Regression standardization with time-to-event data to estimate marginal measures of association and causal effects using the `standsurv` command

Paul C Lambert[1,2]

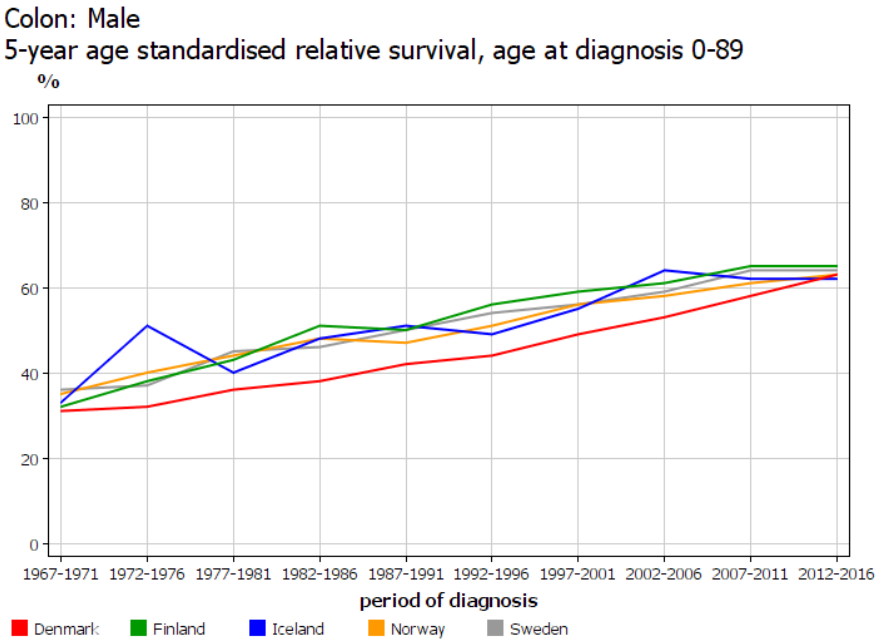
[1]Biostatistics Research Group, Department of Health Sciences, University of Leicester, UK
[2]Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

2021 Stata Biostatistics and Epidemiology Virtual Symposium
18 February 2021

University *of* Leicester

Karolinska Institutet

Standardized Relative/Net Survival



- Relative survival removes differential other cause mortality.
- Age standardization removes differences in age distribution at diagnosis.

Regression Standardization

1. Fit a statistical model incorporating exposure, X, and confounders, Z.
2. Predict outcome for all individuals assuming they are all exposed ($X = 1$).
3. Take mean to give marginal estimate of outcome.
4. Repeat for unexposed ($X = 0$).
5. Take the difference in means to form contrasts.

- Key point is the distribution of confounders, Z, is the same for the exposed and unexposed.
- If model is sufficient for confounding control then such contrasts can be interpreted as causal effects.
- Also known as direct/model based standardization. G-formula (with no time-dependent confounders)[1].

But.... margins does this?

- margins does regression standardization, so why not use this?
- It is an excellent command, but does not do what I wanted for survival data.
- In particular, extensions to competing risks and relative survival.

Marginal survival time

- With survival data

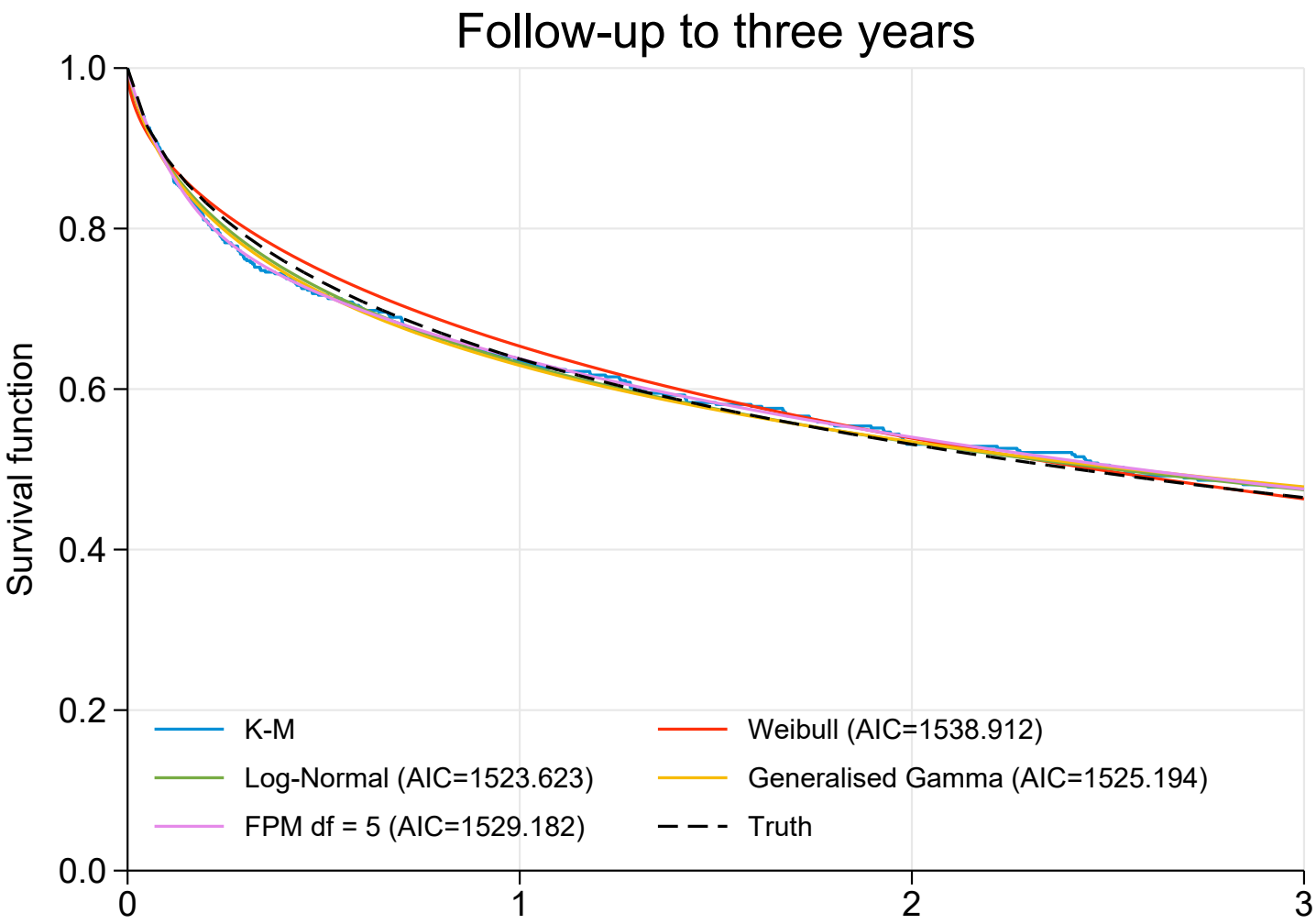
X	-	is a binary exposure: 0 (unexposed) and 1 (exposed).
T	-	is a survival time.
T^0	-	is the potential survival time if X is set to 0.
T^1	-	is the potential survival time if X is set to 1.

- The average causal difference in mean survival time

$$E[T^1] - E[T^0]$$

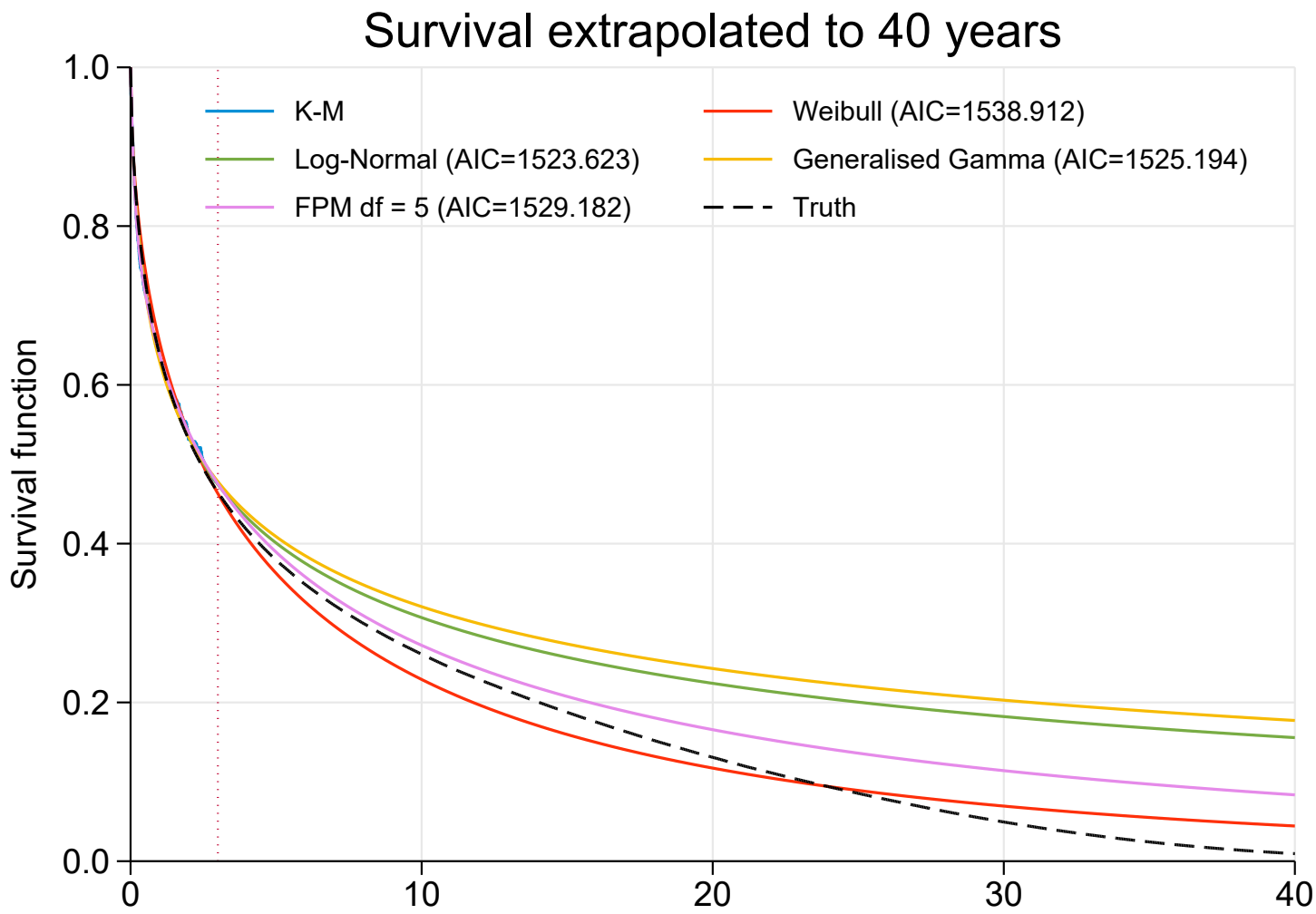
- This is what `stteffects` can estimate.
- We often have limited follow-up and calculating the mean survival requires extrapolation and makes very strong distributional assumptions.

Problems with extrapolation [2]



Need to extrapolate to obtain mean survival

Problems with extrapolation [2]



Need to extrapolate to obtain mean survival

Marginal Survival functions

- Rather than use mean survival we can define our causal effect in terms of the marginal survival function.

$$E[T^1 > t] - E[T^0 > t]$$

- We can limit t within observed follow-up time.
- For confounders, $\boldsymbol{Z}$, we can write this as,

$$E_{\boldsymbol{Z}}[S(t|X = 1, \boldsymbol{Z})] - E_{\boldsymbol{Z}}[S(t|X = 0, \boldsymbol{Z})]$$

- Note that this is the expectation over the distribution of $\boldsymbol{Z}$.

Estimation

- Fit a survival model for exposure X and confounders $\boldsymbol{Z}$.
- Predict survival function for each individual setting $X = x$ and then average.
- Force everyone to be exposed and then unexposed.

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{S}(t|X=1,\boldsymbol{Z}=z_i) - \frac{1}{N}\sum_{i=1}^{N}\widehat{S}(t|X=0,\boldsymbol{Z}=z_i)$$

- Use their observed covariate pattern, $\boldsymbol{Z} = z_i$.
- `standsurv` will perform these calculation.

What can standsurv do?

