Modelling excess mortality comparing to a control population: A combined additive and relative hazards model

Caroline E. Weibull

Therese M-L. Andersson, Michael J. Crowther

2023 Northern European Stata Conference

Population-based cancer patient survival

- We are often interested in estimating *net survival* - the probability of survival where the specific cancer is the only possible cause of death.
- Important measure for the evaluation of cancer patient care.
- Net survival can be estimated in a cause-specific or a relative survival framework [1]. We typically use the latter.

Relative Survival

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

The expected survival, $S^*(t)$, is obtained from national population mortality files (*popmort files*) stratified by age, sex, and calendar year.

Modelling is done on the hazard scale

$$\begin{array}{ccccc} \text{Observed} & & \text{Expected} & & \text{Excess} \\ \text{Mortality Rate} & = & \text{Mortality Rate} & + & \text{Mortality Rate} \end{array}$$

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

The model assumes an additive effect (of cancer) and allows for multiplicative effects (e.g., between stages) among the cancer patients. Expected rate treated as a fixed number (conditional on age, sex, year), with no uncertainty.

Any relative survival analysis requires:

- Cohort of cancer patients, and
- An accurate 'popmort' file.

When we run into problems

What if we wish to estimate excess hazards (i.e., additive effect of cancer), but we don't have an appropriate population mortality rates file?
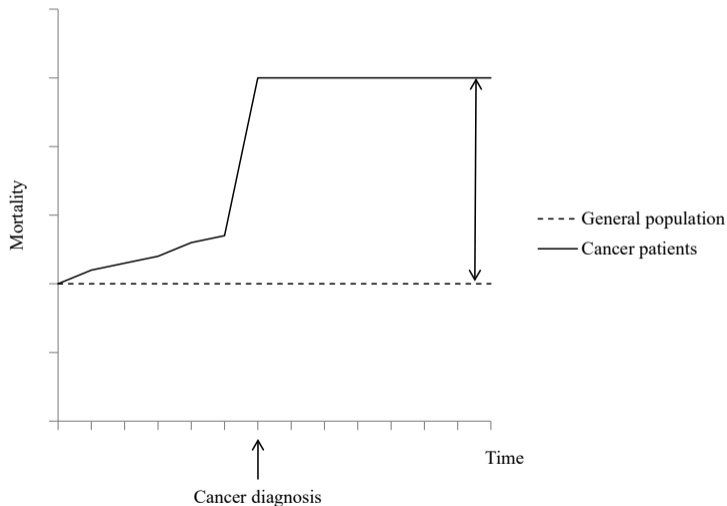
This could be due to:

- It simply doesn't exist.
- Popmort is not stratified finely "enough".
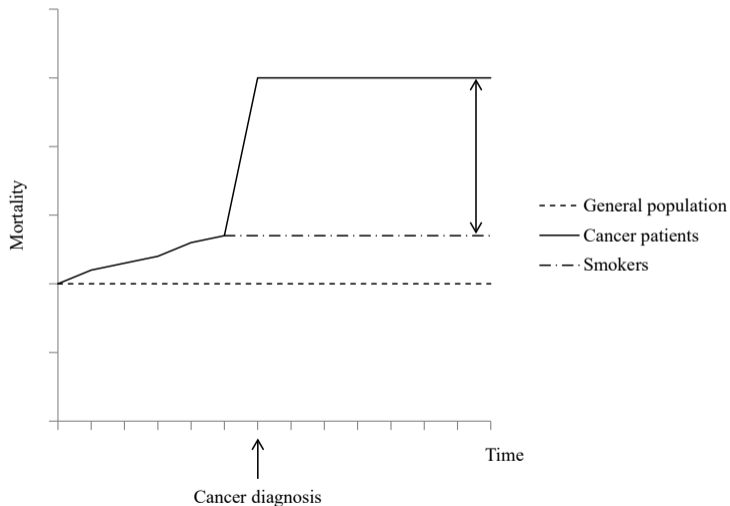- There are additional risk factors that we need to account for.

