Estimating causal effects in the presence of competing events using regression standardisation with the Stata command standsurv

Elisavet Syriopoulou

Joint work with Sarwar I Mozumder, Mark J Rutherford, Paul C Lambert

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet

E-mail: elisavet.syriopoulou@ki.se

9th September 2021

What is the causal effect?

- Often interest lies in examining associations between exposures and an outcome e.g. whether a treatment improves survival time.
- An association does not necessarily imply causality.
- Causal inference methods provide the conceptual framework and algorithmic tools needed for formalising such investigations (including the required identification assumptions).
- Using the counterfactual outcomes framework, we focus on the average causal effect in the total population, e.g. difference in probabilities of death:

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

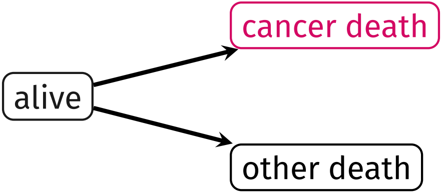
What is the causal effect?

- Often interest lies in examining associations between exposures and an outcome e.g. whether a treatment improves survival time.
- An association does not necessarily imply causality.
- Causal inference methods provide the conceptual framework and algorithmic tools needed for formalising such investigations (including the required identification assumptions).
- Using the counterfactual outcomes framework, we focus on the average causal effect in the total population, e.g. difference in probabilities of death:

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

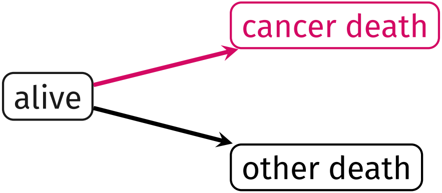
What are competing events?

Events that prevent the occurrence of the event of interest may be present, so-called competing events.



What are competing events?

Events that prevent the occurrence of the event of interest may be present, so-called competing events.

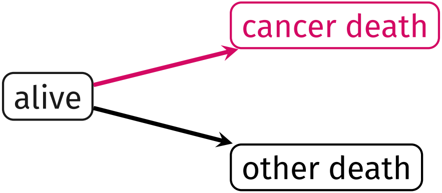


Depending on whether the competing events are accommodated or eliminated, various causal effects can be defined:

- Total effects
- Direct effects
- Separable effects

What are competing events?

Events that prevent the occurrence of the event of interest may be present, so-called competing events.



Depending on whether the competing events are accommodated or eliminated, various causal effects can be defined:

- Total effects
- Direct effects
- Separable effects

Young *et al*. A causal framework for classical statistical estimands in failure-time settings with competing events. Stats Med, 39:1199–1236, 2020
Stensrud *et al*. Separable effects for causal inference in the presence of competing events. J Am Stat Assoc, 2020.

Estimation with regression standardisation

1. Fit a survival model such as flexible parametric model.

Estimation with regression standardisation

1. Fit a survival model such as flexible parametric model.
2. Obtain predictions for each individual in the population by setting $X = x$.

Estimation with regression standardisation

1. Fit a survival model such as flexible parametric model.
2. Obtain predictions for each individual in the population by setting $X = x$.
3. Calculate an average of the individual specific predictions in a population of N patients and form the relevant contrast.

Estimation with regression standardisation

1. Fit a survival model such as flexible parametric model.
2. Obtain predictions for each individual in the population by setting $X = x$.
3. Calculate an average of the individual specific predictions in a population of N patients and form the relevant contrast.

The marginal all-cause probability of death can be estimated by

$$E[\widehat{F}(t|X = x, \boldsymbol{Z})] = \frac{1}{N}\sum_{i=1}^{N}\widehat{F}(t|X = x, \boldsymbol{Z} = \boldsymbol{z_i})]$$

Estimation with regression standardisation

1. Fit a survival model such as flexible parametric model.
2. Obtain predictions for each individual in the population by setting $X = x$.
3. Calculate an average of the individual specific predictions in a population of N patients and form the relevant contrast.

