Dissecting the causal question underlying the association between cancer and ADRD: an opportunity to understand sources of bias

L. Paloma Rojas-Saunero

1. Introduction

Many observational studies have consistently found that individuals with cancer have a lower risk of developing of Alzheimer's disease or related dementias (ADRD) when compared to individuals with no history of cancer (Ma et al. 2014; Hanson et al. 2016; van der Willik, Schagen, and Ikram 2018; Ospina-Romero et al. 2020). These findings have motivated substantial research toward mechanistic explanations, including searching for and hypothesizing that molecular and genetic mechanisms may explain this association(Behrens, Lendon, and Roe 2009; Harris, Tindale, and Cumming 2014; Nudelman et al. 2019; Papin and Paganetti 2020; Jane A. Driver 2014; Olson and Marks 2019; Li et al. 2021; Jane A. Driver, Zhou, and Lu 2015). These research inquiries inevitably lead to discussions of repurposing or augmenting current cancer chemotherapeutics for ADRD (Snyder et al. 2017).

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(Jane A. Driver 2014; Ganguli 2015). However, understanding these or other sources of bias first requires making explicit the causal question. Moreover, making explicit the causal question is one step toward tying a research study to a question that is relevant to decision-making(Didelez 2016; Labrecque and Swanson 2017). 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(Jane A. Driver, Zhou, and Lu 2015; Angelucci and Hort 2017; Li et al. 2021). 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 to connect this particular causal question to the observable data and the assumptions we rely on to study the effect. We exemplify different scenarios with data collected from the Rotterdam Study, a population-based cohort study. We describe the challenges and how they translate into the analytic decisions. Last, we discuss how information on mortality and cause of death can provide insight about the direction of some sources of bias.

2. Overview of the causal structure

If this hypothetical Pin1-targeting drug was developed, the best way to understand its effect on dementia 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 ADRD 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 from (stored) baseline blood samples in a population based-cohort that recruited participants in late midlife.

In the observational setting, confounding could explain the observed association between Pin1 and ADRD. In the directed acyclic graph (Miguel A. Hernán and Robins 2020) of Figure 1, we present Pin1 expression as P and ADRD diagnosis over time as Y_{t+1} . Both nodes may share causes L, and to assess the causal relationship between P and Y_{t+1} we would require adjusting for L. Previous studies have described age, sex, educational level and race/ethnicity as the minimal adjusting set of covariates(Ospina-Romero et al. 2020). However, environmental and behavioral factors such as smoking, which are known to cause microenvironmental changes such as inflammation and changes in tissue remodeling, may translate into Pin1 over-expression and are also related to the development of ADRD[ref].

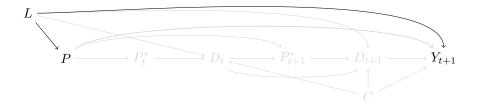


Figure 1: Pin1 as biomarker and risk of dementia

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 ADRD, though Pin1 was not explicitly part of the research question(Ospina-Romero et al. 2020; J. A. Driver et al. 2012; Musicco et al. 2013; Freedman et al. 2016; Bowles et al. 2017; Frain et al. 2017; Schmidt et al. 2017; Sun et al. 2020; Ording et al. 2020). Unlike measuring Pin1 at the same time in all the cohort (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 time-varying process. We depict this feature in Figure 2, where P_t^* and P_{t+1}^* represent cancer diagnosis over time, the measured proxy of P. Although we are measuring 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 P.

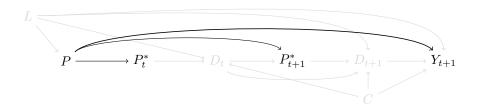


Figure 2: Cancer diagnosis as proxy for Pin1 expression

A challenge that arises by defining cancer diagnosis as the proxy for Pin1 is defining the time zero, the time where follow-up and screening for ADRD would begin after having the exposure measured. In a cohort setting, an unclear definition of time zero could introduce inmortal-time bias. For example, a study performed using data from the Framingham study (J. A. Driver et al. 2012) defined the exposed group with cancer as those participants with prevalent or incident cancer diagnosis (alternatively defined as "ever cancer" (Hanson et al. 2016)). 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" Miguel A. Hernán et al. (2016).