- standsurv will obtain standardized survival curves and related measures over the study population (or a subset).
- Can treat some covariates as fixed (at() option).
- Implemented for streg, stpm2 and strcs models.
- Linear and non-linear function of marginal estimates.
- Weights (useful for external standardization & mediation analysis)
- I will describe the use of standsurv in three frameworks.
  - Standard survival
  - Competing risks
  - Relative survival

Parametric models

- We make a lot of use of flexible parametric survival models[3].
- The flexibility comes from the use of splines to model the effect of time
- Can model on the log cumulative hazard (`stpm2`)[4] or log hazard scale (`strcs`)[5].

$$\ln[H(t)] = s(\ln(t)|\boldsymbol{k_0}) + \boldsymbol{X}\beta$$

$$\ln[h(t)] = s(\ln(t)|\boldsymbol{k_0}) + \boldsymbol{X}\beta$$

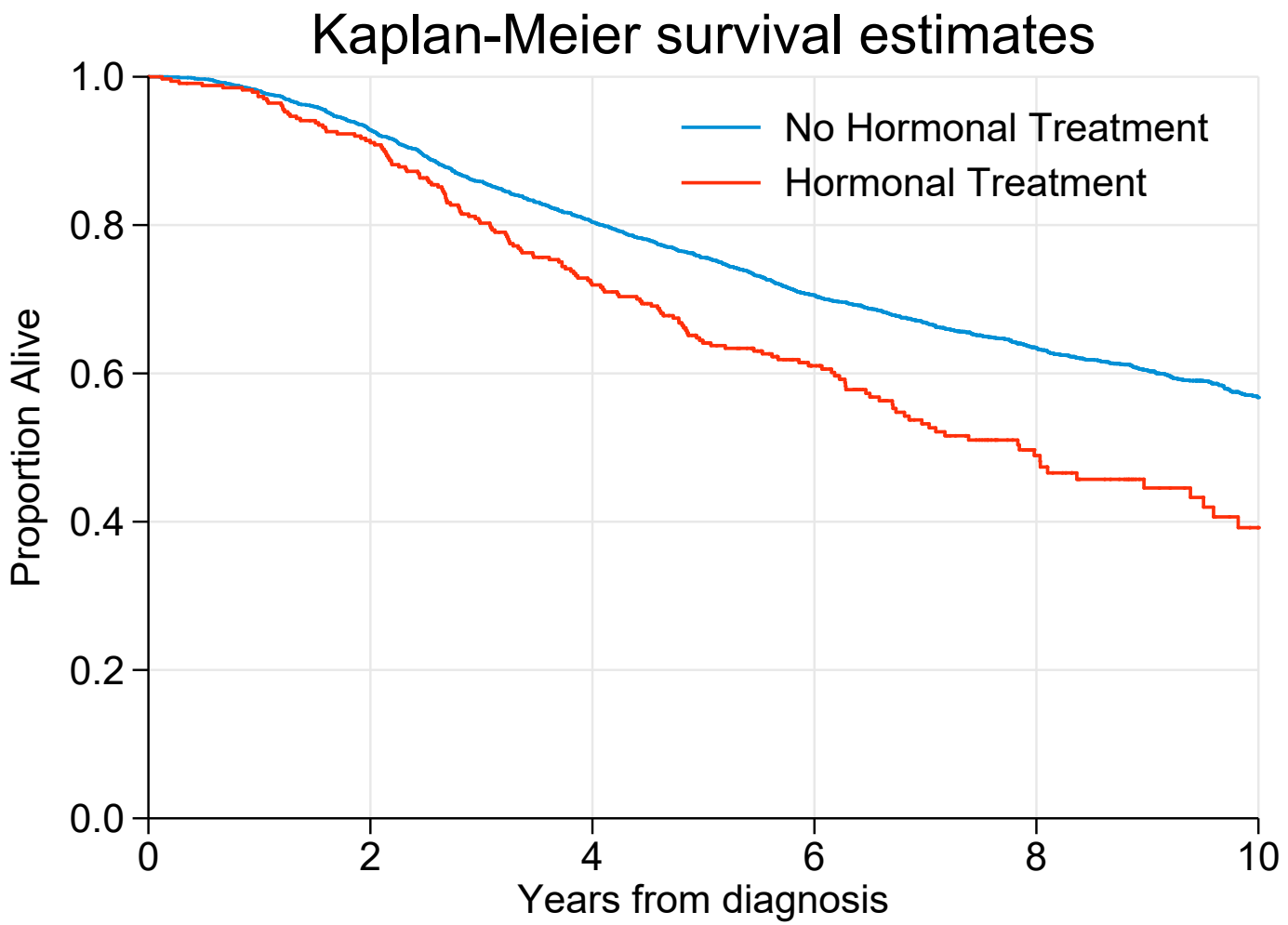
- Shown to capture complex shapes for hazard functions[6, 7].
- Can fit proportional hazards models, but easy to relax this assumption.
- `standsurv` also works with `streg` models.

Example

- I will use the Rotterdam breast cancer data: 2,982 women diagnosed with primary breast cancer.
- Observational study, but interest lies in comparing those taking and not taking hormonal therapy (`hormon`).
- Outcome is all-cause mortality.
- In a simplified analysis I will consider the following confounders.

`age`	Age at diagnosis
`enodes`	Number of positive lymph nodes (transformed)
`pr_1`	Progesterone receptors (fmol/l) (transformed)

Kaplan-Meier curves



Number at risk

No treatment	2643	2436	2083	1668	1188	660
Treatment	339	307	231	141	63	25

Confounders

Confounders

```
. tabstat age nodes pr, by(hormon)
Summary statistics: mean
  by categories of: hormon (Hormonal therapy)
```

hormon	age	nodes	pr
no	54.09762	2.326523	168.706
yes	62.54867	5.719764	108.233
Total	55.05835	2.712274	161.8313

- Those taking treatment tend to be older and have more severe disease.

Hazards Ratios

Unadjusted 1.54 (95% CI 1.30 to 1.82)

Adjusted 0.79 (95% CI 0.66 to 0.94)

- Strong confounding.
- From the adjusted model we can predict the survival for any combination of covariates.

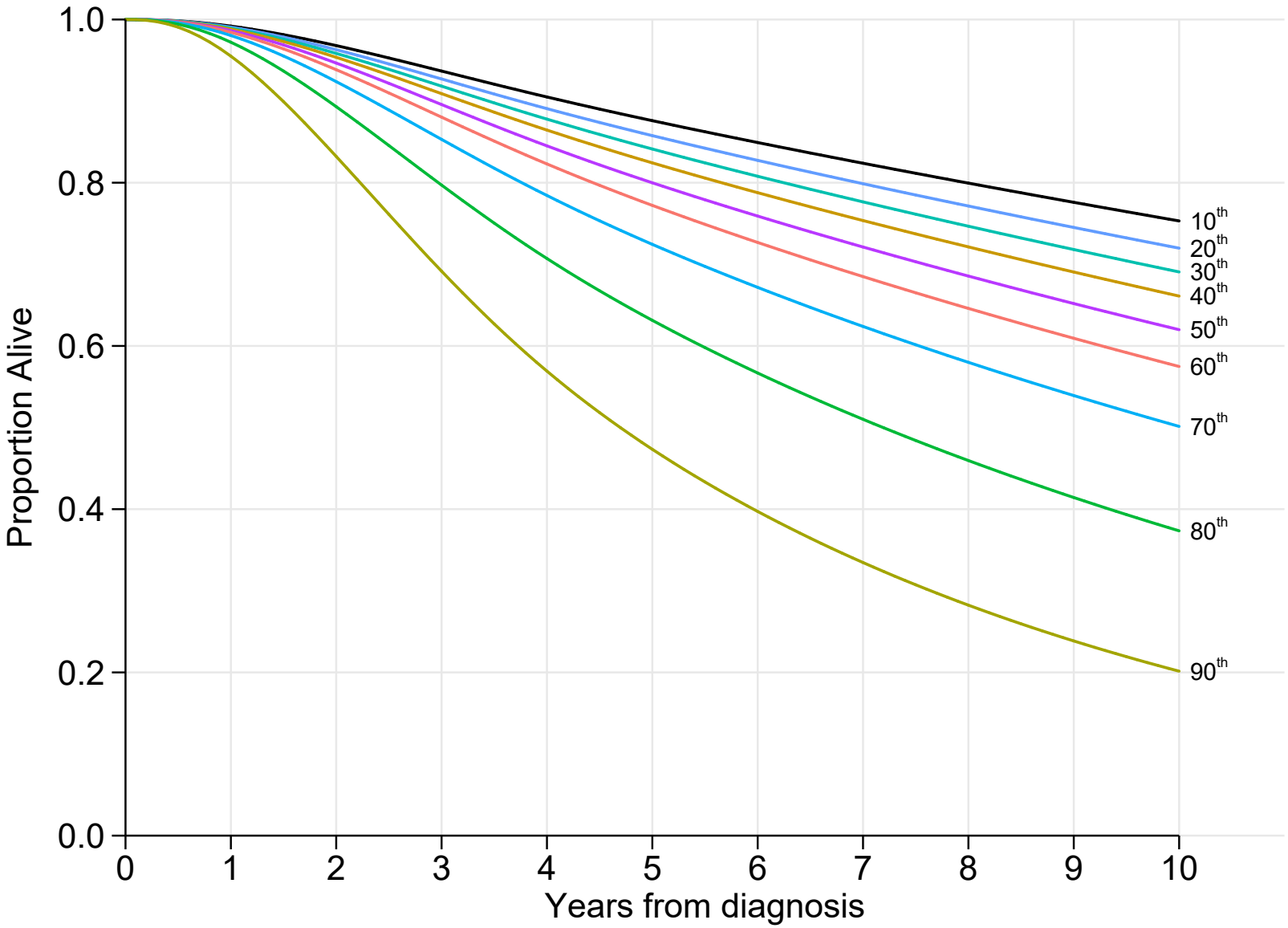
Adjusted Hazard Ratio

```
. stpm2 hormon age enodes pr_1, scale(hazard) df(4) eform nolog cformat(%3.2f)
Log likelihood = -2668.4925                     Number of obs     =      2,982
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
xb						
hormon	0.79	0.07	-2.60	0.009	0.66	0.94
age	1.01	0.00	5.53	0.000	1.01	1.02
enodes	0.11	0.01	-22.40	0.000	0.09	0.14
pr_1	0.91	0.01	-7.46	0.000	0.88	0.93
_rcs1	2.63	0.07	34.67	0.000	2.49	2.78
_rcs2	1.18	0.03	6.08	0.000	1.12	1.25
_rcs3	1.02	0.02	1.36	0.175	0.99	1.05
_rcs4	1.00	0.01	-0.47	0.639	0.98	1.01
_cons	1.10	0.18	0.60	0.546	0.80	1.51

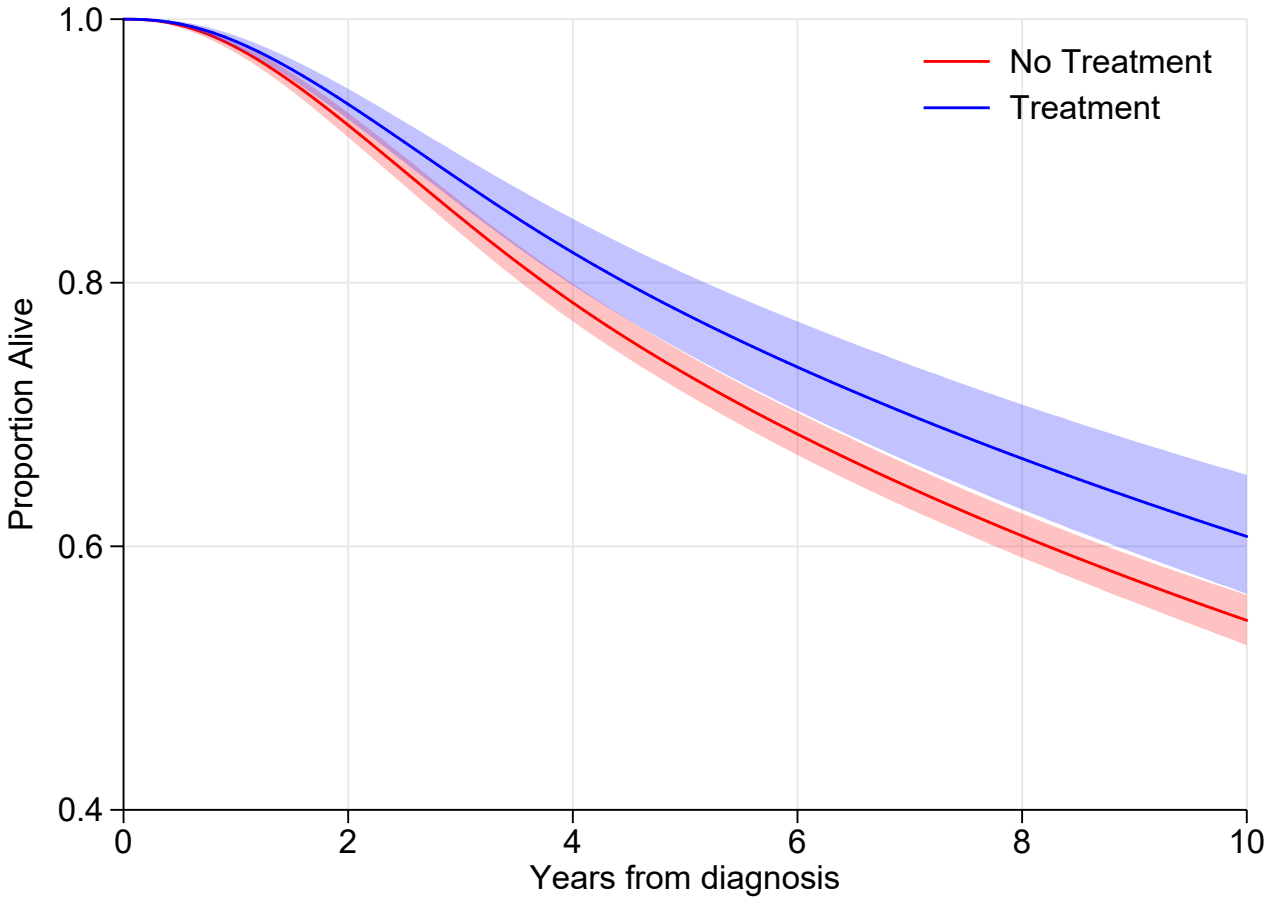
Note: Estimates are transformed only in the first equation.

Predicted survival functions (centiles)



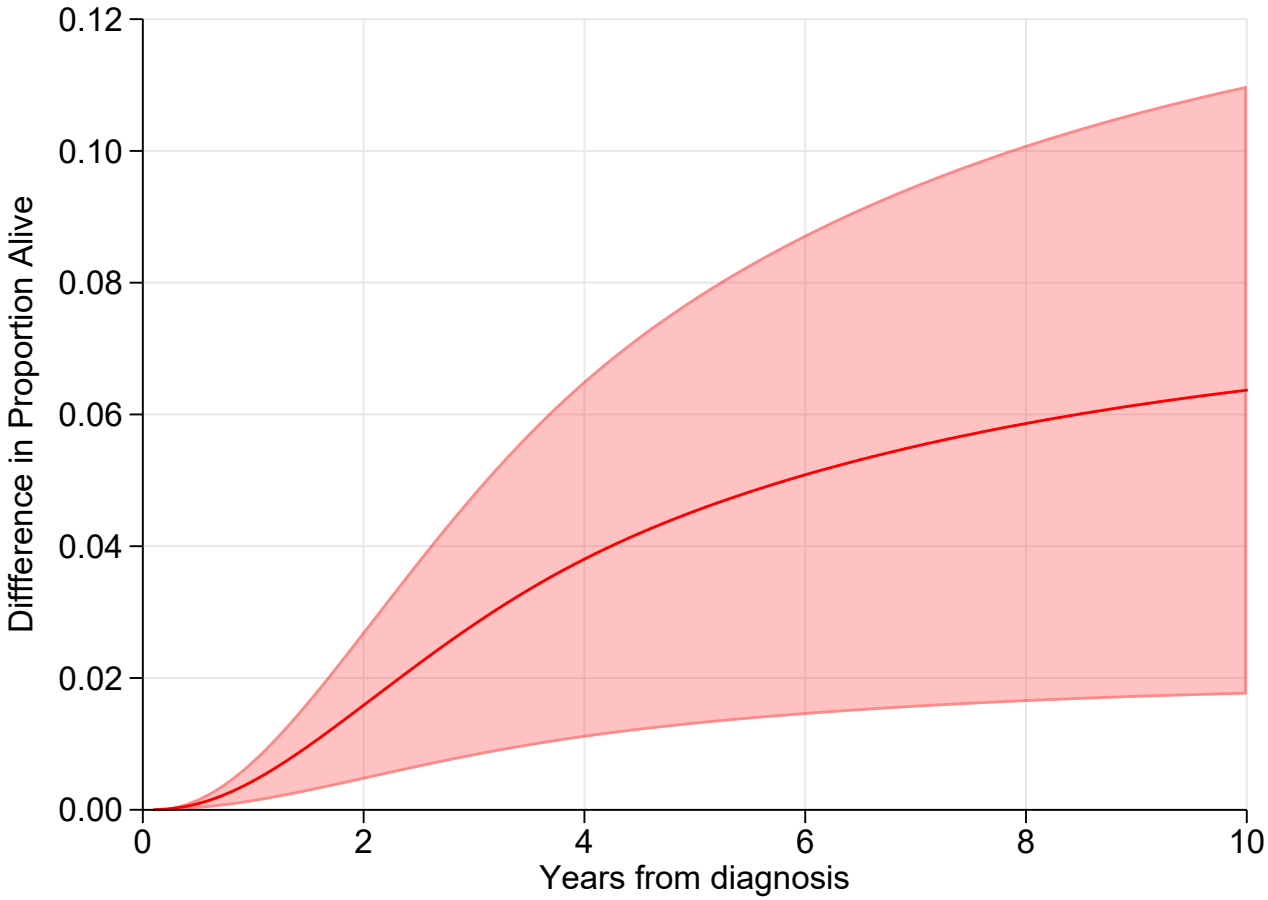
Marginal survival functions using standsurv

