

# Epidemiology

## Towards a clearer causal question underlying the association between cancer and dementia --Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Full Title:</b>	Towards a clearer causal question underlying the association between cancer and dementia
<b>Article Type:</b>	Original Article
<b>Corresponding Author:</b>	L. Paloma Rojas-Saunero, MD Erasmus MC rotterdam, The Netherlands NETHERLANDS
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	Erasmus MC
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	L. Paloma Rojas-Saunero, MD
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	L. Paloma Rojas-Saunero, MD Kimberly D. van der Willik, MD, PhD Sanne B. Schagen, PhD M. Arfan Ikram, MD, PhD Sonja A. Swanson, ScD
<b>Order of Authors Secondary Information:</b>	
<b>Abstract:</b>	Background: Several observational studies have described an inverse association between cancer diagnosis and dementia. Several biological mechanisms and sources of bias have been proposed. Since bias cannot be assessed without a clear causal question, we propose to study the controlled direct effect of the protein Pin1 on the risk of dementia. With this case-study, we will outline the needed assumptions and potential sources of bias and exemplify how these translate into the analytic decisions under the guidance of causal directed acyclic graphs. Methods: We used data from the Rotterdam Study, a population-based cohort. We estimate the association between cancer diagnosis (as proxy for Pin1) and dementia diagnosis using two different proxy methods and with confounding and censoring for death addressed with inverse probability weights. We estimate and compare the complements of a weighted Kaplan-Meier survival estimator at 20-years of follow-up. Results: Out of 3634 participants, 899 (25%) were diagnosed with cancer, of whom 53 (6%) had dementia, and 567 (63%) died. Among those without cancer, 15% (411) were diagnosed with dementia, and 667 (24%) died over follow-up. Depending on the confounding and selection bias control, and the way in which cancer was used as a time-varying proxy, the risk ratio ranged from 0.70 (CI95%: 0.49, 0.93) to 1.05 (CI95%: 0.79, 1.29). Conclusion: Being explicit about the underlying mechanism of interest and defining a clear causal contrast is key to maximize what we can learn about this association given available or readily-collected data, and to defining, detecting, and preventing potential biases.
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Word count, abstract (Enter "0" if not applicable.)	248
Word count, main text. Enter the number	3758

of words in the main text. For research articles, main text typically includes the introduction, methods, results, and discussion. Do not include words in title page, bibliography, tables, figure legends, or figures. Do not paste the text from the document itself into this box.	
Total word count. Enter the total word count for your submission, excluding the title page and abstract. Include the following in the total word count: (a) main text (for research articles, this typically includes the introduction, methods, results, and discussion), (b) bibliography, (c) tables, (d) figure legends, and (e) figures (calculated as 250 words per figure, each panel in a panel figure should be separately counted). Do not count words in Supplemental Digital Content. Do not paste the text from the document itself into this box.	7181
<p>Funding sources All relevant sources of funding should be included here and on the title page of the manuscript with the heading "Source of Funding."</p> <p>-Enter "none" if the work was completed without specific funding support.</p> <p>-If the result reported in the submission corresponds directly to the specific aims of a source (or sources) of funding, then describe that source of funding as: "The results reported herein correspond to specific aims of grant (or other source of support) XXX to investigator YYY from ZZZ", where XXX is a grant or project number, YYY is the Principal Investigator of the grant or project, and ZZZ is the funding agency.</p> <p>-Describe all other sources of support as: "This work was (also) supported by grant(s) (or other source of support) XXX from ZZZ", where '(also)' is inserted only if the listed support is in addition to support corresponding directly to a specific aim, XXX is a grant or project number, and ZZZ is a funding agency. Additional sources of support should be added serially (e.g., grants XXX1 from ZZZ1, XXX2 from ZZZ2, and XXX3 from ZZZ3."</p> <p>Sources of support can include general salary support, which may not have a</p>	This study was partly funded by ZonMW Memorabel (projectnr 73305095005)

<p>grant or project number.</p> <p>Grant or project numbers should be provided in a format that allows interested parties to find the grant in publicly available databases provided by many funding agencies.</p>	
<p>Describe the process by which someone else could obtain the data and computing code required to replicate the results reported in your submission. If one or the other is not available, please write "The [data, computing code, or both] are not available for replication because" and add an explanation for why one or the other is not available. Authors of submissions without original results, such as commentaries or letters, should enter "Not applicable." Your answer to this question will not affect the editors' decision about the submission's suitability for publication. If your paper is published, your response to this question will be included in the manuscript footnotes.</p>	<p>Rotterdam Study can be obtained via requests directed toward the management team of the Rotterdam Study (<a href="mailto:secretariat.epi@erasmusmc.nl">secretariat.epi@erasmusmc.nl</a>), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository. Code is available at: <a href="https://github.com/palolili23/2021_cancer_dementia">https://github.com/palolili23/2021_cancer_dementia</a></p>
<p>RETAINED RIGHTS: Except for copyright, other proprietary rights related to the Work (e.g., patent or other rights to any process or procedure) shall be retained by the author. To reproduce any text, figures, tables, or illustrations from this Work in future works of their own, the author must obtain written permission from Wolters Kluwer Health, Inc. ("WKH").</p> <p>ORIGINALITY: Each author warrants that his or her submission to the Work is original, does not infringe upon, violate, or misappropriate any copyright or other intellectual property rights, or any other proprietary right, contract or other right or interest of any third party, and that he or she has full power to enter into this agreement. Neither this Work nor a similar work has been published nor shall be submitted for publication elsewhere while under consideration by this Publication.</p>	<p>I agree</p>

**AUTHORSHIP RESPONSIBILITY:** Each author warrants that he or she has participated sufficiently in the intellectual content, the analysis of data, if applicable, and the writing of the Work to take public responsibility for it. Each has reviewed the final version of the Work, believes it represents valid work, and approves it for publication. Moreover, should the editors of the Publication request the data upon which the work is based, they shall produce it.

**PREPRINTS:** Upon acceptance of the article for publication, each author warrants that he/she will promptly remove any prior versions of this Work (normally a preprint) that may have been posted to an electronic server.

