Dissecting the causal question underlying the association between cancer and ADRD

L. Paloma Rojas-Saunero

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.0      v dplyr   1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1


## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Motivation

Several observational studies have found that individuals with history or incident cancer that survive over follow-up have a lower risk of Alzheimer´s disease or related dementias (ADRD) when compared to individuals with no history of cancer. Multiple mechanisms related to inflammatory response, immune response, etc etc are postulated to understand this phenomenon. However, the community has raised their concerned about the potential unmeasured confounding, selective survival and misclassification issues that could drive this results.

In the lasts few years there has been an important attention to unveiling the impact of death as a competing event, and the role of survival bias to explain this association. Previous studies alluded to choosing a particular summary measure or modeling approach to account for death (incidence rate ratios (ref), hazard ratios (ref), multistate models (ref)), though the underlying estimand remains unclear.

Overall, although observational studies have measured history of cancer and incident cancer as the key exposure, researchers refer to biological mechanisms which onset may be prior to cancer diagnosis and the effect of cancer treatment.

Pin1 is an enzyme deeply involved in the pathogenesis of cancer and ADRD. One of the roles of Pin1 is regulate mitosis process and apoptosis. Over-expression of Pin1 has been found in several tumors and it's inhibition causes tumor regression. In contrast, since adult neuronal cells do not divide, Pin1 over-expression in brain tissue is used to restore tau and amyloid precursor protein to a functional shape, thus preventing the buildup of the toxic tau and beta-amyloid species associated with AD. For this reason Pin1 regulation is proposed as one of the mechanisms that can explain the inverse association between cancer and AD found in observational studies.

Bias cannot be defined or assessed if the causal question is not explicitly specified. To define the underlying causal question there are two core challenges. First, specifying cancer diagnosis as the proxy for an underlying biological mechanism we would ideally intervene upon.

Second, to rely on the independent censoring assumption in regards to the competing event of death when the inrestes is on the direct effect of cancer in dementia risk.

With these considerations in mind, and illustrated through directed acyclic graphs, we describe a case study using data from the Rotterdam Study, a population-based prospective cohort with over 20 years of follow up. Through this case study we describe how can we connect the causal question to the analytic decisions with real-world data and demonstrate how point estimates may vary substantially depending on the assumptions related to previous points.

How information on mortality and cause of death can provides insight about the direction of some sources of bias.

Our aim is to outline one way to tie the etiological question of Pin1 with the questions that can be answered with observational data where Pin1 measurements are not available. We aim to draw the complexity of the research question by building progressively a directed acyclic graph (DAG) which help us connect the causal question to the observable data and the assumptions we rely on to identify the effect. We describe the challenges and illustrate the assumptions we have to fulfill to identify the effect.

Overview of the causal structure and assumptions related to cancer diagnosis as a proxy for Pin1

Given the previous literature on the topic, we focus the interest in answering the following question: *What is the direct effect of Pin1 over-expression at age of 65 in the risk of ADRD by age 85 compared to regulated expression of Pin1 by age of 65.* To identify this effect with real-world data, we will illustrate the causal question through a directed acyclic graphs (DAGs), spelling out the data generating mechanisms and assumptions tied to the data.

The causal graph of Figure 1 illustrates a situation where Pin1, represented as P, was measured in all participants on their 65 birthday. Based on the measurement of Pin1, participant's exposure was classified as 1) Over-expressed Pin1 or 2) Regulated Pin1. Through out the follow-up each participant followed a strict monitoring for ADRD onset, represented as Y_{85}, and no participant was lost to follow-up nor died during the study period.

was not an exposure or intervention randomized at age 65, therefore we have to consider a set of measured covariates L for which conditional exchangeability holds. Previous studies have described age, sex, educational level and race/ethnicity as the minimal adjusting set of covariates (Ospina). However, environmental and behavioral factors such as smoking, alcohol intake, or chronic diseases which are known to cause microenvironmental changes such as inflammation and changes in tissue remodelling, may translate into Pin1 over-expression and are also related to the development of ADRD. Therefore, had we measured all L we could identify the direct of P in Y_{85} by conditioning on L.
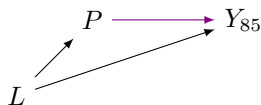

Figure 1: Direct effect of Pin1 in ADRD

Unfortunately, Pin1 expression measurement is not available in real-world settings, thus we can only rely on a proxy of this exposure. Since Pin1 over-expression is present in tumors and tumors are only measured through screening and diagnosis, we considered cancer diagnosis as the proxy for Pin1 over-expression, as most observational studies have defined (all refs). We depict this feature in Figure 2, where P^* represents *incident cancer diagnosis*, the measured proxy of P. In this DAG we colored the path $P^* \leftarrow P \rightarrow Y_{85}$ because although we are measuring the association between P^* and Y_85 in the observed data, we are assuming that the effect is only through P.

