Evaluating mediator interventions for time-to-event outcomes: A causal framework for cancer disparities

Elisavet Syriopoulou

Department of Medical Epidemiology and Biostatistics, Karolinska Institutet

Email: elisavet.syriopoulou@ki.se

10th March 2026

Outline

- About me
- Motivation and background
- Causal mediation analysis
- Future/other work

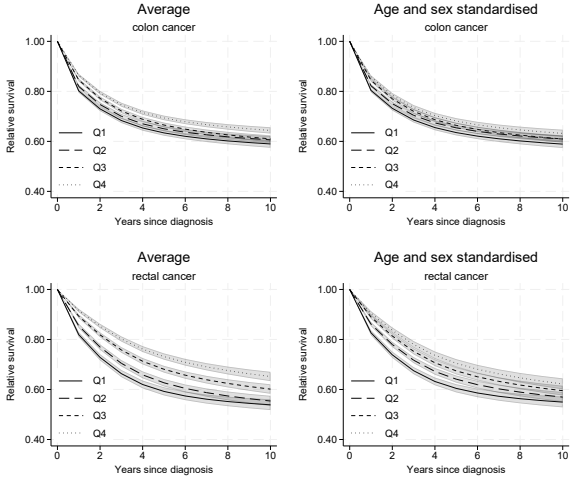
About me

- I have been at MEB for the past 6 years and currently hold a position as Assistant Professor
- My background is in biostatistics, with a focus on survival analysis applied to cancer-registry data
- I am in the process of establishing my own research group:
  - Jonatan Hedberg, PhD student
  - Qiwen He, MSc student
  - Recruitment underway for an additional PhD student
- I am also co-supervisor for two PhD students
- I am a member of the European Special Interest Group in Causal Inference, supported by Statisticians in the Pharmaceutical Industry (PSI) and the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI)

Studying colorectal cancer prognosis using CRCBaSe

- I have been using CRCBaSe to investigate variation in colorectal cancer (CRC) outcomes
- CRCBaSe links the Swedish Colorectal Cancer Registry (SCRCR) with registries at the National Board of Health and Welfare and Statistics Sweden
- It includes information on patient and tumor characteristics, diagnostics, treatment, and follow-up
- Linkage to the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA) provides information on education and income
- For each CRC patient, six comparators from the general population were matched on birth year, sex, year of diagnosis, and county

Differences in survival after colorectal cancer (CRC)[1]

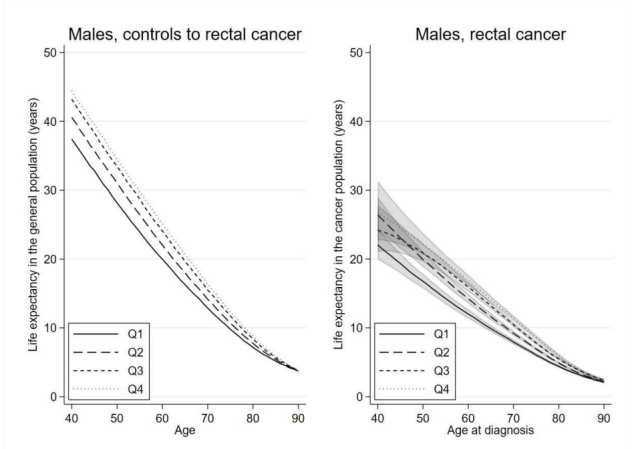


[1]Syriopoulou E, Osterman E, Miething A, et al. Income disparities in loss in life expectancy after colon and rectal cancers: a Swedish register-based study. *J Epidemiol Community Health* 2024.

