UNIVERSITY OF LEICESTER

Understanding inequalities 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]

[1]Biostatistics Research Group, University of Leicester

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

*es303@.le.ac.uk

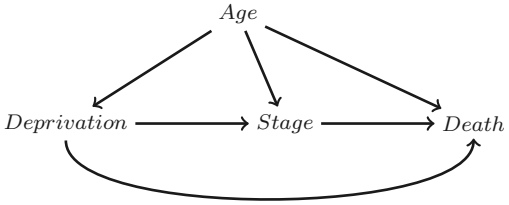
Funding: NIHR Doctoral Research Fellowship Programme

17 July 2019

Motivation

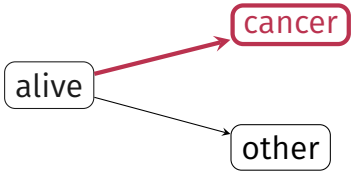
Survival after a cancer diagnosis varies considerably across population groups e.g socioeconomic groups.

Is there a third variable that can partly explain these differences?



Complex mechanisms contribute towards disparities: all-cause survival differences are the result of both cancer-related and other cause factors.

Background



In the presence of competing risks, we can estimate one of the following:

- Cause-specific mortality
- Excess mortality

Information on the cause of death is usually not accurate or not available. Excess mortality does not require information on the cause of death.

Excess mortality and Relative survival

Excess mortality

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

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

Relative survival

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

$$R(t) = \frac{S(t)}{S^*(t)} \qquad S(t) = S^*(t)R(t)$$

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.

- **Appropriate information on the expected survival** of the general population so that the cancer population and the general population are comparable.
- **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.

Marginal estimates

Let us assume we are interested on the effect of an exposure X to the survival time while adjusting for confounders Z.

A summary of the population prognosis can be obtained by the standardised relative survival function $E\left[R(t|Z)\right]$ that is estimated by:

$$E\left[\widehat{R}(t|Z)\right] = \frac{1}{N}\sum_{i=1}^{N}\widehat{R}(t|Z = z_i).$$

Regression standardisation

1. Fit a survival model such as flexible parametric model.
2. Obtain survival predictions for each individual in the population.
3. Calculate an average of the survival predictions.

Forming contrasts

If interested in **relative** survival:

$$E\left[R(t|X=1,Z)\right] - E\left[R(t|X=0,Z)\right]$$

- Refers to a hypothetical world where the cancer of interest is the only possible cause of death.

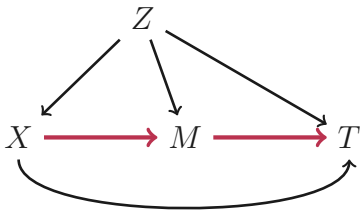
If interested in **all-cause** survival:

$$E\left[S(t|X=1,Z)\right] - E\left[S(t|X=0,Z)\right]$$
$$E\left[S^*(t|X=1,Z)R(t|X=1,Z)\right] - E\left[S^*(t|X=0,Z)R(t|X=0,Z)\right]$$

- Differences may be due to either cancer of interest or other cause mortality or both.

Exploring the effect of a mediator

How much of the differences between exposure groups can be explained by differences at the mediator M distribution?

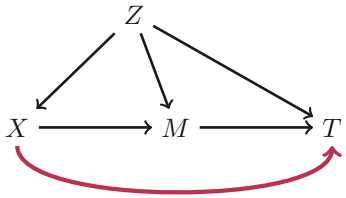


Let M^x denote the counterfactual mediator distribution when intervening to set $X = x$.

Direct & indirect effects - relative survival framework

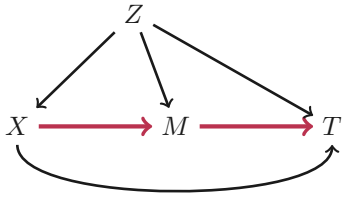
Natural direct effect

$$NDE_{RS} = E\left[R(t|\boldsymbol{X = 1}, Z, M^0)\right] - E\left[R(t|\boldsymbol{X = 0}, Z, M^0)\right]$$



Natural indirect effect

$$NIE_{RS} = E\left[R(t|X = 1, Z, \boldsymbol{M^1})\right] - E\left[R(t|X = 1, Z, \boldsymbol{M^0})\right]$$



Estimation

Step 1. Fit a survival model including X, M, Z. A separate model for the mediator including X, Z should also be fitted.

Estimation

Step 1. Fit a survival model including X, M, Z. A separate model for the mediator including X, Z should also be fitted.

