BMJ Open

Design and methodological characteristics of studies using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases: protocol for a meta-research study

Ferrán Catalá-López ,[1,2,3] Jane A Driver,[4,5,6] Matthew J Page,[7] Brian Hutton ,[3,8] Manuel Ridao ,[9] Clara Berrozpe-Villabona,[10] Adolfo Alonso-Arroyo ,[11,12] Cristina A Fraga-Medín ,[13] Enrique Bernal-Delgado ,[9] Alfonso Valencia ,[14] Rafael Tabarés-Seisdedos [2]

To cite: Catalá-López F, Driver JA, Page MJ, *et al*. Design and methodological characteristics of studies using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases: protocol for a meta-research study. *BMJ Open* 2022;**12**:e058738. doi:10.1136/bmjopen-2021-058738

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-058738).

Received 27 October 2021
Accepted 11 April 2022

For numbered affiliations see end of article.

Correspondence to
Dr Ferrán Catalá-López; ferran_catala@outlook.com

ABSTRACT

Introduction Health services generate large amounts of routine health data (eg, administrative databases, disease registries and electronic health records), which have important secondary uses for research. Increases in the availability and the ability to access and analyse large amounts of data represent a major opportunity for conducting studies on the possible relationships between complex diseases. The objective of this study will be to evaluate the design, methods and reporting of studies conducted using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases.

Methods and analysis This is the protocol for a meta-research study. We registered the study protocol within the Open Science Framework: https://osf.io/h2qjg. We will evaluate observational studies (eg, cohort and case–control) conducted using routinely collected health data for investigating the associations between cancer and neurodegenerative diseases (such as Alzheimer's disease, amyotrophic lateral sclerosis/motor neuron disease, Huntington's disease, multiple sclerosis and Parkinson's disease). The following electronic databases will be searched (from their inception onwards): MEDLINE, Embase and Web of Science Core Collection. Screening and selection of articles will be conducted by at least two researchers. Potential discrepancies will be resolved via discussion. Design, methods and reporting characteristics in each article will be extracted using a standardised data extraction form. Information on general, methodological and transparency items will be reported. We will summarise our findings with tables and graphs (eg, bar charts, forest plots).

Ethics and dissemination Due to the nature of the proposed study, no ethical approval will be required. We plan to publish the full study in an open access peer-reviewed journal and disseminate the findings at scientific conferences and via social media. All data will be deposited in a cross-disciplinary public repository.

Strengths and limitations of this study

- This meta-research study will provide an overview of the methodological and reporting quality of studies conducted using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases.
- This protocol increases transparency and completeness to the methods and definitions used in our planned meta-research study and that are applied to studies conducted using routinely collected health data in this area.
- We do not plan to include grey literature, only research articles published in peer-reviewed journals.
- Use of language restriction to English might exclude additional studies published in other languages.

INTRODUCTION

Cancer and neurological disorders, particularly age-related neurodegenerative diseases, are recognised as major causes of death and disease burden worldwide.[1–3] Multiple epidemiological studies[4–14] and reviews[15–21] have examined the epidemiological associations between cancer and neurodegenerative diseases. A growing body of evidence suggests that neurodegenerative diseases may occur less frequently in cancer survivors, and vice versa.[18 22–24] For example, some studies have found that cancer survivors have a decreased risk of Alzheimer's disease and that people with Alzheimer's disease have lower rates of cancer incidence.[4–7] Other studies have suggested an inverse relation between Parkinson's disease and most cancers.[8–10] A link between cancer and neurodegeneration (the

so-called, 'inverse comorbidity') is plausible as they share several genes and biological pathways.[22–26] Non-biological factors (such as behaviours, diagnostic patterns or medications) might account also for some of these possible connections. It is also probable that spurious associations and inaccurate estimates might arise due to chance, bias and/or confounding in epidemiological studies available in the medical literature.[27–29]

Health services generate large amounts of routine health data (so-called 'real world data', such as administrative databases, disease registries and electronic health records), which have important secondary uses for research and evaluation. Increases in the availability of routine health data, and the ability to store, process, link, access and analyse large amounts of data represent a major opportunity for conducting studies on the possible relationships between complex (serious) diseases and other health events with abundant collected data.[30–34] Using such high-scale data sources often involves challenges for research design, conduct and reporting of studies;[35 36] for example, the description of databases' characteristics, record linkage methodology and any validation of the codes or algorithms used to select the study population.

Unfortunately, poorly conducted or reported studies may be associated with increased potential for biases measures of association, limiting their usefulness. Several methodological research studies[37–40] have previously underscored that the reporting of epidemiological studies is inconsistent. For example, Hemkens *et al*[40] investigated the quality of reporting in studies conducted using routinely collected health data on any clinical or epidemiological topic. A search of PubMed in June 2013 served to include a random sample of 124 articles published in 2012. The majority of studies (73.4%) focused on epidemiology. The reporting quality varied, with only 60.5% reporting the characteristics of data sources, 74.2% providing details of selection criteria of participants, 31.5% using the study design in the title or abstract (eg, 'cohort', 'case–control', 'routinely collected data' or 'registry data'), 29.3% reporting methods of data linkage and 2.4% indicating data availability/sharing.

