

Collider and Reporting Biases Involved in the Analyses of Cause of Death Associations in Death Certificates: an Illustration with Cancer and Suicide

Moussa Laanani (✉ moussa.laanani@assurance-maladie.fr)

French Centre for Epidemiology on Medical Causes of Death (CépiDc-Inserm) <https://orcid.org/0000-0002-7200-1315>

Vivian Viallon

International Agency for Research on Cancer (IARC)

Joël Coste

Paris Descartes University

Grégoire Rey

French Centre for Epidemiology on Medical Causes of Death (CépiDc-Inserm)

Method Article

Keywords: causes of death, causal inference, collider bias, reporting bias

Posted Date: April 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-377726/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Population Health Metrics on December 14th, 2023. See the published version at <https://doi.org/10.1186/s12963-023-00320-y>.

Abstract

Background: Data from death certificates have been studied to explore causal associations between diseases. However, these analyses are subject to collider and reporting biases (selection and information biases, respectively).

Methods: We aimed to assess to what extent associations of causes of death estimated from individual mortality data can be extrapolated to the general population. We used a multistate model to generate populations of individuals and simulate their health states up to death from national health statistics and artificially replicate collider bias. Associations between health states can then be estimated from such simulated deaths by logistic regression and the magnitude of collider bias assessed. Reporting bias can be approximated by comparing the estimates obtained from the observed death certificates (subject to collider and reporting biases) with those obtained from the simulated deaths (subject to collider bias only).

Results: As an illustrative example, we estimated the association between cancer and suicide in French death certificates, and found that cancer was negatively associated with suicide. Collider bias, due to conditioning inclusion in the study population on death, increasingly downwarded the associations with cancer site lethality. Reporting bias was much stronger than collider bias and depended on the cancer site, but not prognosis.

Conclusions: These results argue for an assessment of the magnitude of both collider and reporting biases before performing analyses of cause of death associations exclusively from death certificates. If these biases cannot be corrected, results from these analyses should not be extrapolated to the general population.

Background

National cause-of-death data are widely used to describe the health of populations.[1] These data are exhaustive and collected in a standardized fashion, allowing international comparisons.[2] They are extracted from medical death certificates where certifiers (physicians or coroners) are asked to describe the causal sequence leading to death. These data have been studied to assess associations between diseases in the general population,[3–9] although the difficulties of such study design have long been emphasized.[10–12] For example, the risk of suicide in patients with Parkinson’s disease was estimated in an often-cited study based on death certificate data.[5] The authors found a 10-fold lower risk of suicide in people with Parkinson’s disease than for other individuals who died. However, instead of a decreased risk, prospective studies highlighted a two- to five-fold higher suicide risk in these patients.[13,14] Indeed, the design used in the first study (estimating associations between health states in the general population from death certificates) is subject to two main types of bias, which could explain misleading findings.

Collider bias

Studies based on death certificate data are conducted on non-representative samples of the general population. Indeed, even if all deaths are reported, no information is available on living individuals. This leads to a selection bias, as inclusion in the study population is conditioned on death, which is a common effect of the diseases under study (defined among causes of death), called a collider (Figure 1). This selection bias is called “collider bias”, or “bias due to conditioning on a collider”, and can strengthen or reverse associations between variables of interest.[15,16]

Reporting bias

Studies on death certificate data are also subject to information bias, which we hereafter refer to as “reporting bias”. This bias is the result of (1) the task assigned to the certifier, who has to report diseases and events that effectively contributed to death only, rather than all diseases present prior to death, and (2) possible incompleteness in the filling out of the death certificates (which depends, among other things, on the certifier’s level of knowledge of the deceased patient and his/her medical history).[17]

Aim and organisation of the paper

Seminal literature that warned on the risks of using comprehensive mortality data to assess associations between diseases only provided leads to reduce these risks, without giving a deep insight into the mechanisms of the biases involved. The general purpose of this paper was to assess to what extent associations of causes of death estimated from individual mortality data can be extrapolated to the general population, given collider and reporting biases. As an illustrative example, we estimated the association between cancer and suicide in death certificates depending on the cancer site and assessed the order of magnitude of the collider and reporting biases. In the first section of the paper, we describe how multiple causes-of-death data are constructed, from medical certification to medical coding of causes of death (including the international rules for the selection of the underlying cause of death). We also describe the framework for the assessment of associations between causes of death and the biases involved in such studies. In the second section, we present the methods and results of our illustrative example on cancer and suicide. Finally, we conclude by addressing recommendations for future studies and discussing how to improve the use of multiple causes-of-death data.

Analyses of cause of death associations in death certificates

Data from death certificates

Medical certification of death is mandatory in most industrialized countries and must be performed by a physician or a coroner. The World Health Organization [WHO] has designed the structure of the international medical death certificate with two parts: Part I is dedicated to the description of the causal sequence of events that directly led to death and Part II the reporting of significant morbid conditions that may have contributed to death, but are not involved in the sequence of events that directly led to death (Figure 2).

The WHO defines the *underlying cause of death* as “the disease or injury which initiated the train of morbid events leading directly to death or the circumstances of the accident or violence which produced the fatal injury”. Selection of the underlying cause of death is performed automatically by software (*e.g.*, Iris)[18] or based on the expertise of a mortality medical coder (or “nosologist”) for the most complex cases. This selection is governed by several rules prescribed by the WHO in the tenth revision of the International Statistical Classification of Diseases and Related Health Problems [ICD-10].[19] The main rule, called the “General Principle”, states that “when more than one condition is entered on the certificate, [...] the condition entered alone on the lowest used line of Part I” (*i.e.*, the first condition mentioned in the train of morbid events leading to death) must be selected as the underlying cause of death, “only if it could have given rise to all the conditions entered above it” (*i.e.* to the subsequent conditions of the train of morbid events leading to death). If the General Principle does not apply, Rules 1 and 2 state that the originating cause of the immediate (or final) cause of death, mentioned first in the train of morbid events leading to death, has to be selected as the underlying cause of death. Finally, Rule 3 states that “if the condition selected by the [previous rules] was obviously a direct consequence of another reported condition, whether in Part I or Part II”, this condition has to be selected as the underlying cause of death. For instance, HIV disease and external causes of death can meet Rule 3.

