Towards a clearer causal question underlying the association between cancer and ADRD

## 1. Introduction

Many observational studies have consistently found that individuals with cancer have a lower risk of developing dementia when compared to individuals with no history of cancer [1–4]. These findings have motivated substantial research toward mechanistic explanations, including molecular and genetic pathways that may explain this association[5–12]. These research inquiries inevitably lead to discussions of repurposing or augmenting current cancer chemotherapeutic for ADRD [13].

Nevertheless, inferring any treatment or mechanistic effects from the observed cancer-ADRD inverse association is not straightforward. Researchers have raised concerns related to the competing event of death, unmeasured confounding, and ascertainment error that could explain these results[9,14]. However, understanding these or other sources of bias first requires making explicit the causal question. Moreover, enumerating an explicit causal question is one step toward tying a research study to a question that is relevant to decision-making[15,16].

To illustrate the complexities of inferring hypothetical or available treatments’ effects on ADRD from the observed cancer-ADRD association, we focus on a specific question conceptualizing the Pin1 enzyme as the target of intervention. Previous animal studies have shown that Pin1 enzyme over-expression promotes tumorigenesis, while its down-regulation is attributed to mechanisms that contribute to neurodegeneration and amyloid deposition[11,12,17]. If we one day could develop a drug that increases Pin1 expression specifically in brain tissue in hopes of preventing dementia, we could pose the question as: *What is the effect of this Pin1-targeting drug on the risk of ADRD over time compared to standard treatments?*

To explore how we might learn about this effect using real-world data on cancer and ADRD, we progressively build a causal directed acyclic graph (DAG) to connect this particular causal question to the observable data and the assumptions we rely on to study the effect. We exemplify these challenges and how they translate into the analytic decision using data collected from the Rotterdam Study, a population-based cohort study.

## 2. Methods

### 2.1. Overview of the causal structure

If this hypothetical Pin1-targeting drug was developed, the best way to understand its effect on ADRD risk would be to have a well-conducted randomized trial in which we randomize eligible participants in late midlife (e.g., ages 50-60 years) to receive this drug or not, and closely monitor dementia diagnosis over a lengthy follow-up. Since this drug is not currently available, at best we can use observational data on Pin1 expression measurements. For example, suppose that a biomarker test was available to measure Pin1 and we measured this biomarker in stored baseline blood samples from late-midlife participants recruited for a population based-cohort.

In the observational setting, confounding could explain the observed association between Pin1 and dementia. In the causal diagram [18] in Figure 1, we present Pin1 expression as and dementia diagnosis by time as . Both may share causes , and to assess the causal relationship between and we would thus adjust for . Previous dementia studies have described age, sex, educational level and race/ethnicity as a minimal adjusting set of covariates[4]. However, environmental and behavioral factors such as smoking may translate into Pin1 over-expression[19] and are also related to the development of dementia[20].

< Figure 1 >

Currently, Pin1 expression is not an available biomarker for population-based research, so at best we can only rely on a proxy for it. Because Pin1 over-expression is present in tumors, and tumors are only measured through diagnosis, previous studies have considered Pin1 over-expression as the explanation behind the inverse association between cancer diagnosis and dementia, though Pin1 was not explicitly part of the research question[21–28]. Unlike measuring Pin1 at the same time to all participants (thought this would not necessarily mean this would be the ideal time to measure it, we discuss this point further in the discussion section), cancer diagnosis is a collected over follow-up. We depict this feature in Figure 2, where and represent *cancer diagnosis* over time, the measured proxy of and respectively. Although this means we would measure the association between cancer diagnosis over time and ADRD in the observed data, we are assuming that the captured effect is only through the pathway that involves Pin1 expression over time. Furthermore, even if we have measurements of and and some subset of , our question focused on remains.

< Figure 2 >

A challenge that arises when choosing cancer diagnosis as the proxy for Pin1 is defining the time zero, the time where eligibility criteria is met, where exposure is measured and the time where the screening for dementia would begin after having the exposure measured[29]. Emulating the eligibility criteria to participate in a trial for Pin-targeting drug will not necessarily be possible in a cohort setting recruiting participants for discordant reasons. The latter may not align with the time cancer diagnosis is measured since it happens over follow-up and this situation may introduce inmortal-time bias[29]. For example, a study performed using data from the Framingham study [21] defined the exposed group with cancer as those participants with prevalent or incident cancer diagnosis (alternatively defined as “ever cancer” [2]). This meant that a participant who had cancer diagnosis over follow-up contributed all their person-time to the cancer arm, including the time prior to the cancer diagnosis. By defining the exposure after the time of inclusion to the cohort, only participants who remain alive and free of ADRD diagnosis over follow-up can be defined as “ever cancer” [29].

This problem is partly avoided by recognizing the time-varying nature of cancer diagnosis. Several studies have considered cancer diagnosis as a time-dependent exposure [2,24,31]. The price we pay with this approximation is that implicitly, this means that Pin1 would over-express at the time of cancer diagnosis and not before, which is biologically implausible. Moreover, cancer diagnosis will only be measured in the subset of participants who are alive over follow-up. Thus, in Figure 3 we included death prior to cancer diagnosis as and an arrow between and that represents a deterministic association such as that is only observed if is zero. In addition, we added an arrow between and , since covariates such as smoking may affect Pin1 over-expression but also affect the risk of death due to other causes such as from chronic obstructive pulmonary disease. Although can be time-varying in nature, we only depict at one time-point for readability.

< Figure 3 >

Another challenge to adress is related to death as a competing event for dementia, represented in Figure 4. For interpretability we exclude the time-varying process of cancer diagnosis and focus on Pin1 (and cancer diagnosis) as it had been measured in all participants at time . Given that some participants may die prior to dementia diagnosis, we can only measure dementia over follow-up in the individuals who survive long enough to have a dementia diagnosis. In the causal diagram of Figure 4 we include a node for death after the exposure has been measured, represented as and which acts as a competing event of because if a participant dies by , the participant cannot subsequently develop dementia. Furthermore, since and are events related to aging and its consequences, represents the shared causes of both events such as cardiovascular conditions. We also include an arrow between and following the argument discussed previously.

