SYSTEMATIC REVIEW PROTOCOL

WHO/ILo work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on ischaemic heart disease

Jian Li, MD, PhD 1*
Chantal Brisson, PhD 2
Els Clays, PhD 3
Marco M. Ferrario, MD, PhD 4
Ivan D. Ivanov, MD, PhD 5
Paul Landsbergis, PhD, EdD, MPH 6
Nancy Leppink, JD 7
Frank Pega, PhD 8
Hynek Pikhart, PhD 9
Annette Prüss-Ústün, PhD 10
Reiner Rugulies, PhD 11
Peter L. Schnall, MD, MPH 12
Gretchen Stevens, PhD 13
Akizumi Tsutsumi, MD, DMSc 14
Yuka Ujita, MD, PhD 15
Johannes Siegrist, PhD 16*

* These authors have contributed equally to this work.

1 Institute of Occupational, Social and Environmental Medicine, Centre for Health and Society, Faculty of Medicine, University of Düsseldorf, Universitätsstraße 1, Düsseldorf 40225, Germany, jian.li@uni-duesseldorf.de
2 Centre de Recherche du CHU de Québec-Université Laval, 1050 Chemin Ste-Foy, Quebec City, G1S 4L8, Quebec, Canada, Chantal.Brisson@crchudequebec.ulaval.ca
3 Department of Public Health, Ghent University, Campus University Hospital, 4K3, De Pintelaan 185, B-9000 Gent, Belgium, els.clays@UGent.be
4 Research Centre EPIMED, University of Insubria, Via O Rossi 9, 21100 Varese, Italy, marco.ferrario@uninsubria.it
5 Department of Public Health, Environmental and Social Determinants of Health, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, ivanovi@who.int
6 State University of New York-Downstate School of Public Health, 450 Clarkson Ave., Brooklyn, NY 11238, United States of America, paul.landsbergis@downstate.edu
7 Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International Labour Organization, Route des Morillons 4, 1211 Geneva, Switzerland, leppink@ilo.org
8 Department of Public Health, Environmental and Social Determinants of Health, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, pegaf@who.int
9 Institute of Epidemiology and Health Care, University College London, 1-19 Torrington Place, London WC1E 6BT, United Kingdom, h.pikhart@ucl.ac.uk
10 Department of Public Health, Environmental and Social Determinants of Health, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, pruessa@who.int
11 National Research Centre for the Working Environment, Lersø Parkallé 105; DK-2100 Copenhagen, Denmark, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, DK-1014
Copenhagen, Denmark; and Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen, Denmark, rer@nrcwe.dk

12 Center for Occupational and Environmental Health, University of California-Irvine, 100 Theory Way, Irvine, California, United States of America, pschnall@workhealth.org

13 Department of Information, Evidence and Research, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, stevensg@who.int

14 Department of Public Health, School of Medicine, Kitasato University, 1-15-1 Kitasato, Minami, Sagamihara, 252-0374, Japan, akizumi@kitasato-u.ac.jp

15 Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International Labour Organization, Route des Morillons 4, 1211 Geneva, Switzerland, ujita@ilo.org

16 Life Science Centre, University of Düsseldorf, Merowingerplatz 1a, Düsseldorf 40225, Germany, johannes.siegrist@med.uni-duesseldorf.de

Corresponding author: Dr. Jian Li; Universitätsstraße 1, Düsseldorf 40225, Germany, Phone: +49-211-81-08161.

Running head: Parameters for estimating burden of ischaemic heart disease from long working hours

Abstract: 459 Main text: 8,794 References: 80 Figures: 1 Tables: 2

Appendices: 10
Abstract

Background: The World Health Organization (WHO) and the International Labour Organization (ILO) are developing a joint methodology for estimating the national and global work-related burden of disease and injury (WHO/ILO joint methodology), with contributions from a large network of experts. In this paper, we present the protocol for two systematic reviews of parameters for estimating the number of deaths and disability-adjusted life years of ischaemic heart disease from exposure to long working hours, to inform the development of the WHO/ILO joint methodology.