Framework for the assessment of associations between causes of death and the biases involved

Data from death certificates can be used to assess associations between health states (diseases and/or injuries) mentioned as causes of death. Standardised mortality ratios are a tool to assess such associations.[10,20] Multivariable logistic regression models can also be used, allowing adjustment for potential confounders. Odds ratios [OR], resulting from these models, convey information concerning both the direction of the association (the risk is higher if $OR > 1$ or lower if $OR < 1$) and its magnitude. When the prevalence of the assessed outcome is low, the OR is a good approximation of the relative risk and can be interpreted accordingly.[10]

Assessment of collider bias

Collider bias is due to conditioning the study sample on death. A multistate model can be used to generate populations of individuals and simulate their health states up to their deaths from national health statistics. Associations between health states can then be estimated from such simulated deaths (with logistic regression models, in the same way as with observed deaths) and the collider bias assessed, as these simulated deaths artificially replicate this bias. Collider bias can then be estimated from the following ratio:

$$\text{Collider bias} = \frac{\text{Real (unbiased) association measure}}{\text{Association measure estimated on the simulated deaths}}$$

Multiplicative measures of bias are better suited in this context, in which associations are expressed in the multiplicative scale (ORs).

Assessment of reporting bias

The magnitude of reporting bias can be approximated by the difference between the estimates obtained from observed death certificates (which are subject to both collider and reporting biases) and those obtained from simulated deaths (which are subject to collider bias only). Reporting bias can then be approximated from the following ratio:

$$\text{Collider bias} = \frac{\text{Real (unbiased) association measure}}{\text{Association measure estimated on the simulated deaths}}$$

However, the two sources of reporting bias ((1) the difference of the definition between measuring associations of diseases and measuring associations of causes of death and (2) the incomplete filling out of death certificates by certifiers) cannot be distinguished from one another.

Illustrative example: Association between cancer and suicide in death certificates in France

Suicide is a major public health issue, accounting for 1.4% of all deaths worldwide.[21] The impact of psychiatric diseases (notably, depression, anxiety, and psychotic disorders)[22] is well known, but somatic disorders may also play a role in the occurrence of suicide deaths. Cancer, due to its impact on health, the adverse events of treatments, and stigma, can substantially reduce the quality of life and promote the onset of suicidal ideation and suicide deaths. This phenomenon can vary depending on the cancer site prognosis, notably after receiving the diagnosis.[23]

Our illustrative example is based on French multiple causes-of-death data. Data from death certificates are commonly used to study suicide mortality and its determinants, with various study designs: ecological studies,[24] studies based on disease registries,[25] analyses of cause of death associations.[5,6] Inclusion in our study population was structurally conditioned on death, a common effect of cancer (the exposure) and suicide (the outcome), *i.e.* a collider (Figure 1). We first measured the cancer/suicide association in death certificates, according to cancer site, and then assessed the magnitude of the collider and reporting biases, using simulations.

Methods

Data from death certificates

All deaths of people aged 15 years or older occurring in mainland France between 2000 and 2013 were included in the study, provided that at least one cause was mentioned. Causes of death were coded (throughout the study period) according to the ICD-10.[19] Suicide (ICD-10 codes: X60 to X84 and Y87.0) was defined from the underlying causes of death, as suicide meets Rule 3 criteria: wherever 'suicide' is mentioned on the death certificate, it is almost always selected as the underlying cause of death, even if the certifier indicated that suicide was secondary to depressive disorders or cancer. Cancer (ICD-10 codes: C, see the list of

cancer sites in Additional table 1) was defined from both the underlying cause of death and Part II diagnoses; if cancer was not the first cause in the train of morbid events leading to death declared by the certifier in Part I of the death certificate, it was sought among all other diagnoses, except those mentioned between the immediate cause of death and the underlying cause of death selected by following WHO rules, considered to be consequences of the underlying cause of death. This type of situation is relatively uncommon and concerns exclusively cancer associated with HIV/AIDS.[19] Such a focus on the first cancer site mentioned in the train of morbid events leading to death prevents consideration of secondary cancer sites (including metastases).

Simulation scheme

We performed a simulation study to assess the direction and magnitude of the collider bias involved in this illustrative example. A population of individuals was generated using national statistics of mortality and cancer incidence to simulate the occurrence of cancer as well as death from cancer, suicide, and other causes. The deaths thus simulated were used to artificially replicate the collider bias. A first simulation study was conducted under the null hypothesis of no cancer/suicide association to assess whether collider bias alone could induce high amplitude false associations and determine the direction of such bias. A second simulation study was conducted to approximate the magnitude of the collider bias, using approximations of the real cancer/suicide associations in the French population. To do so, we used relative risks of suicide death for several cancer sites estimated in a recent large cohort study conducted from national Swedish registers (Fang *et al.*'s study).[23]

Deaths from suicide, cancer, and other causes for people aged 15 or older were simulated using a multistate model, with deaths as absorbing states (Figure 3). Transition probabilities to move from one state to another within a year were functions of age and gender. Simulations were performed separately for each gender. Individuals entered the model at age 15 years in the initial healthy state. Individuals could then transit to one of the K cancer states (for K cancer sites listed in Additional table 1) or die from suicide or other causes. Once in one of the K cancer states, individuals could die from the K th cancer, suicide, or other causes, or go back to the healthy state if they did not die within five years. Transition probabilities were derived from national suicide mortality[26] and cancer incidence[27] and survival[28] statistics. Considering individuals in a K th cancer state, net survival was used as the probability of death from the K th cancer and the difference between net and crude survival as the probability of death from other causes.[28] The probability of suicide death for individuals in a K th cancer state was obtained by multiplying the relative risk of suicide corresponding to the K th cancer site by the national suicide mortality rate. In the first simulation study, the relative risks of suicide used were equal to one for every cancer site (to mimic the null hypothesis of no cancer/suicide association); in the second simulation study, those published in the study of Fang *et al.* were applied.[23] For cancer sites not assessed in their study, the mean relative risk of suicide for other cancer sites was used. The simulations were performed using R (V3.4.0).[29]

Statistical analyses

Associations between cancer sites and suicide were estimated for both observed and simulated deaths, with logistic regression models adjusted for age (B-spline with 3 degrees of freedom), gender, and, for observed data, region of death. Analyses were conducted for both genders together for the cancer sites studied by Fang *et al.*, and, in complementary analyses, for men and women separately for the cancer sites listed in Additional table 1, as both cancer epidemiology and suicide epidemiology differ according to gender.[30]

The direction of collider bias was determined using the ORs obtained from the first simulation study (under the null hypothesis). If the OR obtained in the first simulation was lower than 1, the direction of collider bias was considered to be negative, whereas it was considered to be positive if it was higher than 1.

The magnitude of the collider bias was assessed using the second simulation. In the absence of collision, ORs obtained in the second simulation study should be similar to those reported by Fang *et al.* The magnitude of collider bias was then evaluated by computing the ratio between the relative risk of suicide from the study of Fang *et al.* and the OR estimated from the second simulation. As suicide deaths occur rarely in the population, OR and relative risk values can be considered to be relatively similar.