< Figure 4 >

In the setting where represented the targeted-drug for Pin1, and if this drug had no systemic beneficial or harmful side-effects such as that there is no arrow between and , a total effect would quantify the effect of on that does not include any pathway mediated through [32]. However, in the context of cancer diagnosis as the proxy for Pin1 over-expression, we cannot rule out the effect of cancer diagnosis on death, represented as the arrow between and . In this setting, a total effect of in would include the causal pathway mediated by the effect of cancer diagnosis in mortality, which may translate into an inverse association[32].

To isolate the direct effect of in through measurement of we have at least two alternatives of causal questions we can ask. We can either imagine a causal question where we decompose the effect of cancer by the different mechanisms that affect dementia and death separately[33]. Alternatively, we can define an scenario where death could have been prevented. The latter is defined as the controlled direct effect, where death is considered as a censoring event and it relies on the assumption that we have measured all to block the pathway . Previous studies have defined death as a censoring event [25], although failed to explicitly define how to comply with the independent censoring assumption. Moreover, adjusting for confounders for dementia (such as only adjusting for ) will be insufficient to block the pathway mediated by death, time-varying covariates that represent , such as cardiovascular conditions should be considered.

To summarize, the complexity of the causal structure that describes the effect of Pin1 through the proxy of cancer diagnosis in the risk of ADRD illustrate the potential sources of bias, as observed in Figure 5. Even so, this is a simplified version since we omitted additional arrows from to and for brevity, as well as other sources of measurement error and the time-varying nature of all nodes and feedback loops between them, which would further complicate interpretability[18].

< Figure 5 >

We now turn to an application where we show how these challenges translate into analytic decisions. We will show ways to mitigate or better understand bias to the best of the available data’s abilities in an attempt to inform the possible effect of Pin1 on all cause- dementia.

### 2.2. The Rotterdam Study

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, a second wave of recruitment was held between 2000 and 2001. Participants from the first subcohort had follow-up visits between 1993-1995, 1997-1999, 2002-2005, and 2008-2010, while the second subcohort had follow-up visits between 2004 and 2005, and between 2011 and 2012[34]. All participants had data on history of cancer and dementia and incident diagnosis, collected from medical records of general practitioners (including hospital discharge letters) and through linkage with national registries. Date and cause of death was collected via municipal population registries. These ascertainment methods imply that the Rotterdam Study has functionally no loss to follow-up with respect to dementia diagnosis and death

Eligibility criteria included: ages 60 to 70 years at study entry; no history of cancer diagnosis, no history of dementia diagnosis; and free of cognitive decline (defined by a Mini Mental Score <26). Out of 10998 persons who participated at study entry, 3634 were considered eligible. Time to cancer diagnosis, time 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. Given that participants from the second subcohort were followed for 15 years, we assume that they would have had a similar distribution of dementia risk and mortality as the first subcohort, between year 15 and 20.

### 2.3. Statistical Methods

We illustrate the association between cancer and dementia diagnosis under two scenarios, the first of which replicates a common analytic strategy and the second which mitigates some (but not all) the biases described above. Scenario A replicates the setting that defines cancer proxy as *“cancer ever vs. never”* [21], meaning we compare dementia outcomes in individuals who ever develop cancer during follow-up to those who were not observed to develop cancer during follow-up. Scenario B defines cancer diagnosis as time-varying meaning that time prior to cancer diagnosis is allocated to the non-exposed arm, and the time after cancer diagnosis to the exposed arm. To address confounding, we fit inverse probability treatment weights, stabilized and truncated at the 99th percentile. In Scenario A, weights were defined as the inverse of the probability of cancer diagnosis conditional on baseline covariates such as age at study entry, sex, educational attainment, cohort, smoking status. In contrast, for Scenario B, weights were defined to represent the product of the inverse probability of being diagnosed with cancer over time, conditional on the time-varying covariate history prior to cancer diagnosis [35]. Baseline covariates included age at study entry, sex, ApoE4 status, educational attainment and the time-varying covariates such as smoking status, systolic blood pressure, body mass index and prevalent and incident hypertension and diabetes.

Inverse probability censoring weights for death were calculated to relax the independent censoring assumption. In Scenario A, weights represent the inverse of the probability of not dying conditional on cancer diagnosis (ever vs. never) and baseline covariates such as age, educational attainment, ApoE4, and baseline measurements of smoking status, hypertension status, systolic blood pressure, BMI, history of diabetes and cohort. For individuals who died, their censoring weight is zero [18]. In Scenario B time-varying weights represent the product of the inverse probability of surviving in each year prior to *t*, conditional on the measured shared causes of death and dementia. For an individual who has died by time *t*, the year *t* censoring weight is zero [32]. Weights were fitted including the same covariates used to estimate weights for time-varying cancer diagnosis, though we additionally added time-varying cancer, stroke, and heart disease diagnosis as predictors for death.Further details on modeling specifications and weights assessment are presented as **Supplementary material**.

To estimate the controlled direct effect of Pin1 in the risk of ADRD, we compared the complement of a weighted Kaplan-Meier survival estimator for participants who developed cancer versus those who did not, with time indexed in years. The weights are time-varying by follow-up year, defined as a product of the year-specific inverse probability of treatment weights and the inverse probability of censoring by death weights. Estimates of the controlled direct effect are presented as 20-year risk differences (RD) and risk ratios (RR). All 95% confidence intervals were calculated using percentile-based bootstrapping based on 500 bootstrap samples. For illustrative and comparative purposes we calculated hazard ratios (HR). Hazards, unlike risks, inherently condition on surviving both dementia and death, as such a causal interpretation is problematic [32].

Since the conditional independent censoring assumption is untestable, we compute Peterson upper and lower bounds [36] to represent: 1) the extreme scenario of independence, that refers to an scenario were those who died would never develop dementia (lower bound) and 2) complete dependency, that refers to an scenario where those who died would have a dementia prior to death (upper bound). The lower bound is calculated with the Aalen-Johansen estimator treating death as a competing event, and the upper bound is calculated with the Kaplan-Meier estimator for the combined outcome of dementia or death.

All analysis were performed using R; code is provided in supplementary material and available in <https://github.com/palolili23/2021_cancer_dementia>.