Objectives: We aim to systematically review studies on occupational exposure to long working hours (Systematic Review 1) and systematically review and meta-analyse estimates of the effect of long working hours on ischaemic heart disease (Systematic Review 2), applying the Navigation Guide systematic review methodology as an organizing framework. The selection of both, the exposure and the health outcome is justified by substantial scientific evidence on adverse effects of long working hours on ischaemic heart disease risk.

Data sources: Separately for Systematic Reviews 1 and 2, we will search electronic academic databases for potentially relevant records from published and unpublished studies, Medline, EMBASE, Web of Science, CISDOC and PsychINFO. We will also search electronic grey literature databases, Internet search engines and organizational websites; hand-search reference list of previous systematic reviews and included study records; and consult additional experts.

Study eligibility and criteria: We will include working-age (≥15 years) workers in the formal and informal economy in any WHO and/or ILO Member State, but exclude children (<15 years) and unpaid domestic workers. For Systematic Review 1, we will include quantitative prevalence studies of relevant levels of exposure to long working hours (i.e. 35-40, 41-48, 49-54 and ≥55 hours/week) stratified by country, sex, age and industrial sector or occupation. For Systematic Review 2, we will include randomized controlled trials, cohort studies, case-control studies, and other non-randomized intervention studies with an estimate of the relative effect of relevant level(s) of long working hours on the prevalence or incidence of ischaemic heart disease or of mortality from it, compared with the theoretical minimum risk exposure level (i.e. 35-40 hours/week).

Study appraisal and synthesis methods: At least two review authors will independently screen titles and abstracts against the eligibility criteria at a first stage and full texts of potentially eligible records at a second stage, followed by extraction of data from qualifying studies. At least two review authors will assess risk of bias and the quality of evidence, using the most suited tools currently available. For Systematic Review 2, if feasible, we will combine relative risks using meta-analysis. We will report results using the guidelines for accurate and transparent health estimates reporting (GATHER) for Systematic Review 1 and the preferred reporting items for systematic reviews and meta-analyses guidelines (PRISMA) for Systematic Review 2.

PROSPERO registration number: CRD42017084243
BACKGROUND

The World Health Organization (WHO) and the International Labour Organization (ILO) are developing a joint methodology for estimating the work-related burden of disease and injury (WHO/ILO joint methodology) \(^1\). The organizations plan to estimate the numbers of deaths and disability-adjusted life years (DALYs) that are attributable to selected occupational risk factors, in the first place for the year 2015. The WHO/ILO joint methodology will be based on already existing WHO and ILO methodologies for estimating the burden of disease for selected occupational risk factors \(^2,3\). It will expand existing methodologies with estimation of the burden of several prioritized additional pairs of occupational risk factors and health outcomes. For this purpose, population attributable fractions \(^4\) – the proportional reduction in burden from the health outcome achieved by a reduction of exposure to the theoretical minimum risk exposure level – will be calculated for each additional risk factor-outcome pair, and these fractions will be applied to the total disease burden envelopes for the health outcome from the WHO Global Health Estimates \(^5\).

The WHO/ILO joint methodology may include a methodology for estimating the burden of ischaemic heart disease from occupational exposure to long work hours if feasible, as one additional prioritized risk factor-outcome pair. To optimize parameters used in estimation models, a systematic review is required of studies on the prevalence of exposure to long working hours (‘Systematic Review 1’), as well as a second systematic review and meta-analysis of studies with estimates of the effect of exposure to long work hours on ischaemic heart disease (‘Systematic Review 2’). In the current paper, we present the protocol for these two systematic reviews, in parallel to presenting systematic review protocols on other additional risk factor-outcome pairs elsewhere \(^6-13\). To our knowledge, this is the first systematic review protocol of its kind. The WHO/ILO joint estimation methodology and the burden of disease estimates are separate from these systematic reviews, and they will be described and reported elsewhere.