A major challenge related to assigning cancer diagnosis as the proxy of Pin1 is defining time zero. Not everyone who had Pin1 over-expression will be diagnosed with cancer by age 65. For this reason prospective cohort studies have considered cancer as a time-varying or time-dependent exposure. This means that a
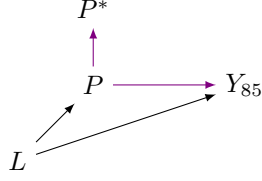
Figure 2: Direct effect of Pin1 in ADRD, with cancer diagnosis as proxy of Pin1

participant who had cancer diagnosis at age 70 contributes to the "regulated Pin1" from age 65 to 70, and from age 70 and onwards to the "Over-expressed Pin1" arm. Alternatively, studies that have included participants with cancer diagnosis at the time of the diagnosis (for example from cancer registries such as SEER) and matched participants by age. If age at cancer diagnosis is considered as the time zero, we should be careful of adjusting for post-baseline covariates of P or mediators between P and P^*.

Since the population is elderly, the risk of death at age 65 is not trivial and it increases over time. This means that some individuals may die prior to cancer diagnosis due to other leading causes of death at this age group, such as cardiovascular disease. As such, we can only measure P^* in the subset of individuals who have survived long enough to have a cancer diagnosis. We zoom in the relationship of $P \to P^*$ in Figure 3. In this figure we include $D = 0$ between P and P^*, this illustrates that to observe P^*, we condition (box around $D =$) on surviving long enough to have a cancer diagnosis.

Several risk factors that increase the risk of cancer might also be shared causes of death prior to cancer diagnosis, for example smoking may cause lung cancer and chronic obstructive pulmonary disease (a leading cause of death in this age group). Therefore, to isolate the effect of P in P^* (the violet arrow), we need to block all shared causes of Pin1 and death from other causes, and the shared causes of death from other causes and cancer diagnosis, both represented in Figure 3 as C_1 and C_2. In this way we assume a hypothetical scenario in which we could prevent death prior to cancer diagnosis, by conditioning on a rich set of covariates. This assumption must hold regardless of whether we use incident cancer as a time-varying exposure, or match cancer patients to participants free of cancer by age.
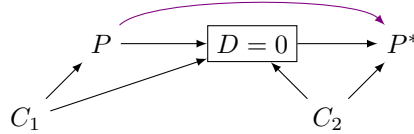

Figure 3: Zoom in the association between P and P star

Up to this point we have outlined the underlying assumptions of using incident cancer diagnosis as a proxy for Pin1 expression. We continue outlining the additional assumptions related to death as a competing event of ADRD.

Overview of the causal question and assumptions related to death as a competing event of ADRD

As illustrated in previous figures, we are interested in the direct effect of Pin1 in ADRD at 85 years, which means a length of follow-up of 20 years. However, mortality in elderly population between 65 and 85 is substancial, and a leading cause of death is cancer. This means that in order to measure ADRD by the age of 85, individuals have to had survive to cancer and to other causes. For this reason death is a competing event of ADRD because if a participant dies prior to ADRD diagnosis, death prevents from observing ADRD at future time-points.

We can visualize this data feature in the causal diagram in Figure 4. In this DAG we include D_{80} as a representation of Death by age 80, an arrow between P and D_{80} since P may increase the mortality risk

irrespective of cancer diagnosis. We also include an arrow between P^* and D_{80} since cancer diagnosis, and subsequent treatment (or lack of treatment) may have an effect on death. Last, the arrow from D_{80} and Y_{85} represents the key feature of a competing events data structure: an individual who dies by age 80 cannot subsequently develop ADRD, and since D_{80} and Y_{85} are events related to aging, C represent the shared causes of both events.

In this DAG we observe that, had we measured and adjusted for L, we could estimate the total effect of P in Y_{85} without further assumptions. However the total effect includes all pathways between P and Y_{85}. This means that if Pin1 has an effect on mortality through cancer or other mechanisms, we will observe a protective effect of Pin1 overexpression in ADRD, partially or fully mediated through death.

Alternatively, there are several ways to conceptualize the direct effect of P in Y_{85} as in Figure 4 (violet arrows) and several causal parameters (estimands) to represent it, such as: the controlled direct effect (CDE), the survivor average causal effect and the natural separable direct effect. Through this section we will discuss the causal question, assumptions related to death and interpretation of the CDE since this estimand connects to frequently used estimators.

THe CDE represents the effect of Pin1 in ADRD in a setting where we could have prevented death over the study period. It relies on the assumption that we have measured all C to block the pathway $Y85 \leftarrow C \rightarrow D_{80} \rightarrow P \rightarrow P*$. This assumption is defined as the independent censoring assumption conditional on covariates. In this setting death is treated as a censoring event and it can be interpreted as those who died would have the same risk of developing dementia if prevented from dying than those who remained alive and free of dementia at a given time point, conditional on shared causes of dementia and death.
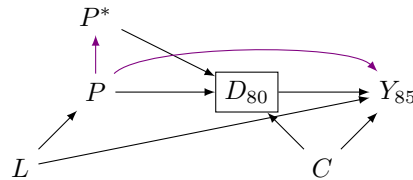

Figure 4: Effect of Pin1 in the risk of ADRD by age 85, with cancer diagnosis as proxy of Pin1

Motivating example with the RS data

In this section we illustrate how previous DAGS may reflect on the analysis performed to unveil the effect of Pin1 in all cause-dementia. We use data collected in the Rotterdam Study, a population-based prospective cohort study among persons living in the Ommoord district in Rotterdam, the Netherlands. Recruitment and initial assessments were held between 1990 and 1993, participants had follow-up visits between 1993-1995, 1997-1999, 2002-2005, and 2008-2010. All participants had data on incident cancer diagnosis and incident dementia diagnosis through follow-up, collected from medical records of general practitioners (including hospital discharge letters) and through linkage with national registries. Date and cause of death was collected on a weekly basis via municipal population registries. Data on clinical outcomes was available until 2015.