```
. range tt 0 10 101
. standsurv, at1(hormon 0) at2(hormon 1) timevar(tt) ci ///
>     contrast(difference) atvar(ms_hor0 ms_hor1) contrastvar(ms_diff)
```

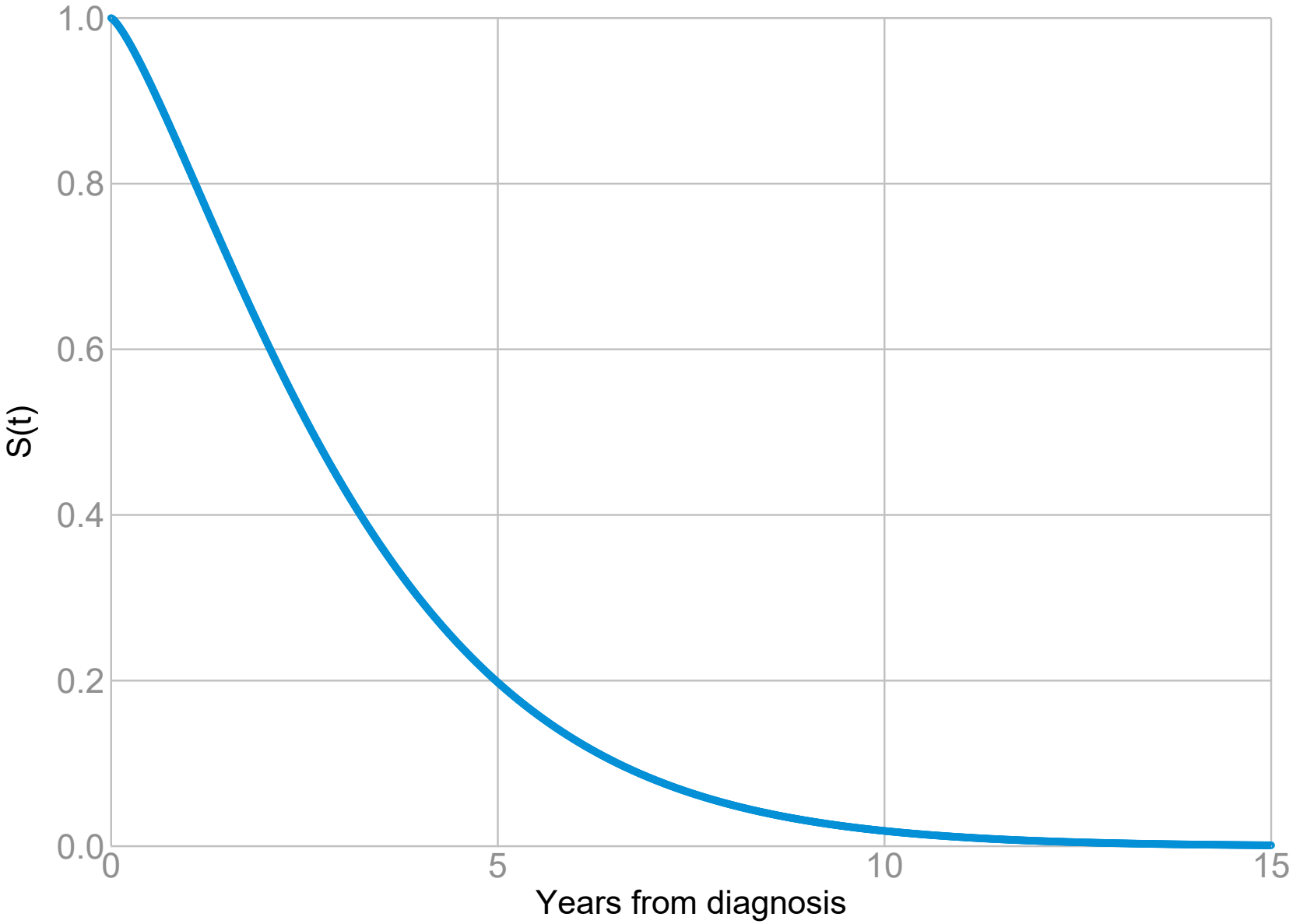


Standardized Survival Difference

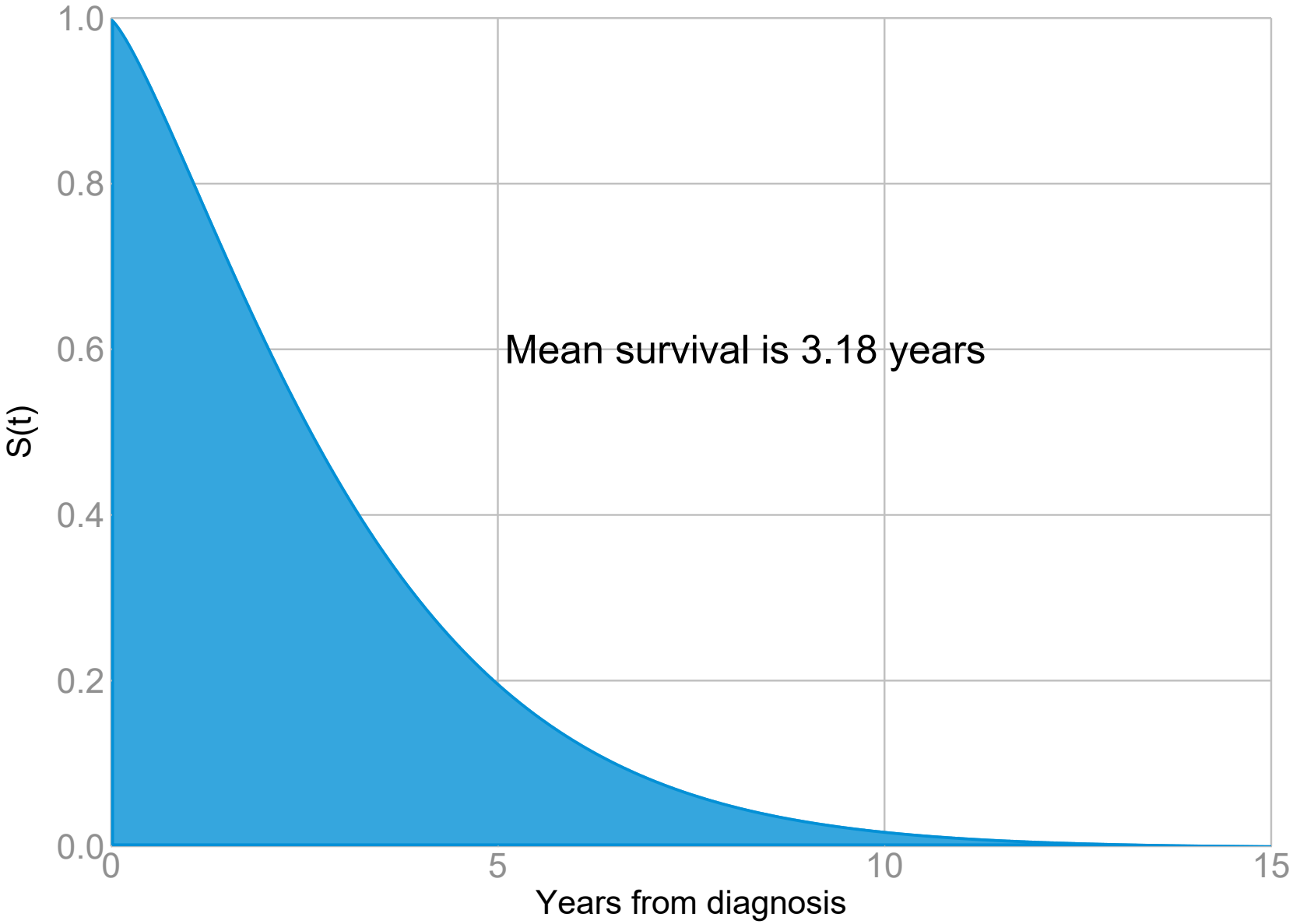
```
. range tt 0 10 101
. standsurv, at1(hormon 0) at2(hormon 1) timevar(tt) ci ///
>     contrast(difference) atvar(ms_hor0 ms_hor1) contrastvar(ms_diff)
```



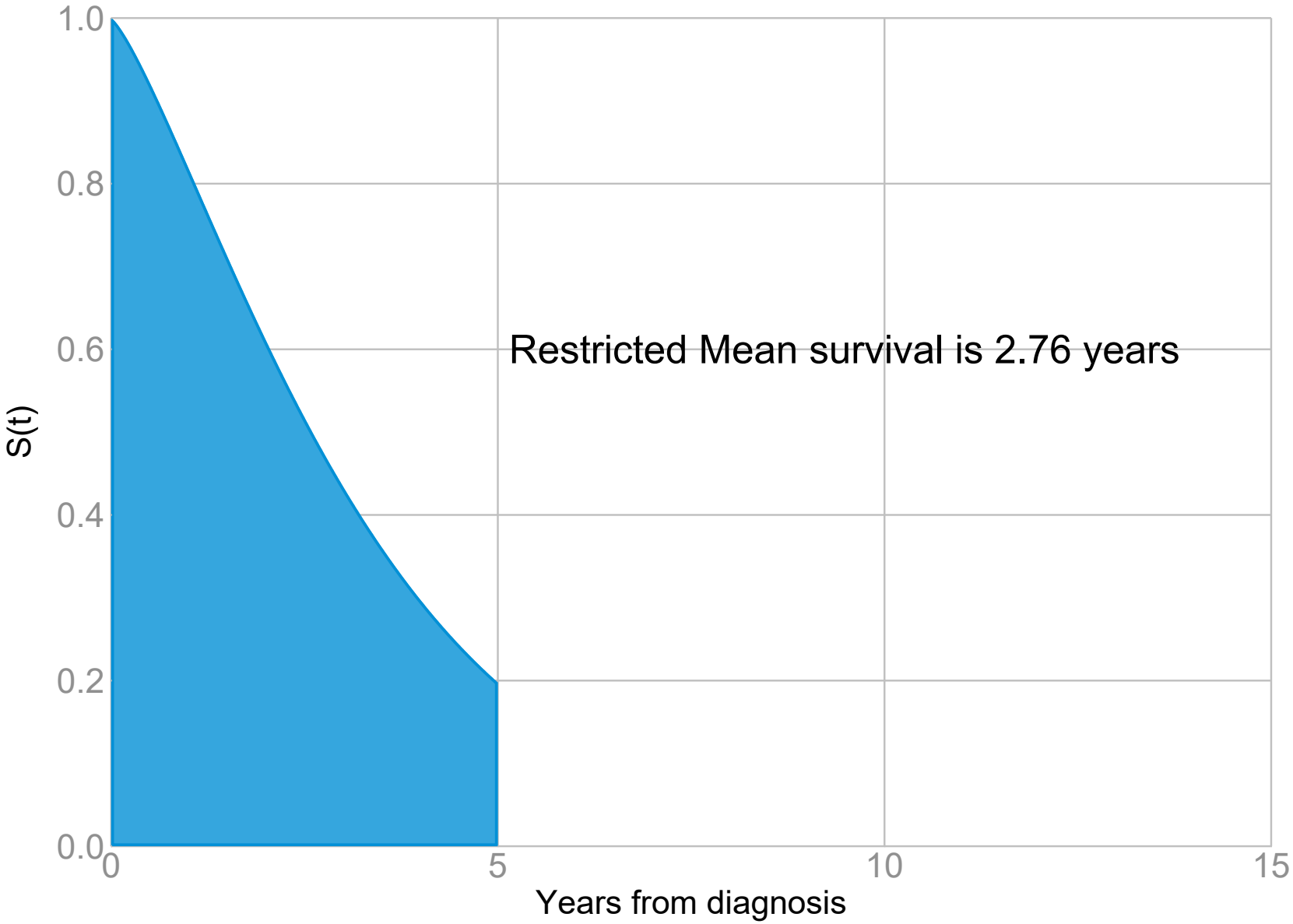
Mean survival time



Mean survival time



Restricted mean survival time (5 years)



Restricted mean survival time (RMST)

restricted mean survival time

$$RMST(t^*) = E\left[min(T, t^*)\right]$$

$$RMST_s(t^*|X = x, Z) = E_Z\left[\int_0^{t^*} S(t|X = x, Z)\right]$$

and is estimated by

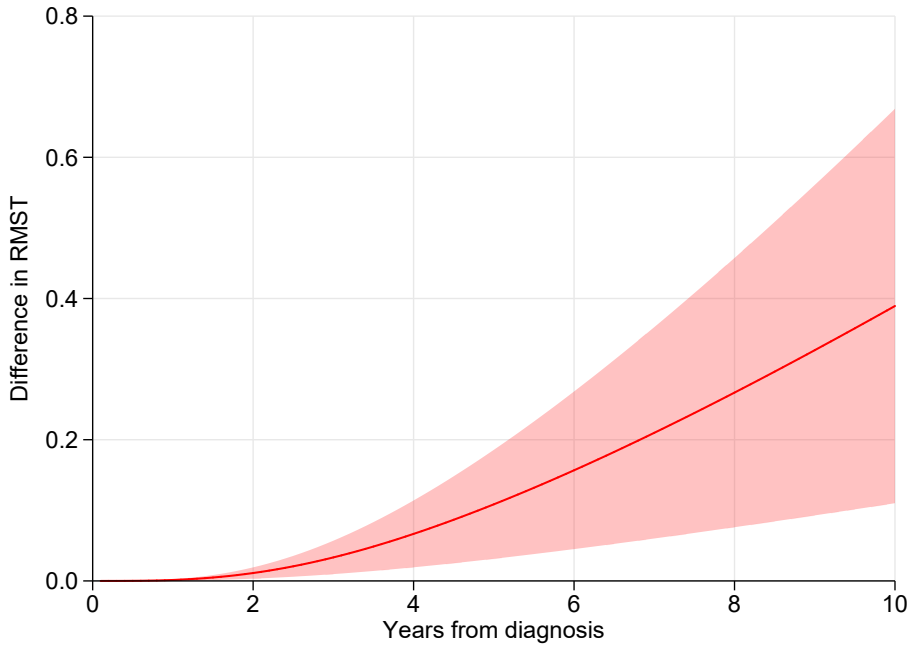
$$\widehat{RMST}_s(t^*|X = x, Z) = \frac{1}{N}\sum_{i=1}^{N}\int_0^{t^*} S(t|X = x, Z = z_i)$$

- we can then take differences or ratios.
- Various authors suggest a better causal effect than HR[11]

Difference in standardized RMST