**DISCLAIMER:** Each author warrants that this Work contains no libelous or unlawful statements and does not infringe or violate the publicity or privacy rights of any third party, libel or slander any third party, contain any scandalous, obscene, or negligently prepared information, or infringe or violate any other personal or proprietary right of others. Each author warrants that the Work does not contain any fraudulent, plagiarized or incorrectly attributed material. Each author warrants that all statements contained in the Work purporting to be facts are true, and any formula or instruction contained in the Work will not, if followed accurately, cause any injury, illness, or damage to the user. If excerpts (e.g., text, figures, tables, illustrations, or audio/video files) from copyrighted works are included, a written release will be secured by the author prior to submission, and credit to the original publication will be properly acknowledged. Each author further warrants that he or she has obtained, prior to submission,

<p>written releases from patients whose names or likenesses are submitted as part of the Work. Should the Editor or WKH request copies of such written releases, the author shall provide them in a timely manner.</p> <p>DISCLOSURES/CONFLICT OF INTEREST</p> <p>Each author must identify any financial interests or affiliations with institutions, organizations, or companies relevant to the manuscript by completing the form below. Additionally, any financial associations involving a spouse, partner or children must be disclosed as well.</p> <p>Note: Some sections below come from the ICMJE Uniform Disclosure Form for Potential Conflicts of Interest at <a href="http://www.icmje.org/downloads/coi_disclosure.pdf">http://www.icmje.org/downloads/coi_disclosure.pdf</a> (dated July 2010).</p>	
<p>Did you or your institution at any time receive payment or support in kind for any aspect of the submitted work (including but not limited to grants, consulting fee or honorarium, support for travel to meetings for the study or other purposes, fees for participation in review activities such as data monitoring boards, statistical analysis, end point committees, and the like, payment for writing or reviewing the manuscript, provision of writing assistance, medicines, equipment, or administrative support, etc...)?</p>	No
<p>Other: Did you or your institution at any time receive additional payments or support in kind for any aspect of the submitted work?</p>	No
<p>Please indicate whether you have financial relationships (regardless of amount of compensation) with entities. You should report relationships that were present during the 36 months prior to submission including board membership, consultancy, employment, expert</p>	No

<p>testimony, grants/grants pending, payment for lectures including service on speakers bureaus, payment for manuscript preparation, patents (planned, pending or issued), royalties, payment for development of educational presentations, stock/stock options, travel/accommodations/meeting expenses unrelated to activities listed (for example, if you report a consultancy above there is no need to report travel related to that consultancy), etc.</p>	
<p>Other (err on the side of full disclosure): Please indicate whether you have any additional financial relationships (regardless of amount of compensation) with entities. You should report relationships that were present during the 36 months prior to submission.</p>	
<p>Other Relationships</p> <p>Are there other relationships or activities that readers could perceive to have influenced, or that give the appearance of potentially influencing, what you wrote in the submitted work?</p>	<p>No other relationships/conditions/circumstances that present potential conflict of interest</p>
<p>AUTHOR'S OWN WORK: In consideration of WKH's publication of the Work, the author hereby transfers, assigns, and otherwise conveys all his/her copyright ownership worldwide, in all languages, and in all forms of media now or hereafter known, including electronic media such as CD-ROM, Internet, and Intranet, to WKH. If WKH should decide for any reason not to publish the Work, WKH shall give prompt notice of its decision to the corresponding author, this agreement shall terminate, and neither the author nor WKH shall be under any further liability or obligation. Each author grants WKH the rights to use his or her name and biographical data (including professional affiliation) in the Work and in its or the journal's promotion. Notwithstanding the foregoing, this paragraph shall not apply, and any transfer made pursuant to this paragraph shall be null and void if (i) the Work has been accepted by WKH for publication, and (ii) the author chooses to have the</p>	<p>I agree</p>

Work published by WKH as an open access publication.

WORK MADE FOR HIRE: If this Work or any element thereof has been commissioned by another person or organization, or if it has been written as part of the duties of an employee, an authorized representative of the commissioning organization or employer must also sign this form stating his or her title in the organization.

GOVERNMENT EMPLOYEES: If the Work or a portion of it has been created in the course of any author's employment by the United States Government, check the "Government" box at the end of this form. A work prepared by a government employee as part of his or her official duties is called a "work of the U.S. Government" and is not subject to copyright. If it is not prepared as part of the employee's official duties, it may be subject to copyright.

INSTITUTIONAL REVIEW BOARD/ANIMAL CARE COMMITTEE APPROVAL: Each author warrants that his or her institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.

WARRANTIES: Each author warranty made in this form is for the benefit of WKH and the Editor; each author agrees to defend, indemnify, and hold harmless those parties for any breach of such warranties.

The journal will permit the author(s) to deposit for display a "final peer-reviewed manuscript" (the final manuscript after peer-review and acceptance for publication but prior to the publisher's copyediting, design,

I agree

<p>formatting, and other services) 12 months after publication of the final article on the author's personal web site, university's institutional repository or employer's intranet, subject to the following:</p> <p>* You may only deposit the final peer-reviewed manuscript.</p> <p>* You may not update the final peer-reviewed manuscript text or replace it with a proof or with the final published version.</p> <p>* You may not include the final peer-reviewed manuscript or any other version of the article on any commercial site or in any repository owned or operated by any third party. For authors of articles based on research funded by the National Institutes of Health ("NIH"), Wellcome Trust, Howard Hughes Medical Institute ("HHMI"), or other funding agency, see below for the services that WKH will provide on your behalf to comply with "Public Access Policy" guidelines.</p> <p>* You may not display the final peer-reviewed manuscript until twelve months after publication of the final article.</p> <p>* You must attach the following notice to the final peer-reviewed manuscript: "This is a non-final version of an article published in final form in (provide complete journal citation)".</p> <p>* You shall provide a link in the final peer-reviewed manuscript to the journal website.</p>	
<p>"Public Access Policy" Funding Disclosure Please disclose below if you have received funding for research on which your article is based from any of the</p>	<p>Other - Please list in the following comments box</p>



following organizations:	
Other (Please list): as follow-up to ""Public Access Policy" Funding Disclosure Please disclose below if you have received funding for research on which your article is based from any of the following organizations:"	ZonMw
Please select:	Author's Own Work
Any additional comments?	
<u>Compliance with RCUK and Wellcome Trust Open Access Policies</u>  Both the Research Councils UK (RCUK) and the Wellcome Trust have adopted policies regarding Open Access to articles that have been funded by grants from the RCUK or the Wellcome Trust. If either "Wellcome Trust" or "Research Councils UK (RCUK)" has been selected above, and the authors of the applicable article choose to have the article published as an open access publication, the following policies will apply:  * If the article is to be published pursuant to the "Gold" route of Open Access, both the RCUK and the Wellcome Trust require that WKH make the article freely available immediately pursuant to the Attribution 4.0 Creative Commons License, currently found at <a href="http://creativecommons.org/licenses/by/4.0/legalcode">http://creativecommons.org/licenses/by/4.0/legalcode</a> (the "CC BY License"). The CC BY License is the most accommodating of the Creative Commons licenses and allows others to distribute, remix, tweak, and build upon the article, even commercially, as long as they credit the authors for the original creation.  * If the article is to be published pursuant to the "Green" route of Open Access, both the RCUK and the Wellcome Trust require that WKH make	I agree

the article freely available within six months pursuant to the Attribution-NonCommercial 4.0 Creative Commons License, currently found at <http://creativecommons.org/licenses/by-nc/4.0/legalcode> (the “CC BY-NC License”). The CC BY-NC License allows others to remix, tweak, and build upon the article non-commercially, and although their new works must also acknowledge the authors for the original creation and be non-commercial, they don't have to license their derivative works on the same terms.