Differences in life expectancy after colon cancer (CRC)[1]

Measure	Q1	Q4
Average estimates		
Life expectancy in the general population	13.64	17.83
Life expectancy in the cancer population	8.81 (8.61 to 9.01)	12.10 (11.88 to 12.32)
Loss in life expectancy	4.83 (4.63 to 5.03)	5.73 (5.51 to 5.94)
Proportion of life lost	35% (34 to 37)	32% (31 to 33)
Total years lost	2714	9320
Age-standardised and sex-standardised estimates		
Life expectancy in the general population	13.80	16.35
Life expectancy in the cancer population	8.86 (8.67 to 9.06)	11.05 (10.83 to 11.28)
Loss in life expectancy	4.94 (4.74 to 5.13)	5.30 (5.08 to 5.53)
Proportion of life lost	36% (34 to 37%)	32% (31 to 34%)
Total years lost	2775	8628

Differences in life expectancy after rectal cancer (CRC)[1]



Does the indicator of socioeconomic position matter?[2]

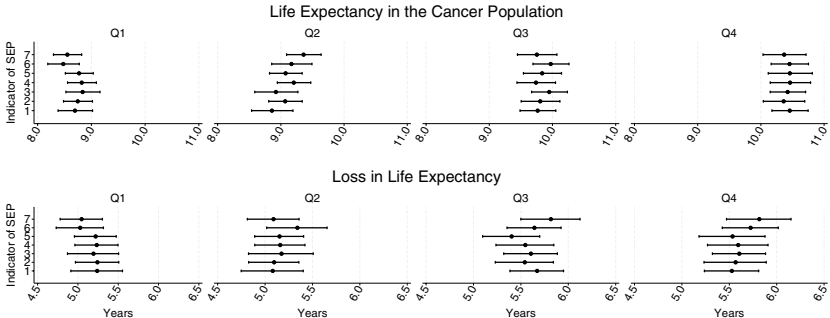
- We explored how different education- and income-based indicators and conceptualisations of SEP influence the estimates of relative survival or life expectancy measures, using data on colon cancer in Sweden.
- For income, we used either:
  - individual disposable income (IDI) or
  - family disposable income per consumption unit (as the sum of the disposable income of all family members divided with their consumption weight) (PHI)
- Key idea: females and older individuals are misclassified to a lower SEP under certain SEP definitions

[2]Syriopoulou E, Miething A, Osterman E, Nordenvall C, Andersson TM-L. Loss in life expectancy after a colon cancer diagnosis by socioeconomic group: does the indicator of socioeconomic position matter? *BMC Public Health* 2025.

Does the indicator of socioeconomic position matter?[2]

- IDI groups using cut-offs based on quartiles using the whole population of controls (overall)
- IDI groups using cut-offs based on quartiles created separately for patients below and above 65 years old at diagnosis
- IDI groups using cut-offs based on quartiles created separately within age-groups (5-year groups apart from <30 years and 90+ years)
- IDI groups using cut-offs based on quartiles created separately by sex
- IDI groups using cut-offs based on quartiles created separately by sex and the 65 years of age cut-off
- PHI groups using cut-offs based on quartiles created using everyone (overall)
- PHI groups using cut-offs based on quartiles created separately for patients above and below 65 years old

Does the indicator of socioeconomic position matter?[2]



1=IDI (overall), 1=IDI (by age cut-off), 3=IDI (by sex), 4=IDI (by age cut-off sex), 5=IDI (by age-groups), 6=PHI (overall), 7=PHI (by age cut-off)

Does the indicator of socioeconomic position matter?[2]

Larger differences were observed for age and sex specific estimates.

For example, for 60-year-old males:

Indicator for SEP	Q1	Q4	Difference Q4–Q1
Loss in life expectancy			
IDI (overall)	8.92 (8.08–9.69)	9.52 (8.86–10.14)	0.60
IDI (by age cut-off)	8.76 (8.05–9.43)	9.56 (8.81–10.27)	0.80
IDI (by sex)	8.82 (8.07–9.53)	9.82 (9.11–10.51)	1.00
IDI (by age cut-off & sex)	8.70 (8.06–9.30)	9.50 (8.70–10.26)	0.80
IDI (by age-groups)	8.79 (8.11–9.44)	9.58 (8.79–10.33)	0.79
PHI (overall)	8.38 (7.62–9.09)	9.64 (8.94–10.30)	1.25
PHI (by age cut-off)	8.43 (7.75–9.07)	9.92 (9.15–10.65)	1.49

Differences in how CRC patients are treated[3]

- Since 2016, almost all patients should be discussed at multidisciplinary meetings
- Are there differences in treatment by income groups?
- Patients in the highest income quartile had
  - a higher chance of being discussed at the preoperative and postoperative MDT
  - a lower risk of emergency surgery
  - a higher chance receiving neoadjuvant and adjuvant treatment
- Despite MDT discussion, differences in treatment remained (when analysing those diagnosed after 2016 separately)

[3]Osterman E, Syriopoulou E, Martling A, Andersson TM-L, Nordenvall C. Despite multi-disciplinary team discussions the socioeconomic disparities persist in the oncological treatment of non-metastasized colorectal cancer. *Brit J Cancer* 2024.