We can overcome this problem by acknowledging the time-varying nature of cancer diagnosis. Several studies have considered cancer diagnosis as a time-dependent exposure. In this way, the time prior to cancer

diagnosis is allocated to the non-exposed arm, and the time after cancer diagnosis to the exposed arm (White et al. 2013; Hanson et al. 2016; Bowles et al. 2017). The price we pay with this approximation is that we implicitely assume that Pin1 rises at the time of cancer diagnosis, which is biologically implausible. Moreover, cancer diagnosis will only be measured in the subset of participants who are alive over follow-up. Thus we included death prior to cancer diagnosis as D_t and an arrow between D_t and P_{t+1}^* that represents a deterministic association such as that P^* is only observed if D_t is zero in Figure 3. We additionally included shared causes between D_t and $P^*t + 1$ as L_t , these shared causes might affect the risk of having a cancer diagnosis or dying from other causes over follow-up.

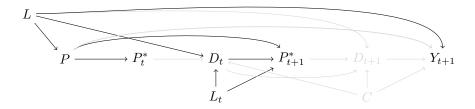


Figure 3: Time-varying cancer diagnosis

We now move to the challenge of death as a competing event for ADRD. Given that some participants may die prior to having a cancer diagnosis over follow-up, we can only have ADRD over follow-up in the individuals who survive long enough to have an ADRD diagnosis. In the causal diagram of Figure 4 we include one more node of death after the exposure P^*t has been measured, represented as D_{t+1} , this acts as a competing event of Y_{t+1} because if a participant dies prior to ADRD diagnosis, death prevents observing ADRD at future time-points. Furthermore, since D_{t+1} and Y_{t+1} are events related to aging and its consequences, C represents the shared causes of both events such as cardiovascular conditions.

In the scenario where P represented the targeted-drug for Pin1, and if this drug has no systemic beneficial or harmful side-effects such as that there is no arrow between P and D_{t+1} , a total effect would quantify the effect of P on Y_{t+1} that does not include any pathway mediated through D_{t+1} (Young et al. 2020). However, in the context of Pin1 as a biomarker or cancer diagnosis as the proxy for Pin1, we cannot rule out their effect on death. In these settings, a total effect of P in Y_{t+1} would include the causal pathway mediated by the effect of cancer diagnosis in mortality, which may translate into an inverse association (Young et al. 2020). To isolate the direct effect of P_t^* in Y_{t+1} through measurement of P_t^* we have at least two alternatives. We can either imagine a causal question where we decompose the effect of cancer by the different mechanisms that affect dementia and death separately (Stensrud et al. 2020). 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 C to block the pathway $Y_{t+1} \leftarrow C \rightarrow D_{t+1} \rightarrow P_t^* \rightarrow P_0$. Previous studies have defined death as a censoring event [ref], although failed to explicitly define how to relax the independent censoring assumption. While adjusting for confounders for dementia may represent partially the adjustment for the competing event of death (Young et al. 2020), adjusting for baseline covariates may not be sufficient.

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 the time varying nature of all nodes in the graph and other sources of measurement error, and thus are ignoring how feedback loops between them further complicate interpretability and identificability.

We now turn to an application where we show how these challenges translate in analytic decisions, and we will show ways to mitigate or better understand to the best of the available data's abilities. Nevertheless, we acknowledge that cancer diagnosis represents a complex and heterogenous health condition that exceeds the representation of Pin1 expression, thus, at best we may test the sharp null hypothesis rather than estimating the causal effect of Pin1. In this section we conduct an analysis of the cancer-ADRD association that is structured to the best of the available data's abilities to inform the possible effect of Pin1 on all

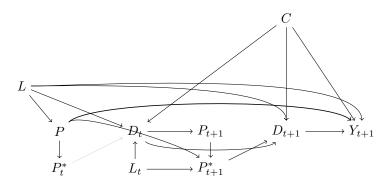


Figure 4: Direct effect

cause-dementia.

4. Application to 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; it was later extended between 2000 and 2001 consisting of individuals who had reached the age of 55 years or who had moved into the study area. Participants from first subcohort had follow-up visits between 1993-1995, 1997-1999, 2002-2005, and 2008-2010, second subcohort had follow-up visits between 2004 and 2005, and between 2011 and 2012(Ikram et al. 2020). 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. Data on clinical outcomes was available until 2015.

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 10998 persons who participated at study entry, 3642 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.

4.1. Methods

We illustrate the association between cancer and dementia diagnosis under two scenarios. For the first scenario (Model A) we defined cancer proxy as "cancer ever vs. never". This scenario mimics the definition used in previous studies where cancer was defined as a time-independent exposure (J. A. Driver et al. 2012; Nudelman et al. 2014). We continue with a second scenario (Model B) where cancer diagnosis is a time-varying exposure. The definition represents the setting where all participants are free of cancer at inclusion and over time some participants will have a cancer diagnosis.

