**Does selection bias explain the inverse relation between cancer and dementia?**

L. Paloma Rojas-Saunero, MDa,\*, Kimberly D. van der Willik, MDa,b,\*, Sanne B. Schagen, PhDb,c, Sonja A. Swanson, ScDa,d, M. Arfan Ikram, PhDa

**Affiliations**

a Department of Epidemiology, Erasmus MC - University Medical Centre Rotterdam, Rotterdam, the Netherlands

b Department of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, the Netherlands

c Brain and Cognition, Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

d Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, United States

\* These authors contributed equally

**Corresponding author**

Prof. M. Arfan Ikram, MD PhD. Department of Epidemiology, Erasmus MC - University Medical Centre Rotterdam, PO Box 2040, 3000CA Rotterdam, the Netherlands. Tel: +31107043489 / Fax: +31107044657. E-mail address: m.a.ikram@erasmusmc.nl.

**ABSTRACT**

**Background** Observational studies have repeatedly shown that cancer patients have a lower risk of dementia than persons without a history of cancer. To illustrate the potential effects of selection bias on this inverse association between cancer and dementia, we replicated and compared previously used study designs. In addition, we presented alternative approaches to account for selection bias.

**Methods** Within the setting of the prospective population-based Rotterdam Study, we followed 8899 participants who were dementia-free at baseline over a period of 15 years. We replicated the following four study designs: (i) cohort study with cancer as time-dependent variable; (ii) cohort study with cancer as time-independent variable; (iii) nested case-control study with cancer as time-independent variable; and (iv) cross-sectional case-control study. For (i-iii) we estimated hazard ratios (HRs) with Cox proportional hazards models and for (iv) odds ratios (ORs) based on logistic regression models. Next, we presented the following three different methods to account for immortal time bias: (i) time-dependent cancer; (ii) inverse probability weighting (IPW); and (iii) cloning and censoring. To deal with the competing risk of death, we compared the risk of dementia among participants with and without cancer, as if (i) we could eliminate death; and (ii) regardless of death. We calculated the risk of dementia at each time point using pooled logistic regression.

**Results** Out of 8899 participants, 1813 (20.4%) were diagnosed with cancer, of whom 68 (3.8%) were subsequently diagnosed with dementia, 183 (10.1%) were lost to follow-up, and 890 (49.1%) died. The risk of dementia in patients with cancer depended on the study design. For instance, when cancer was treated as time-dependent variable the HR for dementia was 0.91 (95% confidence interval [CI] = 0.71 to 1.16]), as time-independent the HR was 0.44 (95% CI 0.35 to 0.56), and in the case-control study the OR for dementia was 0.28 (95% CI = 0.02 to 1.31). When using the alternative methods to deal with immortal time and the competing risk of death, the risk of dementia in participants with cancer was similar to that in participants without cancer.

**Conclusions** This study indicates that selection bias may drive the inverse association between cancer and dementia. Immortal time bias and competing events should be taken into account by using appropriate analytical methods, because these diseases are strongly related to death. In addition, future studies should further disentangle the processes underlying a cancer diagnosis to estimate the causal effect of cancer on dementia.

**INTRODUCTION**

Ageing populations worldwide have resulted in an increased prevalence of non-communicable diseases.1,2 There is a particular interest in the link between the non-communicable diseases cancer and dementia, because these diseases share multiple common risk factors including higher age and smoking. In addition, cancer and dementia have several overlapping pathways such as DNA damage and inflammation, suggesting that these diseases frequently co-occur.3,4 Many clinical studies have indeed found that patients with cancer often have impaired cognitive function.5-7 Observational studies, however, have repeatedly shown that patients with cancer have a lower risk of dementia than persons without a history of cancer.8-21 Several pathophysiological mechanisms underpinning this potential inverse link have been proposed and primarily involve differential expression of cell proliferation and survival pathways.22

In addition to these biological mechanisms, selection bias due to shortcomings in previous study designs and analytical decisions may underlie the inverse association between cancer and dementia.4,23 Selection bias can manifest itself in different ways of which we will highlight three examples, i.e., survival bias, immortal time bias, and bias due to the competing risk of death. Firstly, in cross-sectional studies, participants have to be alive at the moment of assessment to be included in the study. Such conditioning on participants who have survived up to the moment of study assessment can lead to survival bias. Secondly, in longitudinal studies, participants are followed until the date of dementia diagnosis or death. The starting point of follow-up can differ between participants who remain free of cancer during follow-up and those who are diagnosed with cancer during follow-up. Exclusion or misclassification of the time between study entry and cancer diagnosis, i.e., immortal time, may cause differences in baseline characteristics between participants with and without cancer and can induce immortal time bias.24 Thirdly, most longitudinal studies assume that death occurs at random, whereas in fact, risk factors that are related to death are usually the same that are related to cancer and dementia. This results in that the participant with the worst risk factor profile will die first.25,26

In the current study, we illustrate how decisions on study design and statistical analyses may induce selection bias when studying the association between cancer and dementia. In order to do so, we use data from the prospective population-based Rotterdam Study. Firstly, we provide a visual explanation of the potential problem of selection bias due to study design. Secondly, we replicate previous study designs and analytical decisions and show how results vary accordingly. Thirdly, we present three different methods to account for immortal time bias and we two alternative approaches to address the competing event of death.

**METHODS**

**Study population**

This study is embedded in the Rotterdam Study, a prospective population-based cohort study that was designed to determine causes of diseases in the middle-aged and elderly population.27 After the pilot phase in 1989, all inhabitants aged 55 years and over of the Ommoord area in Rotterdam, the Netherlands, were invited to participate between 1990 and 1993. This first subcohort comprised 7983 participants (response of 78%) and was extended with the second subcohort between 2000 and 2001 consisting of 3011 participants (response of 67%) who had reached the age of 55 years or who had moved into the study area.

Participants were interviewed at home by a trained research assistant, followed by two visits at the research centre for different examinations including physical examinations, laboratory assessments, and imaging. Follow-up examinations of the first subcohort took place from 1993 to 1995, from 1997 to 1999, from 2002 to 2004, from 2009 to 2011, and from 2014 to 2015. For the second subcohort, follow-up examinations took place between 2004 and 2005, and between 2011 and 2012.

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus Medical Centre and by the Ministry of Health, Welfare and Sport of the Netherlands. Written informed consent was obtained from all participants.

Of the total of 10 994 participants, we excluded those with a history of dementia (n=515) or who were insufficiently screened for history of dementia (n=349), those without informed consent to access medical records during follow-up (n=135), and participants with incomplete data on baseline characteristics including education, smoking, body mass index (BMI), systolic blood pressure, and hypertension (n=1096), resulting in 8899 participants for analyses. Participants who had incomplete baseline characteristics were older at baseline (median age [interquartile range [IQR]] 73.4 years [63.8 to 82.4] versus 65.7 years [60.4 to 73.2]) and were more often women than included participants (68.2% versus 57.4%).