To the best of our knowledge, no systematic reviews of all relevant studies have specifically examined the methodological or reporting of research evaluating the epidemiological associations between complex diseases using routinely collected data. We present herein the protocol for a case meta-research (also known as 'research of research' or 'meta-science')[41] study of studies conducted using observational routinely collected data, that can help better understanding research concerning the cancer and neurodegeneration 'inverse comorbidity' model.[22–24]

The objective of this study will be to evaluate the design, methods and reporting of epidemiological studies conducted using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases.

METHODS AND ANALYSIS

This is the protocol for a meta-research study. Our study protocol is part of a knowledge synthesis research programme on the epidemiological evidence for the associations between cancer and central nervous system disorders, which includes an ambitious ongoing umbrella review (a systematic collection and assessment of multiple systematic reviews and meta-analyses).[19]

This study protocol has been registered within the Open Science Framework (https://osf.io/h2qjg). Although the protocol is for a meta-research study, and not a systematic review of health interventions, our protocol is reported in accordance with the reporting guidance from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement[42 43] with not applicable indicated for items not pertaining to meta-research studies (see online supplemental appendix 1). Methods and definitions have been chosen in consultation with methodological work,[40 44–52] including guidance on preparing Cochrane Methodology Reviews.[52]

Eligibility criteria

Detailed eligibility criteria have been developed according to the following: participants, study design, types of data, types of exposures and outcomes of interest, setting and language of publication.

Participants

We will include studies examining the human population (regardless of age and sex).

Study design

Eligible studies will include observational epidemiological studies including prospective cohort, retrospective cohort (also known as historical cohort studies) and case–control studies. We will include case–control studies regardless of whether the authors reported clear time frame of when the events occurred. However, when extracting data from these studies we will record whether (or not) the time frame was clear. Randomised controlled trials will be unavailable for our research question. Cross-sectional studies will be excluded because they cannot be used to infer causality due to the temporal link between cancer and neurodegenerative diseases cannot be established. We will also exclude reviews, meta-analyses, case series, case reports, in vitro studies and animal studies.

Types of data

Eligible studies can use any type of routinely collected health data. Routinely collected health data are defined as data collected for purposes other than research or without specific a priori research questions developed before collection.[51 53] These would include a range of resources for research (eg, patient registries, disease registries), health planning (eg, administrative data), clinical management (eg, primary care databases, pharmacy data), documentation of clinical care (eg, electronic health records repositories) or epidemiological

surveillance (eg, cancer registries, and other public health reporting data).[51 53]

Types of exposures and outcomes

Eligible studies must investigate the associations between cancer and neurodegenerative diseases. Neurodegenerative diseases[54] will include: Alzheimer's disease (International Classification of Diseases (ICD)–9: 331.0, 290.1; ICD-10: F00, G30), amyotrophic lateral sclerosis/motor neuron disease (ICD-9: 335.20; ICD-10: G12.2), Huntington's disease (ICD-9: 294.1, 333.4; ICD-10: F02.2, G10), multiple sclerosis (ICD-9: 340–340.9; ICD-10: G35-G35.9) and Parkinson's disease (ICD-9: 332–332.9; ICD-10: G20-G21.0, G21.2-G22.0). All malignant neoplasms (ICD-9: 140–209; ICD-10: C00-C97) and any site-specific cancer will be considered. We will include: (1) studies in which neurodegenerative disease was the exposure of interest and cancer incidence (eg, new case or hospitalisation) was the outcome and (2) studies in which cancer was the exposure of interest and incidence of a neurodegenerative disease (eg, new case or hospitalisation) was the outcome. Prognostic studies studying neurodegenerative diseases and mortality among patients with cancer or cancer and mortality among patients with neurodegenerative diseases will be excluded. Studies not presenting quantitative data on the associations between cancer and neurodegenerative diseases (eg, relative risks (RR) with 95% CIs, numbers of cases/population, observed and expected cases) or sufficient data for an association to be calculated will be excluded.

Setting

There will be no restriction by study setting.

Language of publication

Publications of studies will be limited to peer-reviewed journal articles written in English with an abstract available.

Information sources and search strategy

To provide a reliable summary of the literature, we will search the following electronic databases (from their inception onwards): MEDLINE through PubMed (National Library of Medicine, Bethesda, Maryland, USA), Embase though Elsevier platform (Elsevier B.V., Amsterdam, The Netherlands) and the Web of Science Core Collection (Clarivate Analytics, Philadelphia, Pennsylvania, USA). The initial literature searches in MEDLINE, Embase and the Web of Science will start on 15 November 2022.

Our main literature search will be peer-reviewed by two senior health information specialists using the Peer Review of Electronic Search Strategies (PRESS) checklist.[55] The search strategy will include a broad range of terms and keywords related to 'neurodegenerative diseases', 'cancer', 'epidemiological studies' and 'routine data/electronic health records/administrative data'. The search will integrate a filter for electronic health records provided by the National Library of Medicine.[56] A draft search strategy for MEDLINE is provided online in the online supplemental appendix 2.

