Disentangling health disparities using an extension of mediation analysis to the relative survival framework

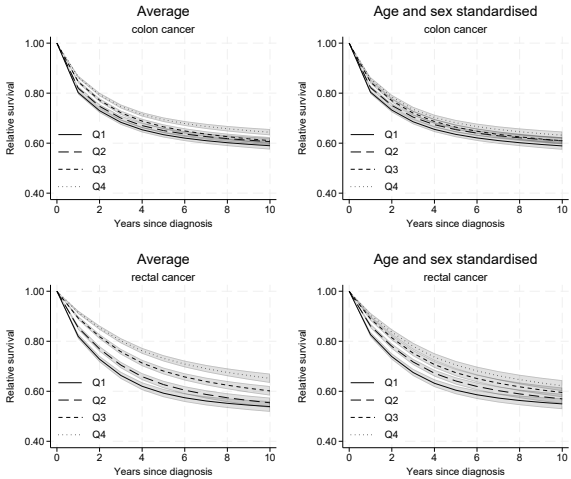
Elisavet Syriopoulou

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet

Email: elisavet.syriopoulou@ki.se

12th June 2024

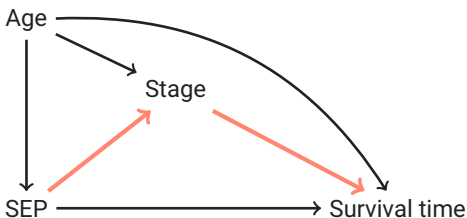
Motivation - Example, Sweden



Syriopoulou E, Osterman E, Miething A, Nordenvall C, Andersson TM-L. Income disparities in loss in life expectancy after colon and rectal cancers: a Swedish register-based study. *J Epidemiol Community Health* 2024;78(6): 402–408.

Investigating other factors

Could stage at diagnosis partly explain the survival differences between the least and most deprived groups?



This is a mediation analysis question!

Mediation analysis

Mediation analysis allows to explore the role of a mediator on an observed association between an exposure - outcome of interest.

Mediation analysis

Mediation analysis allows to explore the role of a mediator on an observed association between an exposure - outcome of interest.

However, we still need to deal with the complex mechanisms that contribute towards cancer disparities:

- Cancer-related factors
- Other cause factors

Relative survival framework

Relative survival allows comparisons of cancer survival between populations without distortions from differential background mortality.

Relative survival framework

Relative survival allows comparisons of cancer survival between populations without distortions from differential background mortality.

We have adapted a formal causal framework to the settings of cancer registry-based epidemiology, extending mediation analysis methods to the relative survival framework[1].

- Main idea: using the relative survival framework allows to isolate cancer-related factors.

[1] Syriopoulou E, Rutherford MJ, Lambert PC. Understanding disparities in cancer prognosis: An extension of mediation analysis to the relative survival framework. *Biometrical Journal*. 2021; 63: 341–353.

Excess mortality rate and relative survival

Excess mortality rate

$$\text{excess mortality} = \text{all-cause mortality} - \text{expected mortality}$$

The survival analog of excess mortality is relative survival.

Relative survival

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

$$R(t) = \frac{S(t)}{S^*(t)} \quad \text{and} \quad S(t) = S^*(t)R(t)$$

Notation

- X denotes the exposure of interest (here SEP), with $x = 1$ if exposed and $x = 0$ if unexposed
- M denotes the mediator of interest (here stage)
- $\boldsymbol{Z}$ denotes the confounding variables (here age)

Notation - potential outcomes

The framework used for formulating the effects of interest is that of potential outcomes: *the outcomes that would be observed if we intervene on X and M to set them on specific values*.

Notation - potential outcomes

The framework used for formulating the effects of interest is that of potential outcomes: *the outcomes that would be observed if we intervene on X and M to set them on specific values*.