**Ascertainment of cancer**

Cancer was diagnosed based on medical records of general practitioners (including hospital discharge letters) and through linkage with the Netherlands Cancer Registry, Dutch Hospital Data, and histology and cytopathology registries in the region. Cancer was defined as any primary malignant tumour, excluding non-melanoma skin cancer, that was confirmed by pathology. Diagnoses were coded independently by two physicians according to the International Classification of Diseases, tenth revision (ICD-10). In case of discrepancy, consensus was sought through consultation with a physician specialised in internal medicine. Date of diagnosis was based on date of biopsy (solid tumours) and laboratory assessment (haematological tumours), or – if unavailable – date of hospital admission or discharge letter. Follow-up was completed up to January 1st, 2015.

**Ascertainment of dementia**

Participants were screened for dementia at baseline and subsequent centre visits with the Mini-Mental State Examination and the Geriatric Mental Schedule organic level.28 Those with a Mini-Mental State Examination score <26 or Geriatric Mental Schedule score >0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. In addition, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study database with medical records from general practitioners and the regional institute for outpatient mental health care. A consensus panel led by a consultant neurologist established the final diagnosis according to standard criteria for dementia (Diagnostic and Statistical Manual of Mental Disorders III-revised). Follow-up was completed up to January 1st, 2016.

**Ascertainment of covariates**

During home interviews, participants provided information on educational level and smoking habits. Educational level was categorised into lower (primary education or lower vocational education), intermediate (lower secondary education, intermediate vocational education, or general secondary education), or higher (higher vocational education or university). Smoking habits were classified as never, former, or current smoking. At the research centre, height and weight were measured to calculate the BMI (kg/m2). Systolic and diastolic blood pressure were measured twice on the right arm using a random-zero sphygmomanometer of which the mean was used. Hypertension was defined as a systolic blood pressure of ≥140 mm Hg, a diastolic blood pressure of ≥90 mm Hg, or use of antihypertensive medication.29 Diabetes mellitus was defined fasting serum glucose level ≥7.1 mmol/L, a random serum glucose level ≥11.1 mmol/L, or use of glucose-lowering medication.30 History of coronary heart disease (myocardial infarction, percutaneous coronary intervention, or coronary artery bypass grafting) and stroke was assessed by interview and verified by reviewing medical records.31,32

**Statistical analyses**

We first visualised the number of years spent in different transition states using raw data to illustrate the potential problem of survival bias when studying the relation between cancer and dementia.

*Replication of previous studies*

Next, we replicated the following study designs and statistical analyses that have previously been used to study the association between cancer and dementia. To understand the difference between these studies, we define the time zero as the moment at which we start observing participants based on their exposure level.24 Given that a cancer diagnosis can occur at any moment in time, previous studies have considered cancer as a time-dependent variable (history of cancer at study entry and incident cancer after cancer diagnosis versus no cancer at study entry and during follow-up) or as a time-independent variable (history of cancer at study entry versus no history of cancer at study entry, or ever versus never cancer). Following these definitions, we replicated (i) cohort study using Cox proportional hazards models and using cancer as time-dependent variable and performing a sensitivity analysis by restricting the study population to participants who survived up to age eighty years during follow-up;10,13,16 (ii) cohort study using Cox proportional hazards models and using cancer as time-independent variable;11,15-17,19 (iii) nested case-control study using Cox proportional hazards models and using cancer as time-independent variable (note that this design was previously used to investigate the risk of cancer in patients with dementia);17 and (iv) cross-sectional case-control study using logistic regression models.9,12 As previous studies corrected for baseline characteristics, we adjusted models for age, sex, educational level, smoking, BMI, systolic blood pressure, hypertension, diabetes mellitus, history of coronary heart disease, and history of stroke. Follow-up time was used as underlying time scale and was rounded to years. Participants were censored at date of loss to follow-up, death, or after 15 years since start of follow-up. For the nested case-control study, we matched participants who were diagnosed with cancer to two participants who were free from cancer during study follow-up based on the age of cancer diagnosis and sex. For the cross-sectional study, we identified participants who visited at the fourth follow-up round of the first subcohort and the second follow-up round of the second subcohort. In addition, we selected participants who were diagnosed within six months after these follow-up rounds, because dementia is often diagnosed as result of examinations performed during the follow-up round. We subsequently used logistic regression models with history of incident cancer as determinant.

*Alternative methods to account for selection bias*

In the abovementioned study designs, depending on how time zero was handled, immortal time bias could be inflicted. Immortal time bias can arise when we fail to align the start of follow-up in participants who develop cancer versus those who remain free of cancer during follow-up.24 Participants who survive longer have a higher probability to be diagnosed with cancer than those who have a shorter survival time. When cancer is treated as a time-independent variable, such as when the time zero for individuals with cancer is the time of cancer diagnosis, persons who died before cancer diagnosis are by definition excluded from this group. In contrast, when the time at study entry is considered as the time zero in both groups and time of cancer diagnosis is a time-dependent variable, immortal time bias can be reduced. To prevent immortal time bias, we must however consider alternative methods that we will discuss later in this section.24,33

For all replications of previous study designs and statistical analyses, death was considered as an uninformative censoring event. A censoring event is an event that prevents observing the true outcome, including the outcome of dementia. For instance, being lost to follow-up is considered as a censoring event, which could be prevented in the study design. We consider a censoring event being uninformative if we assume that persons who are censored are similar to those who remained alive during the study period, given the available covariates. However, by definition, everyone who dies before the outcome, will not develop dementia. This cannot be prevented by the study design, and therefore we can consider death as a competing event.

In the current study, we used the following three alternative approaches to deal with immortal time bias: (i) a naïve approach in which cancer is treated as time-dependent variable; (ii) inverse probability weights (IPW) for the time until cancer diagnosis; and (iii) cloning and censoring.33 These different approaches are implemented to answer two different questions based on how we include death. The first question is ‘what is the risk of dementia among participants who develop cancer versus among those who remain free of cancer during follow-up, considering that we could eliminate death?’. This questions reflects the hypothetical scenario were we could prevent death and therefore treat death as a censoring event. We refer to this question as the direct controlled effect.25 Given that this question relies on the strong assumption of uninformative censoring, we consider time-dependent covariates to simulate a scenario in which censoring for death is uninformative. The second question we propose is ‘what is the risk of dementia among participants who develop cancer versus among those who remain free of cancer during follow-up, regardless of death?’. This question does not require any strong assumption on the competing event of death, but the risk of dementia will be affected by the relationship between cancer and death.34 We refer to this question as the total effect.25 In the following paragraphs, we will describe the technical details of these different approaches.

*IPW for the time until cancer diagnosis*

We computed weights for the time until cancer diagnosis by fitting a pooled logistic model. The product of the estimated conditional probabilities at each time was subsequently used to estimate the time-dependent weight for each participant at each time point, reflecting the time-dependent weight inversely proportional to the probability of not being diagnosed with cancer. Weights were fitted considering time until cancer diagnosis in years, cancer, the interaction between time and cancer, sex, age at study entry, cohort, education, and hypertension, and time-updated covariates including smoking, systolic blood pressure, BMI, history of stroke, history of diabetes, and history of coronary heart disease. Time, systolic blood pressure, and BMI were modelled non-linearly using B-spines with three degrees of freedom.