To ensure literature saturation, the reference lists of studies selected for inclusion will be scanned for additional studies. We will also scan the reference lists of related systematic reviews and meta-analysis identified through the search. In addition, citation searches (eg, Science Citation Index Expanded via the Web of Science) will be carried out for studies selected for inclusion.

Screening

All articles identified from the literature searches will be screened by two researchers independently using the software Rayyan (Rayyan Systems, Cambridge, Massachusetts, USA).[57] First, titles and abstracts of articles returned from initial searches will be screened based on the eligibility criteria outlined earlier. Second, full texts will be examined in detail and screened for eligibility. A form for screening full-text articles will be designed in Microsoft Excel (Microsoft, Seattle, Washington, USA) and pilot tested on a random sample of 10 articles. Third, references of all considered articles will be hand-searched to identify any relevant report missed in the search strategy. Any discrepancies here and throughout will be resolved through consultation to a third researcher, if necessary. A flow chart showing details of studies included and excluded at each stage of the selection process will be provided.

Sample size

We will not perform any sample size calculations since our meta-research study will include all the available studies that would meet the eligibility criteria.

Data extraction

Data for each of the included studies will be abstracted by two researchers, independently, and potential conflicts will be resolved through discussion. We will use predesigned forms that will be piloted initially on a small number of included articles. The data extracted from each article will be comprehensive in scope as we are addressing multiple characteristics of included studies. Full articles and supplementary materials with data and analyses will be examined for general and methodological characteristics, statements of publicly available full protocols and data sets, conflicts of interest and funding disclosures. We will review the final versions of the articles available online. All data will be extracted into Microsoft Excel spreadsheets.

The standardised data extraction form will include the following information of interest:

General characteristics, including study objective(s) and rationale

- First author.
- Year of publication.
- Name of journal, and journal impact factor (eg, according to the latest Journal Citation Report at the time of data extraction).

- Study design (eg, cohort or case–control).
- Country.
- Setting (eg, single-country or multi-country).
- Time frame within which the study took place.
- Study objective(s).
- Main rationale for using routinely collected data (eg, increase study power, validation of findings in a second data source, other, not clearly stated).
- Number of participants.
- Characteristics of participants (eg, proportion of women, mean or median age).
- Selection criteria of participants.
- Details on exposures and outcomes (eg, new cases or hospitalisations).

Methodological characteristics

- Characteristics of the analysed data sets.
- Type of data (eg, administrative data, electronic health records/electronic medical records, registry, other).
- Number of data sources (eg, single data source, two data sources, three data sources, four data sources, five or more data sources).
- Details on methods of study population selection (eg, codes or algorithms used to identify subjects/participants).
- Details on any validation (eg, of the codes or algorithms) used to select the study population (if applicable).
- Type of data linkage across databases (eg, person-level, institutional-level, other, none).
- Methods of record linkage of databases (eg, deterministic, probabilistic, machine learning, other, none).
- Methods of linkage quality evaluation (if applicable).
- Use of any flow diagram or other graphical display to demonstrate the data linkage process (if applicable).
- Variables used for analyses listed and described in sufficient detail.
- A complete list of codes and algorithms used to classify exposures, and outcomes.
- A complete list of codes and algorithms used to classify potential confounders (eg, treatments administered, including chemotherapy).
- Details on the data cleaning methods used in the study.
- Statistical methods (eg, linear regression, logistic regression, Poisson regression, Cox regression, other).
- Confounder control techniques (eg, crude/unadjusted analysis, multivariable analysis, propensity scores, matching, instrumental variables, other).

Main results and limitations

- Unadjusted RR estimates, and if applicable, confounder-maximally adjusted RR estimates with the precision (eg, 95% CIs) from the included studies.
- Discussion of the implications of using data that were not created or collected to answer the specific research question.
- Discussion of potential biases (eg, misclassification, unmeasured confounding, missing data or changing eligibility over time).

Transparency and openness

- Citation of a reporting guideline, such as the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement[51] (no citation, citation without reporting checklist, citation with reporting checklist).
- Open access article or availability of free access in PubMed Central (PMC) based on assignment of a specific ID (PMCID) (yes, no).
- Protocol/registration mentioned (no protocol, indicated that protocol was available on request, full protocol publicly available, full protocol publicly available and preregistered).
- Mention of raw data availability (no data sharing, indicated that raw data were available on request, full access to raw data for reanalysis).
- Mention of access to programming code used to perform analyses (no access, indicated that code was available on request, full access for reanalysis).
- Type of data repository used, if appropriate (eg, Open Science Framework, Mendeley, Zenodo, Dryad, journal repository or other).
- Funding (no statement, no funding, public, private, other, combination of public/private/other).
- Conflicts of interests (no statement, statement no conflicts exist, statement conflicts exist).

Adherence to reporting standards

We will assess reporting quality and completeness of included studies against the RECORD statement (https://www.record-statement.org/).[51] RECORD represents the current best practice reporting standard for studies using observational routinely collected health data. The RECORD statement consists of a checklist of items (see online supplemental appendix 3) that supplement or modify the STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement (https://www.strobe-statement.org/),[58 59] which focused on the reporting of observational studies. The aim will be to assess whether included studies conformed to reporting recommendations included in the RECORD statement.