The marginal all-cause probability of death can be estimated by

$$E[\widehat{F}(t|X=x,\boldsymbol{Z})] = \frac{1}{N}\sum_{i=1}^{N}\widehat{F}(t|X=x,\boldsymbol{Z}=\boldsymbol{z_i})]$$

Command `standsurv` can be used to obtain estimates with regression standardisation and it can be installed by running:

```
ssc install standsurv, replace
```

Illustrative example

- Data from a trial on prostate cancer with individuals that were randomly assigned estrogen therapy.
- We restrict our analysis to high-dose estrogen therapy arm (DES) and placebo
- Data available at `https://hbiostat.org/data`

We fit cause-specific models:

```
// For death due to prostate cancer
stset dtime, failure(eventType==1) exit(time 60)
stpm2 rx normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(4) ///
     tvc(rx) dftvc(2)
estimates store prostate
// For death due to other causes
stset dtime, failure(eventType==2) exit(time 60)
stpm2 rx normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(3)
estimates store other
// Also, create timevar for predictions
range timevar 0 60 121
```

Total effects (or crude)

- Total effects accommodate competing events.
- Refer to a real-world setting where competing events are present.
- They are highly relevant for patients and health professionals.
- Can also aid in policy decisions e.g. on resource allocation.

Examples of total effects are:

- Cause-specific cumulative incidence functions
- Expected loss in life due to a specific cause of death before time t^* (using option `rmft` in `standsurv`)

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X = 0, \boldsymbol{Z})\right]$$

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X=x,\boldsymbol{Z})\right] = E\left[\int_0^t S(u|X=x,\boldsymbol{Z})h_c(u|X=x,\boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X=1,\boldsymbol{Z})\right] - E\left[F_c(t|X=0,\boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X = 0, \boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X = 0, \boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X = 0, \boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

$$E\left[F_c(t|X=x,\boldsymbol{Z})\right] = E\left[\int_0^t S(u|X=x,\boldsymbol{Z})h_c(u|X=x,\boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X=1,\boldsymbol{Z})\right] - E\left[F_c(t|X=0,\boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Cause-specific cumulative incidence function (CIF)

The marginal CIF for death due to prostate cancer in the presence of death due to other causes when setting treatment to $X = x$:

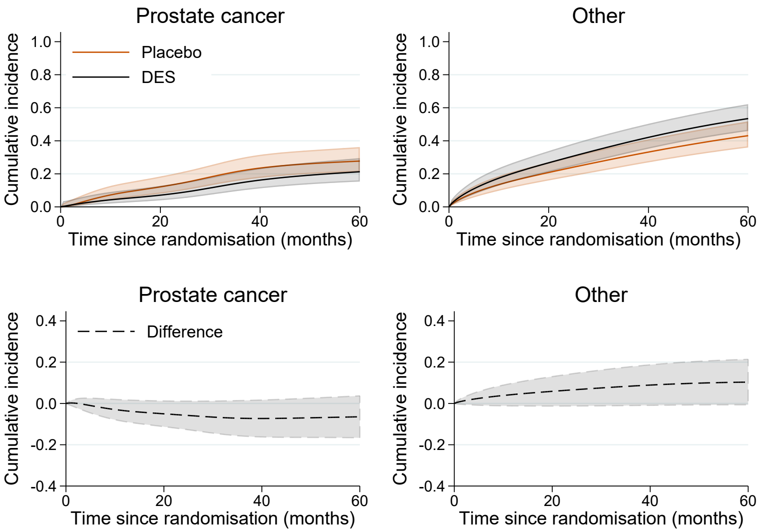
$$E\left[F_c(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E\left[F_c(t|X = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X = 0, \boldsymbol{Z})\right]$$

```
standsurv, crmodels(prostate other) cif ///
   at1(rx 0) at2(rx 1)  timevar(timevar) ///
   contrast(difference) ci ///
   atvars(CIF0 CIF1) contrastvars(CIF_diff)
```

Example - CIFs



Direct effects

- The total effect provides no information about whether part of the treatment effect on the event of interest is due to the treatment effect on the competing event.
- Instead, the direct effect quantifies an effect of treatment on the event of interest that is not mediated by the competing event.
- Direct effects are useful for comparing populations without any possible distortions from competing causes of death.
- They can also be applied to explore temporal trends or to study the aetiology of a disease.

Consider a hypothetical intervention that eliminates the competing deaths due to other causes.