Typical situation when (standard) relative survival is inappropriate
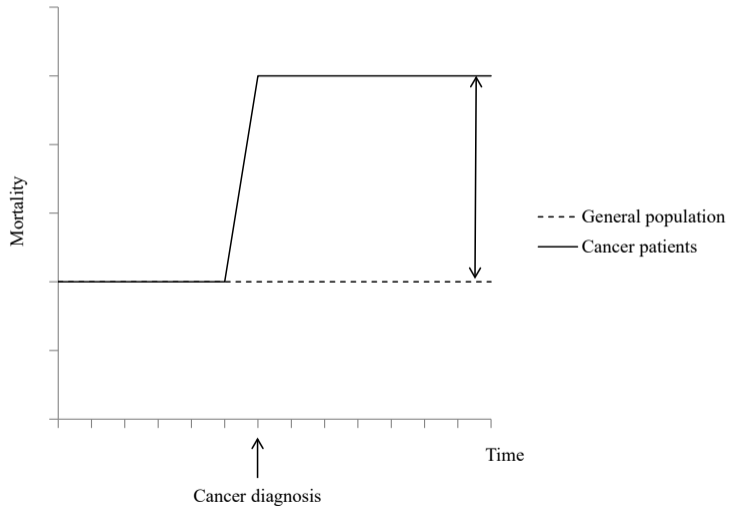
Typical situation when (standard) relative survival is inappropriate

Typical situation when (standard) relative survival is inappropriate

Aim - Develop and implement a spline-based combined additive and relative hazards model for estimation of relative survival using controls

- Solves the problem with inadequate popmort files in relative survival.
- Incorporates uncertainty in the expected mortality.
- Estimation of expected and excess mortality in one single model.
- Enables user-friendly predictions of relative survival and functions thereof.
- Bonus: Extends beyond the relative survival setting.

Model specification

Data consisting of cancer patients and some control population.

$$h(t|\boldsymbol{x}) = h^*(t|\boldsymbol{x_1}) + \mathsf{v} \cdot \lambda(t|\boldsymbol{x_2}), \qquad \text{where } \mathsf{v} = \begin{cases} 1 \text{ for cancer patients} \\ 0 \text{ else} \end{cases}$$

$$h^*(t|\boldsymbol{x_1}) = h_0^*(t) \cdot \exp\left[\boldsymbol{x_1}\boldsymbol{\beta_1}\right]$$

$$\lambda(t|\boldsymbol{x_2}) = \lambda_0(t) \cdot \exp\left[\boldsymbol{x_2}\boldsymbol{\beta_2}\right]$$

Flexible parametric excess hazard model fitted on the log hazard scale.

- Can include time-dependent effects in $h^*(t)$, or $\lambda(t)$, or both.
- Allows for modelling of multiple time-scales.

Model specification

Data consisting of cancer patients and some control population.

$$h(t|\boldsymbol{x}) = h^*(t|\boldsymbol{x_1}) + v \cdot \lambda(t|\boldsymbol{x_2}), \qquad \text{where } v = \begin{cases} 1 \text{ for cancer patients} \\ 0 \text{ else} \end{cases}$$

$$h^*(t|\boldsymbol{x_1}) = h_0^*(t) \cdot \exp\left[\boldsymbol{x_1}\boldsymbol{\beta_1}\right]$$

$$\lambda(t|\boldsymbol{x_2}) = \lambda_0(t) \cdot \exp\left[\boldsymbol{x_2}\boldsymbol{\beta_2}\right]$$

Flexible parametric excess hazard model fitted on the log hazard scale.

- Can include time-dependent effects in $h^*(t)$, or $\lambda(t)$, or both.
- Allows for modelling of multiple time-scales.

Estimation, predictions, and implementation

- Fully specified likelihood, use of numerical integration to calculate the cumulative hazard function.
- Results can easily be quantified through the prediction of the survival, hazard, and relative survival, as well as transformations of these, and crucially with associated confidence intervals on all measures.

Implemented in the Stata package `stexcess`, which can be downloaded here:
github.com/RedDoorAnalytics/stexcess

```
. stexcess  (female, df(5))        /// expected model
            (female, df(5))        /// excess model
            , indicator(v)         // indicator
```

Illustration on real data

- Data from the Swedish Colorectal Cancer Register database (CRCBaSe).
- 6, 315 individuals diagnosed with colorectal cancer between 2007 and 2016.
- 33, 685 comparators matched 1:6 on birthyear, sex, county, and being colorectal cancer free at time of diagnosis.
- Data is linked with several other health registers, offering information on socioeconomy, comorbidities, and more for both patients and comparators.
- Follow-up from date of diagnosis (matching) until date of death (due to any cause) or right censoring.