As a service to our authors, WKH will identify the National Library of Medicine (NLM) articles that require deposit pursuant to the RCUK and Wellcome Trust policies described in this section. This Copyright Transfer Agreement provides the mechanism for identifying such articles.

WKH will transmit the final peer-reviewed manuscript of an article based on research funded in whole or in part by either RCUK or the Wellcome Trust to Pub Med Central.

Upon NIH request, it remains the legal responsibility of the author to confirm with NIH the provenance of his/her manuscript for purposes of deposit. Author will not deposit articles him/herself. Author will not alter the final peer-reviewed manuscript already transmitted to NIH.

With respect to the “Green” route of Open Access, author will not authorize the display of the final peer-reviewed manuscript prior to 6 months following publication of the final article.

<p>Authors of articles that have been funded from grants from the RCUK or the Wellcome Trust are required to sign the WKH Open Access License Agreement prior to publication of the applicable article. Please contact the Editorial Office of the applicable journal to receive the Open Access License Agreement that is to be signed in connection with the publication of the article.</p>	
<p>I am the person in question for this submission or otherwise have approval to complete this agreement.</p>	<p>I agree</p>
<p>CME/CE Disclosure</p> <p>Each author must identify and disclose any financial associations involving a spouse, partner or children by completing the Family Disclosure question below, and whether any off-label uses or unapproved drugs or devices are discussed in his/her manuscript by completing the Off-Label Use/Unapproved Drugs or Products question below. In the event that the Work is published as a continuing education or continuing medical education article, this information will be provided to the accrediting body and may be included in the published article. When applicable, articles accepted for publication may need to comply with additional standards related to CME or CE accreditation. Please refer to guidelines for authors for details.</p> <p>WKH and its affiliates reserve the right to publish the manuscript as a continuing education article.</p>	<p>I agree</p>
<p>Family Disclosure</p> <p>Do your children or your spouse or partner have financial relationships with entities that have an interest in the content of the submitted work?</p>	<p>No other relationships/conditions/circumstances that present potential conflict of interest</p>
<p>Off-Label Use/Unapproved Drugs or Products</p>	<p>I will not discuss unlabeled/investigational uses of any commercial product or device</p>

If your manuscript discusses an unlabeled use of a commercial product or device or an investigational use of a product or device not yet approved by the FDA for any purpose, you must specifically disclose in the manuscript that the product is not labeled for the use under discussion or that the product is still investigational. Please check the item below that applies to you

## Original Research Article

### Towards a clearer causal question underlying the association between cancer and dementia

L. Paloma Rojas-Saunero<sup>1\*</sup>, Kimberly D. van der Willik<sup>1\*</sup>, Sanne B. Schagen<sup>2,3</sup>, M. Arfan Ikram<sup>1</sup>, Sonja A. Swanson<sup>1,4,5</sup>

#### Affiliations:

1. Department of Epidemiology, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, the Netherlands
2. Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, the Netherlands
3. Brain and Cognition, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands
4. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, USA
5. CAUSALab, Harvard T. H. Can School of Public Health, Boston, MA, USA

\* These authors contributed equally to this work

#### Corresponding author:

L. Paloma Rojas-Saunero MD, PhD Candidate,  
Department of Epidemiology, Erasmus University Medical Center  
Mailing address: PO Box 2040, 3000CA Rotterdam, the Netherlands.  
Telephone number: +310626359216  
E-mail address [l.rojassaunero@erasmusmc.nl](mailto:l.rojassaunero@erasmusmc.nl)

**Running head:** A causal question for cancer-dementia association

**Conflicts of interest:** none declared.

**Sources of financial support:** This study was partly funded by ZonMW Memorabel (projectnr 73305095005)

**Data and code access:** Rotterdam Study can be obtained via requests directed toward the management team of the Rotterdam Study ([secretariat.epi@erasmusmc.nl](mailto:secretariat.epi@erasmusmc.nl)), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository. Code is available at: [https://github.com/palolili23/2021\\_cancer\\_dementia](https://github.com/palolili23/2021_cancer_dementia)

**Acknowledgments:** The authors gratefully acknowledge the study participants of the Ommoord district and their general practitioners and pharmacists for their devotion in contributing to the Rotterdam Study. The authors also thank all staff that facilitated assessment of participants in the Rotterdam Study throughout the years. Special thanks to Frank J.A. van Rooij for managing the data.

## Towards a clearer causal question underlying the association between cancer and dementia

### Abstract

**Background:** Several observational studies have described an inverse association between cancer diagnosis and dementia. Several biological mechanisms and sources of bias have been proposed. Since bias cannot be assessed without a clear causal question, we propose to study the controlled direct effect of the protein Pin1 on the risk of dementia. With this case-study, we will outline the needed assumptions and potential sources of bias and exemplify how these translate into the analytic decisions under the guidance of causal directed acyclic graphs. **Methods:** We used data from the Rotterdam Study, a population-based cohort. We estimate the association between cancer diagnosis (as proxy for Pin1) and dementia diagnosis using two different proxy methods and with confounding and censoring for death addressed with inverse probability weights. We estimate and compare the complements of a weighted Kaplan-Meier survival estimator at 20-years of follow-up. **Results:** Out of 3634 participants, 899 (25%) were diagnosed with cancer, of whom 53 (6%) had dementia, and 567 (63%) died. Among those without cancer, 15% (411) were diagnosed with dementia, and 667 (24%) died over follow-up. Depending on the confounding and selection bias control, and the way in which cancer was used as a time-varying proxy, the risk ratio ranged from 0.70 (CI95%: 0.49, 0.93) to 1.05 (CI95%: 0.79, 1.29). **Conclusion:** Being explicit about the underlying mechanism of interest and defining a clear causal contrast is key to maximize what we can learn about this association given available or readily-collected data, and to defining, detecting, and preventing potential biases.

## **Towards a clearer causal question underlying the association between cancer and dementia**

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<sup>1–4</sup>. These findings have motivated substantial research toward mechanistic explanations, including molecular and genetic pathways that may explain this association<sup>5–12</sup>. These research questions have led to discussions of repurposing or augmenting current cancer treatments, including chemotherapy, for dementia<sup>13</sup>.