- $R^x(t)$ is the value that $R(t)$ would have if we intervened on X and set it (possibly counter to fact) to the value x
- M^x is the value that M would take if we intervened on X and set it to x
- $R^{x,M^{x^*}}(t)$ is the the value that $R(t)$ would take if we intervened on X and set it to x and simultaneously intervened on M and set it to M^{x^*}, where x and x^* are not necessarily the same.

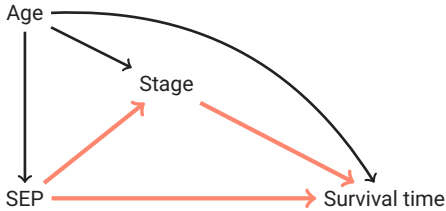
Notation - potential outcomes

The framework used for formulating the effects of interest is that of potential outcomes: *the outcomes that would be observed if we intervene on X and M to set them on specific values*.

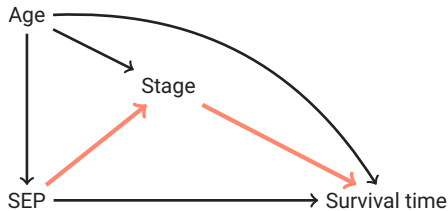
- $R^x(t)$ is the value that $R(t)$ would have if we intervened on X and set it (possibly counter to fact) to the value x
- M^x is the value that M would take if we intervened on X and set it to x
- $R^{x,M^{x^*}}(t)$ is the the value that $R(t)$ would take if we intervened on X and set it to x and simultaneously intervened on M and set it to M^{x^*}, where x and x^* are not necessarily the same.

We will use those to define contrasts of marginal effects of the potential outcomes.

Partitioning the total survival difference



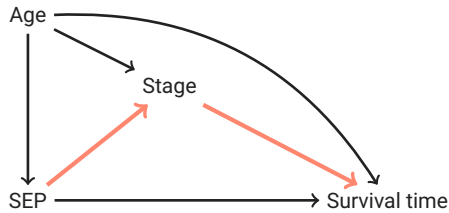
Partitioning the total survival difference



Natural indirect effect: quantifies how much of the observed difference is due to stage differences in the two groups

$$\text{NIE}(t) = R^{1,M^1}(t) - R^{1,M^0}(t) = E[R(t|X=1, \boldsymbol{Z}, M^1)] - E[R(t|X=1, \boldsymbol{Z}, M^0)]$$

Partitioning the total survival difference



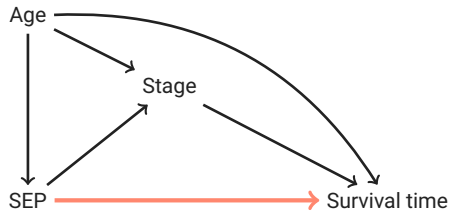
Natural indirect effect: quantifies how much of the observed difference is due to stage differences in the two groups

$$\text{NIE}(t) = R^{1,M^1}(t) - R^{1,M^0}(t) = E[R(t|X=1, \boldsymbol{Z}, M^1)] - E[R(t|X=1, \boldsymbol{Z}, M^0)]$$

This is a comparison of two hypothetical quantities:

- In the first, M is set to M^1 and in the second M is set to M^0, for everyone. In both, X is set to 1.
- X influence $R(t)$ only through its influence on M.
- Thus corresponds to an indirect effect through M.

Partitioning the total survival difference



Natural indirect effect: quantifies how much of the observed difference is due to stage differences in the two groups

$$\text{NIE}(t) = R^{1,M^1}(t) - R^{1,M^0}(t) = E[R(t|X=1,\boldsymbol{Z},M^1)] - E[R(t|X=1,\boldsymbol{Z},M^0)]$$

Natural direct effect: quantifies the differences in relative survival that are *not* due to stage differences

$$\text{NDE}(t) = R^{1,M^0}(t) - R^{0,M^0}(t) = E[R(t|X=1,\boldsymbol{Z},M^0)] - E[R(t|X=0,\boldsymbol{Z},M^0)]$$

Comment

The definitions involve the following term:

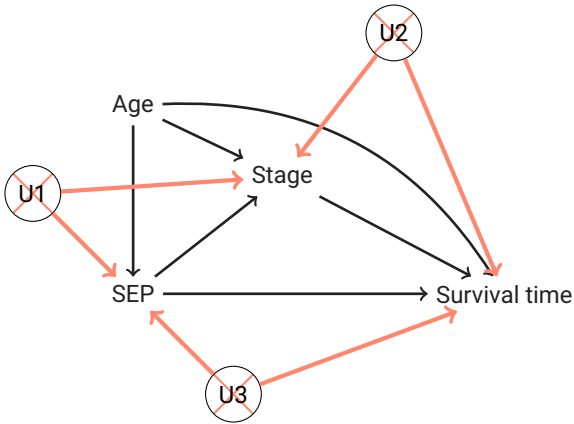
$$R(t|X = 1, \boldsymbol{Z}, M^0)$$

i.e., the relative survival if setting the exposure to the level of the exposed and the mediator to the mediator value if unexposed!

An additional model is required for the mediator.

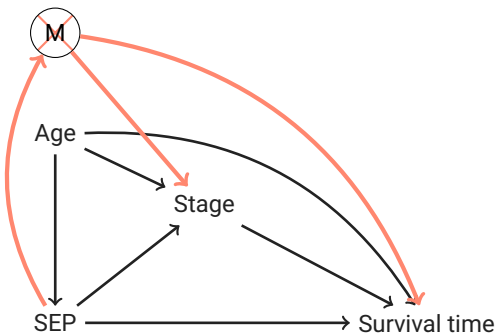
Assumptions for identification

To link the hypothetical quantities with the observed data, we need to assume no interference, consistency and conditional exchangeability.



Assumptions for identification

We also need to make the cross-world independence assumption that implies that there are no intermediate confounders M.



Estimation - regression standardisation

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

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

Identification algorithm

1. Fit a relative survival model including X, M, $\boldsymbol{Z}$.
2. Fit model for the mediator including X, $\mathbf{Z}$.
3. For each individual in the study population obtain estimates for $\widehat{P}(M = m|X = x, \boldsymbol{Z} = \boldsymbol{z_i})$, at each $X = x$.
4. Obtain estimates of standardised $\widehat{R}(t|X = x, \boldsymbol{Z} = \boldsymbol{z_i}, M = m)$ at $X = x$, using the predictions of Step 3 as weights. Form contrasts to obtain the $\widehat{NDE_{RS}}$ and $\widehat{NIE_{RS}}$.

Identification algorithm

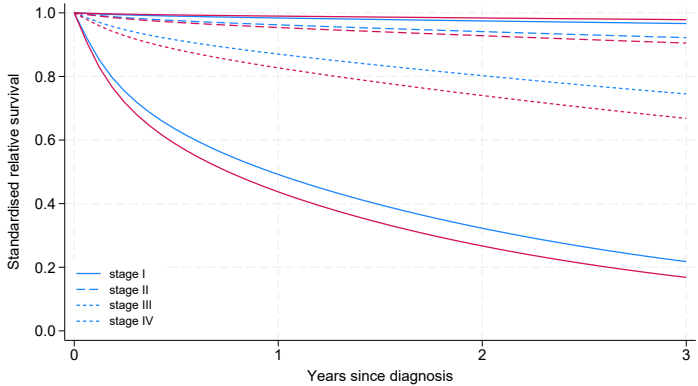
1. Fit a relative survival model including X, M, $\boldsymbol{Z}$.
2. Fit model for the mediator including X, $\mathbf{Z}$.
3. For each individual in the study population obtain estimates for $\widehat{P}(M = m|X = x, \boldsymbol{Z} = \boldsymbol{z_i})$, at each $X = x$.
4. Obtain estimates of standardised $\widehat{R}(t|X = x, \boldsymbol{Z} = \boldsymbol{z_i}, M = m)$ at $X = x$, using the predictions of Step 3 as weights. Form contrasts to obtain the $\widehat{NDE_{RS}}$ and $\widehat{NIE_{RS}}$.
5. Repeat from Step 3, k times, while performing parametric bootstrap for the parameter estimates for both models.
6. 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 in Stata