Data setup

```
+-----------------------------------------------------------------------------+
| id   cancer   sex   dxyr   dxage   _d         _t   _year   _age     rate      rate2 |
|-----------------------------------------------------------------------------|
|  1      0      F    2012    64     0   5.0679006   2017     69   .009929   .009327 |
|  2      0      M    2012    74     1   4.9145767   2016     78   .042893   .044297 |
|  3      1      M    2009    72     0   8.9475413   2017     80     .0495   .053663 |
|  4      0      F    2014    69     0    3.244442   2017     72   .014283   .013329 |
|  5      0      M    2013    69     0   4.6626875   2017     73   .022252   .023235 |
+-----------------------------------------------------------------------------+
```

Popmort file rate based on the whole Swedish population (stratified on age, sex, and year) and modelled rate from the comparators.
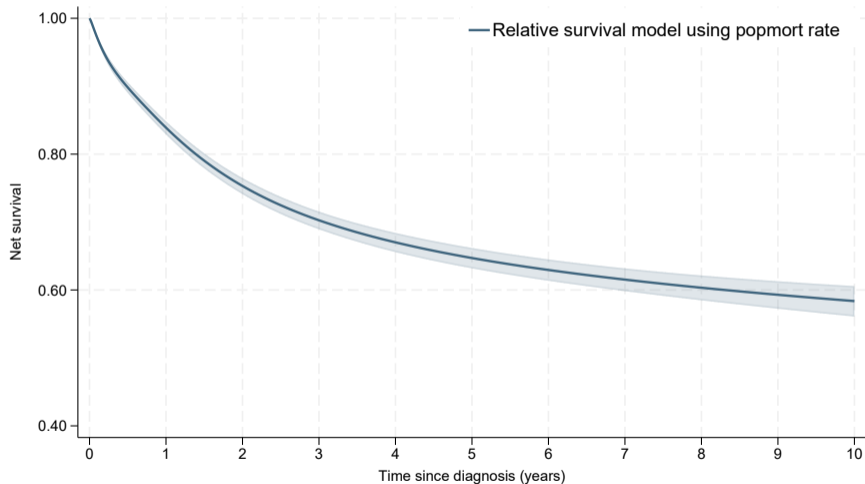
Can we replicate standard relative survival results?

```
// Standard RS model using popmort rate:
. stmerlin if cancer==1, dist(rp) df(6) bhazard(rate)

// Standard RS model using modelled comparator rate:
. stmerlin if cancer==1, dist(rp) df(6) bhazard(rate2)

// Combined additive and relative model:
. stexcess (female                                      /// sex effect
            rcs(_t, df(1) offset(dxyr) log orthog)      /// year timescale
            , df(3) offset(dxage))                      /// age timescale
            (, df(6))                                   /// excess model
            , indicator(cancer)
```
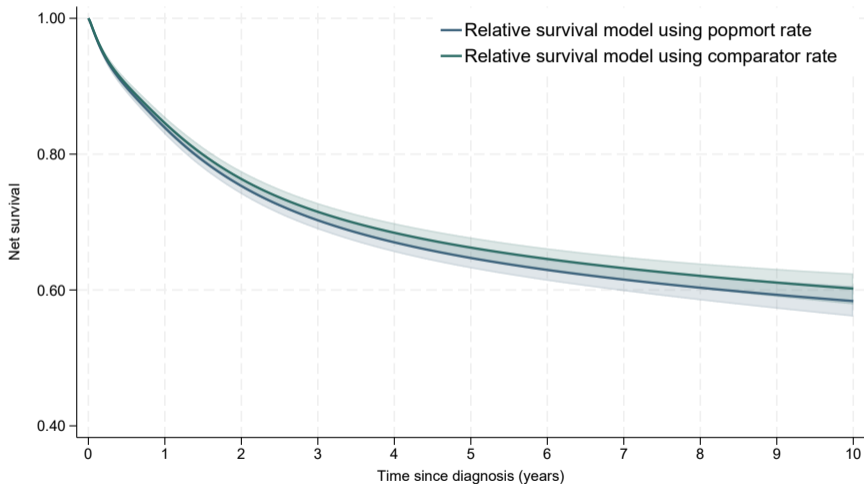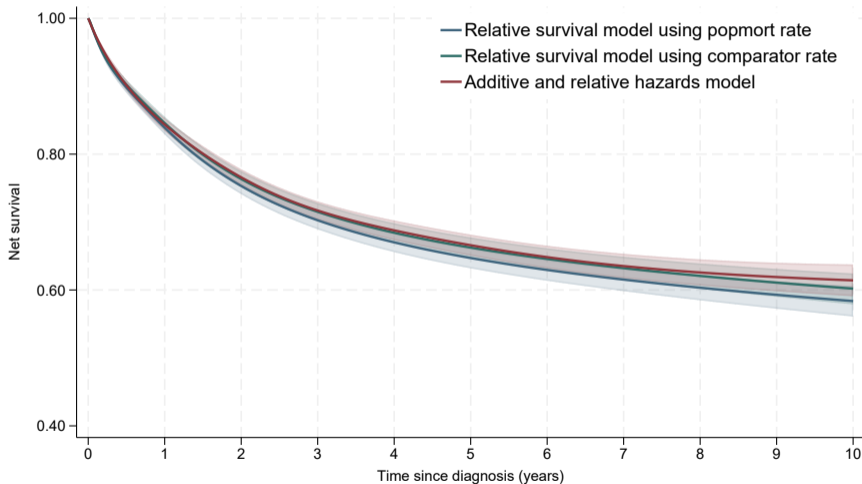
Results: Comparison with standard relative survival