We will operationalise all items of the checklist into dedicated questions that can be answered with 'yes', 'no' or 'partly', indicating adequate ('yes') or inadequate ('no') reporting. We will use the 'partly' answer when not all aspects are adequately reported, for example, when several eligibility criteria existed, but some of them are described, and others are not. This approach is consistent with previous methodological studies.[40] In addition, we will accept references to other publications as adequate descriptions.

Methodological quality (or risk of bias) assessment

The methodological quality (or risk of bias) of included studies will be evaluated using the Newcastle-Ottawa

Scale (NOS) for observational studies.[60] Using the NOS tool, each study is judged on eight items, categorised into three groups: the selection of the study groups (eg, representativeness), the comparability of the groups (eg, matching in the design and/or confounders adjusted for in the analysis) and the ascertainment of either the exposure or outcome of interest for case–control or cohort studies, respectively. Stars are awarded for each item, and the highest methodological quality (or low risk of bias) studies are awarded up to nine stars. We will consider studies with 0–3, 4–6 and 7–9 stars to represent low, moderate and high quality, respectively. The methodological quality (or risk of bias) for each study will be independently assessed by two investigators. Discrepant scores will be resolved by discussion.

Summarising the evidence

We will summarise design, methods and reporting characteristics of the included studies with tables and graphical tools (eg, bar charts, forest plots). This will be done by constructing a clear descriptive summary on the included studies based on a common analytical framework on the study populations, study design, details of exposures and outcomes, key information about the methods, data sources, estimation procedures or accessibility of materials and raw data.

Data will be summarised as frequency for categorical items or median and IQR for continuous items. We will not perform a meta-analysis of pooled estimates since it is out of the scope of the planned meta-research study. Heterogeneity of included studies will be discussed narratively.

Additional analyses/subgroups

If sufficient studies report results separately, we plan to summarise design, methods and reporting characteristics of the included studies by types of exposures (eg, Alzheimer's disease, amyotrophic lateral sclerosis/motor neuron disease, Huntington's disease, multiple sclerosis and Parkinson's disease), outcomes (eg, total cancer vs site-specific cancer), sex (male only vs female only) and study design (cohort vs case–control).

Software considerations

All analyses will be performed using Stata V.17 or higher (StataCorp LP, College Station, Texas, USA).

Patient and public involvement

The draft protocol was revised on receiving feedback from all the research team (including methodologists, scientists and healthcare professionals). Patients or the public were not involved in the setting of the research question, nor in developing plans for design/writing of our protocol. Patients or the public will not be asked to advice on the interpretation or writing up of findings.

Ethics and dissemination

Due to the nature of this study, no ethics approval is required as no human subjects will be involved. We plan to publish the full meta-research study in an open access peer-reviewed journal and disseminate the findings at scientific conferences and via social media (Twitter, and author affiliated websites).

DISCUSSION

Using routinely collected data for research may represent a powerful approach to evaluate the epidemiological associations between complex diseases. However, such applications come with novel challenges and may create novel problems. Some biases are inherent to the observational designs but potential issues such as misclassification, unmeasured confounding or missing data are of particular importance when using routinely collected data. To the best of our knowledge, the planned meta-research study will be the first attempt to investigate the methodology and reporting of all studies conducted using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases. Guidance from the Cochrane Methodology Review Group,[52] and from the Synthesis Without Meta-analysis reporting guideline[61] will be followed during all the research process. The proposed meta-research study will be reported in accordance with the reporting guidance provided in the PRISMA 2020 statement (http://www.prisma-statement.org/).[62 63] Any amendments made to our protocol when conducting the analyses will be outlined and reported in the final manuscript. All data underlying the findings reported in the final manuscript will be deposited in a cross-disciplinary public repository, such as the Open Science Framework (https://osf.io/).

There are several strengths and limitations of our planned methods. We will comprehensively evaluate the methodological and reporting quality of studies conducted using observational routinely collected health data for investigating the link between cancer and neurodegenerative diseases. We anticipate that we will identify knowledge gaps to be filled by new research considering that some methodological and reporting characteristics in studies using routinely collected health data will be poorly covered in the medical literature. A key challenge is that based on knowledge from previous studies,[39 40] we anticipate identifying studies using different study designs, populations, outcomes and analyses with a variable quality of reporting.

Finally, we anticipate the study could be relevant to a variety of audiences (eg, research authors, health professionals, funders, journal editors). Moreover, the proposed meta-research study might offer insight into future research agendas for new studies conducted using routinely collected health data for investigating the epidemiological associations between cancer, neurodegeneration or other medical conditions and risk factors. In our opinion, a better understanding of the links between complex diseases might lead to new or improved forms of prevention and treatment.