## 3. Results

Participants had a mean age of 64.46 years, and 54% (n=1979) were women (Table 1). Over follow-up, 25% (n=899) developed cancer, with a median age of cancer diagnosis at 73 (IQR:69-78). From the total sample, 13% (n=460) had dementia over follow-up with a median age of at 79 (IQR:75-83) years. Among participants with incident cancer, 6% (n=53) had dementia diagnosis, 63% (n=567) died over follow-up, and 31% (n=279) remain alive and dementia-free at 20 years since study entry. In contrast, among participants free of cancer diagnosis over follow-up, 15% (n=411) were diagnosed with dementia, 24% (n=667) died over follow-up, and 61% (n=1657) were alive and dementia-free at the end of follow-up. The proportion of participants in each possible status over follow-up and leading causes of death for both arms are presented as Figure 1 an 2 in **Supplementary Material**.

Results for all scenarios are present in Table 2. Using Scenario A’s analytic design and without adjusting for confounding or selection bias due to conditioning on death, we observe a protective association with a risk ratio (RR) of 0.7 (95% Confidence interval: 0.49,0.93) and a hazard ratio (HR) of 0.52 (95%CI: 0.39,0.69). Though adjusting for measured confounding only minimally changes the observed association, the association is closer to the null after including censoring weights for death [RR: 0.91 (95%CI: 0.65,1.19); HR: 0.72 (95%CI: 0.54,0.98)]. In contrast, using Scenario B’s analytic design, the fully adjusted model results in a RR of: 1.05 (95%CI: 0.79,1.29) and a HR of 1.09 (95%CI: 0.8,1.5), though confidence intervals cross the null. The cumulative incidence curves for both groups, under scenarios A and B, reflect that the difference between groups changes over time and flips direction when considering cancer as a time-varying proxy. ~~In addition, the extreme scenarios of dependence between death and dementia over follow-up represented as bounds are wide and include the null value (RR: 0.39, 2.08).~~

## 4. Discussion

Several observational studies have defined “cancer diagnosis” as exposure, although this does not represent a target for potential intervention or a modifiable risk factor. Instead, this variable is commonly used to represent a mechanism of interest that could not be measured. In this study we describe a particular research question of investigating a potential therapeutic drug-target of Pin1 expression, and discuss using cancer diagnosis as a proxy. By explicitly including Pin1 as part of the research question we connect the unmeasured mechanism of interest to the observed data outlining the data generation process. This practice helps to identify and disentangle potential sources of bias and can guide analytic decisions. We showed how estimates can change substantially according to alternative, yet explicit, assumptions.

For instance, a key challenge of cancer diagnosis as a proxy for Pin1 is the incapacity of defining a clear time zero[29]. In the setting where cancer diagnosis is defined as “ever vs. never,” this definition introduces immortal time bias. All results pertaining to this definition had an inverse association between cancer diagnosis and dementia, while results that did not introduce this particular form of immortal-time bias had point estimates close to the null. Although we attempt to prevent this bias with statistical methods, we can only fully prevent it by having a clear definition of time-zero. This definition does not depend on the collected data nor in analytic decisions. It relies on a deeper discussion related to when would be the best moment to measure this biomarker and to what purpose. Thus, we hope that these unsolved questions guide future discussions and data collection efforts.

On the other hand, death as competing event is a challenge that has some straightforward strategies, beginning first and foremost by choosing the causal parameter (or estimand) of interest [37]. In this study we chose the controlled direct effect, which represents the effect of Pin1 (or cancer) in a setting where death due to cancer and other causes could have been prevented, yet without an explicit intervention, which is yet ambiguous. This is different than conceiving a drug-target that increases Pin1 expression only brain tissue, with no side effects that could increase cancer risk (and thus, death due to cancer). As opposed to prior studies that implicitly address a direct effect, and who define censoring as ignorable[25], we show how point estimates change substantially when we include weights for death to relax the independent censoring assumption [32] when the estimand of interest is the controlled direct effect. Bounds to assess extreme scenarios of dependence between death and dementia[36] illustrate the wide range of possible point estimates that cross the null. This shows that even with the effort of adjusting for time-varying covariates, we may be far from meeting this assumption and thus better efforts to measure shared causes of dementia and death are needed. In addition, presenting the proportion of participants that died prior to ADRD diagnosis in each arm, as well as the proportion of participants in each status over time, improves transparency and puts in evidence the limitations of the data to answer this question.

Pin1 is only one of the several mechanisms proposed about the inverse cancer-dementia relationship. Certainly, cancer diagnosis represents a complex and heterogeneous health condition that exceeds the representation of Pin1 expression. To understand how – if at all – the cancer-dementia association informs the potential effects of any other mechanism or treatment strategy, and its connection to collected data, we may require different causal representations. For example, if researchers used this association to inform the possible effects of different chemotherapeutics on cognitive decline among patients undergoing treatment for cancer, different challenges would arise for mapping the observed association to the hypothetical randomized trial underlying this research question. Notably, that target trial, unlike the one considered here, would include cancer diagnosis as part of the eligibility, and thus researchers would need to instead grapple with how using data on persons without cancer is useful and useable [40]. Because each question requires thinking about bias anew and each question brings its own set of challenges and opportunities, we believe it is crucial to have the discussion about the question or set of questions that are relevant first, as this can also inform efforts to collect more suitable data. Furthermore, we underscore that this is just one case study for how observed associations between two diseases or health states may be disentangled to more transparently unveil possible mechanisms (and sources of bias) behind them.

## References

[1] Ma L-L, Yu J-T, Wang H-F, Meng X-F, Tan C-C, Wang C, et al. Association between Cancer and Alzheimer’s Disease: Systematic Review and Meta-Analysis. Journal of Alzheimer’s Disease 2014;42:565–73. <https://doi.org/10.3233/JAD-140168>.

[2] Hanson HA, Horn KP, Rasmussen KM, Hoffman JM, Smith KR. Is Cancer Protective for Subsequent Alzheimer’s Disease Risk? Evidence From the Utah Population Database. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2016:gbw040. <https://doi.org/10.1093/geronb/gbw040>.

[3] van der Willik KD, Schagen SB, Ikram MA. Cancer and dementia: Two sides of the same coin? European Journal of Clinical Investigation 2018;48:e13019. <https://doi.org/10.1111/eci.13019>.