Patients with a history of mental illness do worse[4]

- Are there differences on treatment and outcomes between patients with and without mental illness history?
- We found smaller differences than in earlier studies, however, patients with a history of mental illnesses still do worse
- Those with severe mental illness had more advanced tumours, more emergency surgeries, and less adjuvant treatment
- 5-year standardised survival was lower for patients with mild (64.6%) and severe (61.8%) mental illness than for those without (69.3%).
- Survival after recurrence was also poorer for patients with severe mental illness.

[4]Osterman E, Syriopoulou E, Martling A, Andersson TM-L, Nordenvall C. Mental illness and non-metastatic colorectal cancer treatment and survival: A nationwide study of almost 70,000 patients. *Acta Oncologica* 2025.

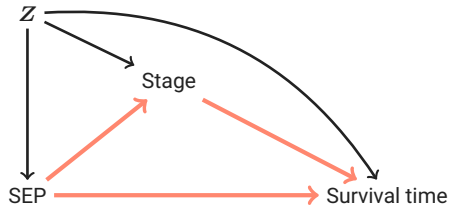
Understanding variation

- There are large differences in the outcomes of cancer patients
- How and why such differences arise?
- Is it possible to reduce/eliminate the differences?
- Can we identify intervention targets to inform policies?

Why is this important?

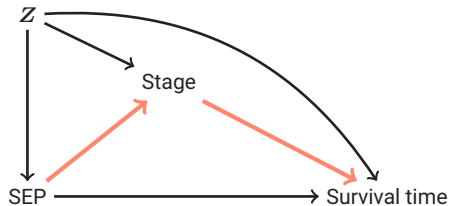
- Understanding how disparities arise is the essential first step to reducing them and improving outcomes for groups with poorer prognosis
- If stage at diagnosis explains much of the survival gap by income, we can promote earlier detection among lower-income groups (through community outreach, improved access to primary care, etc)
- If treatment differences and their long-term effects explain the income-related survival gap, we should ensure equitable treatment allocation and factor late effects into treatment decisions
- In practice this is a challenging task: we have lots of complex mechanisms to untangle, statistical challenges to address, and we are restricted by the data we have
- We can't be perfect but we can can still try to produce meaningful evidence for addressing disparities

Causal mediation analysis - starting simple



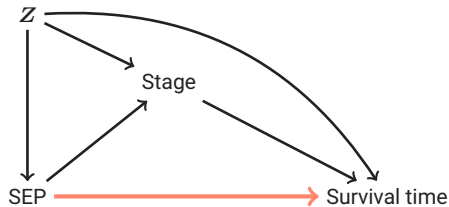
- The framework used for formulating the quantities of interest is that of *potential outcomes*: the outcomes that would be observed if we intervene on the exposure (SEP) and mediator (stage at diagnosis) to set them on specific values
- Causal inference is inherently about hypothetical, what-if questions
- Traditional mediation analysis has focused on the elusive task of discovering pathways

Causal mediation analysis - starting simple



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

Causal mediation analysis - starting simple



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

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

Previous work on causal mediation analysis

- Previously we adapted natural effects to the settings of cancer registry-based epidemiology, extending mediation analysis methods to the relative survival framework
- Main idea: using the relative survival framework allows to isolate cancer-related factors alone as opposed to looking at both cancer and other factors
- You can find an example of this on GitHub: `https://github.com/syriop-elisa/mediation-example-stpm3`


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

DOI: 10.1002/bimj.201900355

Biometrical Journal

RESEARCH PAPER

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

Natural vs interventional effects

- Natural effects have been criticised for their reliance on empirically unverifiable assumptions
- They require untestable assumptions in the common situation of exposure-induced mediator–outcome confounding (e.g. SEP affects readiness to seek diagnosis which in turn has an effect both on stage at diagnosis and survival outcomes)
- Randomised interventional analogues of the natural direct effects have been proposed as alternatives that can be identified under weaker assumptions

