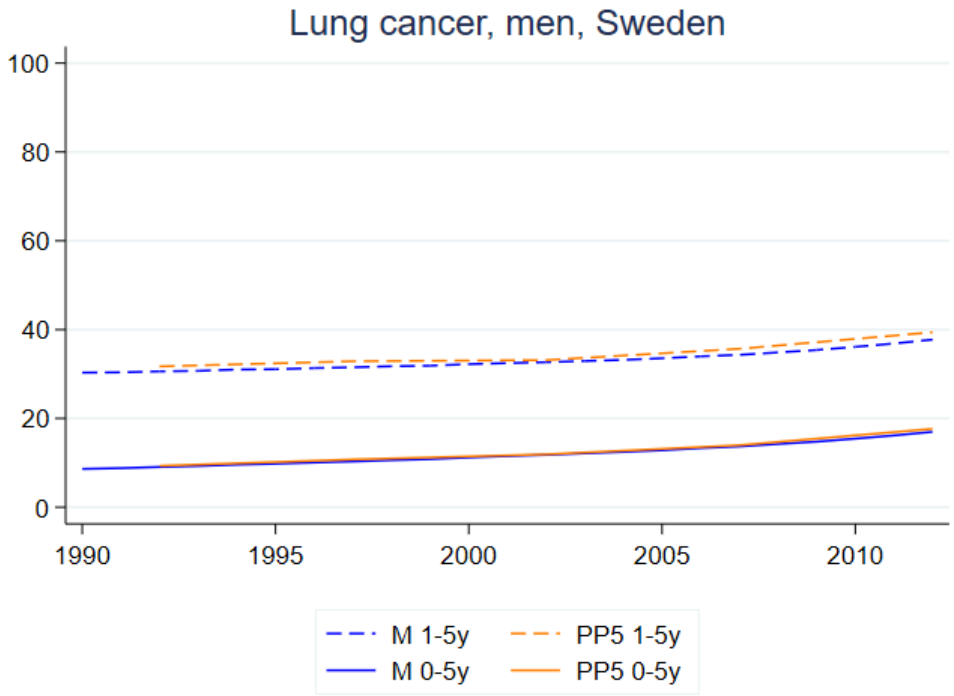
Non-parametric estimates – `strs`

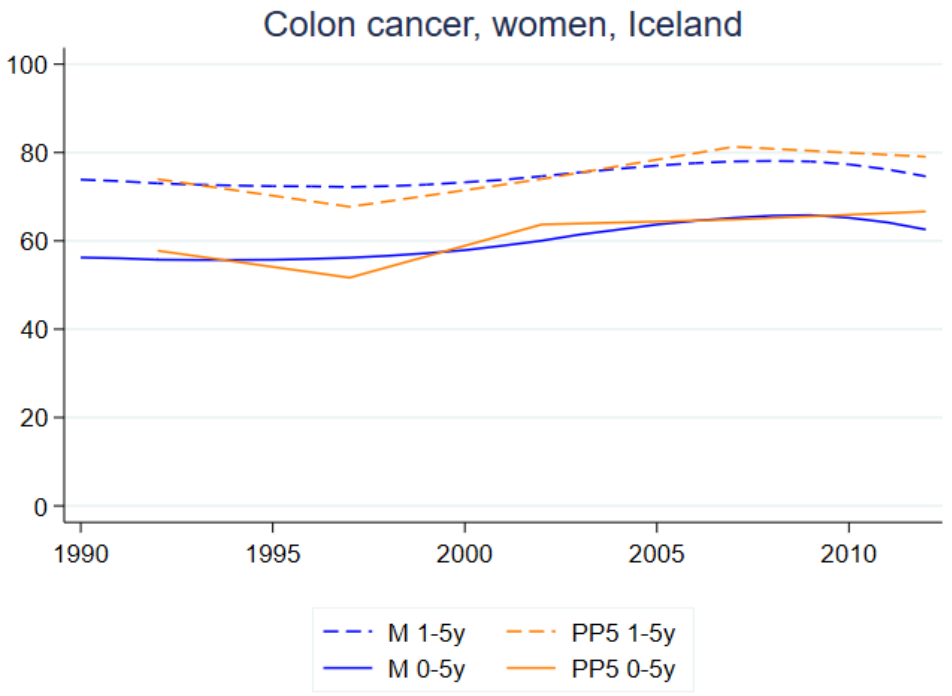
```
bysort yrgrp5 sex: gen totalobs = _N
bysort agegrp yrgrp5 sex: gen totalobs_age = _N
gen standwt = totalobs_age/totalobs
gen indwt = agewt/standwt

stset followup_days, failure(status==2)          ///
   exit(time 10*365.24) scale(365.24) id(id)

strs using popmort_`country'                     ///
 , br(0(0.5)5) mergeby(sex _year _age)           ///
 by(sex yrgrp5) pohar indweight(indwt)
```

Conditional relative survival – `strs`

```
stset followup_days, failure(status==2)          ///
   exit(time 10*365.24) scale(365.24) id(id)     ///
   enter(time 365.24)

strs using popmort_`country'                     ///
 , br(0(0.5)5) mergeby(sex _year _age)           ///
 by(sex yrgrp5) pohar indweight(indwt)
```

Comparisons to non-parametric estimates





Summary

- Advantages of model-based estimation
  - Possible to obtain estimates for specific covariate patterns, e.g. specific ages and years
  - Contrasts for specific covariate patterns
  - Post-estimation of life-years lost and other measures

- Possible limitations
  - Convergence issues especially when data is sparse
  - More work intensive than non-parametric estimation
    - `stpm3` is on the way!

References

- Lundberg FE, et al. Survival trends in patients diagnosed with colon and rectal cancer in the nordic countries 1990–2016: The NORDCAN survival studies. *Eur J Cancer* 2022;172:76–84.
- Lundberg FE, et al. Trends in cancer survival in the Nordic countries 1990-2016: the NORDCAN survival studies. *Acta Oncol* 2020;59(11):1266–74.
- Syriopoulou E, et al. Robustness of individual and marginal model-based estimates: A sensitivity analysis of flexible parametric models. *Cancer Epidemiol* 2019;58:17–24.
- Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J* 2009;9(2):265–90.

Thank you for your attention!

frida.lundberg@ki.se