[4] Ospina-Romero M, Glymour MM, Hayes-Larson E, Mayeda ER, Graff RE, Brenowitz WD, et al. Association Between Alzheimer Disease and Cancer With Evaluation of Study Biases: A Systematic Review and Meta-analysis. JAMA Network Open 2020;3:e2025515. <https://doi.org/10.1001/jamanetworkopen.2020.25515>.

[5] Behrens M, Lendon C, Roe C. A Common Biological Mechanism in Cancer and Alzheimers Disease? Current Alzheimer Research 2009;6:196–204. <https://doi.org/10.2174/156720509788486608>.

[6] Harris RA, Tindale L, Cumming RC. Age-dependent metabolic dysregulation in cancer and Alzheimer’s disease. Biogerontology 2014;15:559–77. <https://doi.org/10.1007/s10522-014-9534-z>.

[7] Nudelman KNH, McDonald BC, Lahiri DK, Saykin AJ. Biological Hallmarks of Cancer in Alzheimer’s Disease. Molecular Neurobiology 2019;56:7173–87. <https://doi.org/10.1007/s12035-019-1591-5>.

[8] Papin S, Paganetti P. Emerging Evidences for an Implication of the Neurodegeneration-Associated Protein TAU in Cancer. Brain Sciences 2020;10:862. <https://doi.org/10.3390/brainsci10110862>.

[9] Driver JA. Inverse association between cancer and neurodegenerative disease: Review of the epidemiologic and biological evidence. Biogerontology 2014;15:547–57. <https://doi.org/10.1007/s10522-014-9523-2>.

[10] Olson B, Marks DL. Pretreatment Cancer-Related Cognitive ImpairmentMechanisms and Outlook. Cancers 2019;11:687. <https://doi.org/10.3390/cancers11050687>.

[11] Li J, Mo C, Guo Y, Zhang B, Feng X, Si Q, et al. Roles of peptidyl-prolyl isomerase Pin1 in disease pathogenesis. Theranostics 2021;11:3348–58. <https://doi.org/10.7150/thno.45889>.

[12] Driver JA, Zhou XZ, Lu KP. Pin1 dysregulation helps to explain the inverse association between cancer and Alzheimer’s disease. Biochimica Et Biophysica Acta (BBA) - General Subjects 2015;1850:2069–76. <https://doi.org/10.1016/j.bbagen.2014.12.025>.

[13] Snyder HM, Ahles T, Calderwood S, Carrillo MC, Chen H, Chang C-CH, et al. Exploring the nexus of Alzheimer’s disease and related dementias with cancer and cancer therapies: A convening of the Alzheimer’s Association & Alzheimer’s Drug Discovery Foundation. Alzheimer’s & Dementia 2017;13:267–73. <https://doi.org/10.1016/j.jalz.2016.11.002>.

[14] Ganguli M. Cancer and Dementia: It’s Complicated. Alzheimer Disease & Associated Disorders 2015;29:177–82. <https://doi.org/10.1097/WAD.0000000000000086>.

[15] Didelez V. Data always be preceded by specifying a target experimental trial? International Journal of Epidemiology 2016;45:3.

[16] Labrecque JA, Swanson SA. Target trial emulation: Teaching epidemiology and beyond. European Journal of Epidemiology 2017;32:473–5. <https://doi.org/10.1007/s10654-017-0293-4>.

[17] Angelucci F, Hort J. Prolyl isomerase Pin1 and neurotrophins: A loop that may determine the fate of cells in cancer and neurodegeneration. Therapeutic Advances in Medical Oncology 2017;9:59–62. <https://doi.org/10.1177/1758834016665776>.

[18] Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.; 2020.

[19] Tan X, Zhou F, Wan J, Hang J, Chen Z, Li B, et al. Pin1 expression contributes to lung cancer prognosis and carcinogenesis. Cancer Biology & Therapy 2010;9:111–9. <https://doi.org/10.4161/cbt.9.2.10341>.

[20] Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. The Lancet 2020;396:413–46. <https://doi.org/10.1016/S0140-6736(20)30367-6>.

[21] Driver JA, Beiser A, Au R, Kreger BE, Splansky GL, Kurth T, et al. Inverse association between cancer and Alzheimer’s disease: Results from the Framingham Heart Study. BMJ 2012;344:e1442–2. <https://doi.org/10.1136/bmj.e1442>.

[22] Musicco M, Adorni F, Di Santo S, Prinelli F, Pettenati C, Caltagirone C, et al. Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study. Neurology 2013;81:322–8. <https://doi.org/10.1212/WNL.0b013e31829c5ec1>.

[23] Freedman DM, Wu J, Chen H, Kuncl RW, Enewold LR, Engels EA, et al. Associations between cancer and Alzheimer’s disease in a U.S. Medicare population. Cancer Medicine 2016;5:2965–76. <https://doi.org/10.1002/cam4.850>.

[24] Bowles EJA, Walker RL, Anderson ML, Dublin S, Crane PK, Larson EB. Risk of Alzheimer’s disease or dementia following a cancer diagnosis. PLOS ONE 2017;12:e0179857. <https://doi.org/10.1371/journal.pone.0179857>.

[25] Frain L, Swanson D, Cho K, Gagnon D, Lu KP, Betensky RA, et al. Association of cancer and Alzheimer’s disease risk in a national cohort of veterans. Alzheimer’s & Dementia 2017;13:1364–70. <https://doi.org/10.1016/j.jalz.2017.04.012>.

[26] Schmidt SAJ, Ording AG, Horváth-Puhó E, Sørensen HT, Henderson VW. Non-melanoma skin cancer and risk of Alzheimer’s disease and all-cause dementia. PLOS ONE 2017;12:e0171527. <https://doi.org/10.1371/journal.pone.0171527>.

[27] Sun M, Wang Y, Sundquist J, Sundquist K, Ji J. The Association Between Cancer and Dementia: A National Cohort Study in Sweden. Frontiers in Oncology 2020;10:73. <https://doi.org/10.3389/fonc.2020.00073>.

[28] Ording AG, Horváth-Puhó E, Veres K, Glymour MM, Rørth M, Sørensen HT, et al. Cancer and risk of Alzheimer’s disease: Small association in a nationwide cohort study. Alzheimer’s & Dementia 2020;16:953–64. <https://doi.org/10.1002/alz.12090>.