There is an example available on this on GitHub:
`https://github.com/syriop-elisa/mediation-example-stpm3`

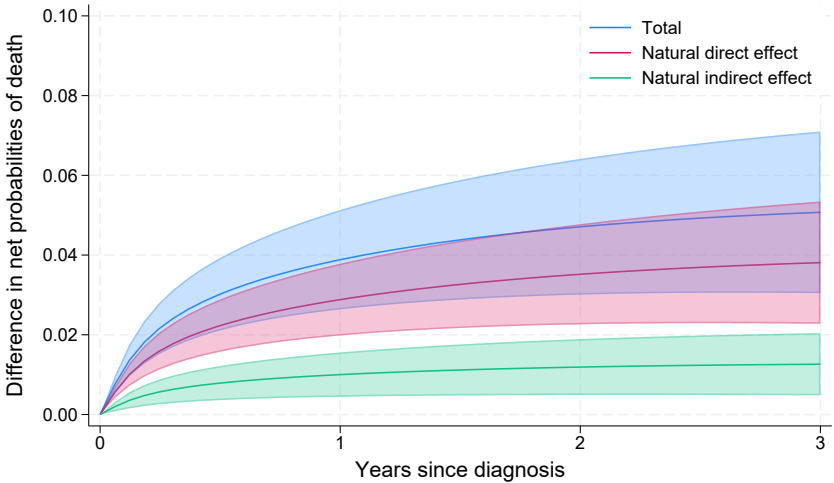
Example - simulated colon cancer data

I will use an example of simulated colon cancer data that is available with the methodological paper.



Blue lines are high and red lines are low socioeconomic position

Plot - with CIs



Standardising among subsets

- We can also define contrasts withing specific subsets such as the NDE and NIE among the explosed (e.g. among the most deprived)

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

- For the estimation, we restrict standardisation to a specific subset of the population.

Moving to a real-world setting

- So far, we only talked about a net setting that requires elimination of competing events and looked at
  - differences in relative survival
- Differences can also be quantified in a real-word setting where no elimination of competing events is required
  - difference in all-cause survival
  - avoidable deaths
- To do so, we need to incorporate the expected mortality rates, $S^*(t)$, in the contrast of interest.
  - There are many different ways to do it.

Direct & indirect effects - all cause setting

- Use the observed distribution of the exposure for $S^*(t)$:

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

Differences in all-cause survival can only be due to the cancer of interest (not due to other-cause differences).

- The difference in all-cause survival between the two SEP groups if we could intervene and shift the stage distribution of the lowest SEP group to that of the highest SEP group, while keeping their background mortality unchanged?

Direct & indirect effects - all cause setting

- Use the observed distribution of the exposure for $S^*(t)$:

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

Differences in all-cause survival can only be due to the cancer of interest (not due to other-cause differences).

- The difference in all-cause survival between the two SEP groups if we could intervene and shift the stage distribution of the lowest SEP group to that of the highest SEP group, while keeping their background mortality unchanged?

A conceptually similar measure has recently been proposed for the standard survival setting, so called separable effects:

Stensrud et al. Separable effects for causal inference in the presence of competing events. J Am Stat Assoc 2022.

Avoidable deaths under hypothetical interventions

How many deaths would be postponed for the lowest SEP group under an intervention that aims to shift the distribution of stage at diagnosis for the lowest SEP to that of the highest SEP group, while keeping other cause mortality rates unchanged/constant?

