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


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. Multiple epidemiological studies and reviews 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. 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. Other studies have suggested an inverse relation between Parkinson’s disease and most cancers. A link between cancer and neurodegeneration (the
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 22 41 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

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.25–29

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 studies37–40 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.
surveillance (eg, cancer registries, and other public health reporting data).53,54

Types of exposures and outcomes
Eligible studies must investigate the associations between cancer and neurodegenerative diseases. Neurodegenerative diseases 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 predefined 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).
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
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/). 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/), 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. 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. 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, and from the Synthesis Without Meta-analysis reporting guideline 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/).

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, 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.
REFERENCES


...the importance of high-quality diagnostic coding to research. CMAJ 2017;189:E1054–5.


# Supplementary Appendix 1. PRISMA-P Checklist.

<table>
<thead>
<tr>
<th>Section and topic</th>
<th>Item</th>
<th>Checklist item</th>
<th>Reported on page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMINISTRATIVE INFORMATION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title: Identification</td>
<td>1a</td>
<td>Identify the report as a protocol of a systematic review</td>
<td>p.1, lines 1-4</td>
</tr>
<tr>
<td>Update</td>
<td>1b</td>
<td>If the protocol is for an update of a previous systematic review, identify as such</td>
<td>NA</td>
</tr>
<tr>
<td>Registration</td>
<td>2</td>
<td>If registered, provide the name of the registry (such as PROSPERO)</td>
<td>p. and registration number</td>
</tr>
<tr>
<td>Authors: Contact</td>
<td>3a</td>
<td>Provide name, institutional affiliation, e-mail address of all protocol authors; provide physical mailing address of corresponding author</td>
<td>p.1, lines 5-34</td>
</tr>
<tr>
<td>Contributions</td>
<td>3b</td>
<td>Describe contributions of protocol authors and identify the guarantor of the review</td>
<td>p.13, lines 435-439</td>
</tr>
<tr>
<td>Amendments</td>
<td>4</td>
<td>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</td>
<td>p.11 and 12, lines 397-400</td>
</tr>
<tr>
<td>Support: Sources</td>
<td>5a</td>
<td>Indicate sources of financial or other support for the review</td>
<td></td>
</tr>
<tr>
<td>Sponsor</td>
<td>5b</td>
<td>Provide name for the review funder and/or sponsor</td>
<td>p.13, lines 441-449</td>
</tr>
<tr>
<td>Role of sponsor or funder</td>
<td>5c</td>
<td>Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rationale</td>
<td>6</td>
<td>Describe the rationale for the review in the context of what is already known</td>
<td>p.4, lines 84-125</td>
</tr>
<tr>
<td>Objectives</td>
<td>7</td>
<td>Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)</td>
<td>p.5, lines 128-130</td>
</tr>
<tr>
<td>METHODS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligibility criteria</td>
<td>8</td>
<td>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</td>
<td>p.5 and 6, lines 147-191</td>
</tr>
<tr>
<td>Information sources</td>
<td>9</td>
<td>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</td>
<td>p.6, lines 193-212</td>
</tr>
<tr>
<td>Search strategy</td>
<td>10</td>
<td>Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated</td>
<td>p.6, lines 203-207</td>
</tr>
<tr>
<td>Study records:</td>
<td>Additional file 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data management</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11a Describe the mechanism(s) that will be used to manage records and data throughout the review</td>
<td>p.7, lines 216-221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>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)</td>
<td>p.7, lines 215-226</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data collection process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>p.7, lines 232-241</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data items</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>p.8 and p.9, lines 244-313</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Outcomes and prioritization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale</td>
<td>p.8 and p.9, lines 316-331 (RECORD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Risk of bias in individual studies</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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</td>
<td>p.9, lines 333-344 (NOS tool)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data synthesis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15a Describe criteria under which study data will be quantitatively synthesised</td>
<td>p.10, lines 346-356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>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 I², Kendall’s τ)</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15c Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)</td>
<td>p.10, lines 359-364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15d If quantitative synthesis is not appropriate, describe the type of summary planned</td>
<td>p.10, lines 346-356</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meta-bias(es)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 Specify any planned assessment of meta-bias(es) (such as publication bias across studies, selective reporting within studies)</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Confidence in cumulative evidence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 Describe how the strength of the body of evidence will be assessed (such as GRADE)</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 **Supplementary Appendix 2.** Draft search for MEDLINE®.


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

Note: Filters: Abstract, English.
1 **Supplementary Appendix 3. RECORD Checklist.**

<table>
<thead>
<tr>
<th>Item No.</th>
<th>STROBE items</th>
<th>Location in manuscript where items are reported</th>
<th>RECORD items</th>
<th>Location in manuscript where items are reported</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title and abstract</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(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</td>
<td></td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background rationale</td>
<td>2</td>
<td>Explain the scientific background and rationale for the investigation being reported</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>3</td>
<td>State specific objectives, including any prespecified hypotheses</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study Design</td>
<td>4</td>
<td>Present key elements of study design early in the paper</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting</td>
<td>5</td>
<td>Describe the setting, locations, and relevant dates, including periods of</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (a) Cohort study</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------------</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up.</td>
<td>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.</td>
<td>For each variable of interest, give sources of data and details of methods of assessment (measurement).</td>
<td></td>
</tr>
<tr>
<td><strong>Case-control study</strong></td>
<td>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.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cross-sectional study</strong></td>
<td>Give the eligibility criteria, and the sources and methods of selection of participants.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(b) Cohort study</strong></td>
<td>For matched studies, give matching criteria and number of exposed and unexposed cases.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Case-control study</strong></td>
<td>For matched studies, give matching criteria and the number of controls per case.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Variables</strong></td>
<td></td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data sources/measurement</strong></td>
<td></td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>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.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Describe comparability of assessment methods if there is more than one group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>9</td>
<td>Describe any efforts to address potential sources of bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study size</td>
<td>10</td>
<td>Explain how the study size was arrived at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative variables</td>
<td>11</td>
<td>Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 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 |
<p>| 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. |</p>
<table>
<thead>
<tr>
<th>Linkage</th>
<th>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.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td>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.</td>
</tr>
</tbody>
</table>

**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

**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 |
<table>
<thead>
<tr>
<th>Interpretation</th>
<th>20</th>
<th>Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalisability</td>
<td>21</td>
<td>Discuss the generalisability (external validity) of the study results</td>
</tr>
<tr>
<td>Other Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>22</td>
<td>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</td>
</tr>
</tbody>
</table>
| 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.