[29] Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of Clinical Epidemiology 2016;79:70–5. <https://doi.org/10.1016/j.jclinepi.2016.04.014>.

[30] Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. n.d.:10.

[31] White RS, Lipton RB, Hall CB, Steinerman JR. Nonmelanoma skin cancer is associated with reduced Alzheimer disease risk. Alzheimer Disease 2013:7.

[32] Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine 2020;39:1199–236. <https://doi.org/10.1002/sim.8471>.

[33] Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA. Separable Effects for Causal Inference in the Presence of Competing Events. Journal of the American Statistical Association 2020:1–9. <https://doi.org/10.1080/01621459.2020.1765783>.

[34] Ikram MA, Brusselle G, Ghanbari M, Goedegebure A, Ikram MK, Kavousi M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. European Journal of Epidemiology 2020;35:483–517. <https://doi.org/10.1007/s10654-020-00640-5>.

[35] Hernán MÁ, Brumback B, Robins JM. Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men: Epidemiology 2000;11:561–70. <https://doi.org/10.1097/00001648-200009000-00012>.

[36] Peterson AV. Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. Proceedings of the National Academy of Sciences 1976;73:11–3. <https://doi.org/10.1073/pnas.73.1.11>.

[37] Rojas-Saunero LP, Young JG, Didelez V, Ikram MA, Swanson SA. Choosing questions before methods in dementia research with competing events and causal goals. 2021. <https://doi.org/10.1101/2021.06.01.21258142>.

[38] van Geloven N, Geskus RB, Mol BW, Zwinderman AH. Correcting for the dependent competing risk of treatment using inverse probability of censoring weighting and copulas in the estimation of natural conception chances: N. VAN GELOVEN *ET AL.* Statistics in Medicine 2014;33:4671–80. <https://doi.org/10.1002/sim.6280>.

[39] Willems S, Schat A, van Noorden M, Fiocco M. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. Statistical Methods in Medical Research 2018;27:323–35. <https://doi.org/10.1177/0962280216628900>.

[40] Huitfeldt A, Hernan MA, Kalager M, Robins JM. Comparative Effectiveness Research Using Observational Data: Active Comparators to Emulate Target Trials with Inactive Comparators. eGEMs (Generating Evidence & Methods to Improve Patient Outcomes) 2016;4:20. <https://doi.org/10.13063/2327-9214.1234>.

## Tables

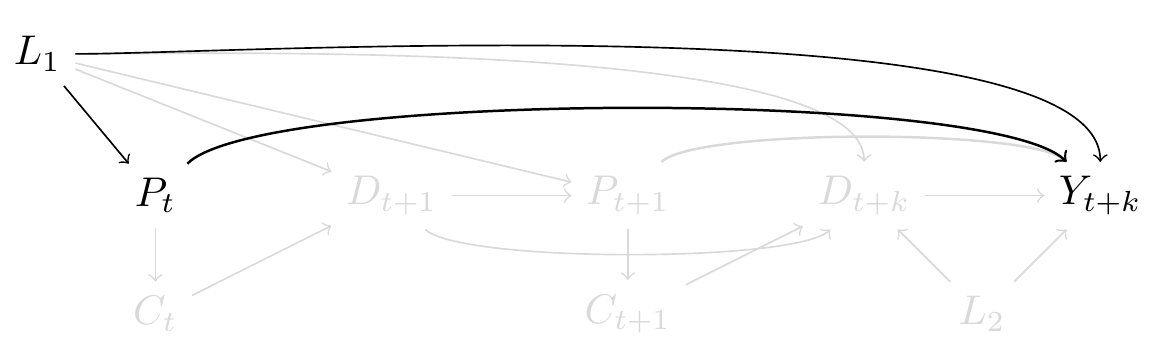
### Table 1: Descriptive characteristics of participants

| Characteristics | Overall |
| --- | --- |
| n | 3634 |
| Male (%) | 1655 (45.5) |
| Age at baseline (mean (SD)) | 64.5 (2.9) |
| Educational attainment (%) |  |
| Higher | 382 (10.5) |
| Intermediate | 1634 (45.0) |
| Lower | 1594 (43.9) |
| Unknown history of diabetes | 24 (0.7) |
| ApoE4 (%) |  |
| Not carrier | 2488 (71.7) |
| One allele carrier | 888 (25.6) |
| Two allele carrier | 96 (2.8) |
| Smoking status (%) |  |
| Current | 922 (25.4) |
| Former | 1720 (47.3) |
| Never | 992 (27.3) |
| Body Mass Index (mean (SD)) | 26.6 (3.7) |
| Systolic blood pressure (mmHg) (mean (SD)) | 138.7 (20.8) |
| History of hypertension at baseline (%) | 2126 (58.5) |
| History of heart disease at baseline (%) | 263 (7.4) |
| Incident heart disease (%) | 1022 (28.1) |
| No history of diabetes at baseline (%) | 2534 (69.7) |
| History of diabetes at baseline (%) | 374 (10.3) |
| Unknown history of diabetes (%) | 726 (20.0) |
| Incident diabetes (%) | 718 (19.8) |
| History of stroke (%) | 62 (1.7) |
| Incident stroke (%) | 472 (13.0) |
| Incident cancer (%) | 899 (24.7) |