The magnitude of the reporting bias was evaluated by comparing the OR estimated from the second simulation and that estimated from observed death certificates. Under the assumptions that our simulations correctly reproduced the French mortality data, that the cancer/suicide associations found by Fang *et al.* are close to those existing in the French population, and that there is no remaining confounders, differences between the results obtained using the data from the second simulation study and the observed deaths are likely to be largely attributable to reporting bias.

Statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina).[31]

Results

French mortality data

Overall, 7.2 million deaths between 2000 and 2013 were considered (3,685,024 of men, of which 107,241 were suicides (3%), and 3,553,707 of women, of which 38,297 were suicides (1%)). The number of deaths (suicide or other causes) according to the presence or not of a cancer diagnosis among causes of death are detailed in Table 1. The analyses performed on data from death certificates showed a highly negative association between suicide and each cancer site (OR adjusted for age, gender, and region of death ranged from 0.01 for central nervous system cancer and cutaneous melanoma, 95% confidence intervals (95% CI) [0.01-0.01] and [0.01-0.02], respectively, to 0.24 for prostate cancer, 95% CI=[0.22-0.26]; see Table 2). The study of Fang *et al.* found a positive association between suicide and each cancer site (with adjusted relative risk from 1.4 for cutaneous melanoma to 4.5 for oesophageal, liver, and pancreatic cancer) (Table 2). Our results were thus inconsistent with theirs.

Estimation of the magnitude of the biases

Each simulation generated 4.7 million deaths for men, of which 2% were suicides, and 4.6 million for women, of which 1% were suicides. The proportion of deaths due to each cause and age distributions at death were similar between the simulated data and that from death certificates (Additional table 2). The first simulation study, conducted under the null hypothesis of no cancer/suicide association, found a negative association for each cancer site, with OR ranging from 0.11 (95% CI=[0.09-0.14]) for central nervous system cancer to 0.71 (95% CI=[0.68-0.75]) for prostate cancer. In the absence of collision, these ORs were expected to be 1 for all cancer sites. The results were thus biased downward by collision.

The second simulation (conducted using the relative risks of suicide published by Fang *et al.*)[23] found a negative cancer/suicide association for all cancer sites (OR from 0.25, 95% CI=[0.22-0.28], for central nervous system cancer to 0.85, 95% CI=[0.78-0.92], for breast cancer), except for prostate cancer (OR=1.14, 95% CI=[1.10-1.18]), although it was lower than the relative risk reported by Fang *et al.*(1.9). In the absence of collider bias, these ORs were expected to be similar to those published by Fang *et al.*

Collider bias was estimated to divide the relative risk of suicide reported by Fang *et al.* by at least 1.7 (for prostate cancer) and up to 9.3 (for central nervous system cancer). The magnitude of collider bias thus varied according to cancer site and appeared to increase with cancer site lethality, as expected. Estimating collider bias from simulation #1 (with the inverse of the obtained OR) produced consistent results. Reporting bias was found to divide the OR of suicide from the second simulation (*i.e.* the relative risk of Fang *et al.* biased by collision) by at least 4.7 (for prostate cancer) and up to 64 (for cutaneous melanoma). Using our approximation, the magnitude of reporting bias was thus much higher than that of collider bias. The magnitude also varied depending on the cancer site, but did not appear to be associated with cancer site lethality. The magnitudes of the collider and reporting biases are presented in Figure 4.

Complementary analyses performed for each gender separately gave similar results for men (Additional table 3). The results were slightly different for women, with a higher overall magnitude of bias. We found the lowest magnitude for collider bias for cutaneous melanoma and the highest for lung cancer, and the lowest magnitude for reporting bias for oesophageal cancer and the highest for liver cancer (Additional table 4).

Discussion

Here, we demonstrated that estimating associations between diseases from death certificate data is exposed to biases and used an illustrative example to assess their direction and magnitude. The cancer/suicide association was inverse when assessed based on data from death certificates (OR ranging from 0.24 for prostate cancer to 0.01 for central nervous system cancer and cutaneous melanoma). However, previous longitudinal studies found positive associations, as notably reported by Fang *et al.*, who found a relative risk of suicide that ranged from 1.4 to 4.5, depending on the cancer site.[23] Part of this discrepancy is attributable to collider bias, which naturally arises when cancer/suicide associations are assessed from death certificates.[15,16] We performed simulations to artificially reproduce collider bias by generating deaths from national statistics of suicide and cancer incidence and mortality. Analyses performed on such simulated deaths showed that conditioning inclusion in the study population on death biased the results towards negative associations, the bias increasing with cancer-site lethality. However, such collider bias was not sufficient to fully explain the discrepancies between the results based on death certificates and those reported by Fang *et al.* Although there are other potential explanations (the two source populations differed, as the study of Fang *et al.* was performed in Sweden), we believe that the remaining bias can be largely attributed to reporting bias.[17,32] Our approximation of reporting bias was much stronger than collider bias and depended on the cancer site, but not the prognosis, as the magnitude of the reporting bias varied between cancer sites, but not according to cancer lethality.

Biases involved in the analyses of cause of death associations in death certificates

Collider bias was first described recently[33] and is of increasing concern among epidemiologists. This type of selection bias has been the source of much scientific debate, such as for the so-called “birth weight paradox”. Let us consider, for example, the risk of neonatal death associated with maternal smoking, which is known to increase the risks of both low birth weight and neonatal mortality. Comparing mortality rates between low birth weight infants born to smokers and those born to non-smokers paradoxically lead to finding lower mortality rates in infants of smokers.[34] Such results “raised doubts” about the pejorative impact of maternal smoking.[35] However, this paradox may be explained by collider bias, as demonstrated by Hernández-Díaz *et al.*[36] Indeed, low birth weight is a collider on which selection in the study sample is conditioned, as it is a common effect of maternal smoking and other unmeasured causes (such as birth defects or malnutrition). The “obesity paradox” is another example of a scientific controversy that may be explained by collider bias. This paradox refers to the lower mortality observed for obese patients, found, for example, among patients with diabetes.[37–40] Collider bias should be considered in all studies conducted with a case-only design,[16] notably those analysing associations of causes of death from death certificates. To our knowledge, our study is the first to consider collider bias in this specific type of studies.