*Cloning and censoring*

Details of this method have been described previously.33 In brief, first, we made two copies of each participant. One of the copies was allocated to a ‘cancer’ arm and the other copy to a ‘cancer-free’ arm. Participants who were diagnosed with cancer during follow-up were censored at date of end of follow-up in the ‘cancer’ arm and at date of cancer diagnosis in the ‘cancer-free’ arm. Those who remained free from cancer during follow-up were censored after a pre-specified period of 15 years in the ‘cancer’ arm, or if their follow-up time was shorter at date of end of follow-up. In the ‘cancer-free’ arm, these participants were censored at date of end of follow-up. Censoring is informative and therefore we accounted for this type of censoring using IPW as described above.

*Risk of dementia with elimination of death*

To “eliminate death”, i.e., to account for potential non-differential death between participants who were diagnosed with cancer and those who remained free of cancer, we computed IPW for death by fitting a pooled logistic model that included time-depended covariates. This resulted in a time-depended weight inversely proportional to the probability of not dying for each participant, considering the time-depended covariates that relate to death and dementia, including sex, age at study entry, cohort, education, smoking status, systolic blood pressure, hypertension, BMI, history of stroke, history of diabetes, and history of coronary heart disease. Follow-up time, systolic blood pressure, and BMI were modelled non-linear using B-spines with three degrees of freedom. In addition, we used these time-depended covariates to calculate IPW for loss to follow-up.

*Risk of dementia regardless of death*

To estimate the risk of dementia regardless of death, we do not rely in the strong assumption of uninformative censoring. We estimated the risk of dementia by the joint probability of incurring either dementia or death at each time point. We calculated the cumulative sum of the probability of surviving both events multiplied by the instantaneous cause-specific hazard of dementia.25,35 In this approach, we only considered censoring and IPW for loss to follow-up.

Combining the approaches for immortal time bias and the competing risk of death, we performed six different analyses. We created standardised cumulative incidence curves using pooled logistic regression. The 95% confidence intervals (CIs) were obtained by bootstrapping. Statistical analyses were performed using R software Version 3.6.1. The code to run the analyses will be made available on GitHub.

**RESULTS**

During a median (IQR) follow-up of 13.0 years (7.0 to 15.0), 1813 (20.4%) out of 8899 participants were diagnosed with cancer. Of the participants who were diagnosed with cancer, 68 (3.8%) were subsequently diagnosed with dementia, 183 (10.1%) were lost to follow-up, and 890 (49.1%) died during follow-up. Of the 7086 participants who remained free of cancer, 781 (11.0%) were diagnosed with dementia, 1330 (18.8%) were lost to follow-up, and 1341 (18.9%) died (**Figure 1**). **Table 1** shows the baseline characteristics of the total study population. An overview of the different transition stages among participants who developed either cancer or dementia and died during follow-up is presented in **Figure 2**. The median (IQR) age at cancer diagnosis of these participants was 75.0 years (69.1 to 80.4), whereas the median (IQR) age at dementia diagnosis of these participants was 82.5 years (77.5 to 86.9).

**Replication of previous studies**

When using cancer as time-dependent variable, the risk of dementia in participants with cancer was similar to that in participants without cancer (hazard ratio [HR] = 0.91, [95% CI = 0.71 to 1.16], **Table 2**). This risk was comparable when restricting the analysis to participants who survived up to at least eighty years (HR = 0.91, [95% CI = 0.67 to 1.26]). Participants with cancer had a lower risk of dementia than participants without cancer when using cancer as time-independent variable (HR = 0.44, [95% CI 0.35 to 0.56]) and in the nested case-control setting (HR = 0.53, [95% CI = 0.41 to 0.69]).

The number of participants that was included in the cross-sectional study design was 5278. Out of these participants, 50 had a dementia diagnosis during the research centre visit. This design is illustrated by **Figure 3**. Participants with a history of cancer had an odds ratio (OR) of 0.28 (95% CI = 0.02 to 1.31) for dementia. When including also participants who were diagnosed with dementia within six months after the research centre visit, the OR was 0.61 (95% CI = 0.26 to 1.24).

**Alternative methods to account for selection bias**

**Figure 4** shows that the risk of death among participants with cancer is higher than among those without cancer. Standardised cumulative incidence curves for the six different analyses are presented in **Figure 5**. When using cancer as time-dependent variable and eliminating death, participants with cancer had a higher risk of dementia than those without cancer, but CIs were overlapping (**Figure 5A**). The risk of dementia regardless of the risk of death was higher in participants with cancer than in those without cancer up to 12 years of follow-up. After 12 years, the difference in the risk of dementia narrows, but CIs largely overlap over the entire follow-up (**Figure 5B**). The risk of dementia was higher when estimating the risk of dementia with elimination of death rather than regardless of death. When we computed IPW for the time until cancer diagnosis, cumulative incidence curves crossed after 12 years of follow-up in both approaches that we used to deal with the competing risk of death (**Figure 5C** and **5D**). Lastly, when using the cloning and censoring method, we found that cumulative incidence curves for participants with and without cancer were completely overlapping, with a slight deviation after 15 years of follow-up when estimating the risk of dementia regardless of death (**Figure 5E** and **5F**).

**DISCUSSION**

This study shows that the direction and magnitude of the risk of dementia in cancer patients is affected by the characteristics of the study design and statistical analyses. When using appropriate methods to account for immortal time bias and the competing event of death, patients with cancer are not at decreased risk of developing dementia. This suggests that the frequently observed inverse relation between cancer and dementia may be based on selection bias. These findings underline the importance of using appropriate study designs and statistical analyses when studying an exposure and outcome that are strongly related to death.

When we replicated study designs and statistical analyses performed in previous literature, hazard ratios for the risk of dementia in cancer patients varied between 0.44 and 0.91. These effect estimates are comparable to those obtained from previous studies on cancer and dementia.8-21 The lowest effect estimates were found for cross-sectional study designs. In cross-sectional studies, participants had to survive up to a certain moment in time to be included in the study. Participants who have survived longer are more likely to be included than those who have a shorter survival time, and are therefore more likely to be included in a cross-sectional study (**Figure 3**). This can result in a selected group of participants who may be healthier and have a longer survival than the general population.

Regarding longitudinal studies, we found a lower risk of dementia in participants with cancer when using cancer as time-independent variable (HR = 0.44 in the cohort setting and HR = 0.53 in the nested case-control setting) than when studying cancer as time-dependent variable (HR = 0.91), which was also observed by Hanson et al.23 The difference between these outcomes may be explained by selection bias due to immortal time.34 We have accounted for immortal time using three different methods: (i) using cancer as time-dependent variable; (ii) using IPW for the time until cancer diagnosis; and (iii) cloning and censoring. These three different methods provided similar results. The time-dependent Cox proportional hazards model also reduces immortal time bias, but cannot completely prevent such bias and has two additional shortcomings, i.e., (i) hazard ratios represent a weighted average of the time-dependent hazard ratios of the total follow-up period and may therefore lose information that is preserved by presenting cumulative incidence curves36; and (ii) the model assumes that censoring of death is uninformative.