Natural vs interventional effects

- Natural effects are defined in terms of individual-level interventions
  - What would happen if *for each patient* in the most deprived group we would set their stage at diagnosis to what it would be if they were in the least deprived group?
  - Here we fix the mediator at a counterfactual value
- Interventional effects rely on population-level interventions
  - What would happen if we *shift the mediator distribution* in the population to that of least deprived?
  - Here we set the mediator for each individual to a random draw from the distribution of the mediator under an exposure level

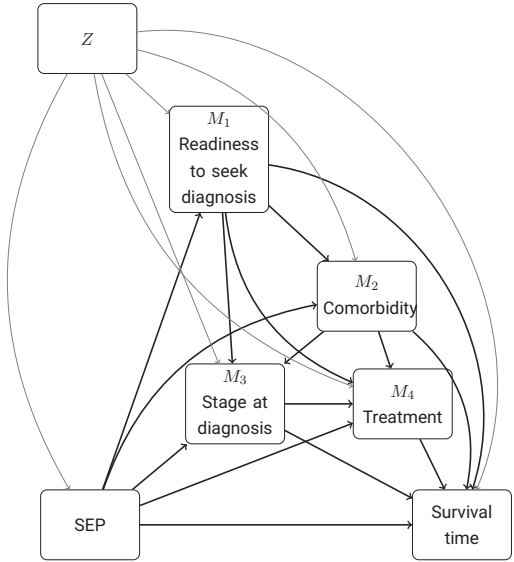
Interventional effects

More recently, another paper introduced interventional effects that map target trials and focus on evaluating shifts in mediator distributions[5]

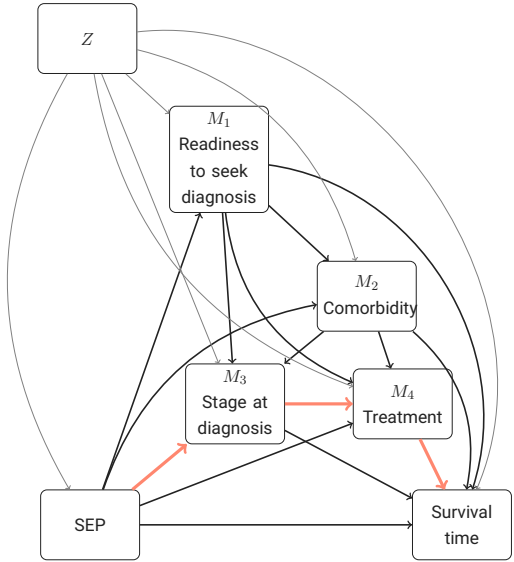
- Emphasis on selecting policy-relevant questions related to mediator-shifting interventions
- Relevant when there are no-well defined interventions
- In the context of my example, the available data do not permit direct assessment of an intervention targeted at shifting the distribution of stage at diagnosis
  - There are many potential interventions for improving stage at diagnosis (e.g. awareness campaigns, invitation letter to screening followed by reminders, etc) and each one could lead to very different conclusions regarding causal effects

[5]Moreno-Betancur et al. Stat Methods Med Res 2021.

Moving to more realistic settings



Moving to more realistic settings



Answering relevant questions

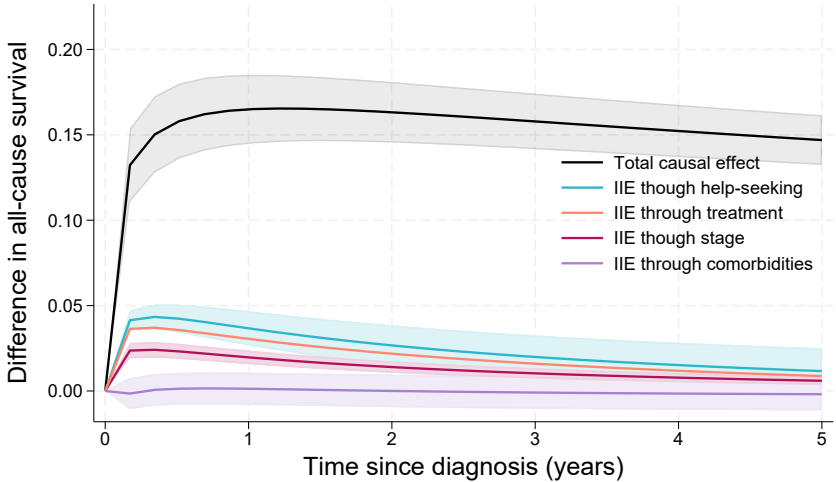
- Move from single-mediator to multi-mediator settings things get complex!
- Intervening on one mediator implicitly changes the others, yet this is typically ignored
- If we “intervene” on stage we implicitly altering downstream mediators like treatment
- Ignoring this creates a hypothetical world that cannot exist
- Interventional effects shift the distribution of a mediator while letting the rest of the mediators “respond”