```
. standsurv, at1(hormon 0) at2(hormon 1) timevar(tt) ci rmst ///
>     contrast(difference) atvar(rmst_hor0 rmst_hor1) contrastvar(rmst_diff)
. list rmst_hor0 rmst_hor1 rmst_diff* if tt==10, noobs
```

rmst_hor0	rmst_hor1	rmst_diff	rms~f_lci	rms~f_uci
7.5505209	7.9399486	.38942772	.11008298	.66877246



Centiles of the marginal survival function

$$E_Z\left[S(t_p|X=x,Z)\right]=\alpha$$

This is done through root finding (using Brent's root finder) by solving for t_p,

$$\frac{1}{N}\sum_{i=1}^{N}S(t_p|X=x,Z=z_i)-\alpha=0$$

- Can perform contrasts, e.g. difference in median of marginal survival functions.
- Use `centile` option.

Hazard of the marginal survival function

- Apply standard transformation from survival to hazard of marginal survival function.

Marginal hazard function

$$h(t) = -\frac{d}{dt}\log\left(E_Z\left[S(t|X=x,Z)\right]\right)$$

and is estimated by,

$$\widehat{h}_s(t) = \frac{1}{N}\frac{\sum_{i=1}^{N} S(t|X=x,Z=z_i)h(t|X=x,Z=z_i)}{\sum_{i=1}^{N} S(t|X=x,Z=z_i)}$$

- Note this is very different from the mean of the hazard functions.
- Can perform contrasts to get marginal hazard ratios (or differences).
- Use the `hazard` option.

User defined functions

- We may need other transformations of standardized functions.
- Use `userfunction()` option for this.
- For example, in survival studies the attributable fraction is defined as,

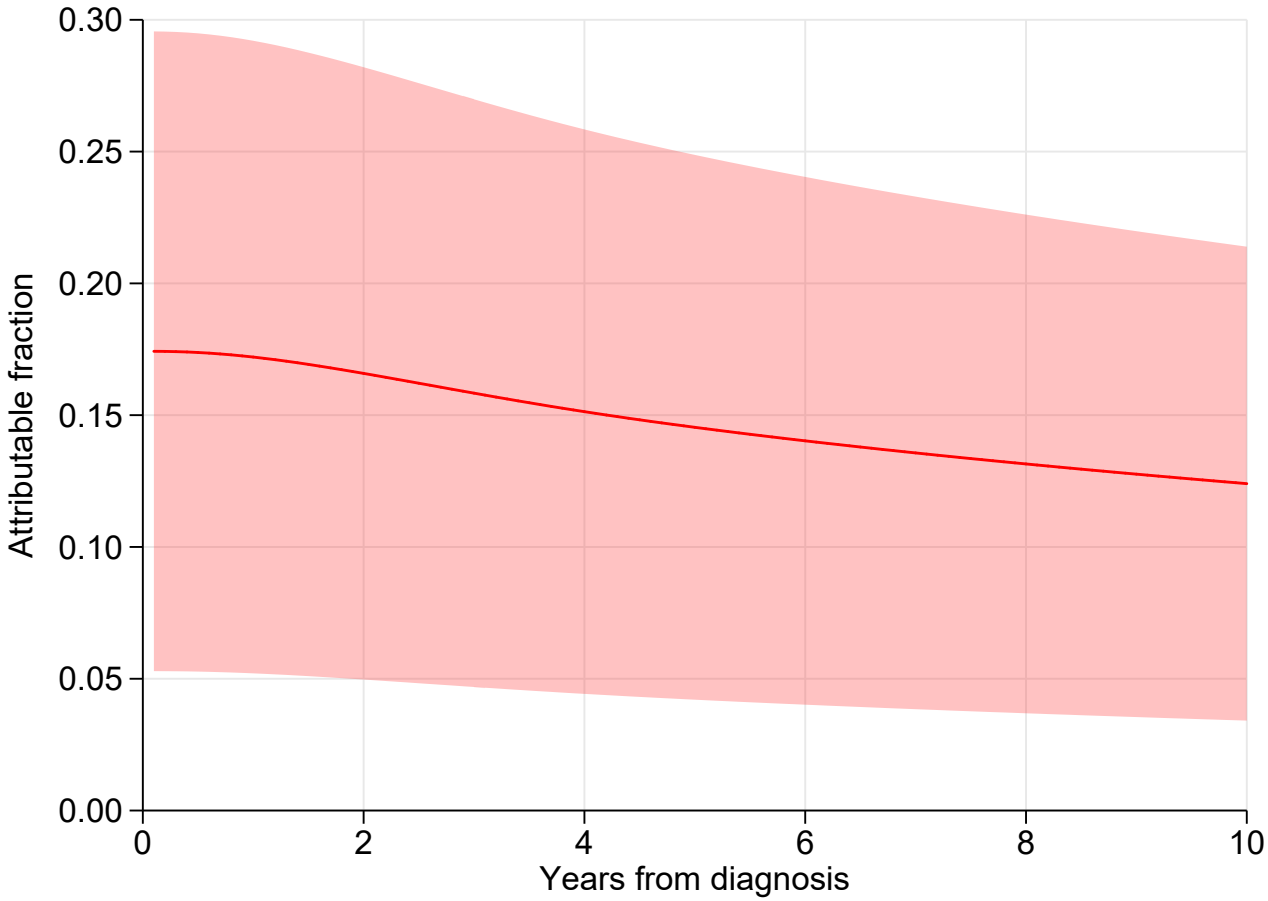
$$AF(t) = \frac{E[F(t|X,Z)] - E[F(t|X=0,Z)]}{E[F(t|X,Z)]}$$

User function

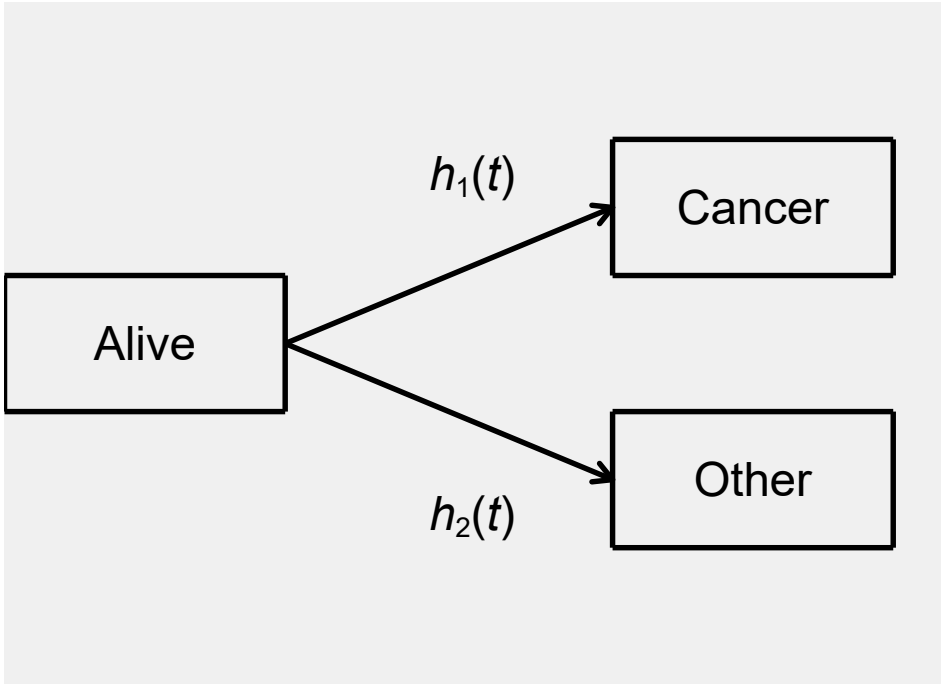
```
mata:
function calcAF(at)
{
  // at2 is F(t|unexposed,Z)
  // at1 is F(t|X,Z)
  return((at[1] - at[2])/at[1])
}
```

Example of user defined function