Author affiliations
[1]Department of Health Planning and Economics, National School of Public Health, Institute of Health Carlos III, Madrid, Spain
[2]Department of Medicine, University of Valencia/INCLIVA Health Research Institute and Centro de Investigación en Red de Salud Mental (CIBERSAM), Valencia, Spain
[3]Knowledge Synthesis Group, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
[4]Geriatric Research Education and Clinical Center, Veterans Affairs Boston Healthcare System, Boston, Massachusetts, USA
[5]Division of Aging, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[6]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
[7]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia
[8]School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[9]Instituto Aragonés de Ciencias de la Salud, Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Zaragoza, Spain
[10]Miguel Servet University Hospital, Zaragoza, Spain
[11]Department of History of Science and Documentation, University of Valencia, Valencia, Spain
[12]Unidad de Información e Investigación Social y Sanitaria, University of Valencia, Spanish National Research Council, Valencia, Spain
[13]National Library of Health Sciences, Institute of Health Carlos III, Madrid, Spain
[14]Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain

Twitter Matthew J Page @mjpages, Brian Hutton @bh_epistat, Cristina A Fraga-Medín @crisfragamedin, Alfonso Valencia @alfons_valencia and Rafael Tabarés-Seisdedos @chromosome8

Contributors All authors contributed to conceptualising and designing the study. FC-L drafted the manuscript. JAD, MJP, BH, MR, CB-V, AA-A, CAF-M, EB-D, AV and RT-S commented for important intellectual content and made revisions. All authors read and approved the final version of the manuscript. FC-L accepts full responsibility for the finished manuscript and controlled the decision to publish.

Funding FC-L and RT-S are supported by the Institute of Health Carlos III/CIBERSAM. MJP is supported by an Australian Research Council Discovery Early Career Researcher Award (DE200101618). MR and EB-D are partially funded by the Spanish Health Services Research on Chronic Patients Network (REDISSEC)/Institute of Health Carlos III.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Ferrán Catalá-López http://orcid.org/0000-0002-3833-9312
Brian Hutton http://orcid.org/0000-0001-5662-8647
Manuel Ridao http://orcid.org/0000-0001-7837-5759
Adolfo Alonso-Arroyo http://orcid.org/0000-0002-5084-2818
Cristina A Fraga-Medín http://orcid.org/0000-0002-4342-3033
Enrique Bernal-Delgado http://orcid.org/0000-0002-0961-3298
Alfonso Valencia http://orcid.org/0000-0002-8937-6789
Rafael Tabarés-Seisdedos http://orcid.org/0000-0002-1089-2204

REFERENCES

1 Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021;71:209–49.
2 GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.
3 GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the global burden of disease study 2016. *Lancet Neurol* 2019;18:459–80.
4 Roe CM, Behrens MI, Xiong C, *et al*. Alzheimer disease and cancer. *Neurology* 2005;64:895–8.
5 Roe CM, Fitzpatrick AL, Xiong C, *et al*. Cancer linked to Alzheimer disease but not vascular dementia. *Neurology* 2010;74:106–12.
6 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.
7 Musicco M, Adorni F, Di Santo S, *et al*. Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. *Neurology* 2013;81:322–8.
8 Driver JA, Kurth T, Buring JE, *et al*. Prospective case-control study of nonfatal cancer preceding the diagnosis of Parkinson's disease. *Cancer Causes Control* 2007;18:705–11.
9 Fois AF, Wotton CJ, Yeates D, *et al*. Cancer in patients with motor neuron disease, multiple sclerosis and Parkinson's disease: record linkage studies. *J Neurol Neurosurg Psychiatry* 2010;81:215–21.
10 Lin P-Y, Chang S-N, Hsiao T-H, *et al*. Association between Parkinson disease and risk of cancer in Taiwan. *JAMA Oncol* 2015;1:633–40.
11 Midgard R, Glattre E, Grønning M, *et al*. Multiple sclerosis and cancer in Norway. A retrospective cohort study. *Acta Neurol Scand* 1996;93:411–5.
12 Kingwell E, Bajdik C, Phillips N, *et al*. Cancer risk in multiple sclerosis: findings from British Columbia, Canada. *Brain* 2012;135:2973–9.
13 Turner MR, Goldacre R, Goldacre MJ. Reduced cancer incidence in Huntington's disease: record linkage study clue to an evolutionary trade-off? *Clin Genet* 2013;83:588–90.
14 Gibson SB, Abbott D, Farnham JM, *et al*. Population-Based risks for cancer in patients with ALS. *Neurology* 2016;87:289–94.
15 Handel AE, Ramagopalan SV. Multiple sclerosis and risk of cancer: a meta-analysis. *J Neurol Neurosurg Psychiatry* 2010;81:1413–4.
16 Bajaj A, Driver JA, Schernhammer ES. Parkinson's disease and cancer risk: a systematic review and meta-analysis. *Cancer Causes Control* 2010;21:697–707.
17 Ma L-L, Yu J-T, Wang H-F, *et al*. Association between cancer and Alzheimer's disease: systematic review and meta-analysis. *J Alzheimers Dis* 2014;42:565–73.
18 Catalá-López F, Suárez-Pinilla M, Suárez-Pinilla P, *et al*. Inverse and direct cancer comorbidity in people with central nervous system disorders: a meta-analysis of cancer incidence in 577,013 participants of 50 observational studies. *Psychother Psychosom* 2014;83:89–105.
19 Catalá-López F, Hutton B, Driver JA, *et al*. Cancer and central nervous system disorders: protocol for an umbrella review of systematic reviews and updated meta-analyses of observational studies. *Syst Rev* 2017;6:69.
20 Ospina-Romero M, Glymour MM, Hayes-Larson E, *et al*. Association between Alzheimer disease and cancer with evaluation of study biases: a systematic review and meta-analysis. *JAMA Netw Open* 2020;3:e2025515.
21 Zhang X, Guarin D, Mohammadzadehhonarvar N, *et al*. Parkinson's disease and cancer: a systematic review and meta-analysis of over 17 million participants. *BMJ Open* 2021;11:e046329.
22 Driver JA. Inverse association between cancer and neurodegenerative disease: review of the epidemiologic and biological evidence. *Biogerontology* 2014;15:547–57.
23 Tabarés-Seisdedos R, Rubenstein JL. Inverse cancer comorbidity: a serendipitous opportunity to gain insight into CNS disorders. *Nat Rev Neurosci* 2013;14:293–304.
24 Tabarés-Seisdedos R, Dumont N, Baudot A, *et al*. No paradox, no progress: inverse cancer comorbidity in people with other complex diseases. *Lancet Oncol* 2011;12:604–8.