Question 1 - one policy premise

If targeting only one mediator, which intervention would provide the “biggest gain”, in terms of reducing disparities between deprivation groups?

1. Intervention B_k would be applied independently of the other mediators
2. Intervention B_k would shift the distribution of mediator M_k to what it would be in the unexposed given $\boldsymbol{Z}$. This is equivalent to setting M_k to a random draw from the distribution it would have *under no exposure* given $\boldsymbol{Z}$.
3. Intervention B_k would sever the dependence on average between M_k and the other mediators, so that the joint distribution of the other mediators is held at what it would be *under exposure* given $\boldsymbol{Z}$ (worst-case scenario)

Example

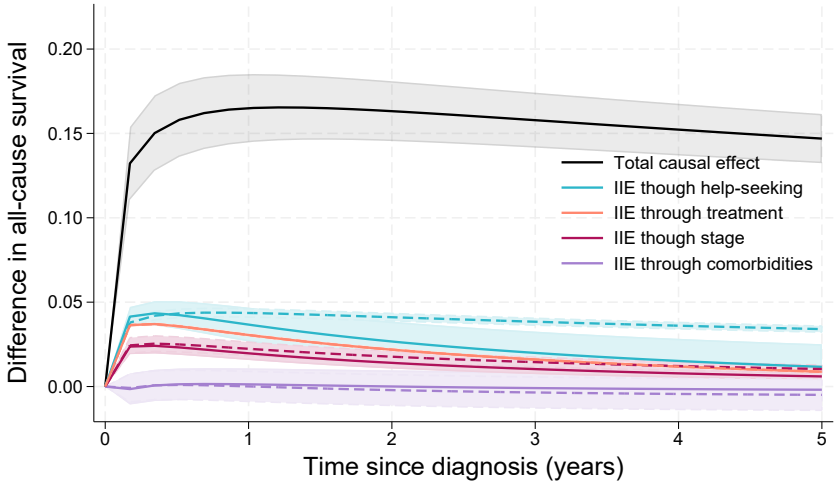

Difference in all-cause survival
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
Time since diagnosis (years)
Total causal effect
IIE though help-seeking
IIE through treatment
IIE though stage
IIE through comorbidities

Question 1 - one policy premise

Relaxing assumption 3 and replacing with:

a. The order of the mediators is $M_1, \ldots M_K$
b. Under intervention B_k, the joint distribution of the causally antecedent mediators of M_k is unaffected, remaining at what it would be under exposure
c. Under intervention B_k, the *conditional* joint distribution of the causally descendent mediators of M_k given $M_1, \ldots M_k$ and $\boldsymbol{Z}$ is what it would be under exposure