To address confounding we fit inverse probability treatment weights, stabilized and truncated at 99th percentile. Weights for model A 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 Model B weights recreate a pseudopopulation in which time-varying covariates do not predict the risk of cancer diagnosis at time t (M. Á. Hernán, Brumback, and Robins 2000). Time varying covariates included: . . . Further details on modeling specifications and weights assessment are presented as **Supplementary material x**.

To estimate the controlled direct effect of Pin1 in the risk of ADRD, that is in the setting where death could have been prevented, we fit inverse probability censoring weights for death.

in time-varying settings we compared the complement of a weighted Kaplan-Meier survival estimator in participants with incident cancer vs. no incident of cancer with time indexed in years. The weights in this case are time-varying by follow-up year, defined as a product of the time-fixed IPT weights above and a year-specific inverse probability of censoring (IPC) by death weights. For an individual still alive in year t, the time t IPC weight is the product of the inverse probability of surviving in each year prior to t, conditional on measured common causes of death and dementia (that is, variables such as C in Figure 4). For an individual who has died by time t, the year t IPC weight is zero. We estimated survival probabilities using a logistic regression model for death as a function of baseline and time-varying covariates. Baseline covariates included age at study entry, sex, apoe4 status, educational attainment and the time-varying covariates smoking status, systolic blood pressure, BMI and prevalent and incident comorbidities such as: cancer, heart disease, stroke and diabetes. We additionally calculated the controlled direct effect considering death as an unconditional independent censoring event (as if there were no arrows from C to D_{15} and Y_{20}) for an illustrative purpose.

Estimates of the controlled direct effect at 20 years of follow-up are presented as risk differences (RD), risk ratios (RR) and hazard ratios (HR). We note that hazards, unlike risks, inherently condition on surviving both dementia and death, as such they will not have a causal interpretation in this case. We present them for comparison against risk ratios.

Since the conditional independent censoring assumption is untestable, we compute Peterson upper and lower bounds 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.

We additionally present the causes of death for participants who developed cancer and for participants free of cancer diagnosis over follow-up. This information provides insights about the association between cancer diagnosis and death and about the potential misclassification of individuals with no cancer diagnosis (cancer as a cause of death).

All 95% confidence intervals were calculated using percentile-based bootstrapping based on 500 bootstrap samples. All analysis were performed using R, code is provided in supplementary material and available in https://github.com/palolili23/2021 cancer dementia.

Anderson, J R, K C Cain, and R D Gelber. n.d. "Analysis of Survival by Tumor Response." 10.

- Angelucci, Francesco, and Jakub Hort. 2017. "Prolyl Isomerase Pin1 and Neurotrophins: A Loop That May Determine the Fate of Cells in Cancer and Neurodegeneration." Therapeutic Advances in Medical Oncology 9 (1): 59–62. https://doi.org/10.1177/1758834016665776.
- Behrens, M., C. Lendon, and C. Roe. 2009. "A Common Biological Mechanism in Cancer and Alzheimers Disease?" Current Alzheimer Research 6 (3): 196–204. https://doi.org/10.2174/156720509788486608.
- Bowles, Erin J. Aiello, Rod L. Walker, Melissa L. Anderson, Sascha Dublin, Paul K. Crane, and Eric B. Larson. 2017. "Risk of Alzheimer's Disease or Dementia Following a Cancer Diagnosis." Edited by Gianluigi Forloni. *PLOS ONE* 12 (6): e0179857. https://doi.org/10.1371/journal.pone.0179857.
- Didelez, Vanessa. 2016. "Data Always Be Preceded by Specifying a Target Experimental Trial?" International Journal of Epidemiology 45 (6): 3.
- Driver, J. A., A. Beiser, R. Au, B. E. Kreger, G. L. Splansky, T. Kurth, D. P. Kiel, K. P. Lu, S. Seshadri, and P. A. Wolf. 2012. "Inverse Association Between Cancer and Alzheimer's Disease: Results from the Framingham Heart Study." *BMJ* 344 (mar12 1): e1442–42. https://doi.org/10.1136/bmj.e1442.
- Driver, Jane A. 2014. "Inverse Association Between Cancer and Neurodegenerative Disease: Review of the Epidemiologic and Biological Evidence." *Biogerontology* 15 (6): 547–57. https://doi.org/10.1007/s10522-014-9523-2.