Nevertheless, inferring any treatment or mechanistic effects from the observed cancer-dementia 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<sup>9,14</sup>. However, understanding these or other sources of bias first requires making explicit a 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<sup>15,16</sup>.

To illustrate the complexities of inferring hypothetical or available cancer treatments' effects on dementia from the observed cancer-dementia 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<sup>11,12,17</sup>. If we – hypothetically speaking – 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 dementia over time compared to standard treatments?*

To explore how we might learn about this effect using real-world data on cancer and dementia, we progressively build a causal directed acyclic graph (DAG) to connect this 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 decisions using data collected from the Rotterdam Study, a population-based cohort study. In our discussion, we describe how other investigators may expand this exercise for other possible causal questions, including repurposing existing chemotherapy regimens or identifying other novel drug targets.

## 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 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 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<sup>18</sup> in Figure 1, we present Pin1 expression as  $P_t$  and dementia diagnosis by time  $t + k$  as  $Y_{t+k}$ . Both may share causes  $L_1$ , and to assess the causal relationship between  $P_t$  and  $Y_{t+k}$  we would thus adjust for  $L_1$ . Previous dementia studies have described age, sex, educational level and race/ethnicity as a minimal adjusting set of covariates<sup>4</sup>. However, environmental and behavioral factors such as smoking may translate into Pin1 over-expression<sup>19</sup> and are also related to the development of dementia<sup>20</sup> and should



therefore be adjusted for. Although  $L_1$  can be time-varying in nature, we only depict  $L_1$  at one time-point for readability.

< 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 diagnosed through biopsies, previous studies have speculated that Pin1 over-expression partly explains the inverse association between cancer diagnosis and dementia, though Pin1 was not explicitly part of the research question<sup>21–28</sup>. Unlike measuring Pin1 at the same time for all participants (though 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 collected during follow-up. We depict this feature in Figure 2, where  $C_t$  and  $C_{t+1}$  represent *cancer diagnosis* over time, the measured proxy of  $P_t$  and  $P_{t+1}$  respectively. Although this means we would measure the association between cancer diagnosis over time and dementia in the observed data, we are assuming that the captured effect is only through the pathway that involves Pin1 expression over time. That is, we only have measurements of  $C_{t+1}$  and  $Y_{t+k}$  and some subset of  $L_1$ , but our question remains focused on the effect of  $P_t$ .

We note that this means we already are changing our question from one of estimating the effect of Pin1 to one that is testing the sharp causal null hypothesis that Pin1 has no effect on dementia in any individuals. That is, an association between cancer and dementia is at best evidence against a null effect, and it would take substantial and unknown assumptions about the relationship between Pin1 and cancer diagnosis to describe how the magnitude of a cancer-dementia association is related to the likely magnitude of a Pin1-dementia effect.

< Figure 2 >

Another challenge that arises when choosing cancer diagnosis as the proxy for Pin1 is defining the time zero, i.e. the time when eligibility criteria are met, “treatment” is assigned, and screening for dementia would begin<sup>29</sup>. Emulating the eligibility criteria to participate in a trial for Pin1-targeting drugs will not necessarily be possible in a cohort setting recruiting participants for discordant reasons. Moreover, the latter may not align with the moment at which cancer diagnosis is measured since it happens continuously over follow-up and this situation may introduce immortal-time bias<sup>29</sup>. For example, a study performed using data from the Framingham Heart Study<sup>21</sup> defined the exposed group with cancer as those participants with prevalent or incident cancer diagnosis (alternatively defined as “ever cancer”)<sup>2</sup>. 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 dementia diagnosis over follow-up can be defined as “ever cancer”<sup>29</sup>.

This problem is partly mitigated by recognizing the time-varying nature of cancer diagnosis. Some studies have considered cancer diagnosis as a time-dependent exposure<sup>2,24,31</sup>. The price we pay for 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. The time between first biological changes that eventually can lead to cancer and cancer manifestation can range between five and forty years<sup>32</sup>. 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  $D_{t+1}$  and an arrow between  $D_{t+1}$  and  $P_{t+1}$  that represents a deterministic association such as that  $P_{t+1}$  is only observed if  $D_{t+1}$  is zero. In addition, we added an arrow between  $L_1$  and  $D_{t+1}$ , 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.

< Figure 3 >

Another challenge to address 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  $t + 1$ . 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  $C_{t+1}$  has been measured, represented as  $D_{t+k}$  and which acts as a competing event of  $Y_{t+k}$  because if a participant dies by  $t + k$ , the participant cannot subsequently develop dementia. Furthermore, since  $D_{t+k}$  and  $Y_{t+k}$  are events related to aging and its consequences,  $L_2$  represents the shared causes of both events such as cardiovascular conditions. We also include an arrow between  $L_1$  and  $D_{t+1}$  following the argument discussed previously.

< Figure 4 >

In the setting where  $P_{t+1}$  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  $P_{t+1}$  and  $D_{t+k}$ , a total effect would quantify the effect of  $P_{t+1}$  on  $Y_{t+k}$  that does not include any pathway mediated through  $D_{t+k}$ <sup>33</sup>. 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  $C_{t+1}$  and  $D_{t+k}$ . In this setting, a total effect of  $P_{t+1}$  in  $Y_{t+k}$  includes the indirect causal pathway mediated by the effect of cancer diagnosis on mortality, which may translate into an inverse association<sup>33</sup>.

To isolate the direct effect of  $P_{t+1}$  in  $Y_{t+k}$  through measurement of  $C_{t+1}$  we have at least two alternatives of causal questions we can ask. One, we can imagine a causal question where we decompose the total effect of cancer by the different pathways that affect dementia and death

separately<sup>34</sup>. Two, we can define a scenario where death could have been prevented. The latter is defined as the controlled direct effect, where death is considered as an independent censoring event by relying on the assumption that we have measured all shared causes  $L_2$  to block the pathway  $Y_{t+k} \leftarrow L_2 \rightarrow D_{t+k} \leftarrow C_{t+1} \leftarrow P_{t+1}$ . Previous studies have defined death as a censoring event<sup>25</sup>, although failed to explicitly define the independent censoring assumption and did not consider its plausibility. Moreover, adjusting for time-fixed shared causes between dementia and death may be insufficient to block this pathway, and therefore time-varying measurements of  $L_2$  should be considered.

To summarize, the complexity of the causal structure that describes the effect of Pin1 on dementia risk while proxying Pin1 with cancer diagnosis illustrates the potential sources of bias, as observed in Figure 5. Even so, this is a simplified version since we omitted additional arrows from  $L_1$  to  $C_t$  and  $C_{t+2}$  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<sup>18</sup>.

< 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 dementia risk.