25 Devine MJ, Plun-Favreau H, Wood NW. Parkinson's disease and cancer: two wars, one front. *Nat Rev Cancer* 2011;11:813–23.
26 Lanni C, Masi M, Racchi M, *et al*. Cancer and Alzheimer's disease inverse relationship: an age-associated diverging derailment of shared pathways. *Mol Psychiatry* 2021;26:280–95.
27 Janiaud P, Agarwal A, Tzoulaki I, *et al*. Validity of observational evidence on putative risk and protective factors: appraisal of 3744 meta-analyses on 57 topics. *BMC Med* 2021;19:157.
28 Ioannidis JPA, Zhou Y, Chang CQ, *et al*. Potential increased risk of cancer from commonly used medications: an umbrella review of meta-analyses. *Ann Oncol* 2014;25:16–23.
29 Kyzas PA, Denaxa-Kyza D, Ioannidis JPA. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007;43:2559–79.
30 Biggin F, Emsley HCA, Knight J. Routinely collected patient data in neurology research: a systematic mapping review. *BMC Neurol* 2020;20:431.
31 Wilkinson T, Ly A, Schnier C, *et al*. Identifying dementia cases with routinely collected health data: A systematic review. *Alzheimers Dement* 2018;14:1038–51.
32 Horrocks S, Wilkinson T, Schnier C, *et al*. Accuracy of routinely-collected healthcare data for identifying motor neurone disease cases: a systematic review. *PLoS One* 2017;12:e0172639.
33 Krumm R, Semjonow A, Tio J, *et al*. The need for harmonized structured documentation and chances of secondary use - results of a systematic analysis with automated form comparison for prostate and breast cancer. *J Biomed Inform* 2014;51:86–99.
34 Moccia M, Annovazzi P, Buscarinu MC, *et al*. Harmonization of real-world studies in multiple sclerosis: retrospective analysis from the rirems group. *Mult Scler Relat Disord* 2020;45:102394.
35 Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ* 2010;341:c4226.
36 Benchimol EI, Manuel DG, To T, *et al*. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.
37 Groenwold RHH, Van Deursen AMM, Hoes AW, *et al*. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol* 2008;18:746–51.
38 Papathanasiou AA, Zintzaras E. Assessing the quality of reporting of observational studies in cancer. *Ann Epidemiol* 2010;20:67–73.
39 Pouwels KB, Widyakusuma NN, Groenwold RHH, *et al*. Quality of reporting of confounding remained suboptimal after the STROBE guideline. *J Clin Epidemiol* 2016;69:217–24.
40 Hemkens LG, Benchimol EI, Langan SM, *et al*. The reporting of studies using routinely collected health data was often insufficient. *J Clin Epidemiol* 2016;79:104–11.
41 Ioannidis JPA, Fanelli D, Dunne DD, *et al*. Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol* 2015;13:e1002264.
42 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
43 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;350:g7647.
44 Mbuagbaw L, Lawson DO, Puljak L, *et al*. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol* 2020;20:226.
45 Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med* 2017;22:139–42.
46 Ioannidis JPA, Greenland S, Hlatky MA, *et al*. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166–75.
47 Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
48 Nosek BA, Alter G, Banks GC, *et al*. Scientific standards. promoting an open research culture. *Science* 2015;348:1422–5.
49 MIT Critical Data. *Secondary analysis of electronic health records*. Cham, CH: Springer Open, 2016.
50 Goldstein N, ed. *Improving population health using electronic health records : methods for data management and epidemiological analysis*. New York, NY: Routledge, 2017.
51 Benchimol EI, Smeeth L, Guttmann A, *et al*. The reporting of studies conducted using observational Routinely-collected health data (record) statement. *PLoS Med* 2015;12:e1001885.
52 Clarke M. Guide to the contents of a Cochrane methodology protocol and review. Cochrane: cochrane methodology review group, 2020. Available: https://methodology.cochrane.org/resources
53 Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. *CMAJ* 2017;189:E1054–5.
54 Williams A. Defining neurodegenerative diseases. *BMJ* 2002;324:1465–6.
55 McGowan J, Sampson M, Salzwedel DM, *et al*. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol* 2016;75:40–6.
56 National Library of Medicine. MEDLINE / PubMed Search Strategy & Electronic Health Record Information Resources. Available: https://www.nlm.nih.gov/services/queries/ehr_details.html
57 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile APP for systematic reviews. *Syst Rev* 2016;5:210.
58 von Elm E, Altman DG, Egger M, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ* 2007;335:806–8.
59 Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297.
60 Wells GA, Shea B, O'Connell D, *et al*. The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses, 2008. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
61 Campbell M, McKenzie JE, Sowden A, *et al*. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ* 2020;368:l6890.
62 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
63 Page MJ, Moher D, Bossuyt PM, *et al*. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.