This latter assumption of the Cox proportional hazards model is often not examined. Given that persons need to survive long enough to develop dementia (median age [IQR] of dementia diagnosis in our study was 82.5 years [77.5 to 86.9]), and that dementia shares risks factors related to cardiovascular disease and cancer, they are at increased risk of death. Censoring for death is therefore informative and may result in biased effect estimates.25 Hanson et al. compared the following two competing risk models to illustrate the effect of the competing risk of death on the relation between cancer and dementia: Fine and Gray and Kalbfleisch and Prentice.23,37,38 The Fine and Gray model calculates the absolute risk of dementia and allows participants to be at risk for dementia after they have experienced the competing risk of death. The subdistribution hazard for dementia was lower than the cause specific hazard for dementia, indicating that mortality was higher in patients with cancer than in persons without cancer. The Kalbfleisch and Prentice method uses internal time-dependent covariates that are strongly related to the competing risk of death. The significant, positive association between these covariates and dementia indicated non-independence between dementia and death. Other studies have tried to account for the effects of the competing risk of death by studying negative control diseases39 such as stroke and automobile injuries,19,40,41 focusing on different cancer types,8,18,19,41,42 and stratifying follow-up time.21 In addition, we have previously used the tumour marker carcinoembryonic antigen as a proxy of preclinical cancer and studied the risk with dementia, given that persons with a preclinical stage of the disease have on average a longer life expectancy than patients with clinically manifested disease.43 Interestingly, higher levels of the tumour marker carcinoembryonic antigen were associated with a higher risk of dementia suggesting that, from a biological perspective, cancer and dementia might even be positively associated rather than inversely. In the current study, we accounted for the competing risk of death by multiplying the instantaneous hazard of dementia by the probability of being free of any event (i.e., total effect of cancer on dementia) and by computing IPW for death (i.e., direct effect of cancer on dementia, not mediated by death).

It must be noted however, that we, like previous studies, defined cancer as a pathology-confirmed cancer diagnosis. A cancer diagnosis can represent multiple causal pathways as will be explained below. We therefore did not examine the causal effect of cancer when using the alternative methods to account for immortal time bias and the competing risk of death, and as such, we did not consider confounders for the association between cancer, dementia and death. If future studies aim to understand the causal relation between cancer and dementia, the research question should be redefined, because cancer diagnosis itself does not cause dementia. An ill-defined research question may result in wrong interpretation of the causal effects.25 Cancer diagnosis may be considered as a proxy for other underlying processes that may cause cognitive impairment and subsequently dementia, see the corresponding directed acyclic graph (DAG) in **Figure 6**. This DAG represents four proposed mechanisms underlying cognitive problems in cancer patients.44-46 Given that cognitive impairment precedes dementia, we hypothesised that these mechanisms may also underlie a causal association between cancer and dementia. Firstly, shared risk factors for cancer and dementia such as higher age, genetics, and smoking may increase the risk of both cancer and dementia. Secondly, the tumour itself can induce different biological processes including inflammation, vascular changes, oxidative stress, and production of extracellular vesicles that can affect cognitive function.44 These processes may differ between different types of cancer and disease stages. Thirdly, cancer treatment can accelerate the ageing process by inducing DNA damage, telomere shortening, oxidative stress, inflammation, and changes in hormonal levels.45,47 In addition, chemotherapy can have direct neurotoxic effects.46,48 Fourthly, psychological distress including depression, anxiety, stress, and fatigue may be caused by cancer diagnosis and cancer treatment. Such factors can also affect cognitive function and the risk of dementia.49

Each of these causal pathways may require adjustment for different confounding structures. For this reason, we recommend future studies to specify the specific pathway of interest. Importantly, measurement error for the proxy of cancer diagnosis may be different for each of these pathways. In addition, even if the causal pathway has been correctly specified, one must assume to have all required information to control for the collider-bias that is induced by condition on surviving (**Figure 6**).

In conclusion, our findings indicate that the type of study design can influence results when studying diseases that are strongly related to death. When taking immortal time bias and the competing risk of death into account, we found that patients with cancer did not have a lower risk of dementia than those without a history of cancer, nor did they have a higher risk. Given the ill-defined definition of cancer diagnosis, we cannot answer the causal effect of cancer on the risk of dementia. Future studies should further disentangle the processes underlying a cancer diagnosis to estimate the causal effect.

**REFERENCES**

1. Licher S, Heshmatollah A, van der Willik KD, et al. Lifetime risk and multimorbidity of non-communicable diseases and disease-free life expectancy in the general population: A population-based cohort study. *PLOS Medicine.* 2019;16(2):e1002741.

2. Bennett JE, Stevens GA, Mathers CD, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *The Lancet.* 2018;392(10152):1072-1088.

3. Nudelman KNH, McDonald BC, Lahiri DK, Saykin AJ. Biological Hallmarks of Cancer in Alzheimer's Disease. *Mol Neurobiol.* 2019.

4. van der Willik KD, Schagen SB, Ikram MA. Cancer and dementia: Two sides of the same coin? *Eur J Clin Invest.* 2018;48(11):e13019.

5. Wefel JS, Kesler SR, Noll KR, Schagen SB. Clinical Characteristics, Pathophysiology, and Management of Noncentral Nervous System Cancer-Related Cognitive Impairment in Adults. *CA: a cancer journal for clinicians.* 2015;65(2):123-138.

6. Koppelmans V, Breteler MM, Boogerd W, Seynaeve C, Gundy C, Schagen SB. Neuropsychological performance in survivors of breast cancer more than 20 years after adjuvant chemotherapy. *J Clin Oncol.* 2012;30(10):1080-1086.

7. Bernstein LJ, McCreath GA, Komeylian Z, Rich JB. Cognitive impairment in breast cancer survivors treated with chemotherapy depends on control group type and cognitive domains assessed: A multilevel meta-analysis. *Neurosci Biobehav Rev.* 2017;83:417-428.

8. Nudelman KNH, Risacher SL, West JD, et al. Association of cancer history with Alzheimer's disease onset and structural brain changes. *Frontiers in Physiology.* 2014;5(423).

9. Yamada M, Sasaki H, Mimori Y, et al. Prevalence and risks of dementia in the Japanese population: RERF's adult health study Hiroshima subjects. Radiation Effects Research Foundation. *J Am Geriatr Soc.* 1999;47(2):189-195.

10. White RS, Lipton RB, Hall CB, Steinerman JR. Nonmelanoma skin cancer is associated with reduced Alzheimer disease risk. *Neurology.* 2013;80(21):1966-1972.