## 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 sub-cohort had follow-up visits between 1993-1995, 1997-1999, 2002-2005, and 2008-2010, while the second sub-cohort had follow-up

visits between 2004 and 2005, and between 2011 and 2012<sup>35</sup>. 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 sub-cohort 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 of follow-up.

### 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*”<sup>21</sup>, 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

1  
2  
3  
4 probability of cancer diagnosis conditional on baseline covariates such as age at study entry,  
5  
6 sex, educational attainment, cohort, smoking status. In contrast, for Scenario B, weights were  
7  
8 defined to represent the product of the inverse probability of being diagnosed with cancer over  
9  
10 time, conditional on the time-varying covariate history prior to cancer diagnosis<sup>36</sup>. Baseline  
11  
12 covariates included age at study entry, sex, ApoE4 status, educational attainment and the time-  
13  
14 varying covariates such as smoking status, systolic blood pressure, body mass index and  
15  
16 prevalent and incident hypertension and diabetes.  
17  
18

19  
20 Inverse probability censoring weights for death were calculated, assuming independent  
21  
22 censoring conditional on measured covariates. In Scenario A, weights represent the inverse of  
23  
24 the probability of not dying conditional on cancer diagnosis (ever vs. never) and baseline  
25  
26 covariates such as age, educational attainment, ApoE4, and baseline measurements of  
27  
28 smoking status, hypertension status, systolic blood pressure, BMI, history of diabetes and  
29  
30 cohort. For individuals who died, their censoring weight is zero<sup>18</sup>. In Scenario B time-varying  
31  
32 weights represent the product of the inverse probability of surviving in each year prior to  $t$ ,  
33  
34 conditional on the measured shared causes of death and dementia. For an individual who has  
35  
36 died by time  $t$ , the year  $t$  censoring weight is zero<sup>33</sup>. Weights were fitted including the same  
37  
38 covariates used to estimate weights for time-varying cancer diagnosis, though we additionally  
39  
40 added time-varying cancer, stroke, and heart disease diagnosis as predictors for death. Further  
41  
42 details on modeling specifications and weights assessment are presented in the eAppendix.  
43  
44  
45  
46  
47

48 To estimate the controlled direct effect of Pin1 on the risk of dementia, we compared the  
49  
50 complement of a weighted Kaplan-Meier survival estimator for participants who developed  
51  
52 cancer versus those who did not, with time indexed in years. The weights are time-varying by  
53  
54 follow-up year, defined as a product of the year-specific inverse probability of treatment weights  
55  
56 and the inverse probability of censoring by death weights. Estimates of the controlled direct  
57  
58 effect are presented as 20-year risk differences (RD) and risk ratios (RR). All 95% confidence  
59  
60  
61  
62  
63  
64  
65

intervals were calculated using percentile-based bootstrapping based on 500 bootstrap samples.

Estimates 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 also calculated hazard ratios (HR). Hazards, unlike risks, inherently condition on surviving both dementia and death, and as such a causal interpretation is problematic<sup>33</sup>.

Since the conditional independent censoring assumption is untestable, we compute Peterson upper and lower bounds<sup>37</sup> to represent: (1) the extreme scenario of independence, that refers to a scenario where 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 diagnosis 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 the corresponding github repository.

### 3. Results

Participants had a mean age of 64.5 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) were diagnosed with 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) remained 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 eFigure 1 and 2 in the eAppendix.

Results for all scenarios are presented 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, 95% Confidence interval) of 0.70 (0.49,0.93) and a hazard ratio [HR, (95% Confidence interval)] of 0.52 (0.39,0.69). Though adjusting for measured confounding only minimally changed the observed association, the association is closer to the null after including censoring weights for death [RR: 0.91 (0.65,1.19); HR: 0.72 (0.54,0.98)]. In contrast, using Scenario B's analytic design, the fully adjusted model results in a RR of RR: 1.05 (0.79,1.29) and a HR of 1.09 (0.80,1.50), though confidence intervals cross the null. The cumulative incidence curves for dementia for participants free of cancer diagnosis and with incident cancer diagnosis, 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. Bounds (rather than point estimates) were wide and covered the null (Lower bound RR: 0.39; Upper bound RR: 2.08).

#### 4. Discussion

Several observational studies have defined "cancer diagnosis" as an 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



1  
2  
3  
4 identify and disentangle potential sources of bias and can guide analytic decisions. We showed  
5  
6 how estimates can change substantially according to alternative, yet explicit, assumptions.  
7  
8

9  
10 For instance, a key challenge of cancer diagnosis as a proxy for Pin1 is the incapacity of  
11 defining a clear time zero<sup>29</sup>. In the setting where cancer diagnosis is defined as “ever vs. never,”  
12 this definition introduces immortal time bias. All results pertaining to this definition had an  
13  
14 inverse association between cancer diagnosis and dementia, while results that did not introduce  
15  
16 this particular form of immortal-time bias had point estimates closer to the null. Although we  
17  
18 attempt to prevent this bias with statistical methods, we can only fully prevent it by having a  
19  
20 clear definition of time-zero. Importantly, this definition does not depend on the collected data  
21  
22 nor in analytic decisions: it relies on a deeper discussion related to when would be the best  
23  
24 moment to potentially intervene on this biomarker and to what purpose. Thus, we hope that  
25  
26 these unsolved questions guide future discussions and data collection efforts.  
27  
28  
29  
30  
31

32  
33 On the other hand, death as a competing event is a challenge that has some straightforward  
34  
35 strategies, beginning first and foremost by choosing the causal parameter (or estimand) of  
36  
37 interest<sup>38</sup>. In this study we chose the controlled direct effect, which represents the effect of Pin1  
38  
39 (or cancer) in a setting where death due to cancer and other causes could have been  
40  
41 prevented, yet without an explicit intervention, which makes it ambiguous. This is different than  
42  
43 conceiving a drug-target that increases Pin1 expression only in brain tissue, with no side effects  
44  
45 that could increase cancer risk (and thus, death due to cancer). As opposed to prior studies that  
46  
47 implicitly address a direct effect, and that define censoring as ignorable<sup>25</sup>, we show how point  
48  
49 estimates change substantially when we include weights for death to relax the independent  
50  
51 censoring assumption<sup>33</sup> when the estimand of interest is the controlled direct effect. Bounds to  
52  
53 assess extreme scenarios of dependence between death and dementia<sup>37</sup> illustrate the wide  
54  
55 range of possible point estimates that cross the null. This shows that even with the effort of  
56  
57 adjusting for time-varying covariates, we may be far from meeting this assumption and thus  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 better efforts to measure shared (time-varying) causes of dementia and death are needed. In  
5  
6 addition, presenting the proportion of participants that died prior to dementia diagnosis in each  
7  
8 arm, as well as the proportion of participants in each status over time, improves transparency  
9  
10 and puts in evidence the limitations of the data to answer this question.  
11  
12