Example

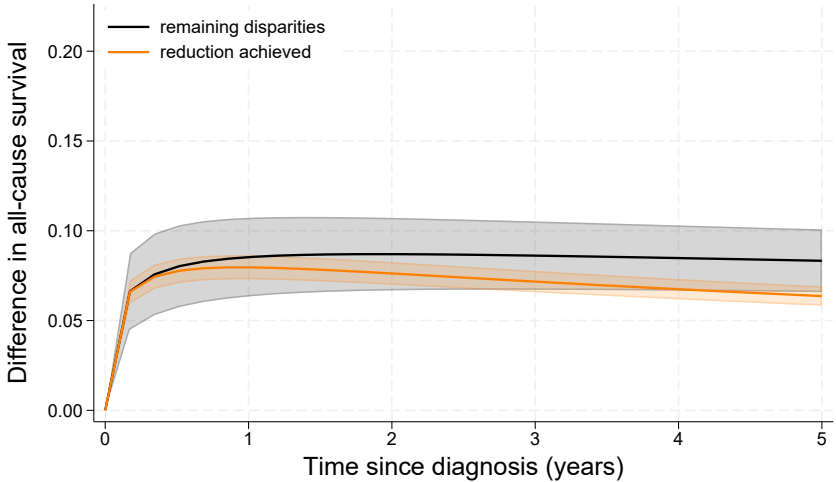

Difference in all-cause survival
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
Time since diagnosis (years)
Total causal effect
IIE though help-seeking
IIE through treatment
IIE though stage
IIE through comorbidities

Question 2 - joint intervention

What would be the remaining disparities between exposure groups if it was possible to jointly target all the mediators?

- B_{all} shifts the joint distribution of the mediators to be as in the unexposed (including the mediators interdependence) given Z
- We consider an intervention where the exposed group is set to be exactly like the unexposed in terms of the joint mediator distribution
- Large remaining disparities after this intervention would suggest that there is a need to investigate additional intermediate processes

Example

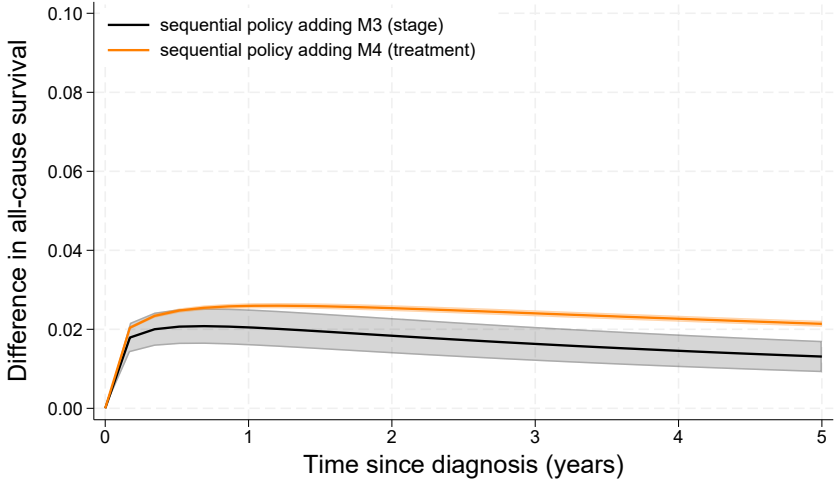

remaining disparities
reduction achieved
Difference in all-cause survival
0.00
0.05
0.10
0.15
0.20
0
1
2
3
4
5
Time since diagnosis (years)

Question 3 - sequential policies

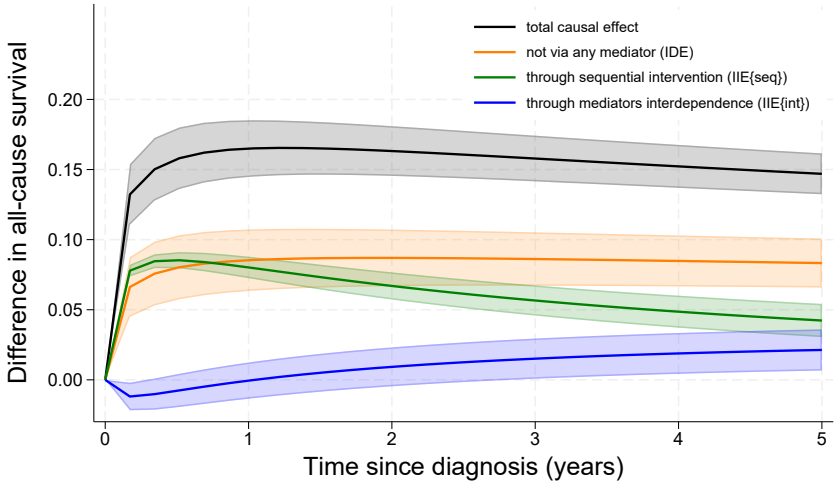
What would be the benefit of sequential policies, applying the separate mediator interventions under Question 1 scenario 1 sequentially?