Step 2. For each individual in the study population obtain predictions for $\widehat{P}(M = m|X = x, Z = z_i)$, at each $X = x$.

Estimation

Step 1. Fit a survival model including X, M, Z. A separate model for the mediator including X, Z should also be fitted.

Step 2. For each individual in the study population obtain predictions for $\widehat{P}(M = m|X = x, Z = z_i)$, at each $X = x$.

Step 3. Obtain predictions of $\widehat{R}(t|X = x, Z = z_i, M = m)$ at $X = x$, using the predictions of Step 2 as weights. Form contrasts to obtain the $\widehat{NDE_{RS}}$ and $\widehat{NIE_{RS}}$.

Estimation

Step 1. Fit a survival model including X, M, Z. A separate model for the mediator including X, Z should also be fitted.

Step 2. For each individual in the study population obtain predictions for $\widehat{P}(M = m|X = x, Z = z_i)$, at each $X = x$.

Step 3. Obtain predictions of $\widehat{R}(t|X = x, Z = z_i, M = m)$ at $X = x$, using the predictions of Step 2 as weights. Form contrasts to obtain the $\widehat{NDE_{RS}}$ and $\widehat{NIE_{RS}}$.

Step 4. Repeat from Step 2, k times, while performing parametric bootstrap for the parameter estimates for both models.

Estimation

Step 1. Fit a survival model including X, M, Z. A separate model for the mediator including X, Z should also be fitted.

Step 2. For each individual in the study population obtain predictions for $\widehat{P}(M = m|X = x, Z = z_i)$, at each $X = x$.

Step 3. Obtain predictions of $\widehat{R}(t|X = x, Z = z_i, M = m)$ at $X = x$, using the predictions of Step 2 as weights. Form contrasts to obtain the $\widehat{NDE_{RS}}$ and $\widehat{NIE_{RS}}$.

Step 4. Repeat from Step 2, k times, while performing parametric bootstrap for the parameter estimates for both models.

Step 5. Calculate 95% confidence intervals either by taking the 2.5% and 97.5% percentiles of the estimates across the bootstrapped samples or by using their standard deviation.

Estimation

$$\widehat{NDE_{RS}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m} \boldsymbol{\widehat{R}(t|X=1,Z=z_i,M=m)}\widehat{P}(M=m|X=0,Z=z_i) - \frac{1}{N}\sum_{i=1}^{N}\sum_{m} \boldsymbol{\widehat{R}(t|X=0,Z=z_i,M=m)}\widehat{P}(M=m|X=0,Z=z_i)$$

$$\widehat{NIE_{RS}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{m} \widehat{R}(t|X=1,Z=z_i,M=m)\boldsymbol{\widehat{P}(M=m|X=1,Z=z_i)} - \frac{1}{N}\sum_{i=1}^{N}\sum_{m} \widehat{R}(t|X=1,Z=z_i,M=m)\boldsymbol{\widehat{P}(M=m|X=0,Z=z_i)}$$

Direct & indirect effects - all cause setting

- Compare $S^*(t|X=1,Z)$ with $S^*(t|X=0,Z)$:

$$NDE_{AC1} = E\left[\boldsymbol{S^*(t|X=1,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X=0,Z)}R(t|X=0,Z,M^0)\right]$$

Differences may be due to either the cancer of interest or other causes or both.

Direct & indirect effects - all cause setting

- Compare $S^*(t|X=1,Z)$ with $S^*(t|X=0,Z)$:

$$NDE_{AC1} = E\left[\boldsymbol{S^*(t|X=1,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X=0,Z)}R(t|X=0,Z,M^0)\right]$$

Differences may be due to either the cancer of interest or other causes or both.

- Use the observed distribution of the exposure, $S^*(t|X,Z)$:

$$NDE_{AC2} = E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=0,Z,M^0)\right]$$

Differences can only be due to the cancer of interest.

Direct & indirect effects - all cause setting

- Compare $S^*(t|X=1,Z)$ with $S^*(t|X=0,Z)$:

$$NDE_{AC1} = E\left[\boldsymbol{S^*(t|X=1,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X=0,Z)}R(t|X=0,Z,M^0)\right]$$

Differences may be due to either the cancer of interest or other causes or both.

- Use the observed distribution of the exposure, $S^*(t|X,Z)$:

$$NDE_{AC2} = E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=0,Z,M^0)\right]$$

Differences can only be due to the cancer of interest.