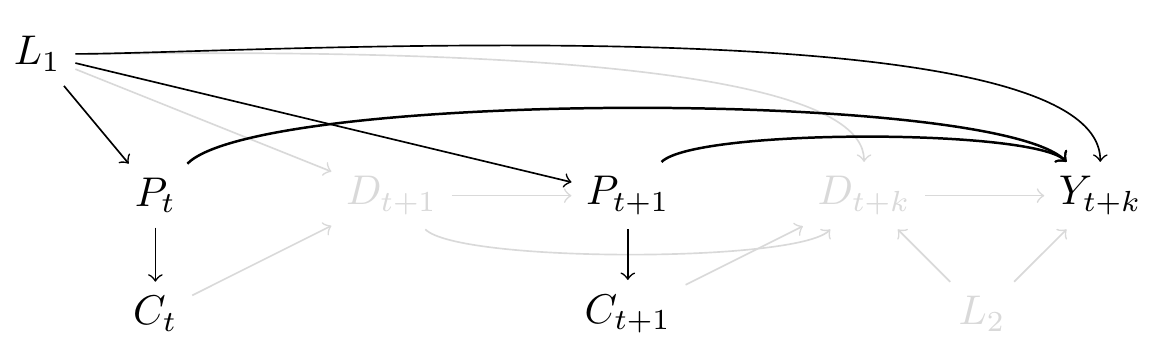
### Table 2. 20-year risk differences, risk ratio, and hazard ratios for dementia risk by cancer proxy and model adjustment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Scenario | Model | Risk Difference | Risk Ratio | Hazard Ratio |
| A | Unadjusted | -5.9 (-10.4,-1.1) | 0.71 (0.49,0.95) | 0.51 (0.39,0.68) |
| B | Unadjusted | -1.4 (-5,3) | 0.94 (0.78,1.15) | 0.95 (0.71,1.27) |
| A | With weights for confounding | -6.0 (-10.6,-1.4) | 0.70 (0.49,0.93) | 0.52 (0.39,0.69) |
| B | With weights for confounding | -0.3 (-6.7,3.8) | 0.98 (0.71,1.19) | 1.03 (0.76,1.39) |
| A | With weights for confounding and for censoring for death | -2.0 (-7.7,3.9) | 0.91 (0.65,1.19) | 0.72 (0.54,0.98) |
| B | With weights for confounding and for censoring for death | 1.2 (-5,6.4) | 1.05 (0.79,1.29) | 1.09 (0.80,1.50) |
| Scenario A represents the setting where cancer is defined as 'ever vs. never'; Scenario B represents the setting where cancer is defined as a time-varying proxy | | | | |

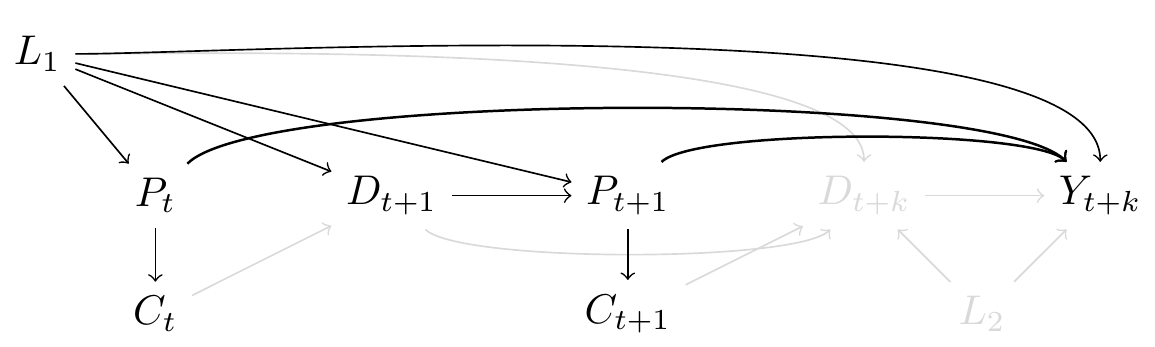
### Figures



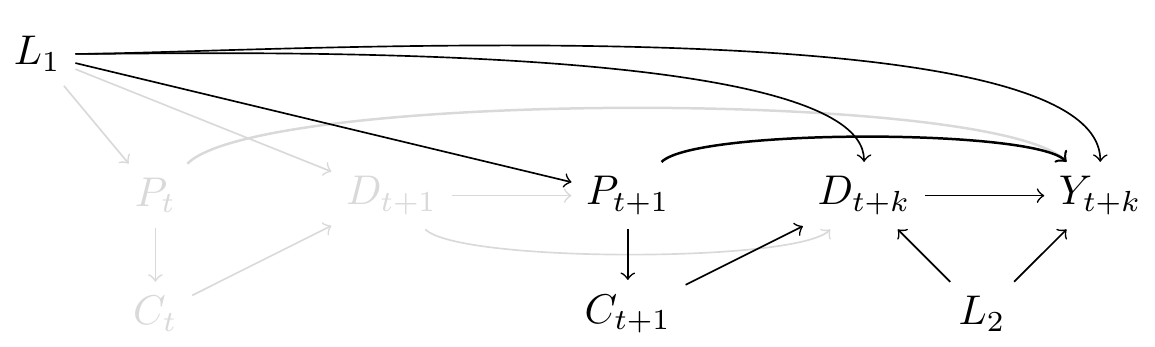
**Figure 1. Pin1 as biomarker and risk of dementia.** Causal directed acyclic graph where Pin1 at time is represented as , Dementia at time is represented as . represents shared causes of Pin1 and Dementia, such as smoking. To isolate the effect of and we need to block the backdoor path .



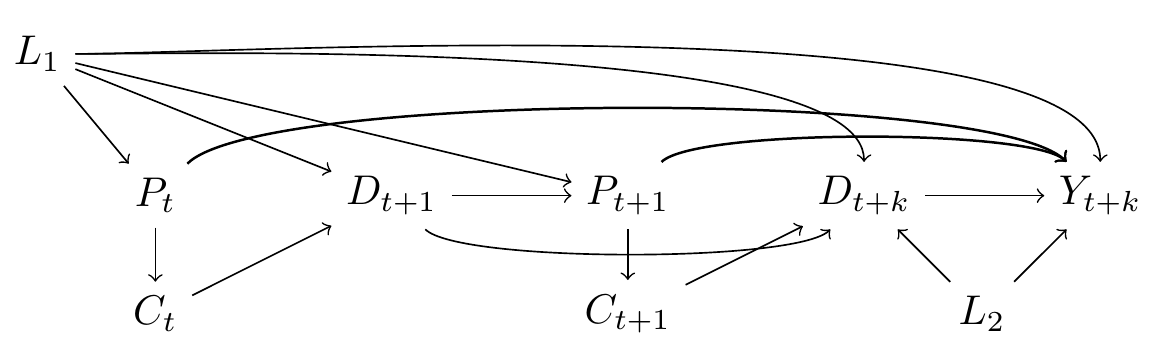
**Figure 2. Incident cancer diagnosis as proxy for Pin1 expression.** This causal graph is an expansion of Figure 1, includes Pin1 expression at time . Incident cancer diagnosis is represented as and . To isolate the effect of and on we need to block the backdoor path and . Although we represent as a single node for readability, is time-varying too.