Interpreting reporting bias is challenging and requires consideration of its two sources. This type of information bias is due (1) to the difference between what is asked of the certifier (*i.e.* reporting a causal sequence of injuries and diseases leading to death) and the information that would be expected for epidemiology (*i.e.* diseases reported regardless of their potential causal link with death).[19] This specificity gives multiple causes-of-death databases their particular interest as they thus provide the opportunity to assess causal relations between diseases or morbid conditions. In return, the information available in causes-of-death data is very conservative. Reporting bias is also due (2) to the incompleteness of certificate filling by certifiers. This depends on the certifier’s knowledge of the deceased patient’s medical history and knowledge (or intuition) of the possibility of a causal association between the underlying cause of death and its comorbidities.[41,42] In our application, without knowledge/intuition of the plausible link between one’s cancer and suicide, the certifier might not mention cancer on the certificate.

Unmeasured confounding is a source of bias we did not address in this paper.[43] We rather focussed on collider and reporting bias for pedagogical reasons to correctly identify them. Unmeasured confounding is often involved when one wants to estimate causal effects. We aimed in our illustrative example to compare our associational ORs with the associational risk ratios of Fang *et al.* In this situation, confounding may be considered to be negligible and is essentially amongst the supplementary factors for which Fang *et al.* adjusted their models.[23] Both our study and that of Fang *et al.* adjusted for age and gender, which are major confounders in the cancer/suicide association. Fang *et al.* also adjusted their models for cohabitation status, socioeconomic status, and educational level, but did not adjust for other major confounders in the cancer/suicide association, such as alcohol consumption.[22,44]

Conclusions

While risks of using comprehensive mortality data to assess associations between diseases have long been highlighted,[10–12] our work aimed to explain the mechanisms of the biases involved in such studies. We used a conceptual framework to demonstrate the impossibility of measuring causal associations from multiple causes-of-death data. We used a simulation study to assess the magnitude of the involved biases, accounting for the specificities of death certificates. Even if we could have tried to correct for collider bias in our illustrative example (by an indicator of cancer-site prognosis, such as survival rate), our results show that reporting bias was of much higher magnitude and heterogeneous across cancer sites. Reporting bias cannot be corrected, as the reason for such heterogeneity could not be clearly linked with the cancer site characteristics. In analyses of cause of death associations exclusively from death certificates, if the reporting bias is too strong, there is little use in correcting for collider bias and results from these analyses should not be extrapolated to the general population. Multiple causes-of-death data are still a remarkably rich source because of their standardized construction and international comparability and because they contain directed causal information, integrating the expert knowledge of the physician or coroner certifying death. The analyses of these data should be performed after full linkage to comprehensive databases, such as registers or medical administrative databases, to take full advantage of these qualities and avoid drawing conclusions based on spurious associations.[45] The issue raised here regarding collider bias can be extended to other case-only designs,[16] including studies on pharmacovigilance databases or disease registries; reporting bias issues are specific to each data type.

List Of Abbreviations

95% CI: 95% confidence interval

CNS: central nervous system

HIV/AIDS: human immunodeficiency virus / acquired immune deficiency syndrome

ICD-10: tenth revision of the International Statistical Classification of Diseases and Related Health Problems

IQR: interquartile range

OR: odds ratio

WHO: World Health Organization

Declarations

Ethics approval and consent to participate

This study was conducted within the framework of law L2223-42, decree 2017-602, and French data protection agency (*Commission Nationale de l'Informatique et des Libertés*) decision number 2017-067.[46–48]

Consent for publication

Not applicable

Availability of data and materials

Anonymized individual data from death certificates used in this work can be shared by the authors under strict security conditions. Applications to access the French mortality data must be submitted to the National Institute for Health Data (*Institut National des Données de Santé*, INDS, www.indsante.fr/).

Aggregated data used for the simulation study are publicly available:

- national suicide mortality rates are available in the CépiDc (French Centre for Epidemiology on Medical Causes of Death) repository, <http://www.cepidc.inserm.fr/inserm/html/index2.htm>[26]
- cancer incidence and survival rates are available in the SPF (Public Health France) repository, <http://invs.santepubliquefrance.fr/Dossiers-thematiques/Maladies-chroniques-et-traumatismes/Cancers/Surveillance-epidemiologique-des-cancers/Estimations-de-l-incidence-de-la-mortalite-et-de-la-survie-stade-au-diagnostic>,[27,28] or on the INCa (French National Cancer Institute) repository, <https://lesdonnees.e-cancer.fr/>,[49]

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no funding.

Authors' contribution

GR and JC designed and supervised the study. ML contributed to the analyses and drafted the manuscript. VV performed the simulation study. All authors contributed to the interpretation of data, and read and approved the final manuscript.

Acknowledgements

The authors would like to thank William Hempel for English editing of the manuscript. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

References

1. WHO Mortality Database [Internet]. WHO. [cited 2018 Aug 6]. Available from: http://www.who.int/healthinfo/mortality_data/en/
2. AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, Lopez AD. Towards universal civil registration and vital statistics systems: the time is now. *Lancet Lond Engl*. 2015;386:1407–18.
3. Goodman RA, Manton KG, Nolan TF, Bregman DJ, Hinman AR. Mortality data analysis using a multiple-cause approach. *JAMA*. 1982;247:793–6.
4. Yang Q, Rasmussen SA, Friedman JM. Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. *Lancet Lond Engl*. 2002;359:1019–25.
5. Myslobodsky M, Lalonde FM, Hicks L. Are Patients with Parkinson's Disease Suicidal? *J Geriatr Psychiatry Neurol*. 2001;14:120–4.
6. Rockett IRH, Wang S, Lian Y, Stack S. Suicide-associated comorbidity among US males and females: a multiple cause-of-death analysis. *Inj Prev*. 2007;13:311–5.
7. Aouba A, Gonzalez Chiappe S, Eb M, Delmas C, de Boysson H, Bienvenu B, et al. Mortality causes and trends associated with giant cell arteritis: analysis of the French national death certificate database (1980-2011). *Rheumatol Oxf Engl*. 2018;57:1047–55.
8. Egidi V, Salvatore MA, Rivellini G, D'Angelo S. A network approach to studying cause-of-death interrelations. *Demogr Res*. 2018;38:373–400.