- Sometimes it might also be of interest to estimate the effect, within subsets of the whole population e.g. NDE among the exposed using $S^*(t|X=1,Z_{X=1})$.

Avoidable deaths under hypothetical interventions

"What if we could eliminate differences in the mediator distribution between exposed and unexposed groups?"

Avoidable deaths under hypothetical interventions

"What if we could eliminate differences in the mediator distribution between exposed and unexposed groups?"

- The predicted number of deaths for the exposed:

$$D_1(t|X=1,M^1) = N^* \times \left(1 - E\left[S^*(t|X=1,Z_{X=1})R(t|X=1,Z,M^1)\right]\right)$$

Avoidable deaths under hypothetical interventions

"What if we could eliminate differences in the mediator distribution between exposed and unexposed groups?"

- The predicted number of deaths for the exposed:

$$D_1(t|X=1,M^1) = N^* \times \left(1 - E\left[S^*(t|X=1,Z_{X=1})R(t|X=1,Z,M^1)\right]\right)$$

- The expected number of deaths if the exposed had the same **mediator distribution** as the unexposed:

$$D_M(t|X=1,M^0) = N^* \times \left(1 - E\left[S^*(t|X=1,Z_{X=1})R(t|X=0,Z,\boldsymbol{M^0})\right]\right)$$

Avoidable deaths under hypothetical interventions

"What if we could eliminate differences in the mediator distribution between exposed and unexposed groups?"

- The predicted number of deaths for the exposed:

$$D_1(t|X=1,M^1) = N^* \times \left(1 - E\left[S^*(t|X=1,Z_{X=1})R(t|X=1,Z,M^1)\right]\right)$$

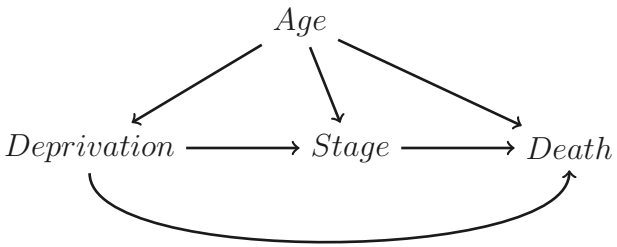
- The expected number of deaths if the exposed had the same **mediator distribution** as the unexposed:

$$D_M(t|X=1,M^0) = N^* \times \left(1 - E\left[S^*(t|X=1,Z_{X=1})R(t|X=0,Z,\boldsymbol{M^0})\right]\right)$$

- The avoidable deaths are:

$$D_1(t|X=1,M^1) - D_M(t|X=1,M^0)$$

Application - colon cancer in England



How much of the differences between deprivation groups can be explained by differences at the stage distribution?

How many deaths could be avoided if the most deprived patients had the same stage distribution as the least deprived? or same relative survival?

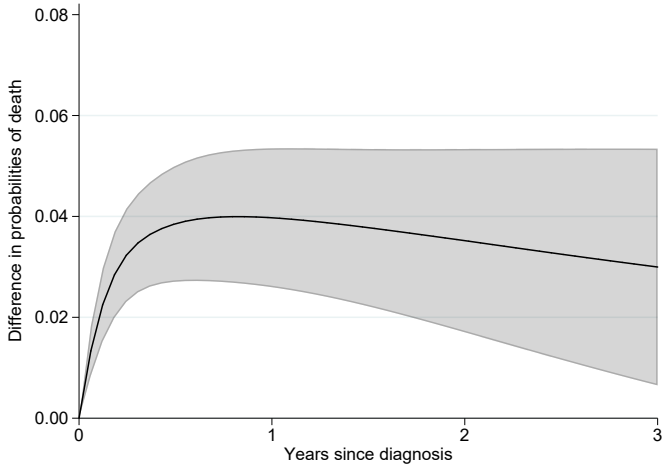
Data and analysis

We conducted a complete case analysis including 15,630 patients diagnosed between 2011-2013 (57.6% in the least deprived group).