```
. standsurv, at1(.) at2(hormon 1) ci failure ///
>   timevar(tt) userfunction(calcAF) userfunctionvar(AF)
```



Competing risks



Separate models for each cause, e.g.

$$h_1(t|Z) = h_{0,1}(t)\exp(\beta_1 Z)$$

$$h_2(t|Z) = h_{0,2}(t)\exp(\beta_2 Z)$$

Two types of probability

- We may be interested in cause-specific survival/failure.

(1) *In the absence of other causes (net)*

$$F_k(t) = 1 - S_k(t) = P(T_k \leq t) = \int_0^t S_k(u)h_k(u)du$$

- We may be interested in cumulative incidence functions.

(2) *In the presence of other causes (crude)*

$$CIF_k(t) = P\left(T \leq t, \text{event} = k\right) = \int_0^t S(u)h_k(u)du$$

- Both are of interest - depends on research question.
- (1) Needs conditional independence assumption to interpret as net probability of death.

Description of Example

- 39,625 patients with bladder cancer in England (2000-2012).
- Death due to cancer and other causes.
- Covariates age, sex and deprivation in five groups.
- Restrict here to most and least deprived.

Models

- Flexible parametric (Royston-Parmar) models[3]
- Separate model for cancer and other causes.
- Age modelled using splines (3 df)
- 2-way interactions
- Time-dependent effects for all covariates.

Two separate cause-specific models

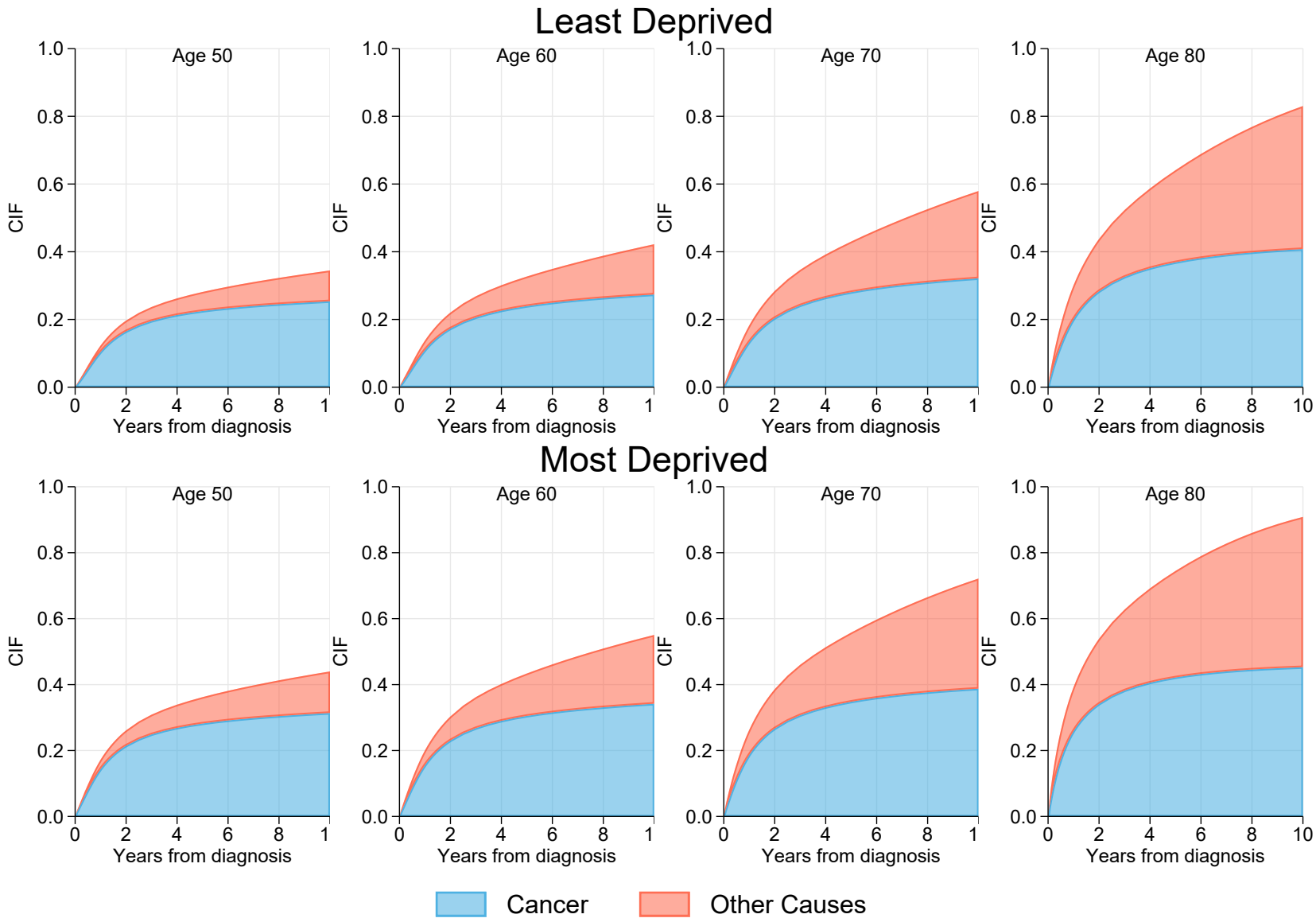
Cancer Model

```
stset dod, failure(status==1) exit(time min(dx+365.24*10,mdy(12,31,2013))) ///
          origin(dx) id(patid) scale(365.24)

stpm2 dep5 male agercs* dep_agercs*, df(5) scale(hazard)                ///
                                     tvc(agercs* male dep5) dftvc(3)
estimates store cancer
```

Other Cause Model

```
stset dod, failure(status==2) exit(time min(dx+365.24*10,mdy(12,31,2013))) ///
          origin(dx) id(patid) scale(365.24)

stpm2 dep5 male agercs* dep_agercs*, df(5) scale(hazard)                ///
                                     tvc(agercs* male dep5) dftvc(3)
estimates store cancer
```

Conditional cause-specific CIFs (Females)



Standardized cause-specific CIF

- Probability of death in the presence of other causes.
- We can standardize the cause-specific CIF in the same way.
- These requires combining K different models

$$E_Z\left[CIF_k(t|X=x,Z)\right]$$

$$\frac{1}{N}\sum_{i=1}^{N}\int_0^t \widehat{S}(u|X=x,\boldsymbol{Z}=z_i)\widehat{h}_k(u|X=x,\boldsymbol{Z}=z_i)du$$

- Calculate for X=1 and X=0 and then obtain contrast.
- Can be interpreted as causal effects under assumptions[12].

Using standsurv

- Take mean of 39,625 CIFs where all individuals forced to be unexposed.
- Take mean of 39,625 CIFs where all individuals forced to be exposed.

Using standsurv

- Take mean of 39,625 CIFs where all individuals forced to be unexposed.
- Take mean of 39,625 CIFs where all individuals forced to be exposed.

```
. standsurv, crmodels(cancer other) timevar(tt) cif ci                          ///
  at1(dep5 0 dep_agercs1 0 dep_agercs2 0 dep_agercs3 0)                          ///
  at2(dep5 1 dep_agercs1=agercs1 dep_agercs2=agercs2 dep_agercs3=agercs3)        ///
  contrast(difference)                                                          ///
  atvar(CIF_s_dep1 CIF_s_dep5))                                                 ///
  contrastvar(CIF_diff)
```

Using standsurv

- Take mean of 39,625 CIFs where all individuals forced to be unexposed.
- Take mean of 39,625 CIFs where all individuals forced to be exposed.