Supplementary Appendix 1. PRISMA-P Checklist.

Section and topic	Item No	Checklist item	Reported on page
ADMINISTRATIVE INFORMATION			
Title:			
Identification	1a	Identify the report as a protocol of a systematic review	p.1, lines 1-4
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	NA
Registration	2	If registered, provide the name of the registry (such as PROSPERO) and registration number	p.
Authors:			
Contact	3a	Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author	p.1, lines 5-34
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	p.13, lines 435-439
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	p.11 and p.12, lines 397-400
Support:			
Sources	5a	Indicate sources of financial or other support for the review	
Sponsor	5b	Provide name for the review funder and/or sponsor	p.13, lines 441-449
Role of sponsor or funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	
INTRODUCTION			
Rationale	6	Describe the rationale for the review in the context of what is already known	p.4, lines 84-125
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	p.5, lines 128-130
METHODS			
Eligibility criteria	8	Specify the study characteristics (such as PICO, study design, setting, time frame) and report characteristics (such as years considered, language, publication status) to be used as criteria for eligibility for the review	p.5 and p.6, lines 147-191
Information sources	9	Describe all intended information sources (such as electronic databases, contact with study authors, trial registers or other grey literature sources) with planned dates of coverage	p.6, lines 193-212
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	p.6, lines 203-207

			Additional file 2
Study records:			
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	p.7, lines 216-221
Selection process	11b	State the process that will be used for selecting studies (such as two independent reviewers) through each phase of the review (that is, screening, eligibility and inclusion in meta-analysis)	p.7, lines 215-226
Data collection process	11c	Describe planned method of extracting data from reports (such as piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	p.7, lines 232-241
Data items	12	List and define all variables for which data will be sought (such as PICO items, funding sources), any pre-planned data assumptions and simplifications	p.8 and p.9, lines 244-313
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	p.8 and p.9, lines 316-331 (RECORD)
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	p.9, lines 333-344 (NOS tool)
Data synthesis	15a	Describe criteria under which study data will be quantitatively synthesised	p.10, lines 346-356
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data and methods of combining data from studies, including any planned exploration of consistency (such as I^2, Kendall's τ)	NA
	15c	Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)	p.10, lines 359-364
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	p.10, lines 346-356
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)	NA
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (such as GRADE)	NA

From: Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;350:g7647. doi: 10.1136/bmj.g7647. Erratum in: BMJ. 2016;354:i4086. PMID: 25555855.

Supplementary Appendix 2. Draft search for MEDLINE®.