Stage at diagnosis	Least Deprived	Most deprived
I	1338(**14.86%**)	912(**13.76%**)
II	2644(29.37%)	1950(29.42%)
III	2435(27.05%)	1716(25.89%)
IV	2585(**28.72%**)	2050(**30.93%**)

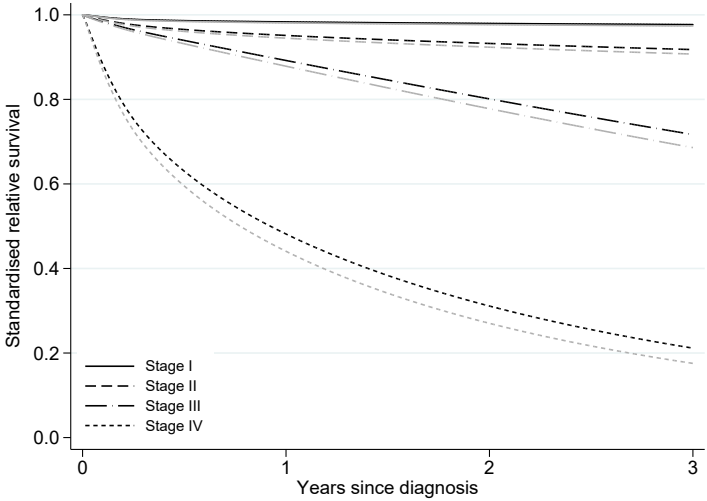
We fitted a flexible parametric survival model that uses restricted cubic splines to model the baseline excess hazard. A multinomial regression model was also fitted for stage. Confidence interval were obtained using the standard deviation of a parametric bootstrap sample with $k = 250$.

Total effect with 95% CIs - all cause setting

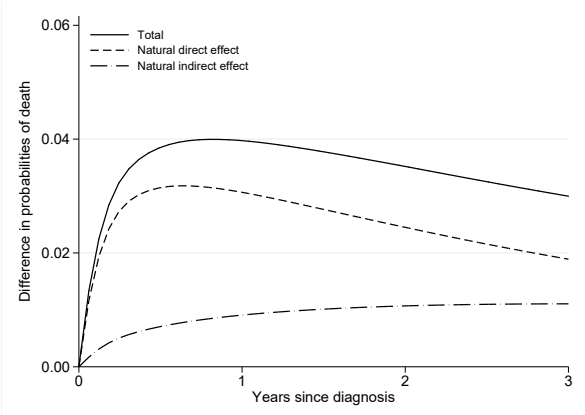


$$TE = E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^1)\right] - E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=0,Z,M^0)\right]$$

Relative survival by stage



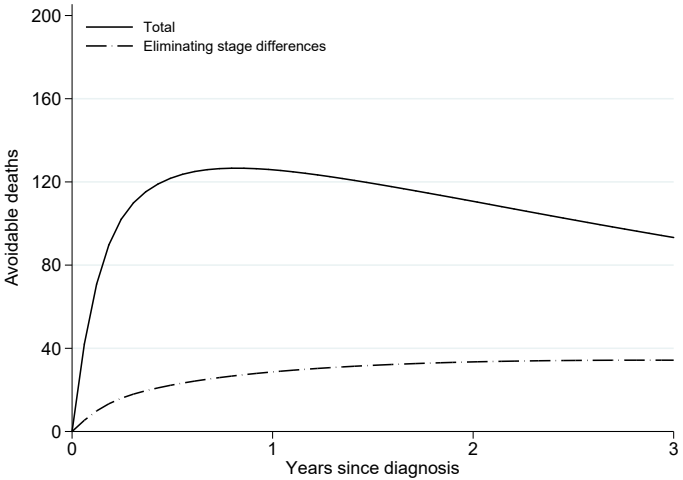
Direct and indirect effects - all cause setting



$$NDE_{AC2} = E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^0)\right] - E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=0,Z,M^0)\right]$$

$$NIE_{AC2} = E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^1)\right] - E\left[\boldsymbol{S^*(t|X,Z)}R(t|X=1,Z,M^0)\right]$$

Avoidable deaths



*Out of 3228 patients ($N*$) from the most deprived group diagnosed in 2013 the most recent year in our data.

Conclusions

- Mediation analysis within the relative survival framework provides a useful tool for understanding differences in cancer survival. It allows to focus on cancer-related factors.
- Need to be careful when interpreting the results as a number of assumption need to hold.
- All predictions can be obtained in Stata using the command `standsurv`.
- The computational time of parametric bootstrap can be decreased significantly by applying M-estimation methods instead.

Selected References I

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

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

Sjölander A.
Regression standardization with the R package stdReg.
European Journal of Epidemiology, 31(6):563-574, 2016.

VanderWeele T.J.
Causal mediation analysis with survival data.
Epidemiology, 22(4):582-585, 2011.

Spika D., Rachet B., Bannon F., Woods L., Maringe C., Bonaventure A., et al.
Life tables for the concord-2 study. Available from:
http://csg.lshtm.ac.uk/life-tables.