```
. standsurv, crmodels(cancer other) timevar(tt) cif ci                          ///
  at1(dep5 0 dep_agercs1 0 dep_agercs2 0 dep_agercs3 0)                          ///
  at2(dep5 1 dep_agercs1=agercs1 dep_agercs2=agercs2 dep_agercs3=agercs3)        ///
  contrast(difference)                                                          ///
  atvar(CIF_s_dep1 CIF_s_dep5))                                                 ///
  contrastvar(CIF_diff)
```

Using standsurv

- Take mean of 39,625 CIFs where all individuals forced to be unexposed.
- Take mean of 39,625 CIFs where all individuals forced to be exposed.

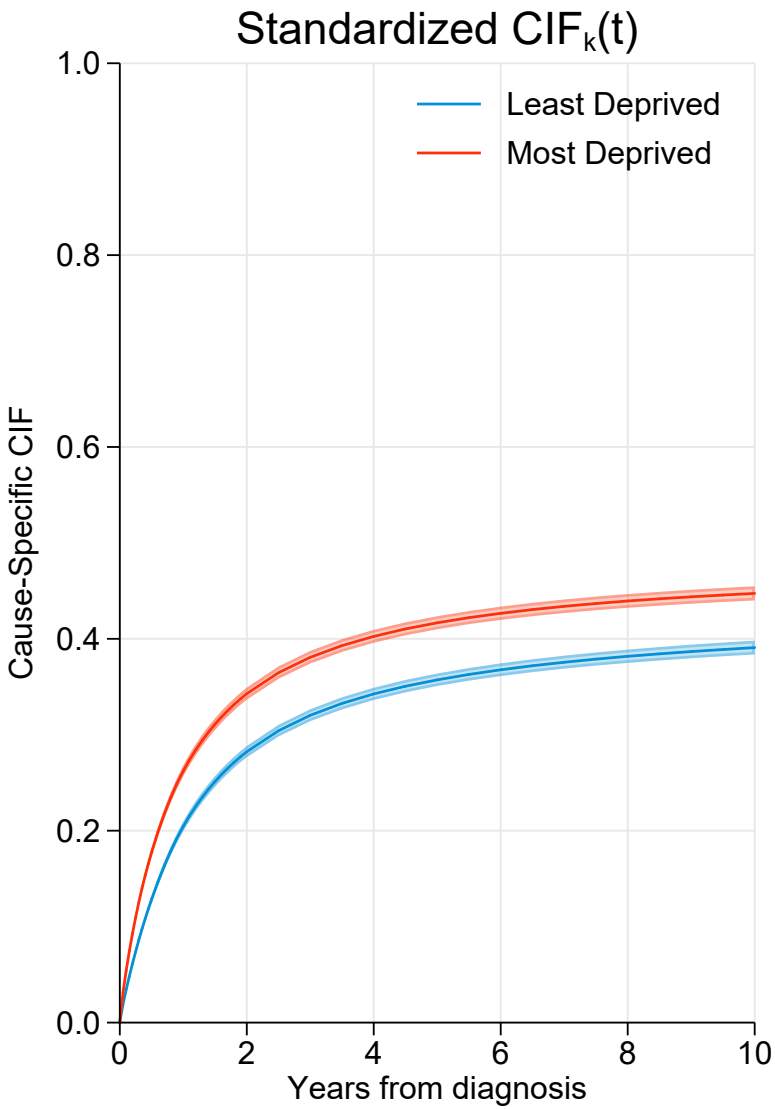
```
. standsurv, crmodels(cancer other) timevar(tt) cif ci                          ///
  at1(dep5 0 dep_agercs1 0 dep_agercs2 0 dep_agercs3 0)                          ///
  at2(dep5 1 dep_agercs1=agercs1 dep_agercs2=agercs2 dep_agercs3=agercs3)        ///
  contrast(difference)                                                          ///
  atvar(CIF_s_dep1 CIF_s_dep5))                                                 ///
  contrastvar(CIF_diff)
```

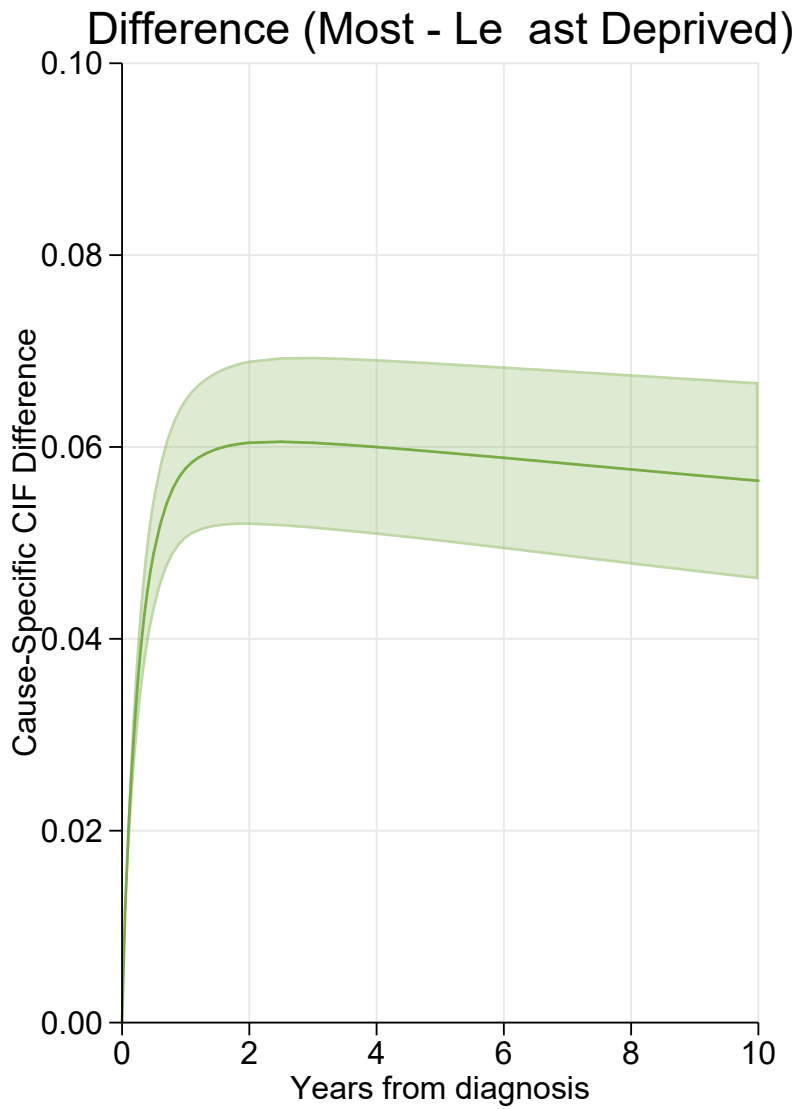
Using standsurv

- Take mean of 39,625 CIFs where all individuals forced to be unexposed.
- Take mean of 39,625 CIFs where all individuals forced to be exposed.

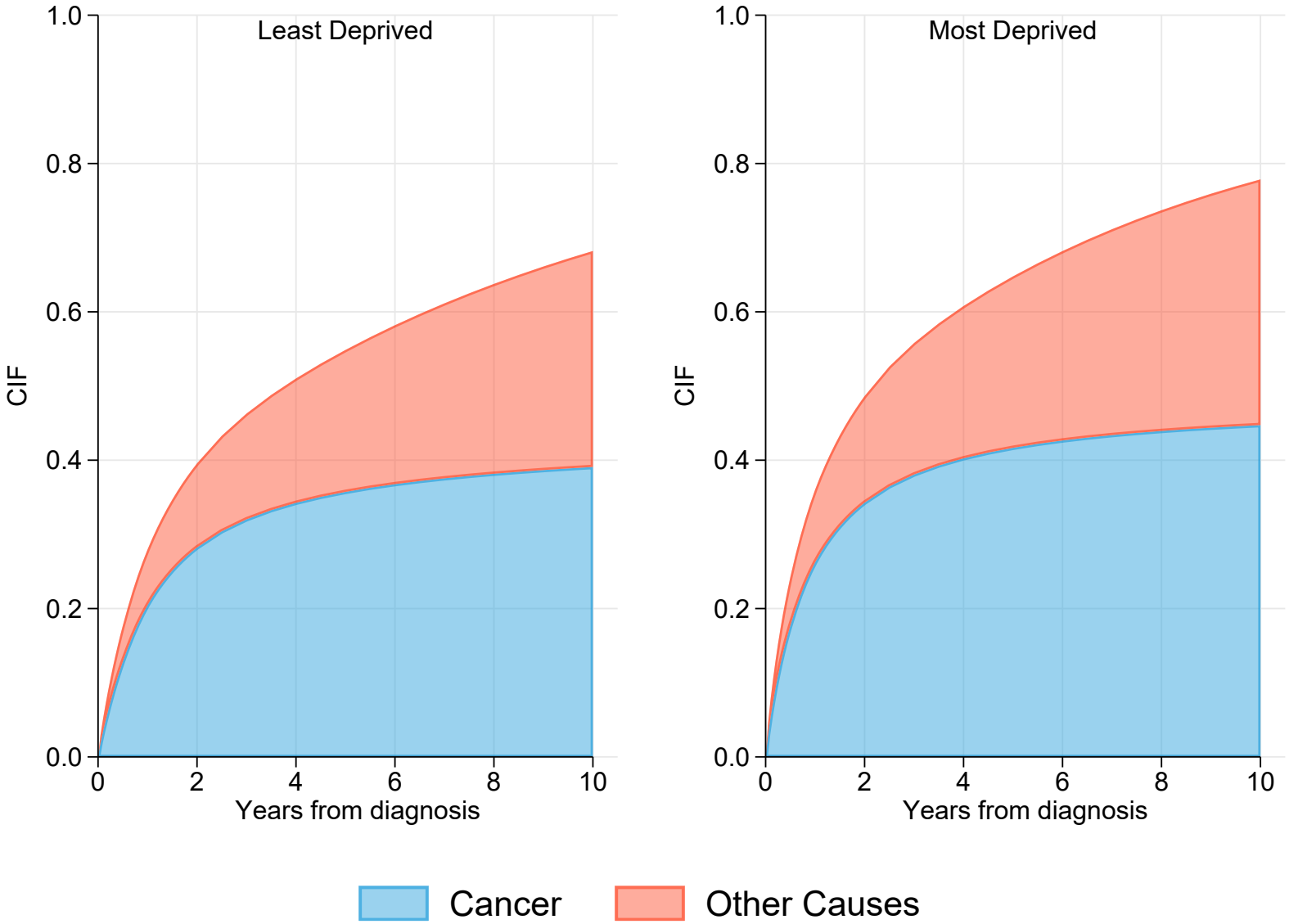
```
. standsurv, crmodels(cancer other) timevar(tt) cif ci                          ///
  at1(dep5 0 dep_agercs1 0 dep_agercs2 0 dep_agercs3 0)                          ///
  at2(dep5 1 dep_agercs1=agercs1 dep_agercs2=agercs2 dep_agercs3=agercs3)        ///
  contrast(difference)                                                          ///
  atvar(CIF_s_dep1 CIF_s_dep5))                                                 ///
  contrastvar(CIF_diff)
```

Standardized cause-specific CIF





Stacked standardized cause-specific CIF



Competing Risks - extensions

- Can also obtain area under standardized CIF which gaves a standardized version of the expected years of life lost (Andersen 2013[13]). Use `cif` and `rmft options`. See Mozumder *et al* 2021 [14].
- Various causal in measures in competing risks described in Young *et al* 2020[15] can be estimated using `standsurv`.
- Separable effects can also be estimated (Stensrud *et al* 2020)[16].
- Can also use user-defined functions, e.g. Standardized attributable fraction in competing risks setting.
- Different parametric models can be used for different causes.
- Different time scales can be used for different causes (e.g. attained age / time from diagnosis).

Relative Survival

- Relative survival models used with large population cancer registry data when cause of death not available or not reliable.

$$h(t|X,Z) = h^*(t|X,Z) + \lambda(t|X,Z)$$

$h(t\|X,Z)$	-	All-cause mortality rate
$h^*(t\|X,Z)$	-	Expected mortality rate
$\lambda(t\|X,Z)$	-	Excess mortality rate

- Expected mortality rates obtained from national lifetables.
- On survival scale.

$$S(t|X,Z) = S^*(t|X,Z)R(t|X,Z)$$

- The equivalent of a CIF is know as a crude probability in the relative survival framework.

Melanoma Example

Relative Survival Model

```
stpm2 dep5 agercs* , scale(hazard) df(5) tvc(dep5 agercs*) dftvc(3) bhazard(rate)
```

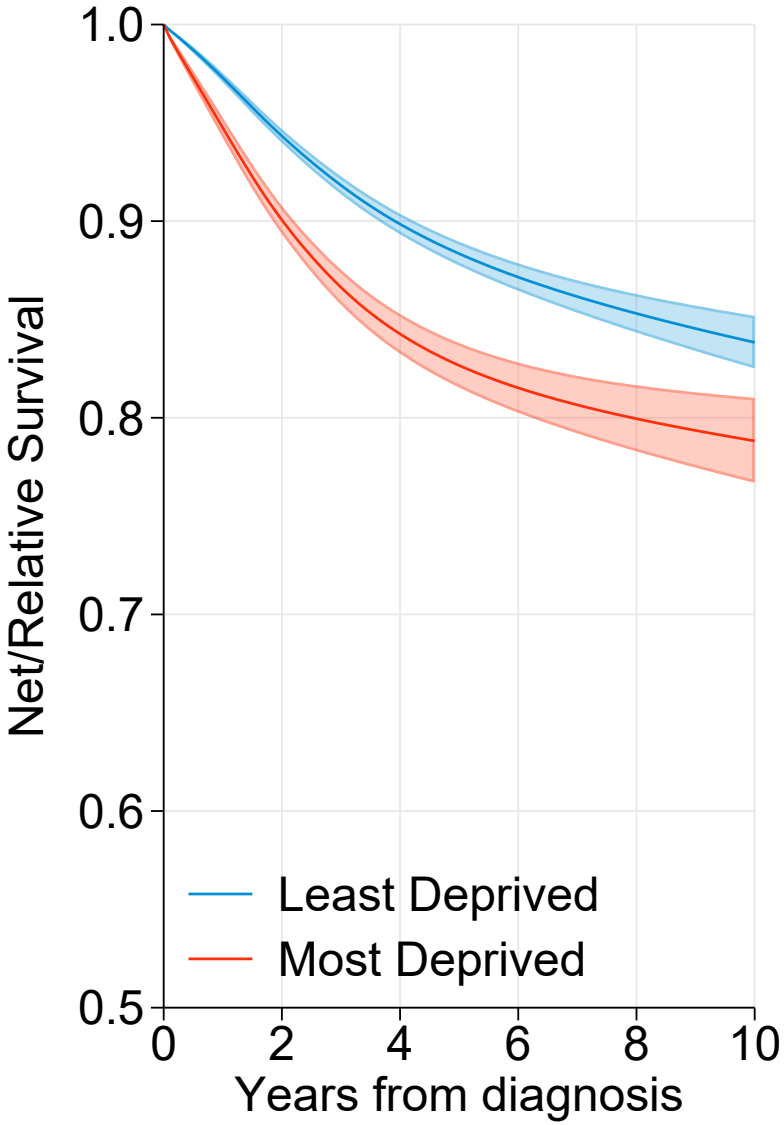
$$\overline{R}(t|X = x, \boldsymbol{Z}) = \frac{1}{N}\sum_{i=1}^{N} w_i R_i(t|X = x, \boldsymbol{Z} = z_i)$$

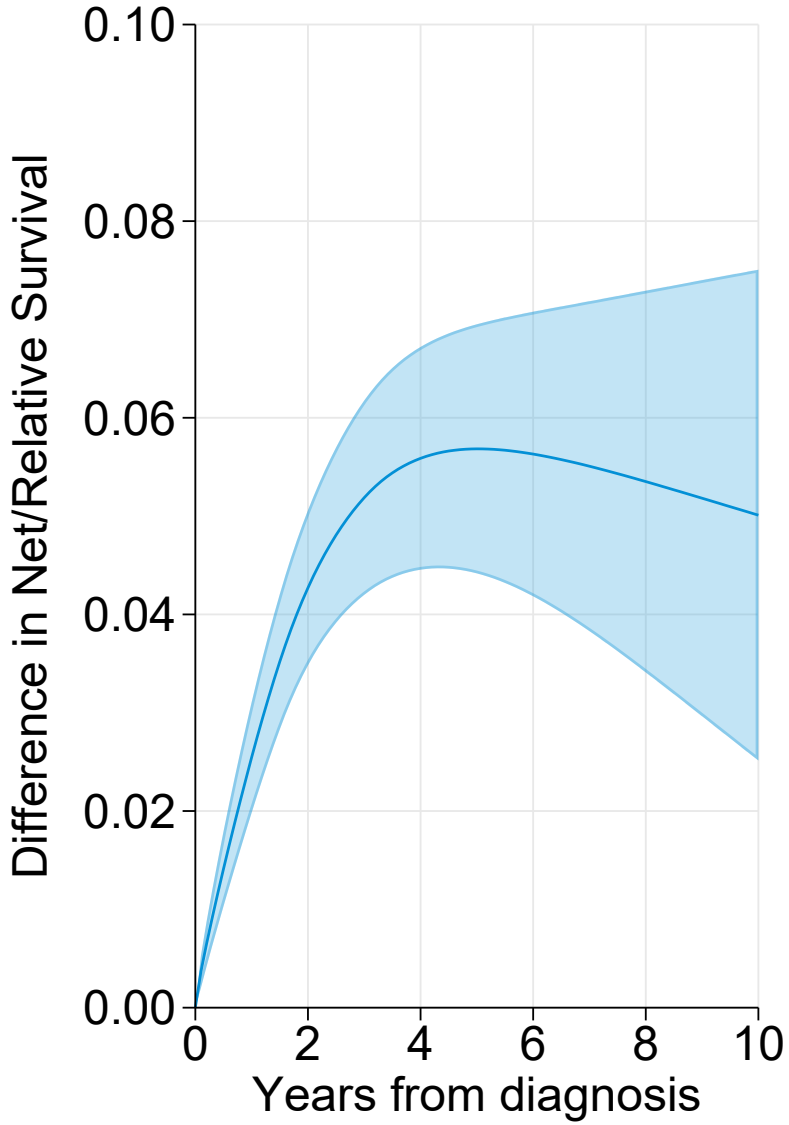
- The weights, w_i, enables standardization to external population through up- or down-weighting relative to a reference population.

Standardized Relative Survival