11. Schmidt SA, Ording AG, Horvath-Puho E, Sorensen HT, Henderson VW. Non-melanoma skin cancer and risk of Alzheimer's disease and all-cause dementia. *PLoS One.* 2017;12(2):e0171527.

12. Realmuto S, Cinturino A, Arnao V, et al. Tumor diagnosis preceding Alzheimer's disease onset: is there a link between cancer and Alzheimer's disease? *J Alzheimers Dis.* 2012;31(1):177-182.

13. Bowles EJA, Walker RL, Anderson ML, Dublin S, Crane PK, Larson EB. Risk of Alzheimer's disease or dementia following a cancer diagnosis. *PLoS One.* 2017;12(6):e0179857-e0179857.

14. Ording AG, Horváth-Puhó E, Veres K, et al. Cancer and risk of Alzheimer's disease: Small association in a nationwide cohort study. *Alzheimers Dement.* 2020.

15. Roe CM, Behrens MI, Xiong C, Miller JP, Morris JC. Alzheimer disease and cancer. *Neurology.* 2005;64(5):895-898.

16. Roe CM, Fitzpatrick AL, Xiong C, et al. Cancer linked to Alzheimer disease but not vascular dementia. *Neurology.* 2010;74(2):106-112.

17. 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:e1442.

18. 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.

19. Freedman DM, Wu J, Chen H, et al. Associations between cancer and Alzheimer's disease in a U.S. Medicare population. *Cancer Med.* 2016;5(10):2965-2976.

20. 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).

21. Frain L, Swanson D, Cho K, et al. Association of cancer and Alzheimer's disease risk in a national cohort of veterans. *Alzheimers Dement.* 2017;13(12):1364-1370.

22. Driver JA. Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence. *Biogerontology.* 2014;15(6):547-557.

23. 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. *J Gerontol B Psychol Sci Soc Sci.* 2017;72(6):1032-1043.

24. 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.

25. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernan MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med.* 2020;39(8):1199-1236.

26. Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer.* 2004;91(7):1229-1235.

27. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. *European Journal of Epidemiology.* 2020.

28. de Bruijn RF, Bos MJ, Portegies ML, et al. The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med.* 2015;13:132.

29. Engberink MF, Bakker SJ, Brink EJ, et al. Dietary acid load and risk of hypertension: the Rotterdam Study. *Am J Clin Nutr.* 2012;95(6):1438-1444.

30. Diabetes mellitus. Report of a WHO Study Group. *World Health Organization technical report series.* 1985;727:1-113.

31. Bos MJ, Koudstaal PJ, Hofman A, Ikram MA. Modifiable Etiological Factors and the Burden of Stroke from the Rotterdam Study: A Population-Based Cohort Study. *PLoS Medicine.* 2014;11(4).

32. Leening MJ, Kavousi M, Heeringa J, et al. Methods of data collection and definitions of cardiac outcomes in the Rotterdam Study. *Eur J Epidemiol.* 2012;27(3):173-185.

33. Maringe C, Benitez Majano S, Exarchakou A, et al. Reflections on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. *International Journal of Epidemiology.* 2020(dyaa057).

34. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ.* 2010;340:b5087.

35. Cole SR, Lau B, Eron JJ, et al. Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to AIDS after initiation of antiretroviral therapy. *Am J Epidemiol.* 2015;181(4):238-245.

36. Stensrud MJ, Hernán MA. Why Test for Proportional Hazards? *JAMA.* 2020;323(14):1401-1402.

37. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc.* 1999;94(446):496-509.

38. Kalbfleisch JD, Prentice RL. Marginal Likelihoods Based on Cox's Regression and Life Model. *Biometrika.* 1973;60(2):267-278.

39. Arnold BF, Ercumen A, Benjamin-Chung J, Colford JM, Jr. Brief Report: Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies. *Epidemiology.* 2016;27(5):637-641.

40. 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:e1442.

41. 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 : the journal of the Alzheimer's Association.* 2017.

42. Bowles EJA, Walker RL, Anderson ML, Dublin S, Crane PK, Larson EB. Risk of Alzheimer's disease or dementia following a cancer diagnosis. *PLoS One.* 2017;12(6):e0179857.

43. van der Willik KD, Schagen SB, Ikram MA. Association Between the Tumor Marker Carcinoembryonic Antigen and the Risk of Dementia. *J Alzheimers Dis.* 2020.

44. Olson B, Marks DL. Pretreatment Cancer-Related Cognitive Impairment-Mechanisms and Outlook. *Cancers (Basel).* 2019;11(5):687.

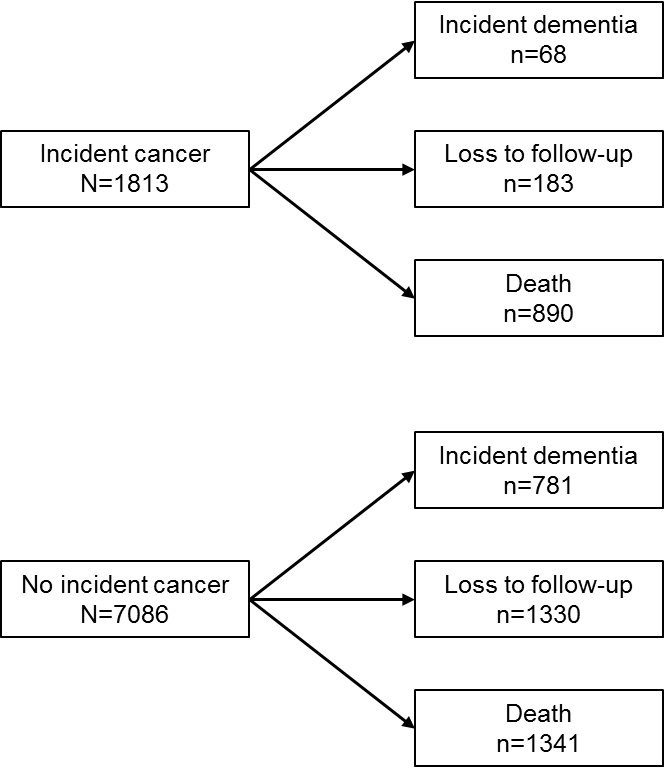
45. Ahles TA, Saykin AJ. Candidate mechanisms for chemotherapy-induced cognitive changes. *Nat Rev Cancer.* 2007;7(3):192-201.

46. Gibson EM, Nagaraja S, Ocampo A, et al. Methotrexate Chemotherapy Induces Persistent Tri-glial Dysregulation that Underlies Chemotherapy-Related Cognitive Impairment. *Cell.* 2019;176(1):43-55.e13.