- Let $B_{\{k\}}$ denote an intervention applying all interventions in the sequence $B_1, \ldots, B_K$ up to B_k
- $B_{\{k\}}$ shifts the distribution of each mediator to what it would be in the unexposed given $\boldsymbol{Z}$, independently of other mediators and severing the dependence on average from the subsequent ones in the sequence

Example


sequential policy adding M3 (stage)
sequential policy adding M4 (treatment)
Difference in all-cause survival
0.00
0.02
0.04
0.06
0.08
0.10
0
1
2
3
4
5
Time since diagnosis (years)

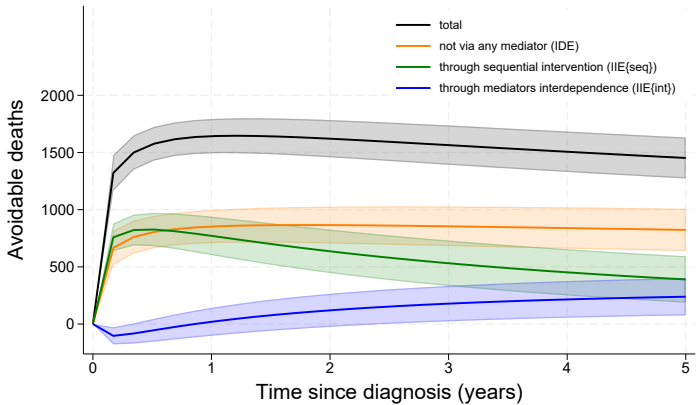
Example


total causal effect
not via any mediator (IDE)
through sequential intervention (IIE{seq})
through mediators interdependence (IIE{int})
Difference in all-cause survival
0.00
0.05
0.10
0.15
0.20
0
1
2
3
4
5
Time since diagnosis (years)

Alternative metrics

- The impact of such interventions can also be visualised in terms of the avoidable/postponables deaths
- The avoidable deaths is a time-specific measure as eventually all deaths will be realised
- Might be easier to communicate
- For example, what is the number of avoidable deaths under the “one policy premise” intervention applied only on the most deprived patients?

Example


total
not via any mediator (IDE)
through sequential intervention (IIE{seq})
through mediators interdependence (IIE{int})
Avoidable deaths
2000
1500
1000
500
0
0
1
2
3
4
5
Time since diagnosis (years)

Extensions to relative survival

- So far, I have mainly focus on differences in all-cause survival probabilities
- All-cause survival differences among cancer patients is the result of complex mechanisms that involve both cancer-related and other-cause factors
- It may easier to focus on cancer-related survival differences (especially useful for comparing subgroups with different background mortality)
- Using the relative survival, we can write all-cause survival as the product of expected and relative survival:

$$S(t) = S^*(t)R(t)$$

- This allow us to consider interventions that target only relative survival differences (cancer-specific)
- Everything we talked about can also be obtained in terms of relative survival differences

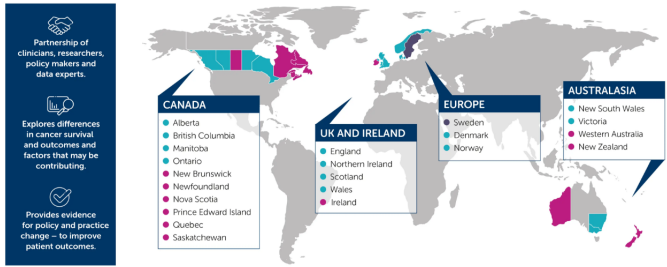
Extensions to relative survival

- When focusing on all-cause survival differences (using relative survival differences), we assume that the intervention has no impact on other cause mortality rates (i.e. other cause mortality remains unchanged after such intervention)
- The two competing events are assumed to be conditionally independent
- Even when referring to an all-cause setting, it is assumed that any potential impact of the intervention on survival is due to changes in cancer mortality