Net probability of death

The marginal counterfactual probability of death from prostate cancer under an intervention of eliminating competing events when setting $X = x$:

$$E\left[F_c^N(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S_c(u|X = x, \boldsymbol{Z})h_c(u|X = x, \boldsymbol{Z})du\right]$$

The average causal difference:

$$E[F_c^N(t|X = 1, \boldsymbol{Z})] - E[F_c^N(t|X = 0, \boldsymbol{Z})]$$

Net probability of death

The marginal counterfactual probability of death from prostate cancer under an intervention of eliminating competing events when setting $X = x$:

$$E\left[F_c^N(t|X = x, \boldsymbol{Z})\right] = E\left[\int_0^t S_c(u|X = x, \boldsymbol{Z}) h_c(u|X = x, \boldsymbol{Z}) du\right]$$

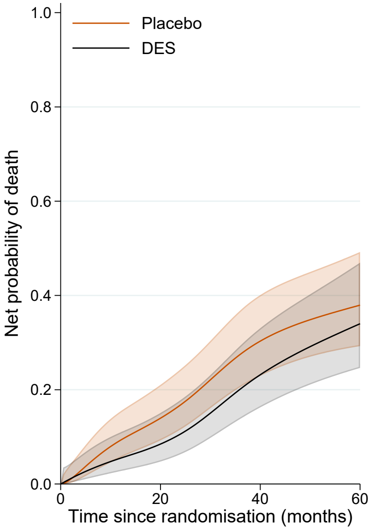
The average causal difference:

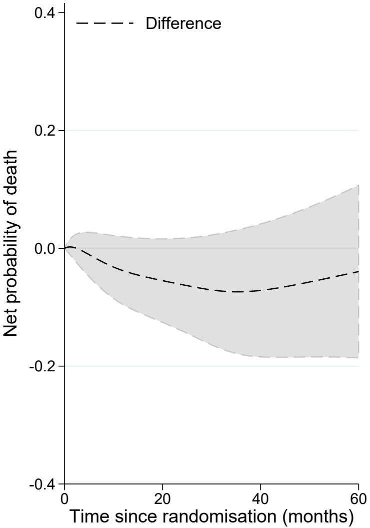
$$E[F_c^N(t|X = 1, \boldsymbol{Z})] - E[F_c^N(t|X = 0, \boldsymbol{Z})]$$

```
estimates restore prostate

standsurv, failure at1(rx 0) at2(rx 1)  ///
   timevar(timevar) contrast(difference) ci ///
   atvars(F_net_prostate0 F_net_prostate1) ///
   contrastvars(F_net_prostate_diff)
```

Example - net probability of death





Separable effects

Suppose that the treatment X can be conceptualised as having two binary components that act through different causal pathways: one component X^c that affects the cancer of interest and one component X^o that affects the competing event.

- The separable direct effect of treatment on the probability of death from cancer is defined as

$$E\left[F_c(t|X^c = 1, X^o = x, \boldsymbol{Z})\right] - E\left[F_c(t|X^c = 0, X^o = x, \boldsymbol{Z})\right]$$

- The separable indirect effect of treatment on the probability of death from cancer as

$$E\left[F_c(t|X^c = x, X^o = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X^c = x, X^o = 0, \boldsymbol{Z})\right]$$

Separable effects

Suppose that the treatment X can be conceptualised as having two binary components that act through different causal pathways: one component X^c that affects the cancer of interest and one component X^o that affects the competing event.

- The separable direct effect of treatment on the probability of death from cancer is defined as

$$E\left[F_c(t|X^c = 1, X^o = x, \boldsymbol{Z})\right] - E\left[F_c(t|X^c = 0, X^o = x, \boldsymbol{Z})\right]$$

- The separable indirect effect of treatment on the probability of death from cancer as

$$E\left[F_c(t|X^c = x, X^o = 1, \boldsymbol{Z})\right] - E\left[F_c(t|X^c = x, X^o = 0, \boldsymbol{Z})\right]$$

Separable effects in Stata

```
gen rx_c = rx
gen rx_o = rx

// Prostate cancer
stset dtime, failure(eventType==1) exit(time 60)
stpm2 rx_c normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(4) ///
     tvc(rx_c) dftvc(2)
estimates store prostate

// Other causes
stset dtime, failure(eventType==2) exit(time 60)
stpm2 rx_o normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(3)
estimates store other

standsurv, crmodels(prostate other) cif ///
   timevar(timevar) contrast(difference) ci ///
   at1(rx_c 1 rx_o 1)  ///
   at2(rx_c 1 rx_o 0)  ///
   at3(rx_c 0 rx_o 0)  ///
   atvars(F_rx11 F_rx10 F_rx00) ////
   contrastvars(F_diff_indirect F_diff_total)
```