Results: Comparison with standard relative survival

Results: Comparison with standard relative survival

How does it look when the popmort file is inadequate?

E.g., comorbidity has an impact on both colorectal cancer risk and mortality, so in a
standard relative survival analysis we would want to stratify the popmort file on (e.g.)
CCI (Charlson comorbidity index, categorised in $0/1/2+$). This is rarely possible.

```
// Adjusting only the excess model:
. stexcess (female                              /// sex effect
           rcs(_t, df(1) offset(dxyr) log orthog)  /// year timescale
           , df(3) offset(dxage))               /// age timescale
           (cci1 cci2plus, df(6))               /// CCI effect
           , indicator(cancer)

// Adjusting both models:
. stexcess (female cci1 cci2plus                /// sex and CCI effect
           rcs(_t, df(1) offset(dxyr) log orthog)  /// year timescale
           , df(3) offset(dxage))               /// age timescale
           (cci1 cci2plus, df(6))               /// CCI effect
           , indicator(cancer)
```
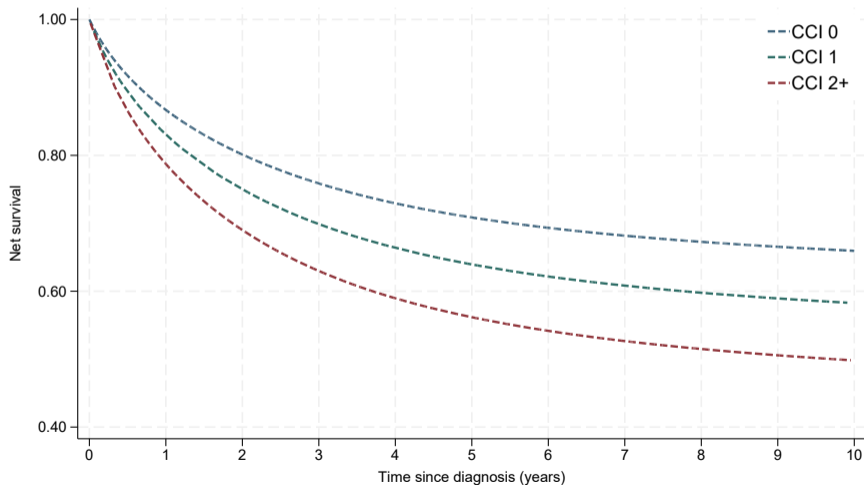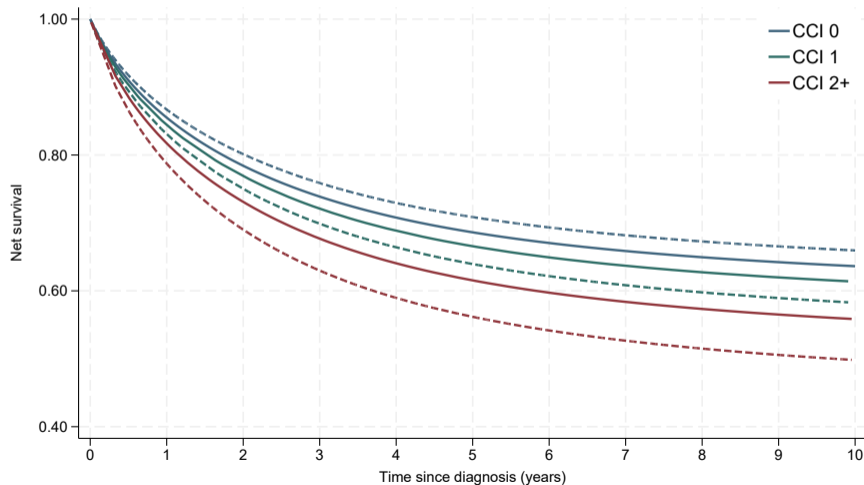
Results: Impact of not accounting appropriately for CCI