We refer separately to Systematic Reviews 1 and 2, because the two systematic reviews address different objectives and therefore require different methodologies. The two systematic reviews will, however, be harmonized and conducted in tandem. This will ensure that – in the later development of the methodology for estimating the burden of disease from this risk factor–outcome pair – the parameters on the risk factor prevalence are optimally matched with the parameters from studies on the effect of the risk factor on the designated outcome. The findings from Systematic Reviews 1 and 2 will be reported in two distinct journal articles. For all four protocols in the series with long working hours as the risk factor \(^6,10,11,14\), one Systematic Review 1 will be published.

**Rationale**

To consider the feasibility of estimating the burden of ischaemic heart disease due to exposure to long working hours, and to ensure that potential estimates of burden of ischaemic heart disease are reported in adherence with the guidelines for accurate and transparent health estimates reporting (GATHER), \(^15\) WHO and ILO require a systematic review of studies on the prevalence of relevant levels of exposure to long working hours (Systematic Review 1), as well as a systematic review and meta-analysis with estimates of the relative effect of exposure to long work hours on the prevalence of and mortality from ischaemic heart disease, compared with the theoretical minimum risk exposure level (Systematic Review 2). The theoretical minimum risk exposure level is the exposure level that would result in the lowest possible population risk, even if it is not feasible to attainable this exposure level in practice \(^4\). These data and effect estimates should be tailored to serve as parameters for estimating the burden of ischaemic heart disease from exposure to long work hours in the WHO/ILO joint methodology.

Our research will substantially extend the current body of systematic review evidence. For instance, a 2012 systematic review and meta-analysis on the effect of exposure with long working hours on cardiovascular disease, which included five cohort studies and six case-control studies published up to
September 2011, reported a pooled odds ratio of 1.37, with a 95% confidence interval (CI) of 1.11-1.70
16. A second systematic review on the effect of long working hours on ischaemic heart disease published
in 2012 included four prospective studies and seven case-control studies published between 1966 and
19 January 2011. For the prospective studies, the authors reported a pooled relative risk of 1.39 (95%
CI: 1.12-1.72), and for the case-control studies a pooled relative risk of 2.43 (95% CI: 1.81-3.26) 17.
Finally, a third systematic review and meta-analysis published in 2015 of 24 cohort studies (including
20 unpublished studies) in Europe, the USA and Australia up to August 20th 2014 found a relative risk
of 1.13 (95% CI: 1.02-1.26) for the effect of long working hours (≥55 hours/week) on ischaemic heart
disease 18. However, our Systematic Review 1 will be the – to the best of our knowledge – first
systematic review of prevalence studies of exposure to long working hours, and Systematic Review 2
will expand the scope of the existing systematic review evidence by covering evidence from studies
published up to 31 May 2018.

Work in the informal economy may lead to different exposures and exposure effects than does work in
the formal economy. The informal economy is defined as “all economic activities by workers and
economic units that are – in law or in practice – not covered or insufficiently covered by formal
arrangements”, but excluding “illicit activities, in particular the provision of services or the production,
sale, possession or use of goods forbidden by law, including the illicit production and trafficking of
drugs, the illicit manufacturing of and trafficking in firearms, trafficking in persons, and money
laundering, as defined in the relevant international treaties” 19. Consequently, formality of work
(informal vs. formal) may be an effect modifier of the effect of long working hours on ischaemic heart
disease. Therefore, we consider in both systematic reviews the formality of the economy reported in
included studies.

Description of the risk factor

The definition of the risk factor, the risk factor levels and the theoretical minimum risk exposure level
are presented in Table 1. Long working hours are defined as any working hours exceeding standard
working hours, i.e. working hours of ≥41 hours/week. Based on results from earlier studies on long
working hours and health endpoints 18,20,21, the preferred four exposure level categories for our review
are 35-40, 41-48, 49-54 and ≥55 hours/week. This will allow calculating estimates both for large
exposure contrast (i.e. comparing the theoretical minimal exposure to ≥55 hours/week) and for potential
dose-response associations (i.e. comparing the theoretical minimal exposure to all other exposure
categories). If the studies provide the preferred exposure level categories, we will use these categories,
but if they provide other exposure categories, we will use the other exposure categories, as long as
exposure exceeds 40 hours/week.