Separable effects in Stata

```
gen rx_c = rx
gen rx_o = rx

// Prostate cancer
stset dtime, failure(eventType==1) exit(time 60)
stpm2 rx_c normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(4) ///
     tvc(rx_c) dftvc(2)
estimates store prostate

// Other causes
stset dtime, failure(eventType==2) exit(time 60)
stpm2 rx_o normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(3)
estimates store other

standsurv, crmodels(prostate other) cif ///
   timevar(timevar) contrast(difference) ci ///
   at1(rx_c 1 rx_o 1)  ///
   at2(rx_c 1 rx_o 0)  ///
   at3(rx_c 0 rx_o 0)  ///
   atvars(F_rx11 F_rx10 F_rx00) ////
   contrastvars(F_diff_indirect F_diff_total)
```

Separable effects in Stata

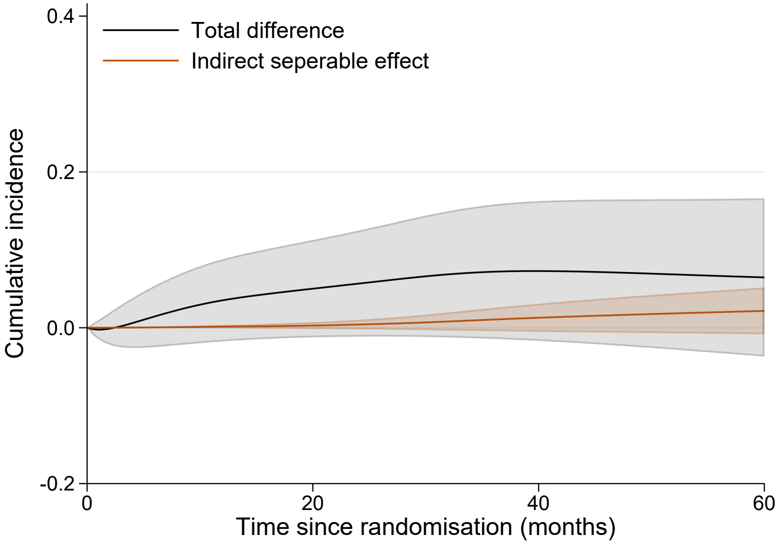
```
gen rx_c = rx
gen rx_o = rx

// Prostate cancer
stset dtime, failure(eventType==1) exit(time 60)
stpm2 rx_c normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(4) ///
     tvc(rx_c) dftvc(2)
estimates store prostate

// Other causes
stset dtime, failure(eventType==2) exit(time 60)
stpm2 rx_o normalAct ageCat2 ageCat3 hx hgBinary, scale(hazard) df(3)
estimates store other

standsurv, crmodels(prostate other) cif ///
   timevar(timevar) contrast(difference) ci ///
   at1(rx_c 1 rx_o 1)  ///
   at2(rx_c 1 rx_o 0)  ///
   at3(rx_c 0 rx_o 0)  ///
   atvars(F_rx11 F_rx10 F_rx00) ////
   contrastvars(F_diff_indirect F_diff_total)
```

Example - separable indirect effect



Final points

- Estimates of interest can also be obtained after fitting more complex FPMs e.g. non-linear effects and interactions.

Final points

- Estimates of interest can also be obtained after fitting more complex FPMs e.g. non-linear effects and interactions.
- In principle, any contrast can be obtained with `standsurv` by creating a user-defined mata function which can be called in the option `userfunction()` instead of the `contrast()`.

Final points

- Estimates of interest can also be obtained after fitting more complex FPMs e.g. non-linear effects and interactions.
- In principle, any contrast can be obtained with `standsurv` by creating a user-defined mata function which can be called in the option `userfunction()` instead of the `contrast()`.
- `standsurv` can also be used to obtain non-marginalised estimates: by specifying the entire covariate pattern so that the predictions are not averaged over any covariate distribution.

Final points

- Estimates of interest can also be obtained after fitting more complex FPMs e.g. non-linear effects and interactions.
- In principle, any contrast can be obtained with `standsurv` by creating a user-defined mata function which can be called in the option `userfunction()` instead of the `contrast()`.
- `standsurv` can also be used to obtain non-marginalised estimates: by specifying the entire covariate pattern so that the predictions are not averaged over any covariate distribution.
- Preprint available at: `https://arxiv.org/abs/2109.03628`