13  
14 Pin1 is only one of the several mechanisms proposed by other investigators as an explanation  
15  
16 for the inverse cancer-dementia relationship. Certainly, cancer diagnosis represents a complex  
17  
18 and heterogeneous health condition that exceeds the representation of Pin1 expression. To  
19  
20 understand how – if at all – the cancer-dementia association informs the potential effects of any  
21  
22 other mechanism or treatment strategy, and its connection to collected data, we may require  
23  
24 different causal representations. For example, if researchers used this association to inform the  
25  
26 possible effects of different chemotherapeutics on cognitive decline among patients undergoing  
27  
28 treatment for cancer, different challenges would arise for mapping the observed association to  
29  
30 the hypothetical randomized trial underlying that new research question. Notably, that target  
31  
32 trial, unlike the one considered here, would include cancer diagnosis as part of the eligibility,  
33  
34 and thus researchers would need to instead grapple with how using data on persons without  
35  
36 cancer is useful and useable<sup>41</sup>. Each question requires thinking about bias anew and each  
37  
38 question brings its own set of challenges and opportunities.  
39  
40  
41  
42

43  
44 We underscore that our study is just one case study for how observed associations between  
45  
46 two diseases or health states may be disentangled to more transparently unveil possible  
47  
48 mechanisms (and sources of bias) behind them, and how to maximize what we can learn about  
49  
50 the potential mechanisms given available or readily-collected data. We hope to stimulate  
51  
52 researchers first have a discussion about the exact research question, rather than trying to  
53  
54 replicate the cancer and dementia relation in their dataset. Such discussion is needed to obtain  
55  
56 meaningful conclusions.  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Table 1. Descriptive characteristics of individuals included in the study.**

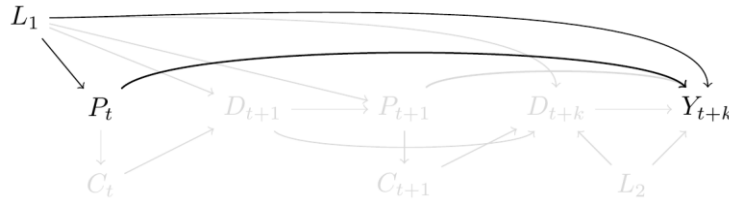
Characteristics	Overall (n = 3634)
Male, N (%)	1655 (45.5)
Age at baseline (mean (SD))	64.46 (2.86)
Educational attainment, N (%)	
Higher	382 (10.5)
Intermediate	1634 (45.0)
Lower	1594 (43.9)
Unknown history of diabetes	24 (0.7)
ApoE4, N (%)	
Not carrier	2488 (71.7)
One allele carrier	888 (25.6)
Two allele carrier	96 (2.8)
Smoking status, N (%)	
Current	922 (25.4)
Former	1720 (47.3)
Never	992 (27.3)
Body Mass Index (mean (SD))	26.59 (3.74)
Systolic blood pressure (mmHg) (mean (SD))	138.68 (20.85)
History of hypertension, N (%)	2126 (58.5)
History of heart disease, N (%)	263 (7.4)
Incident heart disease, N (%)	1022 (28.1)
No history of diabetes	2534 (69.7)
History of diabetes	374 (10.3)
Unknown history of diabetes	726 (20.0)
Incident diabetes, N (%)	718 (19.8)
History of stroke, N (%)	62 (1.7)
Incident stroke, N (%)	472 (13.0)
Incident cancer, N (%)	899 (24.7)
SD: Standard deviation	

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

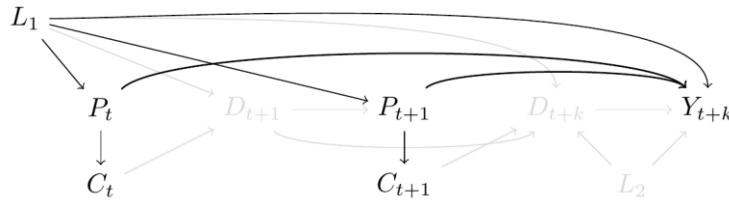
Scenario	Model	Risk Difference (95%CI)	Risk Ratio (95%CI)	Hazard Ratio (95%CI)
A	Unadjusted	-5.9 (-10.4,-1.1)	0.71 (0.49,0.95)	0.51 (0.39,0.68)
B	Unadjusted	-1.4 (-5.0,3.0)	0.94 (0.78,1.15)	0.95 (0.71,1.27)
A	With weights for confounding	-6.0 (-10.6,-1.4)	0.70 (0.49,0.93)	0.52 (0.39,0.69)
B	With weights 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 (-5.0,6.4)	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

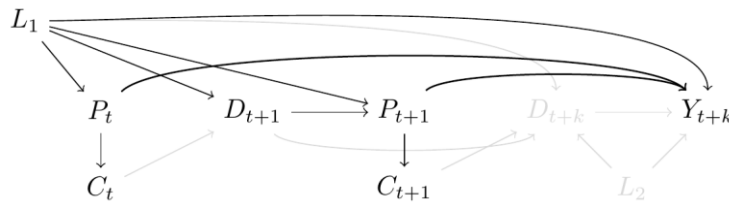
## Figures



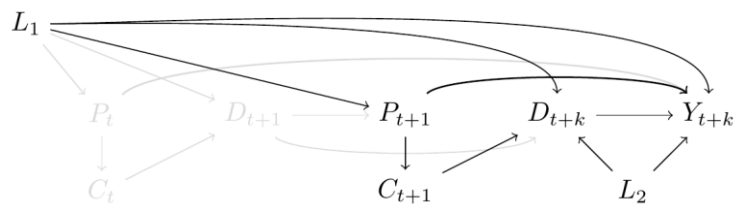
**Figure 1. Causal directed acyclic graph highlighting confounding of the potential effect of Pin1 on risk of dementia.** Pin1 at time  $t$  is represented as  $P_t$ ; dementia at time  $t + k$  is represented as  $Y_{t+k}$ ;  $L_1$  represents shared causes of Pin1 and dementia, such as smoking. To isolate the effect of  $P_t$  on  $Y_{t+k}$  we need to block the backdoor path  $Y_{t+k} \leftarrow L_1 \rightarrow P_t$ . Gray nodes and arrows are described progressively in Figures 1-5.