Avoidable deaths

- The predicted number of deaths for the exposed $X = 1$ in a typical calendar year with N^* diagnoses:

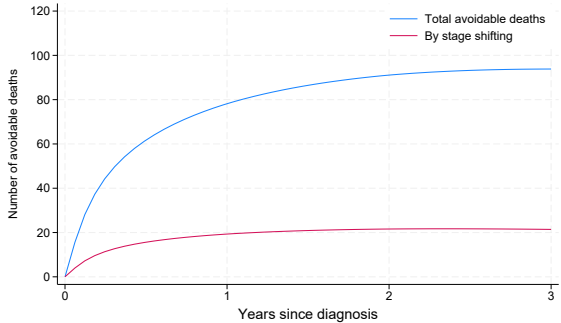
$$N^* \times \left(1 - E\left[S^*(t|X, \boldsymbol{Z}^{*X=1}) R(t|X = 1, \boldsymbol{Z}^{X=1}, M^1)\right]\right)$$

- The expected number of deaths under the intervention of shifting the mediator distribution of the exposed to the one of the unexposed (setting M to M^0):

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

- The avoidable deaths is given by their difference.

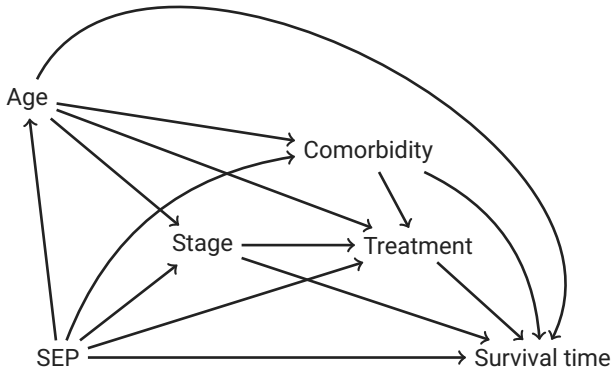
Avoidable deaths for colon cancer



- For total ADs, we shifted the relative survival and stage at diagnosis distribution of the lowest SEP to that of the highest SEP.
- For both scenarios, we kept the expected survival of the lowest SEP unchanged.
- 3 years after diagnosis there would be approx. 94 avoidable deaths in total, out of 2170 patients from the lowest SEP diagnosed in 2013 the most recent year in our cohort study.
- Partitioning that further, we found that approx. 22 deaths of the total deaths would be from eliminating stage differences and the remaining 72 would be from removing the remaining relative survival differences.

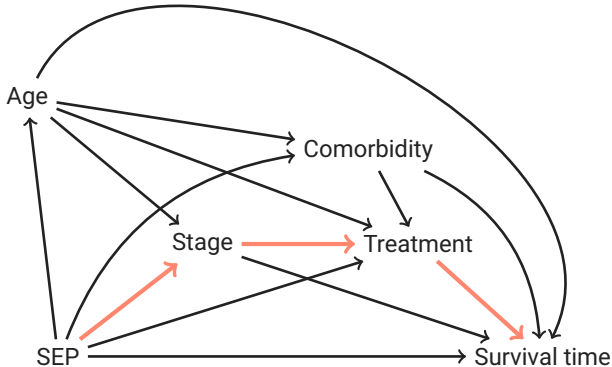
It will get trickier!

There will be multiple mediators (stage, treatment, comorbidity).



It will get trickier!

There will be multiple mediators (stage, treatment, comorbidity).



We are working to extend randomised interventional analogues of the NDE and the NIE into the relative survival framework.

Vansteelandt S, Daniel RM. Interventional Effects for Mediation Analysis with Multiple Mediators. Epidemiology 2017.

Conclusions

- There are differences in the prognosis of cancer patients.

Conclusions

- There are differences in the prognosis of cancer patients.
- Understanding mechanisms driving disparities is important.

Conclusions

- There are differences in the prognosis of cancer patients.
- Understanding mechanisms driving disparities is important.
- Understanding mechanisms driving disparities is difficult!

Conclusions

- There are differences in the prognosis of cancer patients.
- Understanding mechanisms driving disparities is important.
- Understanding mechanisms driving disparities is difficult!
- Causal mediation analysis using relative survival can be a valuable tool for exploring such settings.