Extensions to relative survival

- The motivation for this distinction is that the effect of certain interventions would be separable across the two competing causes; for instance, increased cancer screening engagement would likely impact only on the mortality for the cancer cause specifically, whilst only indirectly impacting on other-cause death probabilities but not the underlying other-cause mortality rates
- This assumption is easy to justify for some mediators but not others
- For instance, a hypothetical intervention targeting comorbidities would likely impact also the non-cancer mortality
- So there is lots more to consider still about how to go ahead with this

International Cancer Benchmarking Partnership (ICBP) project

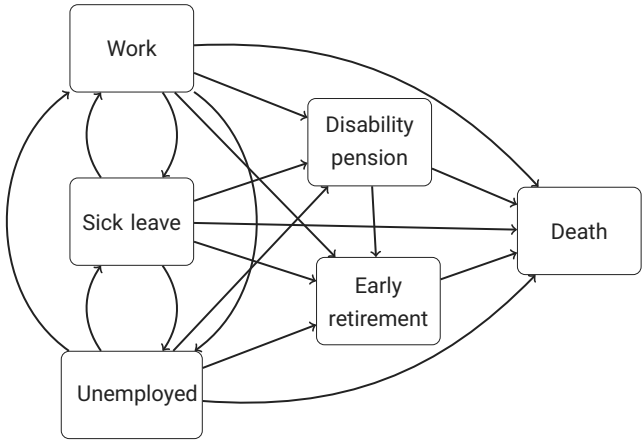


- This project involves an international comparison of cancer outcomes
- I collaborate closely with researchers from University College London (UCL) and the International Agency for Research on Cancer (IARC)
- A key component of the phase 3 work is to understand how variations in diagnostic pathways and access to treatment contribute to survival differences
- We plan to incorporate mediation analysis to explore these relationships

Exploring work-related outcomes after a cancer diagnosis

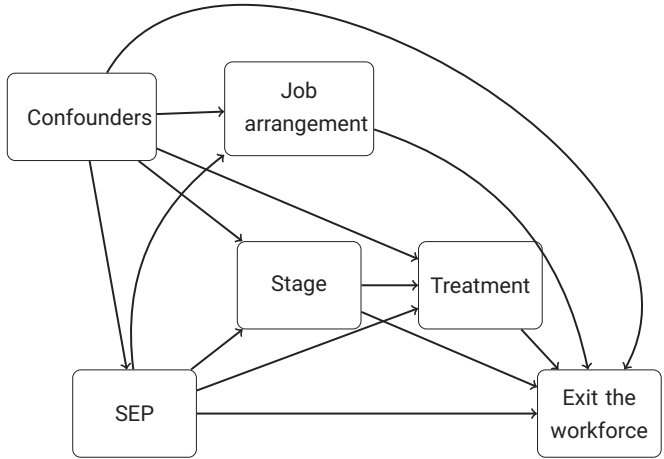
- Although survival differences among cancer patients are well-documented, we know little about outcomes related to returning to work after sick leave
- Most studies have only looked at specific time points and the lifetime perspective is limited
- I plan to explore such variations for colorectal cancer using data from CRCBaSe and focus on metrics such as working years lost and excess number of recurrent events (e.g. sick leave)

Explore trajectories with multi-state modelling



I plan to quantify measures such as loss in working years and excess number of recurrent events.

Apply causal mediation analysis



This also involves methods development for mediation analysis using pseudo-observations.

Other work: missing data

- Often we have a small proportion of missing values for variables such as stage at diagnosis
- We usually handle this with multiple imputation (MI)
- For MI to work well, the imputation model and the analysis model should rely on the same assumptions (i.e., be compatible)
- My work looks at several open questions around MI (part of Jonatan's PhD):
  - How well does MI behave with time-to-event data, especially when hazards aren't proportional?
  - There's been very little testing of MI in the relative survival setting
  - Stage I–IV is only a proxy for TNM staging—so should we be imputing stage I–IV or TNM?

Other work: alternative ways to quantify cancer prognosis

- I have been looking at different ways to describe cancer prognosis, not just the usual survival probabilities
- One idea is to estimate how many years of life, on average, someone loses after being diagnosed
- To do this, we often need to extrapolate survival beyond the data we actually have
- Qiwen is working on a project comparing two related measures
  - *loss in life expectancy (LLE)*
  - *years of life lost (YLL)*

  which aim to capture the same quantity but rely on different assumptions

Thank you