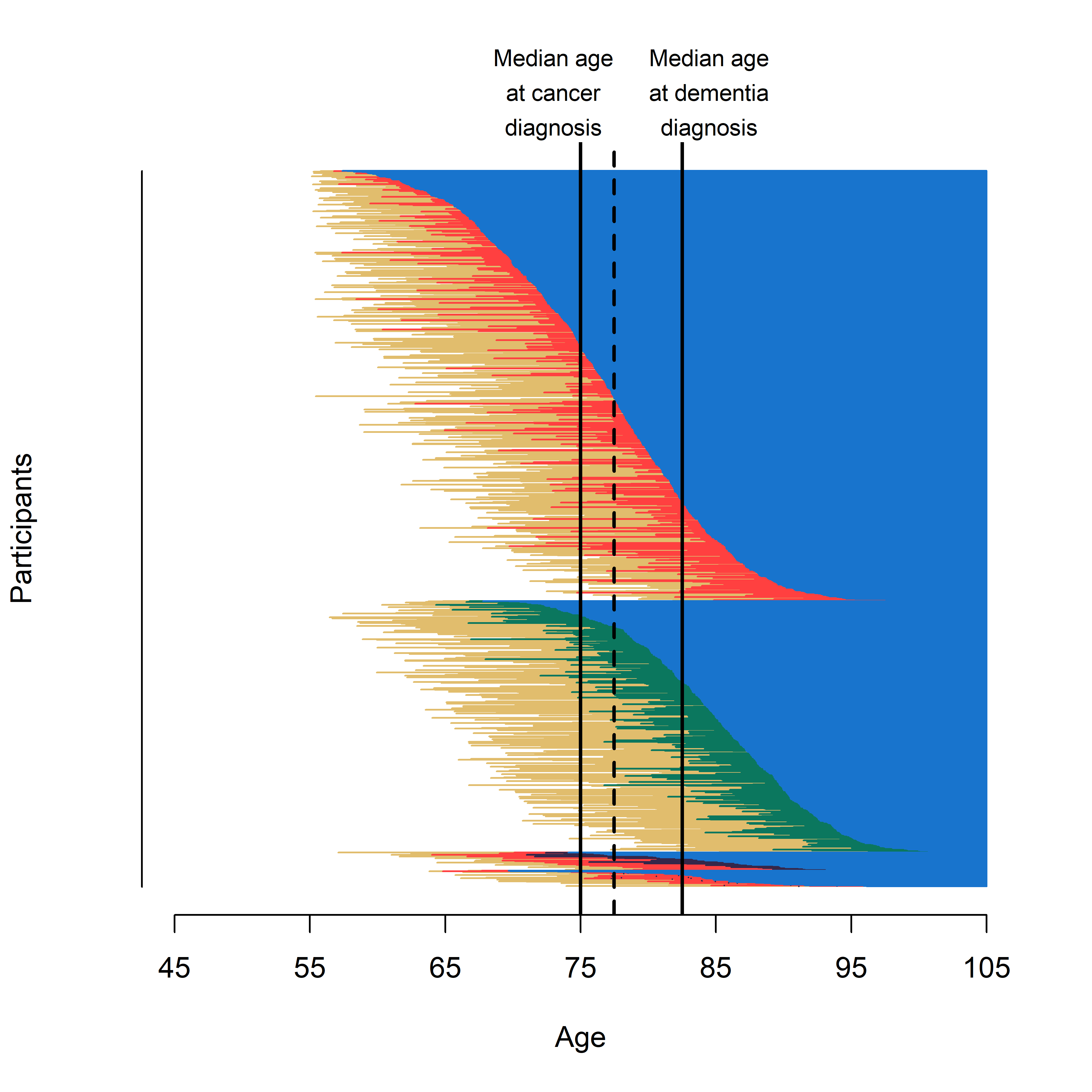
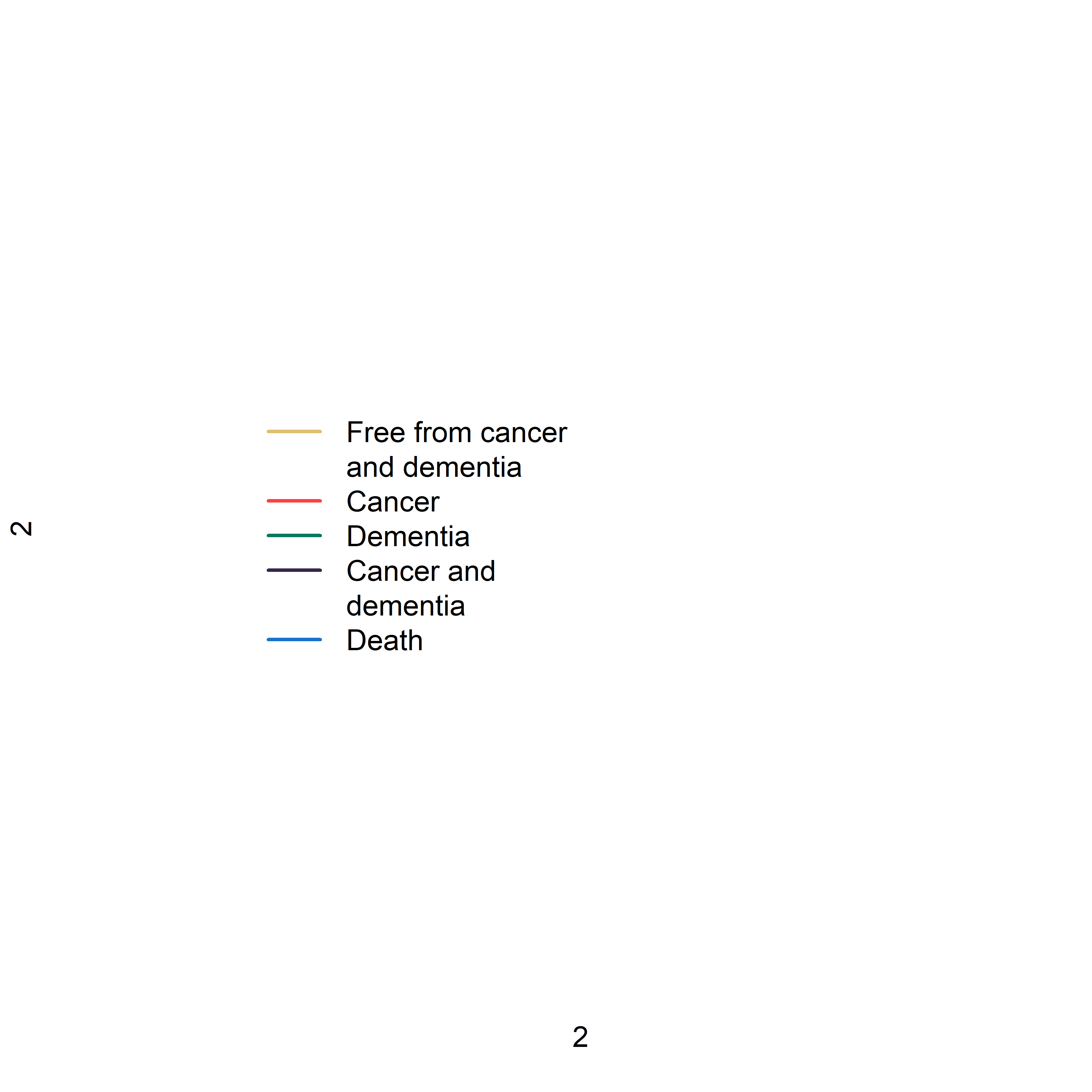
47. Maccormick RE. Possible acceleration of aging by adjuvant chemotherapy: a cause of early onset frailty? *Med Hypotheses.* 2006;67(2):212-215.