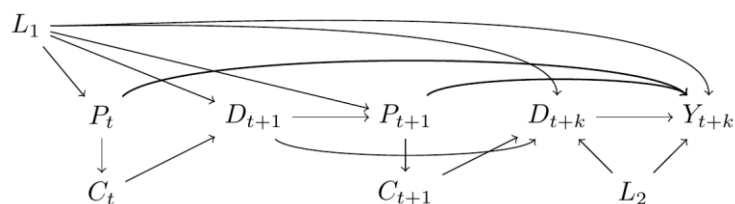
**Figure 2. Causal directed acyclic graph highlighting the use of incident cancer diagnosis as a proxy for Pin1 expression.** Pin1 at time  $t$  and  $t + 1$  are represented as  $P_t$  and  $P_{t+1}$ , respectively; dementia at time  $t + k$  is represented as  $Y_{t+k}$ ;  $L_1$  represents shared causes of Pin1 and dementia; incident cancer diagnosis at time  $t$  and  $t + 1$  are represented as  $C_t$  and  $C_{t+1}$ . To isolate the effect of  $P_t$  and  $P_{t+1}$  on  $Y_{t+k}$  we need to block the backdoor path  $Y_{t+k} \leftarrow L_1 \rightarrow P_t$  and  $Y_{t+k} \leftarrow L_1 \rightarrow P_{t+1}$ . Although we represent  $L_1$  as a single node for readability,  $L_1$  is time-varying too. Gray nodes and arrows are described progressively in Figures 1-5.



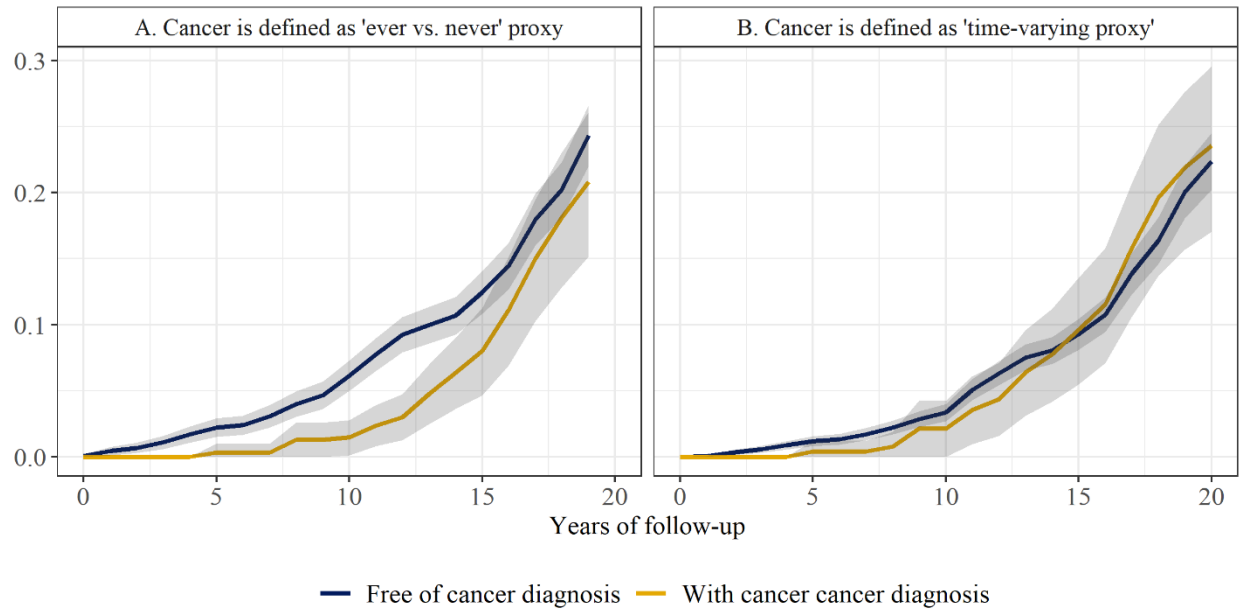
**Figure 3. Causal directed acyclic graph highlighting the time-varying nature of cancer diagnosis and immortal time bias.** Pin1 at time  $t$  and  $t + 1$  are represented as  $P_t$  and  $P_{t+1}$ , respectively; dementia at time  $t + k$  is represented as  $Y_{t+k}$ ;  $L_1$  represents shared causes of Pin1 and dementia; incident cancer diagnosis at time  $t$  and  $t + 1$  are represented as  $C_t$  and  $C_{t+1}$ .  $D_{t+1}$  represents death at time  $t + 1$ , cancer diagnosis at  $t + 1$  can only be measured among those who are alive at  $t + 1$ . Gray nodes and arrows are described progressively in Figures 1-5.



**Figure 4. Causal directed acyclic graph highlighting death as a competing event of dementia.** Pin1 at time  $t$  and  $t + 1$  are represented as  $P_t$  and  $P_{t+1}$ , respectively; dementia at time  $t + k$  is represented as  $Y_{t+k}$ ;  $L_1$  represents shared causes of Pin1 and dementia; incident cancer diagnosis at time  $t$  and  $t + 1$  are represented as  $C_t$  and  $C_{t+1}$ ;  $D_{t+1}$  represents death at time  $t + 1$ . In this graph we only focus attention to the exposure as if it was measured for all at time  $P_{t+1}$ . We include  $D_{t+k}$  as death at time  $t + k$  since  $Y_{t+1}$  is only observable when participants are alive at  $t + k$ .  $L_2$  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  $C_{t+1}$  and  $D_{t+k}$ . To isolate the direct effect of  $P_{t+1}$  and  $Y_{t+k}$ , we have to block the backdoor pathway from  $Y_{t+k} \leftarrow L_2 \rightarrow D_{t+k} \leftarrow C_{t+1} \leftarrow P_{t+1}$  and  $Y_{t+k} \leftarrow D_{t+k} \leftarrow L_1 \rightarrow P_{t+1}$ . Gray nodes and arrows are described progressively in Figures 1-5.



**Figure 5. Causal directed acyclic graph depicting multiple challenges to using the proxy of cancer diagnosis to study the effect of Pin1 on risk of dementia.** Pin1 at time  $t$  and  $t + 1$  are represented as  $P_t$  and  $P_{t+1}$ , respectively; dementia at time  $t + k$  is represented as  $Y_{t+k}$ ;  $L_1$  represents shared causes of Pin1 and dementia; incident cancer diagnosis at time  $t$  and  $t + 1$  are represented as  $C_t$  and  $C_{t+1}$ ;  $D_{t+1}$  and  $D_{t+k}$  represents death at time  $t + 1$  and  $t + k$ ;  $L_2$  represents possible shared causes of dementia and death. The distinct challenges were highlighted separately in Figures 1-4.



