UNIVERSITY OF LEICESTER

Inverse probability weighting and doubly robust standardisation in the relative survival framework

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

[1]Biostatistics Research Group, University of Leicester

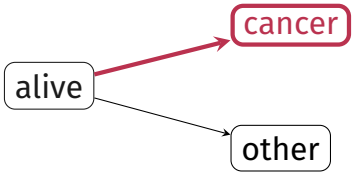
[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet

*es303@.le.ac.uk

Funding: NIHR Doctoral Research Fellowship Programme

August 2020

Population-based cancer data



In the presence of competing risks, we can estimate:

- Cause-specific mortality
- Excess mortality

Information on the cause of death is usually not accurate.

Why focus on excess mortality?

Excess mortality does not require cause of death information:

- It compares the all-cause mortality of the cancer population to the expected mortality of a comparable group in the general population.
- The expected mortality is considered to be known and is obtained by available population lifetables.

Let X denote an exposure taking values $x = 0$ for unexposed and $x = 1$ for exposed.

Let $\boldsymbol{Z}$ denote the set of all confounders, with

- $\boldsymbol{Z_1}$ the confouders for expected mortality and
- $\boldsymbol{Z_2}$ the confounders for excess mortality

Excess mortality and Relative survival

Excess mortality

$$\begin{array}{c}\text{excess}\\ \text{mortality}\end{array} = \begin{array}{c}\text{all-cause}\\ \text{mortality}\end{array} - \begin{array}{c}\text{expected}\\ \text{mortality}\end{array}$$

$$\lambda(t|X = x_i, \boldsymbol{Z_2} = \boldsymbol{z_{2i}}) = h(t|X = x_i, \boldsymbol{Z} = \boldsymbol{z_i}) - h^*(t|X = x_i, \boldsymbol{Z_1} = \boldsymbol{z_{1i}})$$

Relative survival

$$\text{relative survival} = \frac{\text{all-cause survival}}{\text{expected survival}}$$

$$R(t|X = x_i, \boldsymbol{Z_2} = \boldsymbol{z_{2i}}) = \frac{S(t|X = x_i, \boldsymbol{Z} = \boldsymbol{z_i})}{S^*(t|X = x_i, \boldsymbol{Z_1} = \boldsymbol{z_{1i}})}$$

Mortality rates and survival probabilities vary between individuals with different characteristics.

Assumptions

Relative survival estimates survival in a hypothetical world where the only possible cause of death is the cancer of interest (net survival).

- **The competing risks are conditionally independent** i.e. there are no other factors to affect both competing events than the factors we have adjusted for.
- **Appropriate information on the expected survival** of the general population so that the cancer population and the general population are comparable.

Marginal measures and the average causal effect

- The marginal relative survival for $X = x$:

$$\theta(t|X = x) = E\left[R(t|X = x, \boldsymbol{Z_2})\right]$$

with the expectation over the marginal distribution of $\boldsymbol{Z_2}$.

Marginal measures and the average causal effect

- The marginal relative survival for $X = x$:

$$\theta(t|X = x) = E\left[R(t|X = x, \boldsymbol{Z_2})\right]$$

with the expectation over the marginal distribution of $\boldsymbol{Z_2}$.

- The average causal difference:

$$E\left[R(t|X = 1, \boldsymbol{Z_2})\right] - E\left[R(t|X = 0, \boldsymbol{Z_2})\right]$$

with the first term setting $X = 1$ for everyone in the study population and the second term setting $X = 0$.

Identification

- Similar assumptions as for a standard survival setting (conditional exchangeability, consistency, positivity) but this time they are extended to both competing events: death due to cancer and death due to other causes.
- Conditional exchangeability for the other cause mortality can only be achieved by adjusting the available population lifetables of the general population for sufficient variables.