9. Viallon V, Banerjee O, Jouglu E, Rey G, Coste J. Empirical comparison study of approximate methods for structure selection in binary graphical models. *Biom J Biom Z.* 2014;56:307–31.
10. Rothman KJ, Lash TL, Greenland S. *Modern Epidemiology. Third, Mid-cycle revision.* Philadelphia, PA, USA: Lippincott Williams and Wilkins; 2012.
11. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol.* 1992;135:1029–41.
12. McLaughlin JK, Blot WJ, Mehl ES, Mandel JS. Problems in the use of dead controls in case-control studies. I. General results. *Am J Epidemiol.* 1985;121:131–9.
13. Lee T, Lee HB, Ahn MH, Kim J, Kim MS, Chung SJ, et al. Increased suicide risk and clinical correlates of suicide among patients with Parkinson's disease. *Parkinsonism Relat Disord.* 2016;32:102–7.
14. Kostić VS, Pekmezović T, Tomić A, Ječmenica-Lukić M, Stojković T, Špica V, et al. Suicide and suicidal ideation in Parkinson's disease. *J Neurol Sci.* 2010;289:40–3.
15. Hernán MA, Robins JM. *Causal Inference.* Boca Raton, FL, USA: Chapman & Hall/CRC; forthcoming.
16. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol.* 2010;39:417–20.
17. Richaud-Eyraud E, Rondet C, Rey G. [Transmission of death certificates to CepiDc-Inserm related to suspicious deaths, in France, since 2000]. *Rev Epidemiol Sante Publique.* 2018;
18. Johansson LA, Pavillon G. IRIS: A language-independent coding system based on the NCHS system MMDS. Tokyo, Japan; 2005.
19. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. WHO; 2008.
20. Israel RA, Rosenberg HM, Curtin LR. Analytical potential for multiple cause-of-death data. *Am J Epidemiol.* 1986;124:161–79.
21. Disease burden and mortality estimates [Internet]. WHO. [cited 2018 May 23]. Available from: http://www.who.int/healthinfo/global_burden_disease/estimates/en/
22. Hawton K, van Heeringen K. Suicide. *Lancet.* 2009;373:1372–81.
23. Fang F, Fall K, Mittleman MA, Sparén P, Ye W, Adami H-O, et al. Suicide and cardiovascular death after a cancer diagnosis. *N Engl J Med.* 2012;366:1310–8.
24. Chang S-S, Stuckler D, Yip P, Gunnell D. Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. *BMJ.* 2013;347:f5239–f5239.
25. Schairer C, Brown LM, Chen BE, Howard R, Lynch CF, Hall P, et al. Suicide After Breast Cancer: an International Population-Based Study of 723 810 Women. *JNCI J Natl Cancer Inst.* 2006;98:1416–9.
26. French national causes of death register (Centre for Epidemiology on Medical Causes of Death) [Internet]. CépiDc-INSERM. [cited 2017 Nov 3]. Available from: <http://www.cepidc.inserm.fr/inserm/html/index2.htm>
27. Binder-Foucard F, Belot A, Delafosse P, Remontet L, Woronoff A-S, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Partie 1 - Tumeurs solides. Saint-Maurice, France: Institut de veille sanitaire; 2013 p. 122.
28. Cowppli-Bony A, Uhry Z, Remontet L, Guizard A-V, Voirin N, Monnereau A, et al. Survie des personnes atteintes de cancer en France, 1989-2013. Etude à partir des registres des cancers du réseau Francim. Partie 1 - tumeurs solides. Saint-Maurice, France: Institut de veille sanitaire; 2016 Feb p. 274.
29. R. Vienna, Austria: The R Foundation for Statistical Computing;
30. Kendal WS. Suicide and cancer: a gender-comparative study. *Ann Oncol Off J Eur Soc Med Oncol.* 2007;18:381–7.
31. SAS. Cary, NC, USA: Statistical Analysis System;
32. Aouba A, Péquignot F, Camelin L, Jouglu E. [Quality assessment and improvement in the knowledge of suicide mortality data, metropolitan France, 2006]. *Bull Epidémiol Hebd.* 2011;497–500.
33. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiol Camb Mass.* 2003;14:300–6.

34. Wilcox AJ. Birth Weight and Perinatal Mortality: The Effect of Maternal Smoking. *Am J Epidemiol.* 1993;137:1098–104.
35. Yerushalmy J. The relationship of parents' cigarette smoking to outcome of pregnancy—implications as to the problem of inferring causation from observed associations. *Am J Epidemiol.* 1971;93:443–56.
36. Hernández-Díaz S, Schisterman EF, Hernán MA. The birth weight “paradox” uncovered? *Am J Epidemiol.* 2006;164:1115–20.
37. Carnethon MR, De Chavez PJD, Biggs ML, Lewis CE, Pankow JS, Bertoni AG, et al. Association of weight status with mortality in adults with incident diabetes. *JAMA.* 2012;308:581–90.
38. Banack HR, Kaufman JS. The obesity paradox: understanding the effect of obesity on mortality among individuals with cardiovascular disease. *Prev Med.* 2014;62:96–102.
39. Sperrin M, Candlish J, Badrick E, Renehan A, Buchan I. Collider Bias Is Only a Partial Explanation for the Obesity Paradox. *Epidemiol Camb Mass.* 2016;27:525–30.
40. Viallon V, Dufournet M. Re: Collider Bias Is Only a Partial Explanation for the Obesity Paradox. *Epidemiol Camb Mass.* 2017;28:e43–5.
41. Smith Sehdev AE, Hutchins GM. Problems with proper completion and accuracy of the cause-of-death statement. *Arch Intern Med.* 2001;161:277–84.
42. Mieno MN, Tanaka N, Arai T, Kawahara T, Kuchiba A, Ishikawa S, et al. Accuracy of Death Certificates and Assessment of Factors for Misclassification of Underlying Cause of Death. *J Epidemiol.* 2016;26:191–8.
43. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health.* 2001;22:189–212.
44. Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Bouvard V, et al. Carcinogenicity of alcoholic beverages. *Lancet Oncol.* 2007;8:292–3.
45. Rey G, Bounebach K, Rondet C. Causes of deaths data, linkages and big data perspectives. *J Forensic Leg Med.* 2018;57:37–40.
46. Code général des collectivités territoriales - Article L2223-42. Code Général Collectiv. Territ.
47. Décret n° 2017-602 du 21 avril 2017 relatif au certificat de décès [Internet]. Available from: <https://www.legifrance.gouv.fr/eli/decret/2017/4/21/AFSP1705016D/jo/texte>
48. Délibération n° 2017-067 du 16 mars 2017 portant avis sur un projet de décret relatif au certificat de décès modifiant le code général des collectivités territoriales (demande d'avis n° 16023949).
49. Cancer data - French National Cancer Institute [Internet]. [cited 2019 Apr 16]. Available from: <https://lesdonnees.e-cancer.fr/>

Tables

Table 1. Characteristics of the study population (data observed from death certificates, France, 2000-2013)