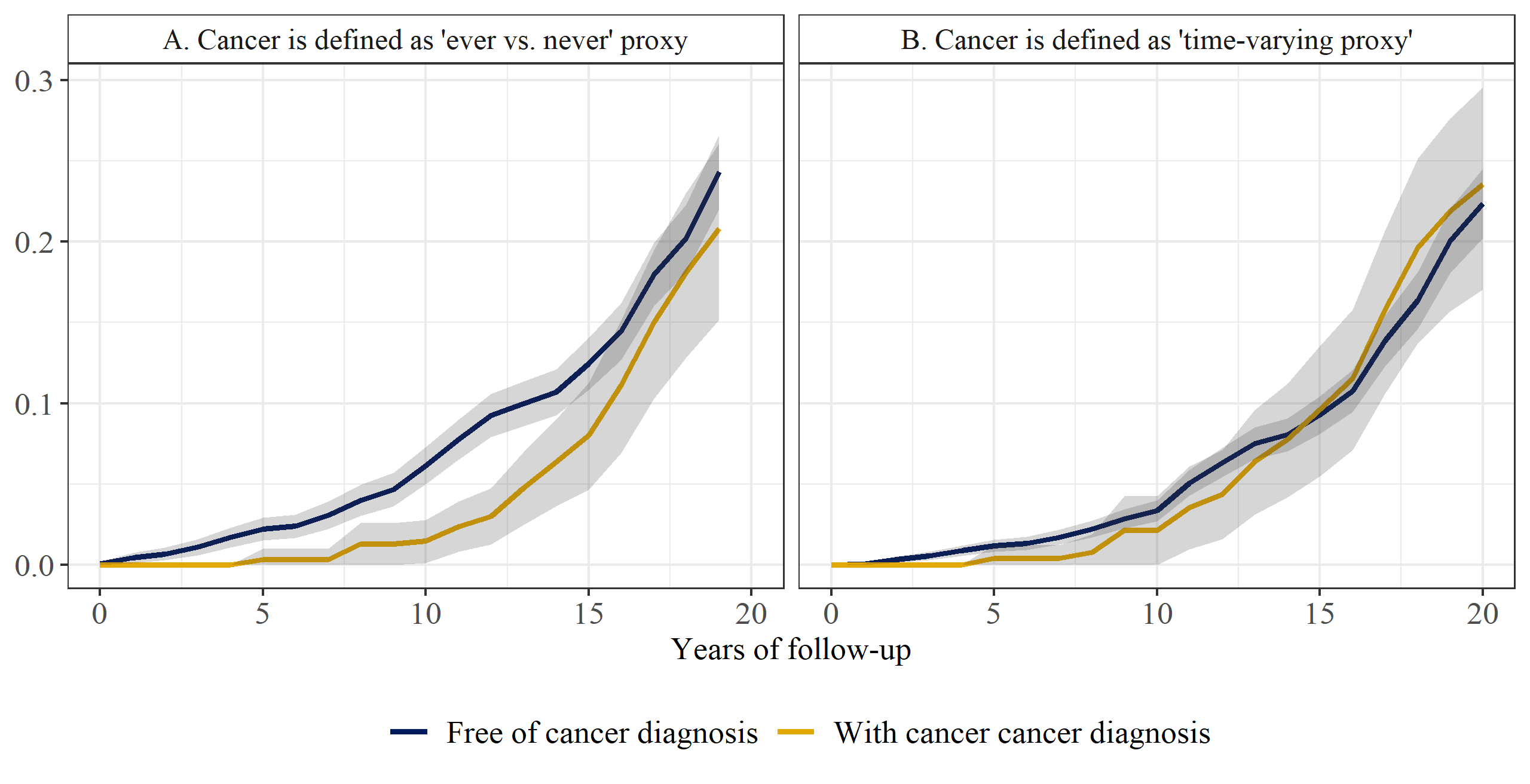
**Figure 3. Time-varying cancer diagnosis and inmortal-time bias.** This causal graph is an expansion of Figure 2, where represents death at time . Cancer diagnosis at can only be measured among those who are alive at .



**Figure 4. Death as competing event of dementia.** In this graph we only focus attention to the exposure as if it was measured for all at time . We include as death at time , represents possible shared causes of dementia and death (such as cardiovascular comorbidities). Pin1 may affect the risk of death through cancer diagnosis, represented as an arrow between and . To isolate the direct effect of and , we have to block the backdoor pathway from and



**Figure 5. Causal graph for effect of Pin1 through the proxy of cancer diagnosis in the risk of dementia, including all challenges discussed through Figures 1 to 4.**



**Figure 6. Risk of dementia under elimination of death over 20 years of follow-up by cancer diagnosis status.** Panel A. represents the scenario where cancer was considered as “ever vs. never”. Panel B. represents the scenario where cancer was considered as a time-varying proxy.