Estimation

Regression standardisation

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

The average causal difference can be estimated as:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{R}(t|X=1,\boldsymbol{Z_2}=\boldsymbol{z_{2i}}) - \frac{1}{N}\sum_{i=1}^{N}\hat{R}(t|X=0,\boldsymbol{Z_2}=\boldsymbol{z_{2i}})$$

Estimation

Regression standardisation yields an estimator that consistently estimates the causal effect if the correct model has been fitted for the survival outcome conditional on exposure and confounders.

Estimation

Regression standardisation yields an estimator that consistently estimates the causal effect if the correct model has been fitted for the survival outcome conditional on exposure and confounders.

Other approaches:

- inverse probability weighting (IPW)
  $\longrightarrow$*requires a correct model for the exposure conditional on the confounders (propensity score)*
- doubly robust standardisation
  $\longrightarrow$*requires that at least one of the propensity score model or the survival model is correctly specified*

Inverse probability weighting

On a standard survival setting:

1. Fit a propensity score model for the exposure given confounders $\boldsymbol{Z_2}$.
2. Obtain predictions for each exposure level given confounders to calculate the weights

$$w_i^s = \frac{\hat{P}(X = x)}{\hat{P}(X = x | \boldsymbol{Z_2} = \boldsymbol{z_{2i}})}$$

3. Fit a survival model using the weights and including only the exposure of interest.

The IPW approach requires a marginal structural model to be fitted. However, this cannot be directly extended to a relative survival setting!

Issues

When no confounders are included in the survival model:

- For a standard survival model, the estimates should be in good agreement with the estimates of a non-parametric approach.
- For a relative survival model, the all-cause mortality can be written as:

$$h(t|X = x, \boldsymbol{Z_1} = \boldsymbol{z_{1i}}) = \boldsymbol{h^*(t|X = x, Z_1 = z_{1i})} + \lambda(t|X = x)$$

The excess mortality remains constant across individuals but the expected mortality varies for individuals with different confounders Z_1 (such as sex, age and calendar year)

Marginal expected mortality rates

Need to incorporate the marginal expected mortality rates at time t, $\bar{h}^*(t|X = x)$, rather than individual expected mortality rates.

$$h_m(t|X = x) = \bar{h}^*(t|\boldsymbol{X = x}) + \lambda_m(t|X = x)$$

The mean expected hazard at risk time t can be written as:

$$\bar{h}^*(t|X = x) = \frac{\sum_{j \in \mathcal{R}(t)} w_i^*(t) h^*(t|X = x, \boldsymbol{Z_1 = z_{1j}})}{\sum_{j \in \mathcal{R}(t)} w_i^*(t)}$$

with weights $w_i^*(t)$ varying by individual and time and being equal to the inverse of the expected survival at time t:

$$w_i^*(t) = \frac{1}{S^*(t|X = x, \boldsymbol{Z_1 = z_{1i}})}$$

See talk by Paul Lambert for more details on marginal models for relative survival.

Inverse probability weighting in relative survival

1. Fit a propensity score model for the exposure given confounders $\boldsymbol{Z_2}$.
2. Obtain predictions for each exposure level given confounders $\hat{P}(X = x|\boldsymbol{Z_2} = \boldsymbol{z_{2i}})$ to calculate the weights w_i^s.
3. **Obtain the marginal expected mortality, $\bar{h}^*(t|X = x)$, by replacing weights $w_i^*(t)$ with weights $w_i(t)$:**

$$\boldsymbol{w_i(t) = w_i^*(t) \times w_i^s}$$

4. Fit a survival model using the weights, including only the exposure of interest **and by incorporating the marginal expected mortality instead of the individual rates.**

Doubly robust standardisation in relative survival

1. Fit a propensity score model for the exposure given confounders $\boldsymbol{Z_2}$.
2. Obtain predictions for each exposure level given confounders $\hat{P}(X = x|\boldsymbol{Z_2} = \boldsymbol{z_{2i}})$ to calculate the weights.
3. Fit a relative survival model using the weights **and including exposure and confounders.**
4. Obtain survival predictions for each individual in the study population by setting $X = 1$ and $X = 0$.
5. The individual predictions are then averaged and the relevant contrasts are formed.