Gender	Cause of death	Number (%)	Age at death, median [IQR]
Men	Overall	3,685,024 (100%)	76 [64-84]
	Suicide		
	Bladder cancer	150 (0%)	78 [69-83]
	CNS cancer	20 (0%)	56 [45-64]
	Colorectal cancer	331 (0%)	75 [68-81]
	Cutaneous melanoma	13 (0%)	76 [58-80]
	Head and neck cancer	157 (0%)	65 [55-75]
	Kidney cancer	77 (0%)	71 [65-77]
	Larynx cancer	103 (0%)	67 [57-75]
	Liver cancer	82 (0%)	72 [63-76]
	Lung cancer	448 (0%)	70 [62-77]
	Oesophageal cancer	104 (0%)	71 [61-79]
	Pancreatic cancer	110 (0%)	73 [65-80]
	Prostate cancer	643 (0%)	79 [72-84]
	Stomach cancer	100 (0%)	75 [65-81]
	Testis cancer	9 (0%)	52 [44-55]
	Thyroid gland cancer	6 (0%)	76 [47-78]
	Other or no cancer ^a	104,888 (2.9%)	50 [39-66]
	Other cause		
	Bladder cancer	53,867 (1.5%)	77 [69-84]
	CNS cancer	24,204 (0.7%)	64 [53-73]
	Colorectal cancer	130,632 (3.5%)	76 [67-82]
	Cutaneous melanoma	12,298 (0.3%)	68 [56-78]
	Head and neck cancer	46,689 (1.3%)	62 [54-71]
	Kidney cancer	30,515 (0.8%)	73 [63-81]
	Larynx cancer	19,431 (0.5%)	66 [57-76]
	Liver cancer	80,578 (2.2%)	71 [63-78]
Lung cancer	309,349 (8.4%)	68 [59-77]	
Oesophageal cancer	45,215 (1.2%)	68 [58-77]	
Pancreatic cancer	61,217 (1.7%)	71 [62-79]	
Prostate cancer	154,378 (4.2%)	82 [75-87]	
Stomach cancer	43,501 (1.2%)	74 [64-81]	
Testis cancer	1,503 (0%)	45 [33-63]	
Thyroid gland cancer	2,227 (0.1%)	72 [62-80]	
Other or no cancer ^a	2,562,179 (70%)	78 [67-86]	
Women	Overall	3,553,707 (100%)	84 [76-90]

Gender	Cause of death	Number (%)	Age at death, median [IQR]	
<i>Suicide</i>	Bladder cancer	10 (0%)	76 [57-87]	
	Breast cancer	264 (0%)	63 [52-74]	
	CNS cancer	4 (0%)	62 [50-77]	
	Colorectal cancer	53 (0%)	73 [63-83]	
	Corpus uteri cancer	7 (0%)	70 [62-77]	
	Cutaneous melanoma	6 (0%)	77 [74-79]	
	Head and neck cancer	17 (0%)	61 [57-76]	
	Kidney cancer	11 (0%)	73 [62-81]	
	Larynx cancer	4 (0%)	54 [50-59]	
	Liver cancer	4 (0%)	68 [63-72]	
	Lung cancer	37 (0%)	67 [56-78]	
	Oesophageal cancer	10 (0%)	73 [51-80]	
	Ovary cancer	8 (0%)	66 [58-70]	
	Pancreatic cancer	19 (0%)	71 [68-81]	
	Stomach cancer	9 (0%)	64 [53-70]	
	Thyroid gland cancer	5 (0%)	79 [67-81]	
	Other or no cancer ^a	37,829 (1.1%)	53 [42-68]	
	<i>Other cause</i>	Bladder cancer	16,694 (0.5%)	82 [75-88]
		Breast cancer	180,821 (5.1%)	74 [60-84]
		CNS cancer	18,831 (0.5%)	68 [57-77]
Colorectal cancer		116,793 (3.3%)	80 [71-87]	
Corpus uteri cancer		10,105 (0.3%)	75 [67-82]	
Cutaneous melanoma		10,472 (0.3%)	73 [57-83]	
Head and neck cancer		9,737 (0.3%)	67 [56-80]	
Kidney cancer		17,161 (0.5%)	78 [69-85]	
Larynx cancer		2,125 (0.1%)	67 [57-79]	
Liver cancer		28,184 (0.8%)	77 [69-84]	
Lung cancer		89,479 (2.5%)	69 [57-79]	
Oesophageal cancer		10,597 (0.3%)	75 [63-83]	
Ovary cancer		47,964 (1.4%)	73 [63-81]	
Pancreatic cancer		57,766 (1.6%)	78 [69-84]	
Stomach cancer		25,717 (0.7%)	80 [70-87]	
Thyroid gland cancer		3,896 (0.1%)	79 [71-86]	
Other or no cancer ^a		2,869,068 (81%)	86 [79-91]	

CNS: central nervous system; IQR: interquartile range.

^a Includes other cancer sites, multiple cancers, and haematological malignancies.

Table 2. Suicide ORs by cancer site in observed and simulated mortality data, and estimated bias magnitudes

Cancer site	French mortality data		Simulation #1: Independence		Simulation #2: RR from Fang <i>et al.</i> [4]		Fang <i>et al.</i> study[4]	Collider bias ^b	Reporting bias ^c
	OR	[95% CI]	OR	[95% CI]	OR	[95% CI]	RR		
No cancer ^a	1.00		1.00		1.00		1.0	1.0	1.0
Prostate	0.24	[0.22;0.26]	0.71	[0.68;0.75]	1.14	[1.10;1.18]	1.9	1.7	4.7
Cutaneous melanoma	0.01	[0.01;0.02]	0.61	[0.53;0.70]	0.77	[0.68;0.86]	1.4	1.8	64
Breast	0.04	[0.03;0.04]	0.58	[0.52;0.64]	0.85	[0.78;0.92]	1.6	1.9	24
Colorectal	0.05	[0.05;0.06]	0.35	[0.33;0.37]	0.50	[0.48;0.52]	1.6	3.2	9.4
Oesophageal, Liver, Pancreatic	0.03	[0.03;0.03]	0.17	[0.16;0.19]	0.55	[0.53;0.58]	4.5	8.1	20
Lung	0.02	[0.02;0.02]	0.14	[0.13;0.15]	0.36	[0.35;0.38]	3.3	9.1	17
Central nervous system	0.01	[0.01;0.01]	0.11	[0.09;0.14]	0.25	[0.22;0.28]	2.3	9.3	35

Logistic regression models adjusted for age (B-spline of degree 3), gender, and region of death; mainland France, 2000-2010.

OR: odds ratio; RR: relative risk; 95% CI: 95% confidence interval.

^a Excluding other cancer sites, multiple cancers, and haematological malignancies.