- Driver, Jane A., Xiao Zhen Zhou, and Kun Ping Lu. 2015. "Pin1 Dysregulation Helps to Explain the Inverse Association Between Cancer and Alzheimer's Disease." *Biochimica Et Biophysica Acta (BBA) General Subjects* 1850 (10): 2069–76. https://doi.org/10.1016/j.bbagen.2014.12.025.
- Frain, Laura, David Swanson, Kelly Cho, David Gagnon, Kun Ping Lu, Rebecca A. Betensky, and Jane Driver. 2017. "Association of Cancer and Alzheimer's Disease Risk in a National Cohort of Veterans." *Alzheimer's & Dementia* 13 (12): 1364–70. https://doi.org/10.1016/j.jalz.2017.04.012.
- Freedman, Daryl Michal, Jincao Wu, Honglei Chen, Ralph W. Kuncl, Lindsey R. Enewold, Eric A. Engels, Neal D. Freedman, and Ruth M. Pfeiffer. 2016. "Associations Between Cancer and Alzheimer's Disease in a U.S. Medicare Population." *Cancer Medicine* 5 (10): 2965–76. https://doi.org/10.1002/cam4.850.
- Ganguli, Mary. 2015. "Cancer and Dementia: It's Complicated." Alzheimer Disease & Associated Disorders 29 (2): 177–82. https://doi.org/10.1097/WAD.000000000000086.
- Hanson, Heidi A., Kevin P. Horn, Kelli M. Rasmussen, John M. Hoffman, and Ken R. Smith. 2016. "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, April, gbw040. https://doi.org/10.1093/geronb/gbw040.
- Harris, Richard A., Lauren Tindale, and Robert C. Cumming. 2014. "Age-Dependent Metabolic Dysregulation in Cancer and Alzheimer's Disease." *Biogerontology* 15 (6): 559–77. https://doi.org/10.1007/s10522-014-9534-z.
- Hernán, Miguel A., Brian C. Sauer, Sonia Hernández-Díaz, Robert Platt, and Ian Shrier. 2016. "Specifying a Target Trial Prevents Immortal Time Bias and Other Self-Inflicted Injuries in Observational Analyses." Journal of Clinical Epidemiology 79 (November): 70–75. https://doi.org/10.1016/j.jclinepi.2016.04.014.
- Hernán, Miguel A, and James M Robins. 2020. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
- Hernán, Miguel Ángel, Babette Brumback, and James M. Robins. 2000. "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men:" *Epidemiology* 11 (5): 561–70. https://doi.org/10.1097/00001648-200009000-00012.
- Ikram, M. Arfan, Guy Brusselle, Mohsen Ghanbari, André Goedegebure, M. Kamran Ikram, Maryam Kavousi, Brenda C. T. Kieboom, et al. 2020. "Objectives, Design and Main Findings Until 2020 from the Rotterdam Study." *European Journal of Epidemiology* 35 (5): 483–517. https://doi.org/10.1007/s10654-020-00640-5.
- Labrecque, Jeremy A., and Sonja A. Swanson. 2017. "Target Trial Emulation: Teaching Epidemiology and Beyond." European Journal of Epidemiology 32 (6): 473–75. https://doi.org/10.1007/s10654-017-0293-4.
- Li, Jingyi, Chunfen Mo, Yifan Guo, Bowen Zhang, Xiao Feng, Qiuyue Si, Xiaobo Wu, et al. 2021. "Roles of Peptidyl-Prolyl Isomerase Pin1 in Disease Pathogenesis." *Theranostics* 11 (7): 3348–58. https://doi.org/10.7150/thno.45889.
- Ma, Ling-Li, Jin-Tai Yu, Hui-Fu Wang, Xiang-Fei Meng, Chen-Chen Tan, Chong Wang, and Lan Tan. 2014. "Association Between Cancer and Alzheimer's Disease: Systematic Review and Meta-Analysis." *Journal of Alzheimer's Disease* 42 (2): 565–73. https://doi.org/10.3233/JAD-140168.
- Musicco, M., F. Adorni, S. Di Santo, F. Prinelli, C. Pettenati, C. Caltagirone, K. Palmer, and A. Russo. 2013. "Inverse Occurrence of Cancer and Alzheimer Disease: A Population-Based Incidence Study." Neurology 81 (4): 322–28. https://doi.org/10.1212/WNL.0b013e31829c5ec1.
- Nudelman, Kelly N. H., Brenna C. McDonald, Debomoy K. Lahiri, and Andrew J. Saykin. 2019. "Biological Hallmarks of Cancer in Alzheimer's Disease." *Molecular Neurobiology* 56 (10): 7173–87. https://doi.org/10.1007/s12035-019-1591-5.
- Nudelman, Kelly N. H., Shannon L. Risacher, John D. West, Brenna C. McDonald, Sujuan Gao, Andrew J. Saykin, and Alzheimer's Disease Neuroimaging Initiative. 2014. "Association of Cancer History