Monte Carlo simulation study

How sensitive are point estimates obtained from regression standardisation, IPW and doubly robust standardisation to model misspecification?

Monte Carlo simulation study - generating data

- 2000 observations and 1000 replications
- A binary exposure variable from a binomial distribution.
- Three confounders L_1, L_2, L_3 with varying degree of correlation between them (high, medium, none).
- All effects were assumed to be proportional with time.
- Time to death was taken as the minimum out of:
  - Time to death from cancer (simulated from a Weibull distribution) and
  - Time to death from other causes (simulated from exponential distributions using rates from lifetables in England)

Monte Carlo simulation study - Estimands

Interested in relative survival both at 1 and 5 years.

- marginal relative survival of the exposed
- marginal relative survival of the unexposed
- Difference in marginal relative survival between exposure groups

Monte Carlo simulation study - methods compared

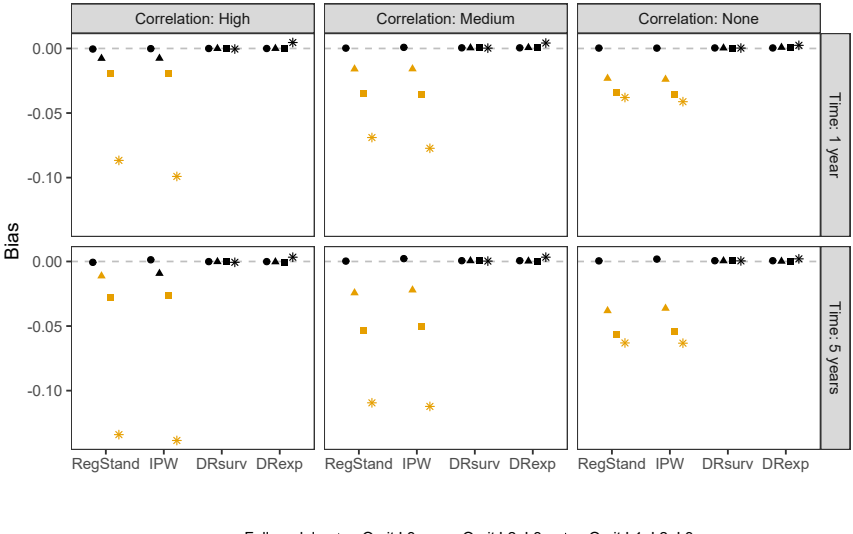
The four methods compared are:

- Regression standardisation (RegStand)
- Inverse probability weighting (IPW)
- Doubly robust standardisation - assuming a correct model for the survival outcome (DRsurv)
- Doubly robust standardisation - assuming a correct model for the exposure outcome (DRexp)

Confounders were gradually omitted from the relevant model:

- Scenario 1: All confounders L_1, L_2, L_3 are included.
- Scenario 2: Omit confounder L_3.
- Scenario 3: Omit confounders L_2 and L_3.
- Scenario 4: Omit all confounders.