```
standsurv, timevar(tt) ci                                                    ///
  at1(dep5 0 agercs1_dep5 0 agercs2_dep5 0 agercs3_dep5 0)                   ///
  at2(dep5 1 agercs1_dep5=agercs1 agercs2_dep5=agercs2 agercs3_dep5=agercs3) ///
  indweights(wt)                                                             ///
  contrast(difference)                                                       ///
  atvar(R_dep5 R_dep1)                                                       ///
  contrastvar(R_diff)
```

Standardized Relative Survival



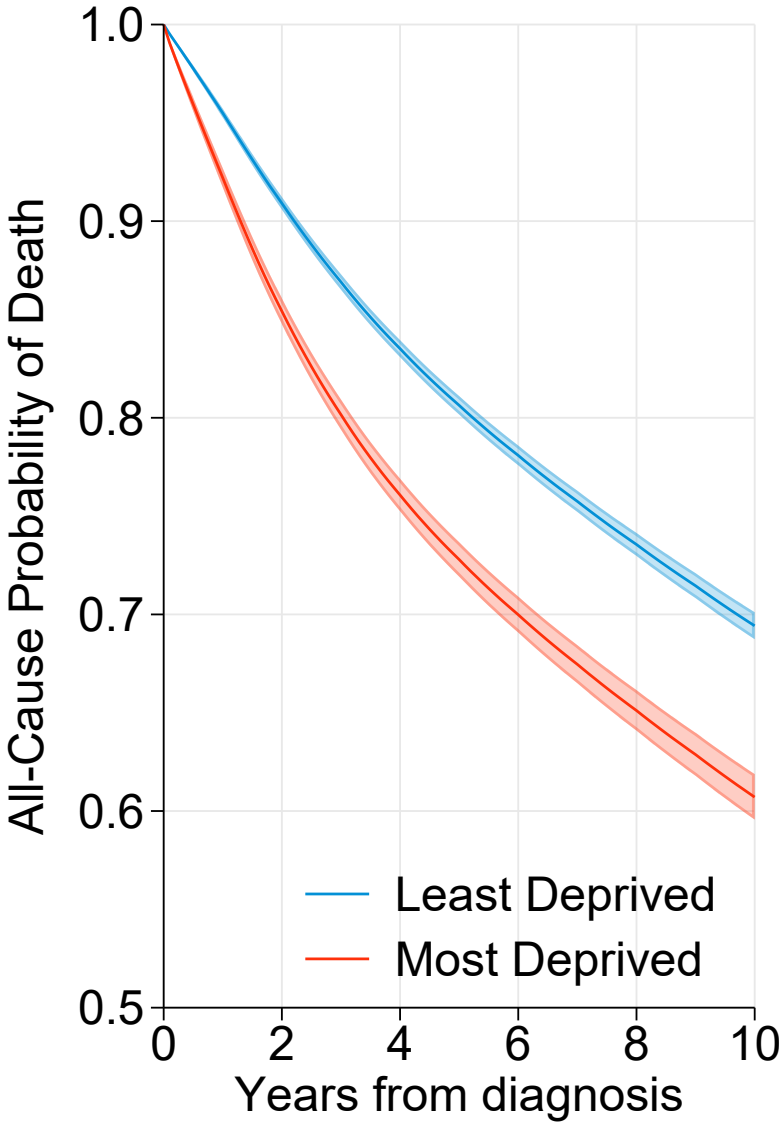


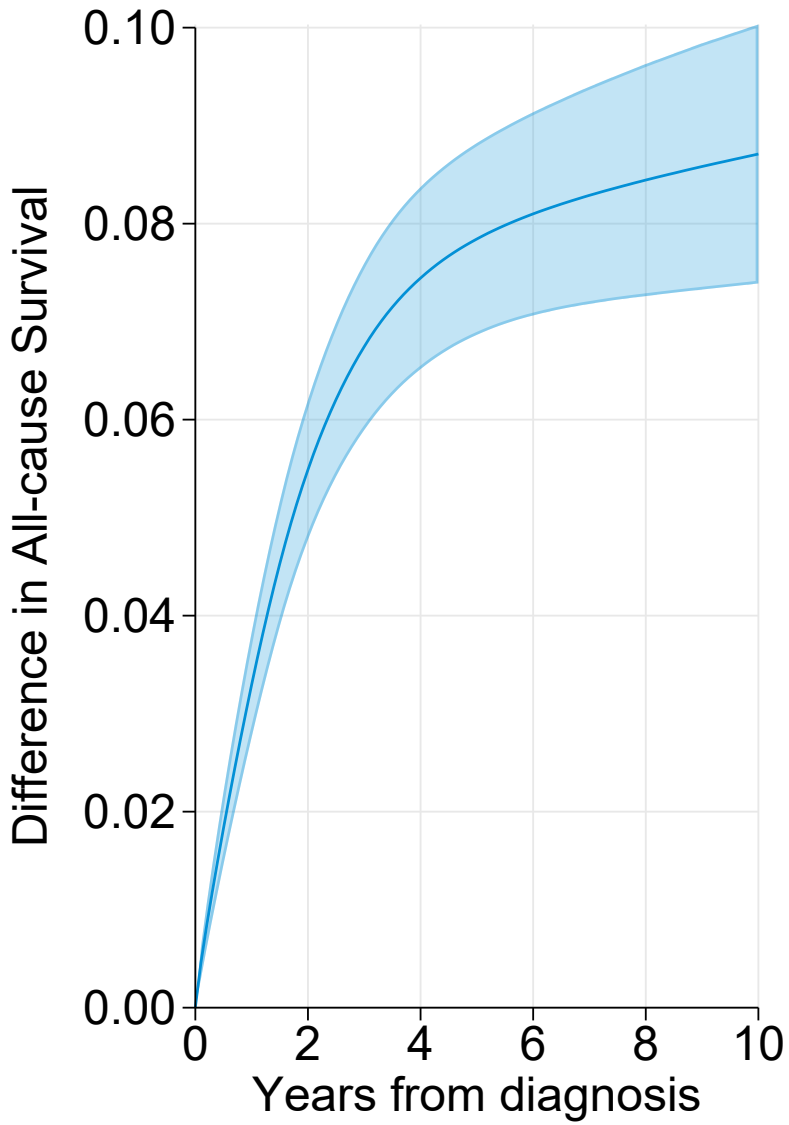
All-cause Survival

$$\overline{S}(t|X=x,\boldsymbol{Z}) = \frac{1}{N}\sum_{i=1}^{N} S^*(t|X=x,\boldsymbol{Z}=z_i)R_i(t|X=x,\boldsymbol{Z}=z_i)$$