## Supplementary material

### Modeling description

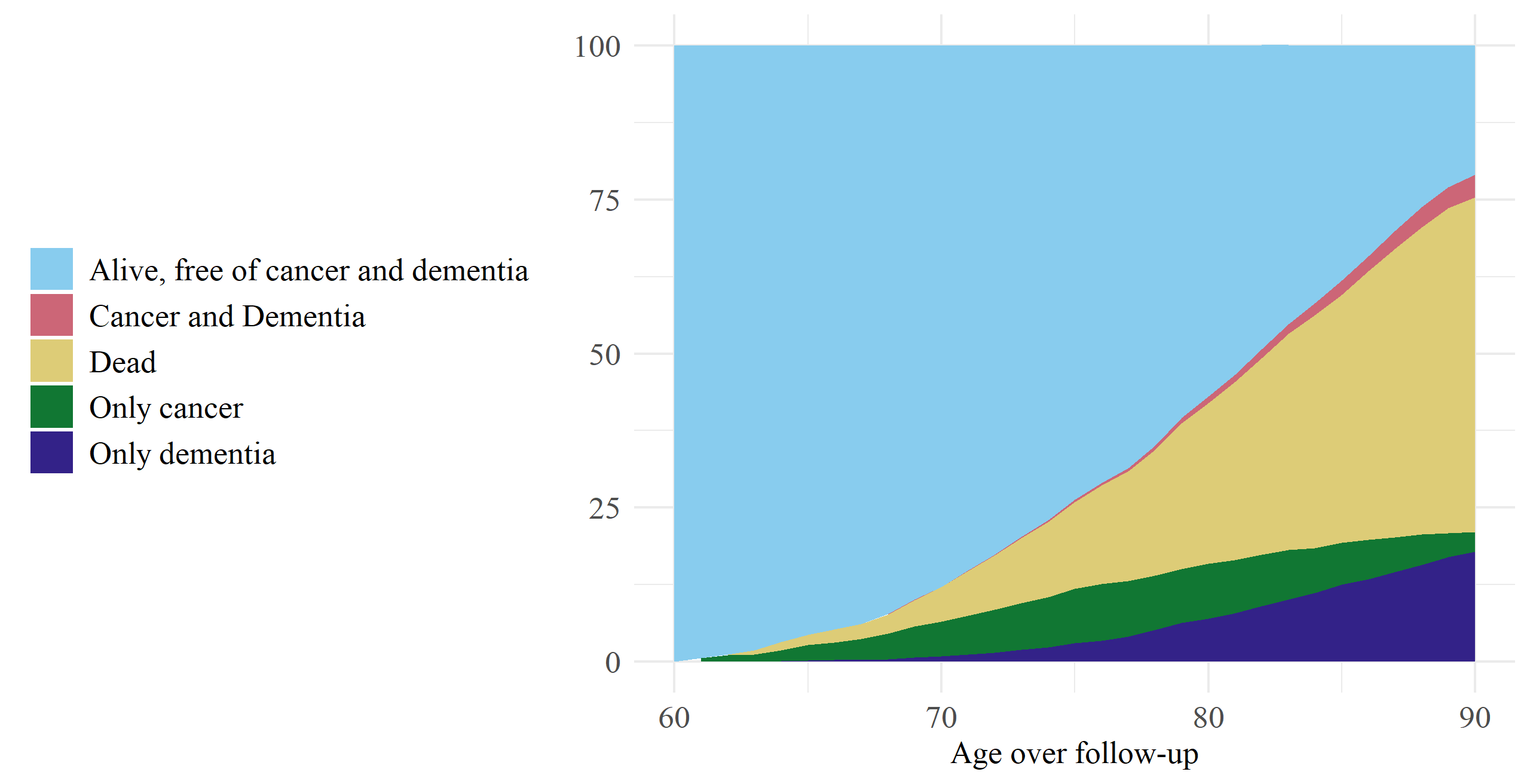
**Scenario A: Cancer diagnosis as “ever vs. never” proxy for Pin1**

|  |  |  |
| --- | --- | --- |
| **Weights** | **Numerator** | **Denominator** |
| **Inverse probability of treatment weights** | Pr[At+k = 1|1]  Where:  *A* = cancer diagnosis (0 = never, 1 = ever) | Pr[At+k = 1|V]  Where:  *A*= cancer diagnosis (0 = never, 1 = ever); *V* = age at study entry with natural cubic splines, sex (women vs. men), education (three categories), apoe4 (three categories), smoking status at study entry(three categories), cohort (two categories); product terms between covariates |
| **Inverse probability of censoring weights for death** | Pr[Dt+k = 0| Yt+k = 0, 1]  Where:  *D* = died prior to dementia diagnosis over follow-up (0 = no, 1 = yes); Y = dementia diagnosis | Pr[Dt+k = 0|Yt+k = 0, A t+k,L]  Where:  *D* = died prior to dementia diagnosis over follow-up (0 = no, 1 = yes); Y = dementia diagnosis (0 = no, 1 = yes); *A* = cancer diagnosis (0 = never, 1 = ever); *V* = age at study entry with natural cubic splines, sex (women vs. men), education (three categories), apoe4 (three categories), cohort (two categories), smoking status at study entry(three categories), hypertension status at study entry (two categories), history of diabetes at study entry (three categories), systolic blood pressure and body mass index at study entry (continuous variables); product terms between covariates |

**Scenario B. Cancer diagnosis as a time-varying proxy for Pin1**

|  |  |  |
| --- | --- | --- |
| **Inverse probability of treatment weights** | Pr[At+1 = 1| At = 0, Dk = 0, Yk = 0, T]  Where:  *A*= cancer diagnosis (0 = no, 1 = yes); *T* = year with natural cubic splines | Pr[A*t+1* = 1| At = 0, Dt = 0, Yt = 0, T, V, Lt]  Where *A* = cancer diagnosis (0 = no, 1 = yes); *T* = year with natural cubic splines; *V* = age at study entry with cubic splines, sex (women vs. men), education (three categories), apoe (three categories), cohort (two categories); *L*= time-varying smoking status (three categories), hypertension status (two categories), hypertension medication (two categories), diabetes status (three categories), systolic blood pressure and body mass (continuous variables); no product terms between covariates |
| **Inverse probability of censoring weights for death** | Pr[D*t+1* = 0| D*t* = 0, Y*t* = 0, A*t*, T, V]  Where:  *D* = death (0 = no, 1 = yes); *Y =* dementia diagnosis (0 = no, 1 = yes); *A* = cancer diagnosis (0 = no, 1 = yes), *Y* = year with natural cubic splines; *V* = cohort (two categories); no product terms between covariates | Pr[D*k+1* = 0| D*k* = 0, Y*t* = 0, A*t*, T, V, L*t*]  Where *D* = death (0 = no, 1 = yes); *Y* = dementia diagnosis (0 = no, 1 = yes); ; *A*= cancer diagnosis (0 = no, 1 = yes); *T =* year with natural cubic splines; *V* = age at study entry with cubic splines, sex (women vs. men), education (five categories), apoe (three categories), cohort (two categories); *L*t = time-varying smoking status (three categories), hypertension status (two categories), hypertension medication (two categories), diabetes status (three categories), heart disease condition (yes, no), incident stroke (yes, no), systolic blood pressure and body mass (continuous variables) and no product terms between covariates |

### Figure 1. Distribution of participants under each health status, by age over follow-up



### Figure 2. Causes of death for participants who died prior to dementia diagnosis

C:\Users\040609\AppData\Local\Microsoft\Windows\INetCache\Content.Word\causes_death.tiff