Adjusting only the excess model for CCI

Results: Impact of not accounting appropriately for CCI

Adjusting only the excess model (dashed) vs. adjusting both models (solid)

In summary

We propose a combined additive and relative hazards model to supplement the existing toolbox in cancer patients survival research, which can be especially useful in case of insufficient or missing background (expected) mortality.

- Currently in the proof of concept and implementation stage; the model performs well in terms of bias when the data generating mechanism model is accurate, but a full simulation study to evaluate the model in various biologically plausible scenarios (coverage and beyond) is initiated.
- Upcoming work on increasing the speed of the model in more complex situations.

Thank you and some references below

1. Dickman, P. & Coviello, E. Estimating and modeling relative survival. *Stata Journal* **15,** 186–215 (2015).

2. Bower, H. *m. fl.* Adjusting Expected Mortality Rates Using Information From a Control Population: An Example Using Socioeconomic Status. *Am J Epidemiol* **187,** 828–836 (2018).

3. Rubio, F., Rachet, B., Giorgi, R., Maringe, C. & Belot, A. On models for the estimation of the excess mortality hazard in case of insufficiently stratified life tables. *Biostatistics* **22,** 51–67 (2021).

4. Aalen, O. A model for no-parametric regression analysis of counting processes. *Lecture notes on Mathematical Statistics and Probability* **2,** 1–25 (1980).

5. Andersen, P. & Vaeth, M. Simple parametric and nonparametric models for excess and relative mortality. *Biometrics* **45,** 523–35 (1989).

6. Elie, C., Landais, P. & De Rycke, Y. A model combining excess and relative mortality for population-based studies. *Stat Med* **33,** 275–88 (2014).

Why excess mortality due to cancer?

All-cause mortality for males with colon cancer and general population
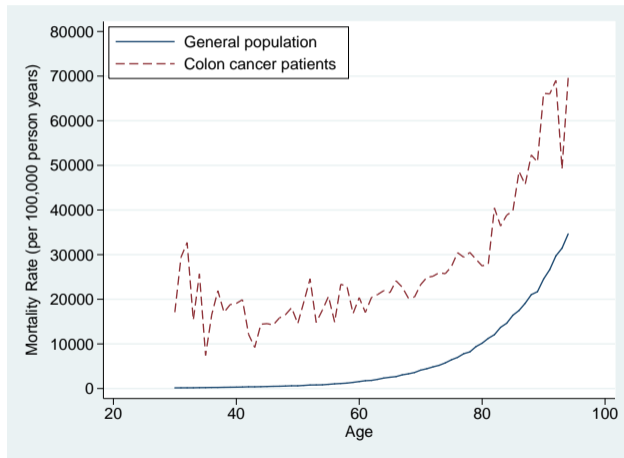
Previous work

- Adjusted life tables (Bower, 2018 [2])
- Corrections to excess models accounting for mismatch in life tables (Rubio, 2021 [3])
- Additive linear regression model (Aalen, 1980 [4])
- Special case of the additive linear regression model, but not allowing for covariates (Andersen, 1989 [5])
- Extension of (step-wise) constant rate model allowing for covariates (Elie, 2014 [6])

To the best of our knowledge, no spline-based approach allowing for smooth estimation of both baseline expected and excess rate, and easy inclusion of time-dependent effects, has been defined and implemented.

Likelihood contribution

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

$$= d_i \cdot \ln\left[h^*(t_i) + v_i \cdot \lambda(t_i)\right] + \ln\left[\exp\{-\int_0^{t_i} \exp\{\ln\left[h^*(t_i) + v_i \cdot \lambda(t_i)\right]\}du\}\right]$$