To match the analysis to our initial question, we considered as inclusion criteria being between 60 and 70 years old at study entry, without history of cancer diagnosis, free of cognitive decline or with previous history of dementia. Out of 7983 persons who participated at baseline, 3642 were considered eligible. Time to cancer diagnosis, to dementia diagnosis and death status was measured for all participants. All participants were followed from study entry until dementia diagnosis, death or 20 years after their individual baseline date, whichever occurred first.

Participants had a mean age of 64.46 (SD: 2.86), and 54% (n = 1981) were women. Further details on participants are presented in Table 1. Over follow up, 24% (n = 878) developed cancer and 76% (n = 2764) remain free of cancer diagnosis, the median age of cancer diagnosis was 73 (IQR: 69-77). From the total sample, 12% (n = 431) had dementia over follow-up and median time to dementia was 79 (IQR: 75-83).

Among participants with incident cancer, 6% (n = 50) had dementia diagnosis and 63% (n = 549) died over follow-up, 32% (n = 279) remain alive at 20 years since study entry. In contrast, among participants free of cancer diagnosis over follow-up, 14% (n = 385) were diagnosed with dementia and 23% (n = 624) died over follow-up, 63% (n = 1755) were alive at the end of follow-up.

We present the following analysis:

The difference in dementia risk treating cancer as ever vs. never (As if Pin1 and cancer dx. happened at the same time if looked retrospectively).

We begin illustrating the most simple scenario (as Figure 1), were Pin1 overexpression is defined as "ever vs. never" and we consider death as an independent censoring mechanism. We consider baseline covariates for adjustment such as age at study entry, sex, educational attainment, smoking status, alcohol intake status.

The difference in dementia risk considering cancer as a time-varying exposure independent of death. For example. at five years from follow-up the risk of cancer was 6.5 (95%CI: 5.8, 7.4) and the risk of death was 2.6 (95%CI: 2.2, 3.2), thus we assume that those who died would have the same risk of cancer had they not died, unconditionally.

Discussion

We need to figure out what the key aging-dependent changes are and how to modulate these factors safely

Causes such as aging, smoking, irradiation etc are causing micro environmental changes, like inflamation and changes in tissue remodelling. This promotes selection for adaptive mutations.

Efforts to prevent and treat cancer should converge with similar efforts to prevent other aging- associated diseases.

We need better means to

Youthful tissues

Somewhere

Knowing the cause of death provides information about the direction of missclassification. In patients with cancer we observed a slight % that died with dementia as a primary cause. On the other hand, among individuals free of cancer, we observed % of individuals who died with cancer as a primary cause.

In cancer you are afraid of missing dementia in cancer free you are afraid of missing cancer dx

Make a table of definition of weights based on cancer def. and

Table 1

	0	1
n	2764	878
sex = 1 (%)	1623 (58.7)	358 (40.8)
age_0 (mean (SD))	64.42 (2.86)	64.61 (2.87)
education (%)		
0	1251 (45.3)	347 (39.5)
1	1225 (44.3)	412 (46.9)
2	269 (9.7)	114 (13.0)
Unknown	19 (0.7)	5 (0.6)
apoe4 (%)		
0	1874 (67.8)	622 (70.8)
1	685 (24.8)	203 (23.1)
2	78 (2.8)	18 (2.1)
Unknown	127 (4.6)	35 (4.0)
smoke1 (%)		
0	799 (28.9)	193 (22.0)
1	1301 (47.1)	425 (48.4)
2	664 (24.0)	260 (29.6)
bmi1 (mean (SD))	26.65 (3.80)	26.39 (3.50)
oh1 (mean (SD))	10.13 (15.50)	13.25 (17.97)
sbp1 (mean (SD))	138.67 (20.80)	138.74 (20.98)
ht1 = 1 (%)	1623 (58.7)	509 (58.0)
hd_prev (%)		
0	2503 (90.6)	800 (91.1)
1	203 (7.3)	64 (7.3)
2	58 (2.1)	14 (1.6)
hd_v = 1 (%)	797 (28.8)	213 (24.3)
diabetes_prev (%)		
0	1981 (71.7)	553 (63.0)
1	284 (10.3)	91 (10.4)
2	499 (18.1)	234 (26.7)
diab_v = 1 (%)	549 (19.9)	165 (18.8)
stroke_prev = 1 (%)	47 (1.7)	15 (1.7)
stroke_v = 1 (%)	361 (13.1)	101 (11.5)
cancer_v = 1 (%)	0 (0.0)	878 (100.0)