```
standsurv,  timevar(tt) ci                                                  ///
  at1(dep5 0 agercs1_dep5 0 agercs2_dep5 0 agercs3_dep5 0)                  ///
  at2(dep5 1 agercs1_dep5=agercs1 agercs2_dep5=agercs2 agercs3_dep5=agercs3) ///
  expsurv(using(popmort_uk_regions_2017.dta)                                ///
          datediag(dx)                                                      ///
          agediag(agediag)                                                  ///
          pmrate(rate)                                                      ///
          pmage(age)                                                        ///
          pmyear(year)                                                      ///
          pmother(sex dep region)                                           ///
          pmmaxyear(2016)                                                   ///
          at1(dep 1)                                                        ///
          at2(dep 5))                                                       ///
  contrast(difference)                                                      ///
  atvar(S_dep5 S_dep1)                                                      ///
  contrastvar(S_diff)
```

Standardized All-cause Survival



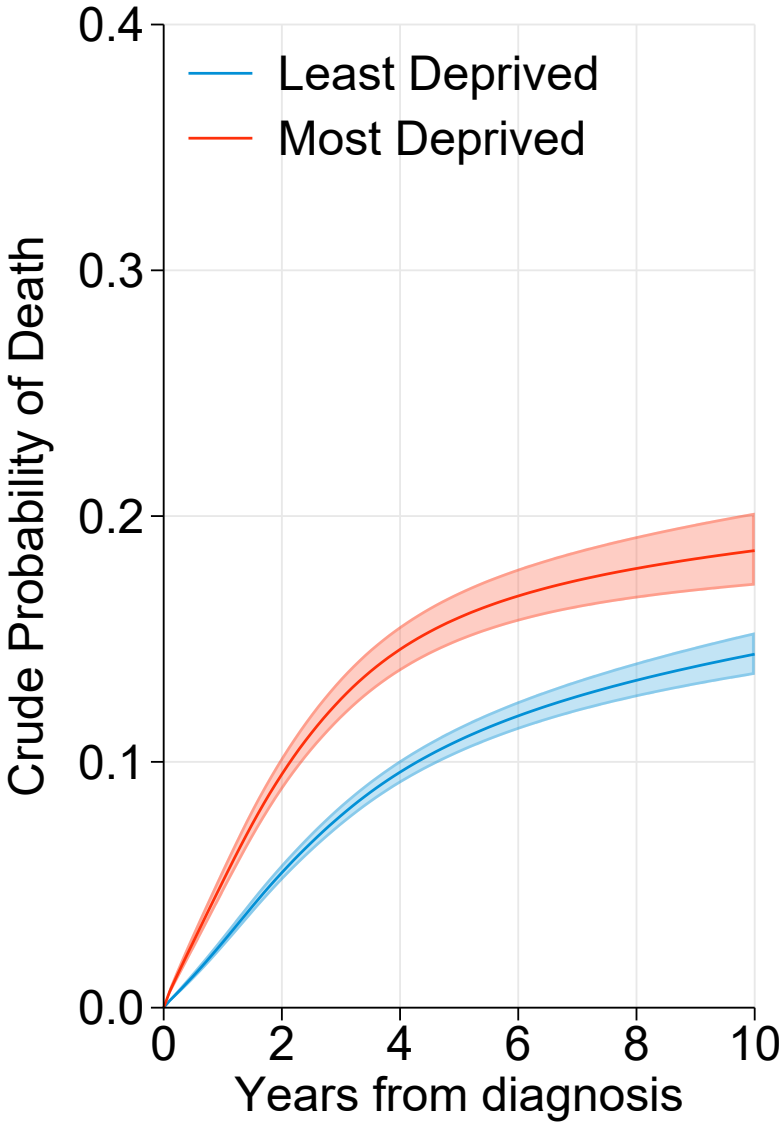


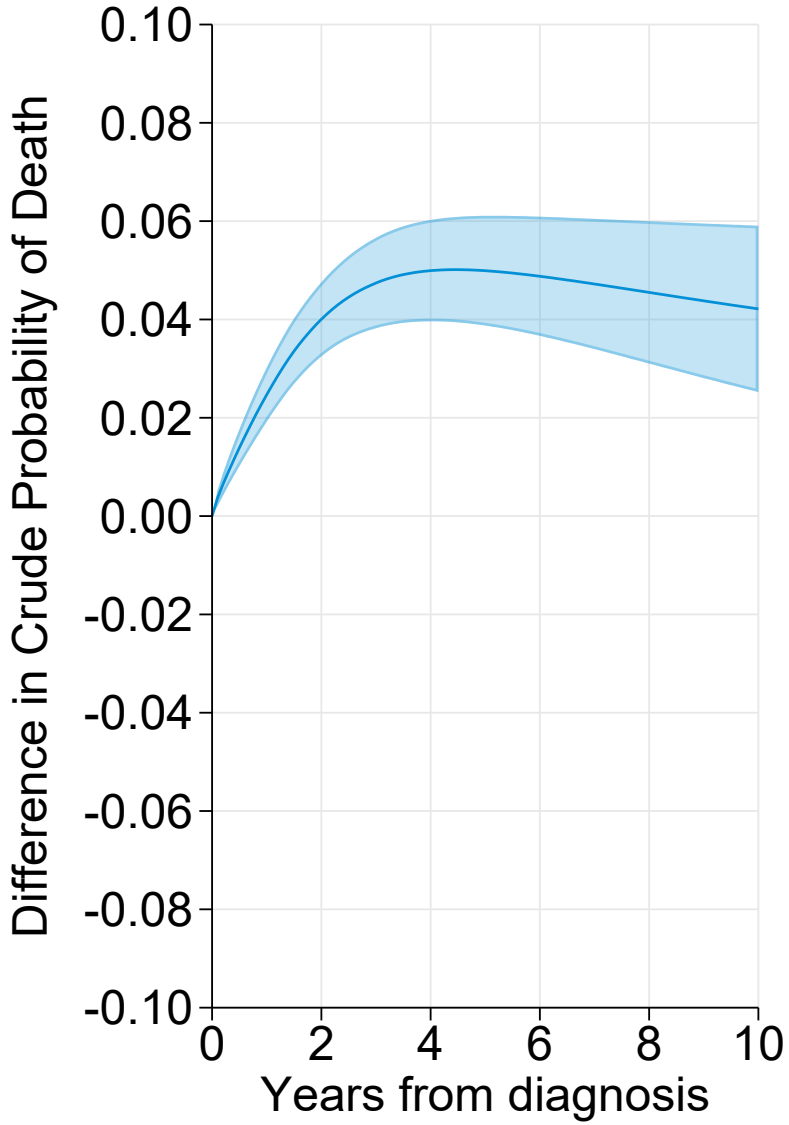
Standardized Crude Probabilities

$$\overline{F}_c(t|X=x,\boldsymbol{Z}) = \frac{1}{N}\sum_{i=1}^{N}\int_0^t S^*(u|X=x,\boldsymbol{Z}=\boldsymbol{z}_i)R(u|X=x,\boldsymbol{Z}=\boldsymbol{z}_i)\lambda(u|X=x,\boldsymbol{Z}=\boldsymbol{z}_i),$$

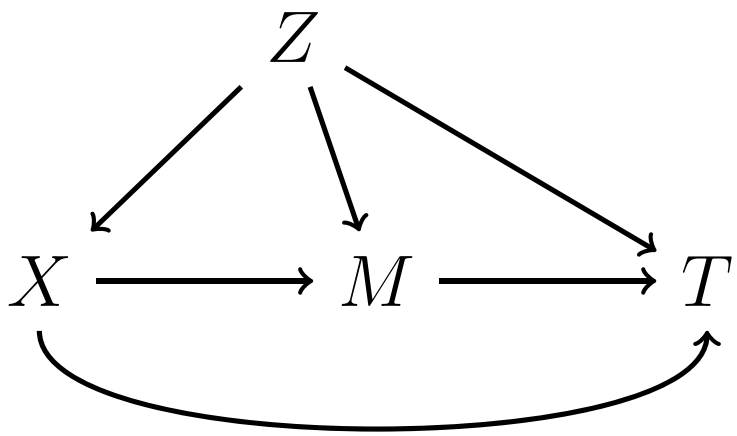
```
standsurv,  crudeprob timevar(tt) ci                                          ///
       at1(dep5 0 agercs1_dep5 0 agercs2_dep5 0 agercs3_dep5 0)               ///
       at2(dep5 1 agercs1_dep5=agercs1 agercs2_dep5=agercs2 agercs3_dep5=agercs3)  ///
        expsurv(using(popmort_uk_regions_2017.dta)                            ///
                datediag(dx)                                                  ///
                agediag(agediag)                                              ///
                pmrate(rate)                                                  ///
                pmage(age)                                                    ///
                pmyear(year)                                                  ///
                pmother(sex dep re)                                           ///
                pmmaxyear(2016)                                               ///
                at1(dep 1)                                                    ///
                at2(dep 5))                                                   ///
         contrast(difference)                                                 ///
         atvar(CP_dep5 CP_dep1)                                               ///
         contrastvar(CP_diff)
```

Standardized Crude Probabilities of Death





Mediation Analysis in Relative Survival Framework



$$\widehat{\text{NDE}_{RS}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m}\hat{R}(t|X=1,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i},\ M=m)\,\hat{P}(M=m|X=0,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i})$$

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{m}\hat{R}(t|X=0,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i},\ M=m)\,\hat{P}(M=m|X=0,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i})$$

$$\widehat{\text{NIE}_{RS}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m}\hat{R}(t|X=1,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i},\ M=m)\,\hat{P}(M=m|X=1,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i})$$

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{m}\hat{R}(t|X=1,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i},\ M=m)\,\hat{P}(M=m|X=0,\ \boldsymbol{Z}_2=\boldsymbol{z}_{2i})$$

Mediation Analysis in Relative Survival Framework

```
// Natural Indirect Effect
standsurv, failure timevar(tt)   ///
  at1(dep5 1 stage2 0 stage3 0 stage4 0 st2dep5 0 st3dep5 0 st4dep5 0, atindweights(p11)) ///
  at2(dep5 1 stage2 1 stage3 0 stage4 0 st2dep5 1 st3dep5 0 st4dep5 0, atindweights(p12)) ///
  at3(dep5 1 stage2 0 stage3 1 stage4 0 st2dep5 0 st3dep5 1 st4dep5 0, atindweights(p13)) ///
  at4(dep5 1 stage2 0 stage3 0 stage4 1 st2dep5 0 st3dep5 0 st4dep5 1, atindweights(p14)) ///
  at5(dep5 1 stage2 0 stage3 0 stage4 0 st2dep5 0 st3dep5 0 st4dep5 0, atindweights(p01)) ///
  at6(dep5 1 stage2 1 stage3 0 stage4 0 st2dep5 1 st3dep5 0 st4dep5 0, atindweights(p02)) ///
  at7(dep5 1 stage2 0 stage3 1 stage4 0 st2dep5 0 st3dep5 1 st4dep5 0, atindweights(p03)) ///
  at8(dep5 1 stage2 0 stage3 0 stage4 1 st2dep5 0 st3dep5 0 st4dep5 1, atindweights(p04)) ///
  lincom(1 1 1 1 -1 -1 -1 -1) lincomvar(t_nie) ci
```

Mediation Analysis in Relative Survival Framework

```
// Natural Indirect Effect
standsurv, failure timevar(tt)   ///
  at1(dep5 1 stage2 0 stage3 0 stage4 0 st2dep5 0 st3dep5 0 st4dep5 0, atindweights(p11)) ///
  at2(dep5 1 stage2 1 stage3 0 stage4 0 st2dep5 1 st3dep5 0 st4dep5 0, atindweights(p12)) ///
  at3(dep5 1 stage2 0 stage3 1 stage4 0 st2dep5 0 st3dep5 1 st4dep5 0, atindweights(p13)) ///
  at4(dep5 1 stage2 0 stage3 0 stage4 1 st2dep5 0 st3dep5 0 st4dep5 1, atindweights(p14)) ///
  at5(dep5 1 stage2 0 stage3 0 stage4 0 st2dep5 0 st3dep5 0 st4dep5 0, atindweights(p01)) ///
  at6(dep5 1 stage2 1 stage3 0 stage4 0 st2dep5 1 st3dep5 0 st4dep5 0, atindweights(p02)) ///
  at7(dep5 1 stage2 0 stage3 1 stage4 0 st2dep5 0 st3dep5 1 st4dep5 0, atindweights(p03)) ///
  at8(dep5 1 stage2 0 stage3 0 stage4 1 st2dep5 0 st3dep5 0 st4dep5 1, atindweights(p04)) ///
  lincom(1 1 1 1 -1 -1 -1 -1) lincomvar(t_nie) ci
```

Received: 25 November 2019 | Revised: 7 August 2020 | Accepted: 21 September 2020

DOI: 10.1002/bimj.201900355

RESEARCH PAPER

Biometrical Journal

Understanding disparities in cancer prognosis: An extension of mediation analysis to the relative survival framework

Elisavet Syriopoulou[1] | Mark J. Rutherford[1] | Paul C. Lambert[1,2]

Code available as online Appendix

Relative Survival Framework Extensions

- Common to use weights (`indweights()`)to standardize to external population.
- Reference adjusted measures - using expected mortality rates from external population[17].
- Incorporate inverse probability weights into model to get doubly robust standardization.
- Marginal measures of life expectancy.

Summary

- Regression standardisation is a simple and underused tool with survival data.
- As long as we can predict survival function, models can be as complex as we like (non-linear effects, non-proportional hazards, interactions with exposure etc.)
- Marginal estimates also used in validation of prognostic models

References

[1] Vansteelandt S, Keiding N. Invited commentary: G-computation–lost in translation? *Am J Epidemiol* 2011;**173**:739–742.

[2] Rutherford M, Lambert P, Sweeting M, Pennington B, Crowther M, Abrams K, Latimer N. Nice dsu technical support document 21:flexible methods for survival analysis 2020;.

[3] Royston P, Lambert PC. *Flexible parametric survival analysis in Stata: Beyond the Cox model*. Stata Press, 2011.

[4] Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *The Stata Journal* 2009;**9**:265–290.

[5] Bower H, Crowther MJ, Lambert PC. strcs: A command for fitting flexible parametric survival models on the log-hazard scale. *The Stata Journal* 2016;**16**:989–1012.

[6] Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2015;**85**:777–793.

[7] Bower H, Crowther M, Rutherford MJ, Andersson TML, Clements M, Liu X, *et al.*. Capturing simple and complex time-dependent effects using flexible parametric survival models:. *Communications in Statistics Simulation and Computation* 2019;.

References 2

[8] Aalen OO, Cook RJ, RÅ¸ysland K. Does cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime data analysis* 2015;**21**:579–593.

[9] Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010;**21**:13–15.

[10] Martinussen T, Vansteelandt S, Andersen PK. Subtleties in the interpretation of hazard ratios. *ArXivorg* 2018;.

[11] Stensrud MJ, Aalen JM, Aalen OO, Valberg M. Limitations of hazard ratios in clinical trials. *European heart journal* 2018;.

[12] Young JG, Tchetgen Tchetgen EJ, Hernan MA. The choice to define competing risk events as censoring events and implications for causal inference. *arXiv preprint* 2018;.

[13] Andersen PK. Decomposition of number of life years lost according to causes of death. *Statistics in Medicine* 2013;**32**:5278–85.

[14] Mozumder SI, Rutherford M, Lambert P. Estimating restricted mean survival time and expected life-years lost in the presence of competing risks within flexible parametric survival models 2021 (in press);.

[15] Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in medicine* 2020;**39**:1199–1236.

References 3

[16] Stensrud MJ, Young J, Didelez V, Robins J, Hernán M. Separable effects for causal inference in the presence of competing events 2020;.

[17] Lambert PC, Andersson TML, Rutherford MJ, Myklebust TÅ, Møller B. Reference-adjusted and standardized all-cause and crude probabilities as an alternative to net survival in population-based cancer studies. *International Journal of Epidemiology* 2020;.