Bias for the average difference



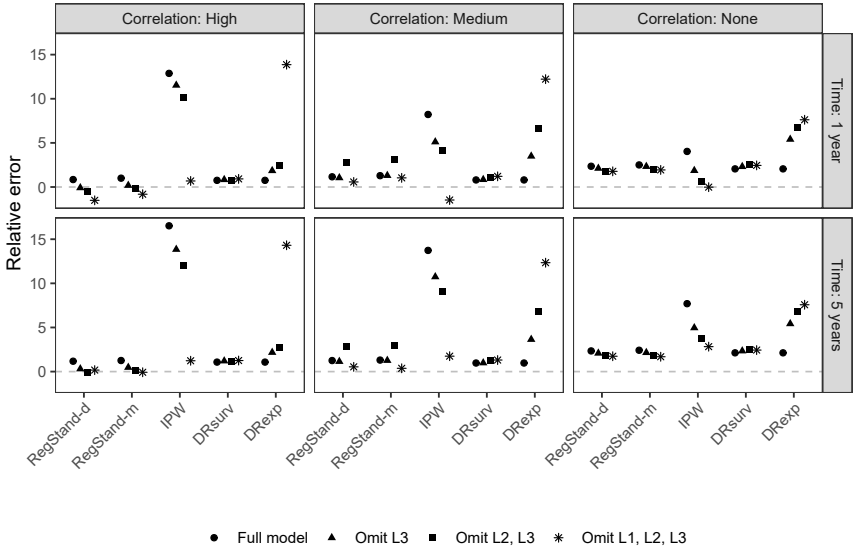
Absolute bias larger than 0.01 is shown in orange.

Obtaining standard errors

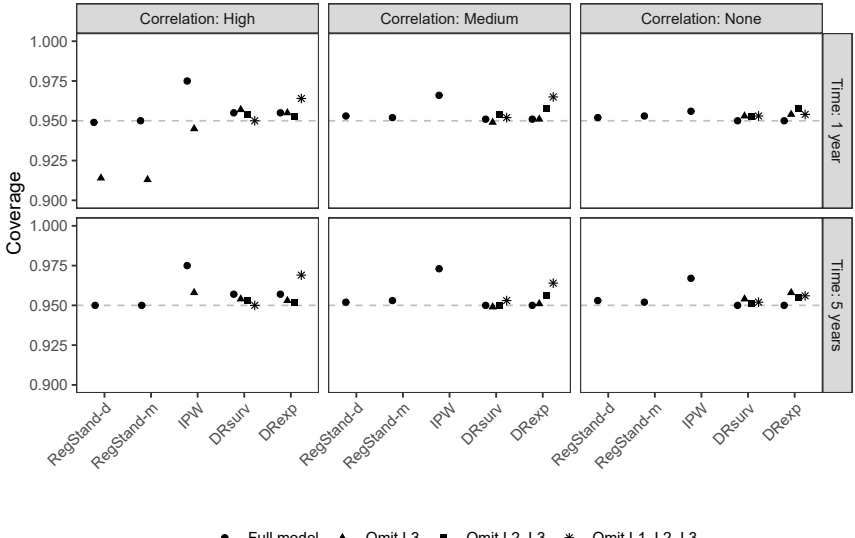
- The standard errors obtained from each method were also compared.
- Standard errors for the regression standardisation were obtained by applying either the delta method (RegStand-d) or the M-estimation (RegStand-m).
- The standard errors of the IPW, DRsurv and DRexp were obtained with the delta method while using robust clustered standard errors.
- Next figure shows the relative errors which are defined as

$$100\left(\frac{\text{modSE}}{\text{empSE}} - 1\right)$$

Relative errors for the average difference



Coverage for the average difference



Conclusions

- All methods performed well when correctly specified models were fitted.
- In practice, model misspecification is very common. Doubly robust standardisation might be preferable when this is applicable.
- To quantify survival in a real-world setting in which both cancer and other causes of death are present, the marginal all-cause survival and marginal crude probabilities of death can be obtained.

Selected References I

Pohar Perme M, Stare J & Estève J
On Estimation in Relative Survival.
Biometrics,, 68, 113-120, 2012.

Royston P & Lambert P C
Flexible parametric survival analysis in Stata: Beyond the Cox model.
Stata Press, 2011.

Stefanski L A & Boos D D
The Calculus of M-Estimation.
The American Statistician, 56, 29-38, 2002.

Vansteelandt S & Keiding N
Invited commentary: G-computation-lost in translation?
American Journal of Epidemiology, 173, 739-742, 2011.

Syriopoulou E, Rutherford M J & Lambert, P C
Marginal measures and causal effects using the relative survival framework.
International Journal of Epidemiology, 49, 619–628, 2020.