The magnitude of the biases was estimated from the following ratios:

^b Collider bias = RR from Fang *et al.* / OR estimated from the data of simulation #2

^c Reporting bias = OR estimated from the data of simulation #2 / OR estimated from observed deaths

Figures

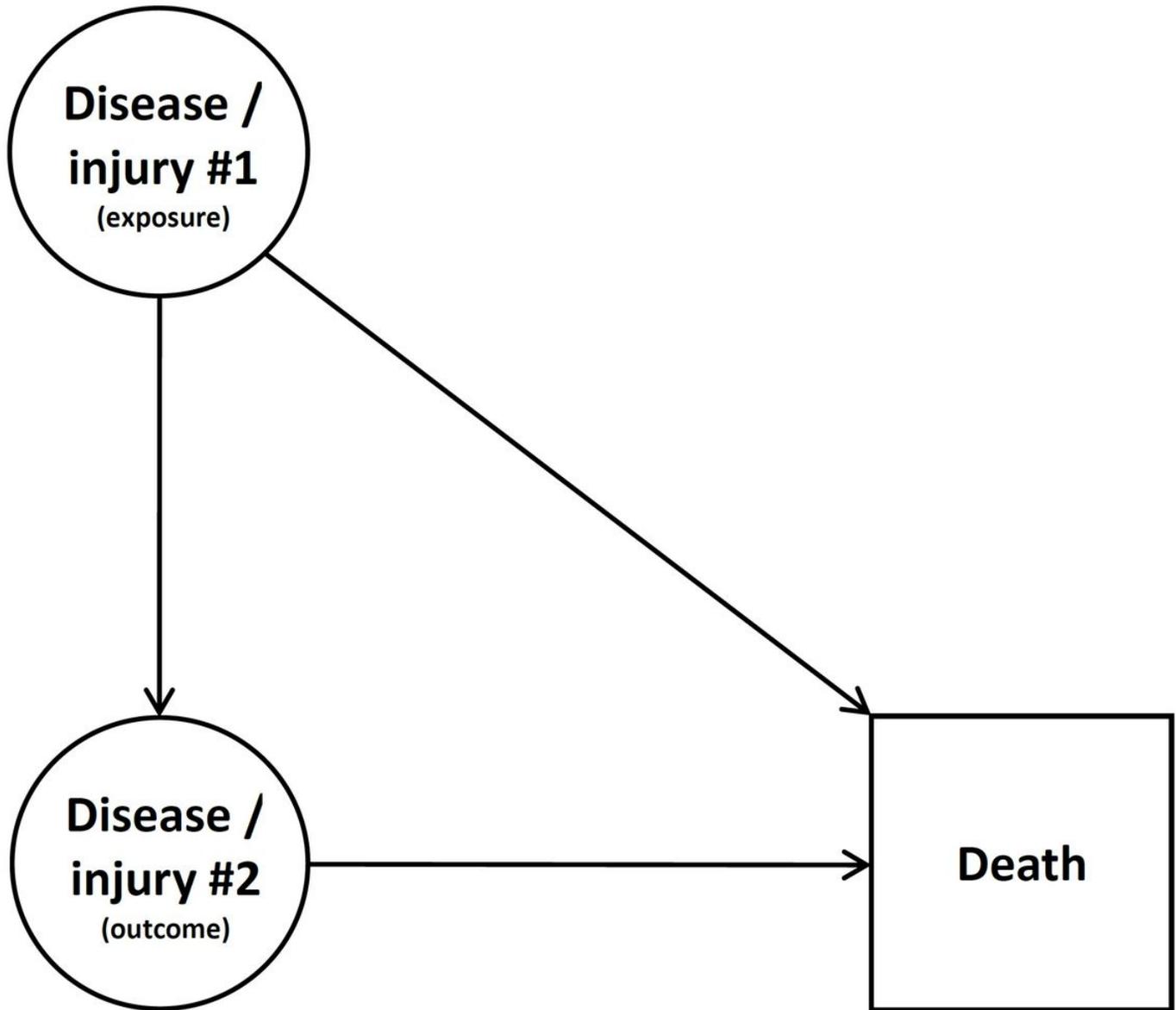


Figure 1

Directed acyclic graph representing the causal process underlying studies based on causes-of-death data. Collider bias emerges from conditioning inclusion in the study population on death, which is a consequence (descendant) of both exposure and outcome, which are the two factors defined among the diseases and injuries coded as causes of death for which the association is assessed. In our illustrative example, cancer is "disease #1" (the exposure), and suicide is "injury #2" (the outcome). Death is the collider on which inclusion in the study population is conditioned.

	Cause of death	Approximate interval between onset and death
I Disease or condition directly leading to death*	(a)
	due to (or as a consequence of)	
	(b)
	due to (or as a consequence of)	
Antecedent causes Morbid conditions, if any, giving rise to the above cause, stating the underlying condition last	(c)
	due to (or as a consequence of)	
	(d)
	due to (or as a consequence of)	
II Other significant conditions contributing to the death, but not related to the disease or condition causing it

<i>*This does not mean the mode of dying, e.g. heart failure, respiratory failure. It means the disease, injury, or complication that caused death.</i>		

Figure 2

International form of medical certificate of cause of death (WHO, ICD-10, 1993)