- with Alzheimer's Disease Onset and Structural Brain Changes." Frontiers in Physiology 5 (October). https://doi.org/10.3389/fphys.2014.00423.
- Olson, Brennan, and Daniel L. Marks. 2019. "Pretreatment Cancer-Related Cognitive ImpairmentMechanisms and Outlook." *Cancers* 11 (5): 687. https://doi.org/10.3390/cancers11050687.
- Ording, Anne G., Erzsébet Horváth-Puhó, Katalin Veres, M. Maria Glymour, Mikael Rørth, Henrik T. Sørensen, and Victor W. Henderson. 2020. "Cancer and Risk of Alzheimer's Disease: Small Association in a Nationwide Cohort Study." *Alzheimer's & Dementia* 16 (7): 953–64. https://doi.org/10.1002/alz.12090.
- Ospina-Romero, Monica, M. Maria Glymour, Eleanor Hayes-Larson, Elizabeth Rose Mayeda, Rebecca E. Graff, Willa D. Brenowitz, Sarah F. Ackley, John S. Witte, and Lindsay C. Kobayashi. 2020. "Association Between Alzheimer Disease and Cancer With Evaluation of Study Biases: A Systematic Review and Meta-Analysis." *JAMA Network Open* 3 (11): e2025515. https://doi.org/10.1001/jamanetworkopen. 2020.25515.
- Papin, Stéphanie, and Paolo Paganetti. 2020. "Emerging Evidences for an Implication of the Neurodegeneration-Associated Protein TAU in Cancer." Brain Sciences 10 (11): 862. https://doi.org/10.3390/brainsci10110862.
- Schmidt, Sigrun A. J., Anne G. Ording, Erzsébet Horváth-Puhó, Henrik T. Sørensen, and Victor W. Henderson. 2017. "Non-Melanoma Skin Cancer and Risk of Alzheimer's Disease and All-Cause Dementia." Edited by Kristel Sleegers. *PLOS ONE* 12 (2): e0171527. https://doi.org/10.1371/journal.pone.0171527.
- Snyder, Heather M., Tim Ahles, Stuart Calderwood, Maria C. Carrillo, Honglei Chen, Chung-Chou H. Chang, Suzanne Craft, et al. 2017. "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 13 (3): 267–73. https://doi.org/10.1016/j.jalz.2016.11.002.
- Stensrud, Mats J., Jessica G. Young, Vanessa Didelez, James M. Robins, and Miguel A. Hernán. 2020. "Separable Effects for Causal Inference in the Presence of Competing Events." *Journal of the American Statistical Association*, June, 1–9. https://doi.org/10.1080/01621459.2020.1765783.
- Sun, Ming, Youxin Wang, Jan Sundquist, Kristina Sundquist, and Jianguang Ji. 2020. "The Association Between Cancer and Dementia: A National Cohort Study in Sweden." Frontiers in Oncology 10 (February): 73. https://doi.org/10.3389/fonc.2020.00073.
- van der Willik, Kimberly D., Sanne B. Schagen, and M. Arfan Ikram. 2018. "Cancer and Dementia: Two Sides of the Same Coin?" European Journal of Clinical Investigation 48 (11): e13019. https://doi.org/10.1111/eci.13019.
- White, Robert S, Richard B Lipton, Charles B Hall, and Joshua R Steinerman. 2013. "Nonmelanoma Skin Cancer Is Associated with Reduced Alzheimer Disease Risk." *Alzheimer Disease*, 7.
- Young, Jessica G., Mats J. Stensrud, Eric J. Tchetgen Tchetgen, and Miguel A. Hernán. 2020. "A Causal Framework for Classical Statistical Estimands in Failure-Time Settings with Competing Events." *Statistics in Medicine* 39 (8): 1199–1236. https://doi.org/10.1002/sim.8471.