1. "Alzheimer disease"[mh] OR "alzheimer"[tiab] OR "alzheimer s"[tiab] OR "amyotrophic lateral sclerosis"[mh] OR "amyotrophic lateral sclerosis"[tiab] OR "motor neuron disease"[mh] OR "motor neuron disease*"[tiab] OR "Huntington disease"[mh] OR "huntington disease*"[tiab] OR "huntington s disease*"[tiab] OR "multiple sclerosis"[mh] OR "multiple sclerosis"[tiab] OR "Parkinson disease"[mh] OR "parkinson"[tiab] OR "parkinson s"[tiab]
2. "neoplasms"[mh] OR "neoplasm*"[tiab] OR "neoplasia*"[tiab] OR "cancer*"[tiab] OR "carcinoma"[mh] OR "carcinoma*"[tiab] OR "adenocarcinoma"[mh] OR "adenocarcinoma*"[tiab] OR "tumor*"[tiab] OR "tumour*"[tiab] OR "maligna*"[tiab]
3. "epidemiolog*"[tiab] OR "cohort studies"[mh] OR "cohort stud*"[tiab] OR "cohort s"[tiab] OR "cohorte"[tiab] OR "longitudinal stud*"[tiab] OR "case control studies"[mh] OR "case control*"[tiab]
4. "Administrativ*"[tiab] OR "Claims"[tiab] OR "routine data*"[tiab] OR "routinely collected"[tiab] OR "Registries"[mh] OR register*[tiab] OR registr*[tiab] OR "Databases as Topic"[mh] OR "database*" [tiab] OR "healthcare data*"[tiab] OR "health care data*"[tiab]
5. (("health information exchange"[tw] OR "hie"[tw] OR "rhio"[tw] OR "regional health information organization"[tw] OR "hl7"[tw] OR "health level seven"[tw] OR "unified medical language system"[majr] OR "umls"[tw] OR "loinc"[tw] OR "rxnor"[tw] OR "snomed"[tw] OR "icd9 cm"[ti] OR "icd 9 cm"[ti] OR "icd10"[ti] OR "icd 10"[ti] OR "metathesaurus"[tw] OR "patient card"[tw] OR "patient cards"[tw] OR "health card"[tw] OR "health cards"[tw] OR "electronic health data"[tw] OR "personal health data"[tw] OR "personal health record"[tw] OR "personal health records"[tw] OR "Health Records, Personal"[majr] OR "Health Record, Personal"[majr] OR "ehealth"[tw] OR "e-health"[tw] OR "medical informatics application"[mh] OR "medical informatics applications"[mh] OR "medical records system, computerized"[mh] OR "computerized patient medical records"[tw] OR "automated medical record system"[tw] OR "automated medical record systems"[tw] OR "automated medical records system"[tw] OR "automated medical records systems"[tw] OR "computerized medical record"[tw] OR "computerized medical records"[tw] OR "computerized patient records"[tw] OR "computerized patient record"[tw] OR "computerized patient medical record"[tw] OR "electronic health record"[tw] OR "electronic health records"[tw] OR "Electronic Health Record"[majr] OR "Electronic Health Records"[majr] OR "electronic patient record"[tw] OR "electronic patient records"[tw] OR "electronic medical record"[tw] OR "electronic medical records"[tw] OR "electronic healthcare records"[tw] OR "electronic healthcare record"[tw] OR "electronic health care record"[tw] OR "electronic health care records"[tw] OR "archives"[majr] OR "ehr"[tw] OR "ehrs"[tw] OR "phr"[tw] OR "phrs"[tw] OR "emr"[tw] OR "emrs"[tw] OR "Health Information Systems"[majr] OR "health information interoperability"[mh] OR "health information interoperability"[tw]) AND ("medical record"[ti] OR "medical records"[mh] OR "medical records"[ti] OR "patient record"[ti] OR "patient records"[ti] OR "patient health record"[ti] OR "patient health records"[ti] OR "patient identification system"[mh] OR "patient identification systems"[ti] OR "Patient Outcome Assessment"[majr] OR "Patient Discharge Summaries"[majr] OR "healthcare record"[ti] OR "healthcare records"[ti] OR "health care record"[ti] OR "health care records"[ti] OR "health record"[ti] OR "health records"[ti] OR "hospital information system"[tw] OR "hospital information systems"[tw] OR "umae"[ti] OR "attitude to computers"[mh] OR "medical informatics"[ti] OR "Information Technology"[mh] OR

"Information Technology"[tw])) OR (("medical records systems, computerized"[majr] OR "medical records systems, computerized"[mh] OR "computerized patient medical record"[tw] OR "computerized patient medical records"[tw] OR "automated medical record system"[tw] OR "automated medical record systems"[tw] OR "automated medical records system"[tw] OR "automated medical records systems" [tw] OR "computerized medical record"[tw] OR "computerized medical records"[tw] OR "computerized patient records"[tw] OR "computerized patient record"[tw] OR "patient generated health data"[mh] OR "patient generated health data"[tw] OR "electronic health record"[tw] OR "electronic health records"[tw] OR "electronic patient record"[tw] OR "electronic patient records"[tw] OR "electronic medical record"[tw] OR "electronic medical records"[tw] OR "electronic healthcare records"[tw] OR "electronic healthcare record"[tw] OR "electronic health care record"[tw] OR "electronic health care records"[tw] OR "unified medical language system"[majr] OR "unified medical language system"[tw] OR "umls"[tw] OR "loinc"[tw] OR "rxnorm"[tw] OR "snomed"[tw] OR "icd9 cm"[ti] OR "icd 9 cm"[ti] OR "icd10"[ti] OR "icd 10"[ti] OR "metathesaurus"[tw] OR "ehr"[tw] OR "ehrs"[tw] OR "phr"[tw] OR "phrs"[tw] OR "emr"[tw] OR "emrs"[tw] OR "meaningful use"[tiab] OR "meaningful use"[tw] OR "Meaningful Use"[majr]) AND ("j ahima"[ta] OR "j am med inform assoc"[ta] OR "amia annu symp proc"[ta] OR "health data manag"[ta] OR "int j med inform"[ta] OR "yearb med inform"[ta] OR "telemed j e health"[ta] OR "stud health technol inform" [ta]))

6. #1 AND #2 AND #3
7. #4 OR #5
8. #6 AND #7

Note: Filters: Abstract, English.

Supplementary Appendix 3. RECORD Checklist.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			
Objectives	3	State specific objectives, including any prespecified hypotheses			
Methods					
Study Design	4	Present key elements of study design early in the paper			
Setting	5	Describe the setting, locations, and relevant dates, including periods of			

		recruitment, exposure, follow-up, and data collection			
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement).			

		Describe comparability of assessment methods if there is more than one group			
Bias	9	Describe any efforts to address potential sources of bias			
Study size	10	Explain how the study size was arrived at			
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			

Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			
Discussion					
Key results	18	Summarise key results with reference to study objectives			
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured	

				confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	

From: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12(10):e1001885. doi: 10.1371/journal.pmed.1001885. PMID: 26440803; PMCID: PMC4595218.