48. Dietrich J, Han R, Yang Y, Mayer-Pröschel M, Noble M. CNS progenitor cells and oligodendrocytes are targets of chemotherapeutic agents in vitro and in vivo. *J Biol.* 2006;5(7):22-22.

49. Mirza SS, Wolters FJ, Swanson SA, et al. 10-year trajectories of depressive symptoms and risk of dementia: a population-based study. *Lancet Psychiatry.* 2016;3(7):628-635.

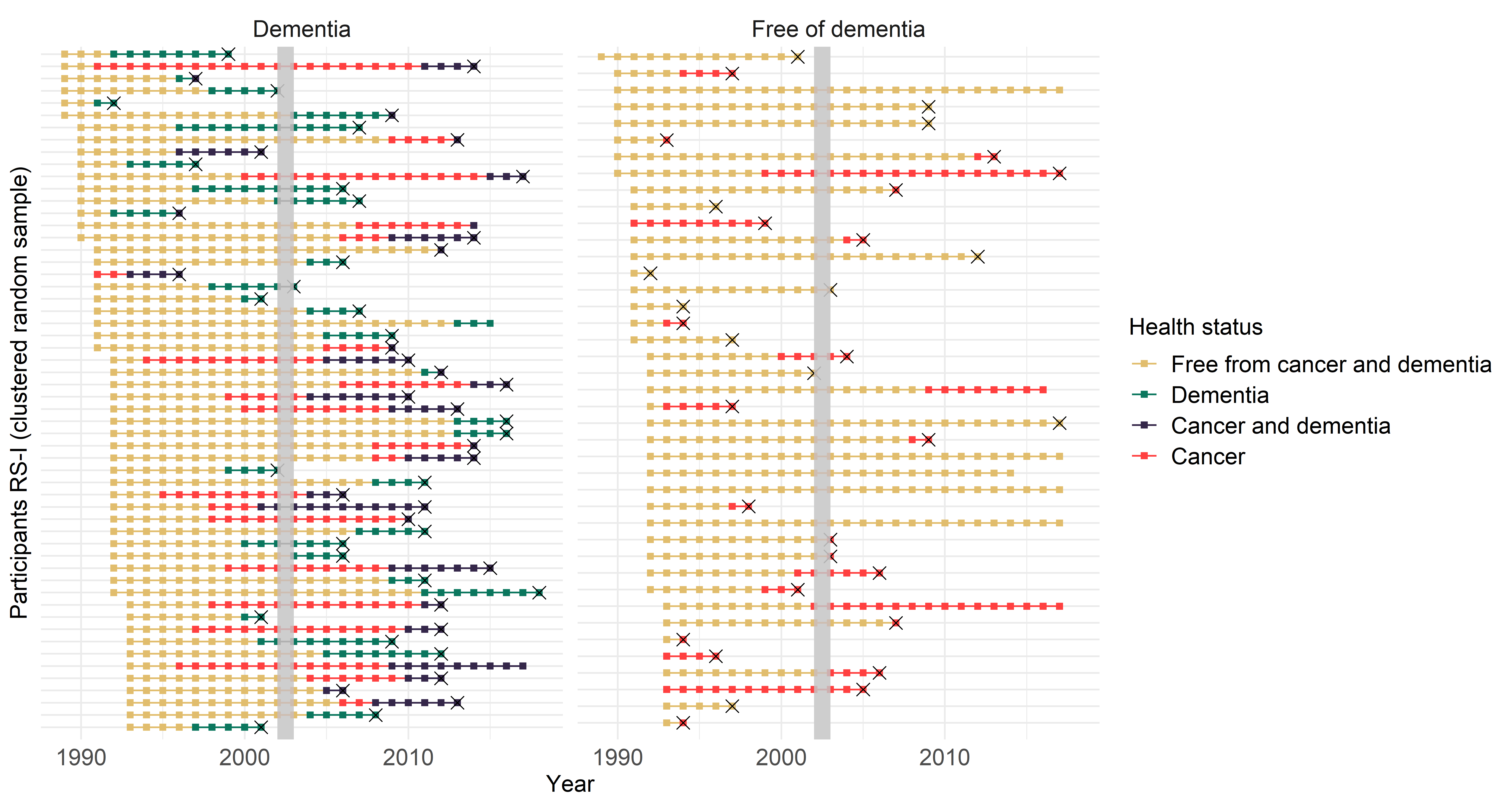


**Figure 1 Flowchart of outcomes stratified by incident cancer diagnosis.**



**Figure 2 Graphic overview of transition states for participants who were free of cancer and dementia at study entry and who developed cancer or dementia during follow-up and died.**

This graph consists of raw data (i.e., participants were not censored after dementia diagnosis and follow-up was not censored after 15 years) of 1556 participants. Out of these participants, 935 (60.1%) were diagnosed with cancer, 546 (36.2%) with dementia, and 75 (4.8%) with both cancer and dementia. The median (interquartile range) age at cancer diagnosis of these participants was 75.0 years (69.1 to 80.4), whereas the median (interquartile range) age at dementia diagnosis of these participants was 82.5 years (77.5 to 86.9).



**Figure 3 Graphic overview of cross-sectional study design when the assessment took place in 2002 and 2003.**

This graph consists of raw data (i.e., participants were not censored after dementia diagnosis and follow-up was not censored after 15 years) of a random sample of participants from the first subcohort of the Rotterdam Study. When assessing participants during the fourth visit round in 2002 and 2003, only those who have survived up to 2002 and 2003 will be included in the study population. Therefore, cross-sectional study designs may result in selection bias.

RS = Rotterdam Study.