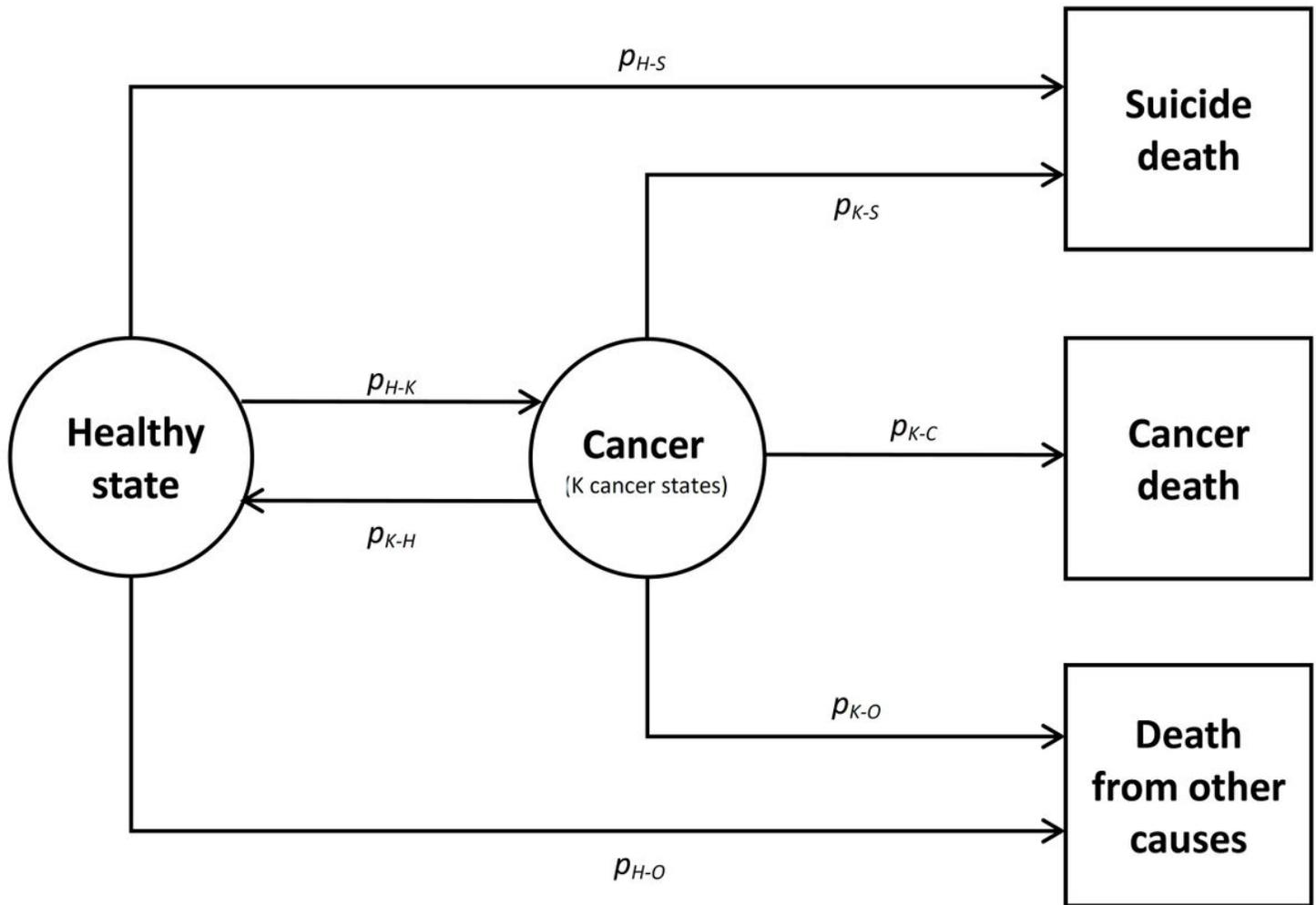


Figure 3

Multistate model used for the simulation of death data in people aged 15 or older Transition probabilities were obtained from national cancer incidence and survival and suicide mortality statistics:

- p_{H-S} = transition probability from the initial healthy state to the absorbing suicide death state
- p_{H-K} = transition probabilities from the initial healthy state to the Kth cancer state
- p_{H-O} = transition probability from the initial healthy state to the absorbing other causes of death state
- p_{K-H} = transition probabilities from the Kth cancer state to the initial healthy state
- p_{K-S} = transition probabilities from the Kth cancer state to the absorbing suicide death state
- p_{K-C} = transition probabilities from the Kth cancer state to the absorbing Kth cancer death state
- p_{K-O} = transition probabilities from the Kth cancer state to the absorbing other causes of death state

The probability of suicide death for individuals in a Kth cancer state p_{K-S} was obtained by multiplying the relative risk of suicide corresponding to the Kth cancer site by the national suicide mortality rate. In the first simulation study, the relative risks of suicide used were equal to one for every cancer site (to mimic the null hypothesis of no cancer/suicide association); in the second simulation study, those published in the study of Fang et al. were applied.[23]

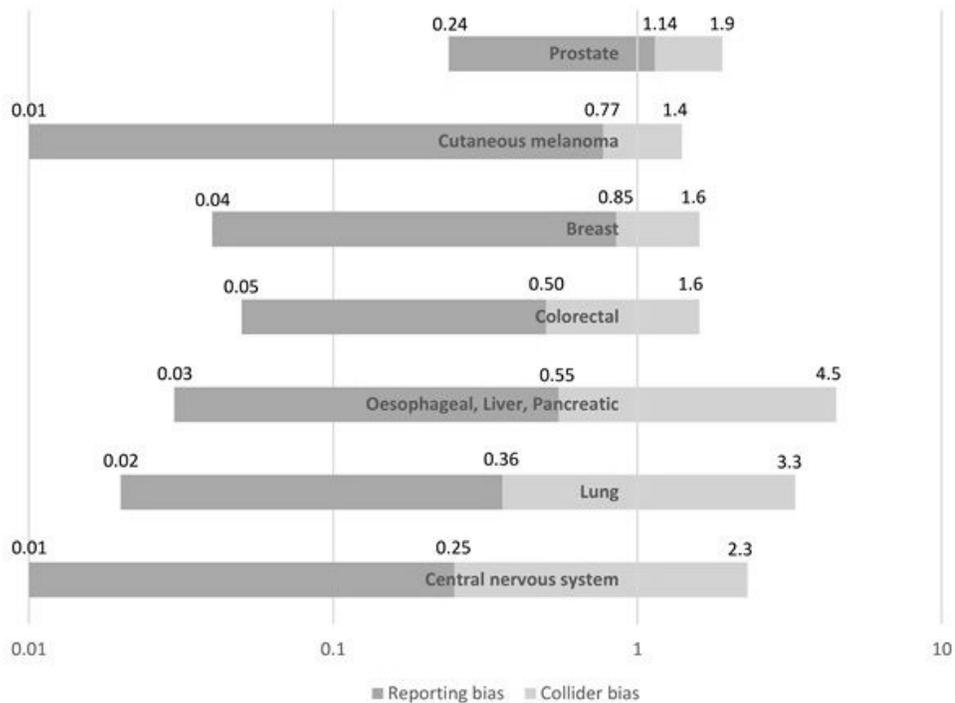


Figure 4

Magnitude of collider and reporting biases, according to cancer site The figure is interpreted as follows: The “real” relative risk (approximated from that of the study of Fang et al.) is at the right end of the bar. The light grey part of the bar represents the magnitude of the collider bias. The odds ratio from simulation #2 is at the junction between the light and the dark grey parts of the bar. The dark grey part of the bar represents the magnitude of the reporting bias. The observed odds ratio (obtained from French mortality data) is at the left end of the bar. For example, for breast cancer, the “real” relative risk of suicide is 1.6. The collider bias divides this relative risk by 1.9. The odds ratio from simulation #2 is 0.85. The reporting bias divides this odds ratio by 24. The odds ratio observed from French mortality data is 0.04. The scale of the x-axis is logarithmic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile.docx](#)