**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.

## References

1. Ma L-L, Yu J-T, Wang H-F, et al. Association between Cancer and Alzheimer's Disease: Systematic Review and Meta-Analysis. *Journal of Alzheimer's Disease*. 2014;42(2):565-573. doi: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*. Published online April 2016:gbw040. doi: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(11):e13019. doi:10.1111/eci.13019
4. Ospina-Romero M, Glymour MM, Hayes-Larson E, et al. Association Between Alzheimer Disease and Cancer With Evaluation of Study Biases: A Systematic Review and Meta-analysis. *JAMA Network Open*. 2020;3(11):e2025515. doi: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(3):196-204. doi:10.2174/156720509788486608
6. Harris RA, Tindale L, Cumming RC. Age-dependent metabolic dysregulation in cancer and Alzheimer's disease. *Biogerontology*. 2014;15(6):559-577. doi: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(10):7173-7187. doi: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(11):862. doi:10.3390/brainsci10110862
9. Driver JA. Inverse association between cancer and neurodegenerative disease: Review of the epidemiologic and biological evidence. *Biogerontology*. 2014;15(6):547-557. doi:10.1007/s10522-014-9523-2
10. Olson B, Marks DL. Pretreatment Cancer-Related Cognitive Impairment Mechanisms and Outlook. *Cancers*. 2019;11(5):687. doi:10.3390/cancers11050687
11. Li J, Mo C, Guo Y, et al. Roles of peptidyl-prolyl isomerase Pin1 in disease pathogenesis. *Theranostics*. 2021;11(7):3348-3358. doi: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(10):2069-2076. doi:10.1016/j.bbagen.2014.12.025
13. Snyder HM, Ahles T, Calderwood S, 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(3):267-273. doi:10.1016/j.jalz.2016.11.002



14. Ganguli M. Cancer and Dementia: It's Complicated. *Alzheimer Disease & Associated Disorders*. 2015;29(2):177-182. doi:10.1097/WAD.0000000000000086
15. Didelez V. Should the analysis of observational data always be preceded by specifying a target experimental trial? *International Journal of Epidemiology*. 2016;45(6):3.
16. Labrecque JA, Swanson SA. Target trial emulation: Teaching epidemiology and beyond. *European Journal of Epidemiology*. 2017;32(6):473-475. doi: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(1):59-62. doi: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, et al. Pin1 expression contributes to lung cancer prognosis and carcinogenesis. *Cancer Biology & Therapy*. 2010;9(2):111-119. doi:10.4161/cbt.9.2.10341
20. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. 2020;396(10248):413-446. doi:10.1016/S0140-6736(20)30367-6
21. Driver JA, Beiser A, Au R, et al. Inverse association between cancer and Alzheimer's disease: Results from the Framingham Heart Study. *BMJ*. 2012;344(mar12 1):e1442-e1442. doi:10.1136/bmj.e1442
22. Musicco M, Adorni F, Di Santo S, et al. Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study. *Neurology*. 2013;81(4):322-328. doi:10.1212/WNL.0b013e31829c5ec1
23. Freedman DM, Wu J, Chen H, et al. Associations between cancer and Alzheimer's disease in a U.S. Medicare population. *Cancer Medicine*. 2016;5(10):2965-2976. doi: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. Forloni G, ed. *PLOS ONE*. 2017;12(6):e0179857. doi:10.1371/journal.pone.0179857
25. Frain L, Swanson D, Cho K, et al. Association of cancer and Alzheimer's disease risk in a national cohort of veterans. *Alzheimer's & Dementia*. 2017;13(12):1364-1370. doi: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. Sleegers K, ed. *PLOS ONE*. 2017;12(2):e0171527. doi: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. doi:10.3389/fonc.2020.00073
28. Ording AG, Horváth-Puhó E, Veres K, et al. Cancer and risk of Alzheimer's disease: Small association in a nationwide cohort study. *Alzheimer's & Dementia*. 2020;16(7):953-964. doi: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;79:70-75. doi:10.1016/j.jclinepi.2016.04.014
30. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. :10.
31. White RS, Lipton RB, Hall CB, Steinerman JR. Nonmelanoma skin cancer is associated with reduced Alzheimer disease risk. *Alzheimer disease*. Published online 2013:7.
32. Nadler DL, Zurbenko IG. Developing a Weibull Model Extension to Estimate Cancer Latency. *ISRN Epidemiology*. 2013;2013:1-6. doi:10.5402/2013/750857
33. 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(8):1199-1236. doi:10.1002/sim.8471
34. 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*. Published online June 2020:1-9. doi:10.1080/01621459.2020.1765783
35. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European Journal of Epidemiology*. 2020;35(5):483-517. doi:10.1007/s10654-020-00640-5
36. 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(5):561-570. doi:10.1097/00001648-200009000-00012
37. 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(1):11-13. doi:10.1073/pnas.73.1.11
38. 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. doi:10.1101/2021.06.01.21258142
39. 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(26):4671-4680. doi:10.1002/sim.6280
40. 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(2):323-335. doi:10.1177/0962280216628900
41. 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(1):20. doi:10.13063/2327-9214.1234

November 3rd, 2021

Dear Prof. Lash,

On behalf of my coauthors, I am submitting the manuscript entitled “Towards a clearer causal question underlying the association between cancer and dementia?”, to be considered for publication within the category of Original Article, at the corresponding journal.

Over the last decades, several observational studies have studied the association between cancer and dementia incidence. Given that most results found an inverse association, several explanations have been proposed, from biological mechanisms to bias. However, no study has questioned why “cancer diagnosis” or “history of cancer” is an interesting exposure in the first place. To illustrate the complexities of inferring hypothetical or available cancer treatments’ effects on dementia from the observed cancer-dementia association, we focus on a specific question conceptualizing the Pin1 enzyme as the target of intervention. To explore how we might learn about this effect using real-world data on cancer and dementia, we progressively build a causal directed acyclic graph and outline the needed assumptions and potential sources of bias. We exemplify how these translate into the analytic decisions using data collected from the Rotterdam Study, a population-based cohort study.

This case study is just one example of how observed associations between two diseases or health states may be disentangled to more transparently unveil possible mechanisms (and sources of bias). We hope our work encourages researchers to be explicit about the causal question of interest to maximize what we can learn given available data.

All authors have reviewed and approved the content of the manuscript. This manuscript has not been published previously in any journal and is not under revision at any other journal.

We look forward to your thoughts, and thank you very much for your consideration.

L. Paloma Rojas-Saunero MD,  
PhD Candidate, Department of Epidemiology, Erasmus University Medical Center  
Mailing address: PO Box 2040, 3000CA Rotterdam, the Netherlands  
E-mail address [l.rojassaunero@erasmusmc.nl](mailto:l.rojassaunero@erasmusmc.nl)



Click here to access/download  
**Supplemental Digital Content**  
supplemental.docx