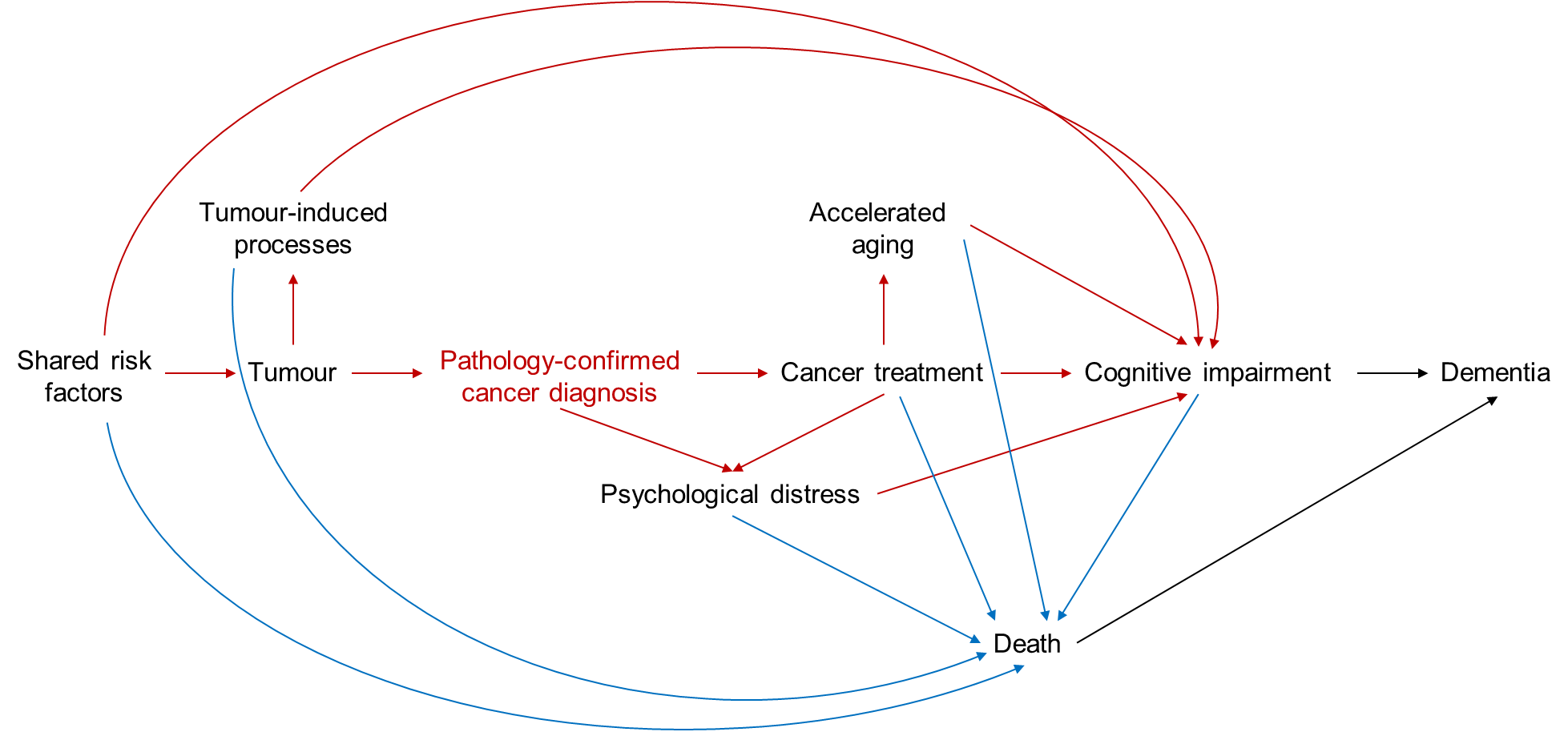
**Figure 4 Risk of death for participants who are diagnosed with cancer and those who remain free of cancer during follow-up.**

**V:\HomeDir\960006(K. Willik)\PhD dissertation\Figures\Solution 123 AI.tif**

**Figure 5 Standardised survival curves for dementia.**

Curves for participants with cancer are presented in yellow and for participants without cancer in blue. Panels on the left (A, C, and E) represent the risk of dementia when we eliminated death, i.e., the controlled direct effect. Panels on the right (B, D, and F) represent the risk of dementia regardless of death, i.e., the total effect. Panel A and B show the curves when immortal time is handled by treating cancer as time-depended variable. Panel C and D are the curves obtained after we computed IPW for the time until cancer diagnosis. Panel E and F show the curves after cloning and censoring.

IPW = inverse probability weights.



**Figure 6 Directed acyclic graph of the relation between cancer and dementia.**

When studying the relation between cancer and dementia, cancer is most often defined as cancer diagnosis that is confirmed by pathology. If the question is aetiological, i.e., does cancer cause dementia, then we should redefine this definition of cancer, because the diagnosis itself does not cause dementia. To estimate the causal effect of cancer on dementia, we should take the following four processes into account: (i) shared risk factors for both cancer and dementia such as age, genetics, smoking, and alcohol use; (ii) biological processes induced by the tumour itself including inflammation, vascular changes, oxidative stress, and production of extracellular vesicles; (iii) cancer treatment that may accelerate the ageing process by causing DNA damage, changes in hormonal levels, and oxidative stress; and (iv) psychological distress, including depression, anxiety, stress, and fatigue caused by cancer and cancer treatment. Given that multiple arrows point towards death, conditioning on survival may induce collider bias.

**Table 1 Baseline characteristics of study population**

|  |  |
| --- | --- |
| **Characteristic** | **Participants**  **(N=8899)** |
| Age, years, median (IQR) | 65.7 (60.4 to 73.2) |
| Women, No. (%) | 5112 (57.4) |
| Education, No. (%) |  |
| Lower | 4220 (47.4) |
| Intermediate | 3695 (41.5) |
| Higher | 984 (11.1) |
| Smoking, No. (%) |  |
| Never | 2914 (32.7) |
| Former | 3931 (44.2) |
| Current | 2054 (23.1) |
| Body mass index, kg/m2, mean (SD) | 26.6 (3.8) |
| Systolic blood pressure, mm Hg, mean (SD) | 140.5 (22.2) |
| Hypertension, No. (%) | 5514 (62.0) |
| Diabetes mellitus, No. (%) | 852 (9.6) |
| History of coronary heart disease, No. (%) | 694 (7.8) |
| History of stroke, No. (%) | 254 (2.9) |

IQR = interquartile range, N = number of participants, SD = standard deviation.

**Table 2 Results of replication of previous study designs and analyses.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Analysis** | **N**  **cancer** | **N dementia** | **N**  **total** | **HR (95% CI)** |
| Cancer as time-dependent variable | 1813 | 875 | 8899 | 0.91 (0.71 to 1.16) |
| Survived up to at least 80 years | 619 | 523 | 3423 | 0.91 (0.67 to 1.26) |
| Cancer as time independent variable | 1813 | 875 | 8899 | 0.44 (0.35 to 0.56) |
| Nested case-control setting | 1805\* | 409 | 5414 | 0.53 (0.41 to 0.69) |
|  |  |  |  | **OR (95% CI)** |
| Dementia at ERGO4 | 359 | 50 | 5278 | 0.28 (0.02 to 1.31) |
| Dementia <6 months after ERGO4 | 405 | 142 | 5278 | 0.61 (0.26 to 1.24) |

Model is adjusted for baseline measurements of age, sex, education, smoking status, body mass index, systolic blood pressure, hypertension, diabetes mellitus, history of coronary heart disease, and history of stroke.

\* Eight out of 1813 participants with cancer could not be matched to a control. One participant with cancer was matched to only one control.

ERGO4 = fourth visit of the first subcohort and the second visit of the second subcohort, HR = hazard ratio, N = number of participants, OR = odds ratio.