The theoretical minimum risk exposure are standard working hours defined as 35-40 hours/week. We
acknowledge that it is possible that the theoretical minimum risk exposure might be lower than standard
working hours, but we have to exclude working hours ≤35 hours/week, because studies indicate that a
proportion of individuals working less than standard hours do so because of existing health problems
17,20. Thus, this exposure concerns full-time workers in the formal and informal economy. In other words,
individuals working less than standard hours might belong to a health-selected group or a group
concerned with family care and therefore cannot serve as comparators. Consequently, if a study used as
the reference group individuals working less than standard hours or a combination of individuals
working standard hours and individuals working less than standard hours, it will be excluded from the
systematic review and meta-analysis. The category 35-40 hours/week is the reference group used in
many large studies and previous systematic reviews 15-17. Since the theoretical minimum risk exposure
level is usually set empirically based on the causal epidemiological evidence, we will change the
assumed level as evidence suggests.

If several studies report exposure levels differing from the standard levels we define here, then, if
possible, we will convert the reported levels to the standard levels and, if not possible, we will report
analyses on these alternate exposure levels as supplementary information in the systematic reviews. In the latter case, our protocol will be updated to reflect our new analyses.

Table 1: Definitions of the risk factor, risk factor levels and the minimum risk exposure level

<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Risk factor</strong></td>
</tr>
<tr>
<td><strong>Risk factor levels</strong></td>
</tr>
<tr>
<td><strong>Theoretical minimum risk exposure level</strong></td>
</tr>
</tbody>
</table>

Description of the outcome

The WHO *Global Health Estimates* group outcomes into standard burden of disease categories,\(^5\) based on standard codes from the *International Statistical Classification of Diseases and Related Health Problems 10th Revision* (ICD-10) \(^2\). The relevant WHO *Global Health Estimates* category for this systematic review is “II.H.2 Ischaemic heart disease”\(^3\). In line with the WHO *Global Health Estimates*, we define the health outcome covered in Systematic Review 2 as ischaemic heart disease, defined as conditions with ICD-10 codes I120 to I125 (Table 2). We will consider prevalence of, incidence of and mortality from ischaemic heart disease. Table 2 presents for each disease or health problem included in the WHO *Global Health Estimates* category the inclusion in this review. This review covers all the relevant WHO *Global Health Estimates* categories.

Table 2: ICD-10 codes and disease and health problems covered by the WHO burden of disease category *II.H.2 Ischaemic heart disease* and their inclusion in this review

<table>
<thead>
<tr>
<th>ICD-10 code</th>
<th>Disease or health problem</th>
<th>Included in this systematic review</th>
</tr>
</thead>
<tbody>
<tr>
<td>I20</td>
<td>Angina pectoris</td>
<td>Yes</td>
</tr>
<tr>
<td>I21</td>
<td>Acute myocardial infarction</td>
<td>Yes</td>
</tr>
<tr>
<td>I22</td>
<td>Subsequent myocardial infarction</td>
<td>Yes</td>
</tr>
<tr>
<td>I23</td>
<td>Certain current complications following acute myocardial infarction</td>
<td>Yes</td>
</tr>
<tr>
<td>I24</td>
<td>Other acute ischaemic heart diseases</td>
<td>Yes</td>
</tr>
<tr>
<td>I25</td>
<td>Chronic ischaemic heart disease</td>
<td>Yes</td>
</tr>
</tbody>
</table>

How the risk factor may impact the outcome

Figure 1 presents the logic model for our systematic review of the causal relationship between exposure to long working hours and ischaemic heart diseases. This logic model is an *a priori*, process-orientated one \(^2\) that seeks to capture the complexity of the risk factor–outcome causal relationship.\(^2\)
7

Theoretically, distinct social contexts in labour market are likely to exacerbate or mitigate the effect of exposure to long working hours on ischaemic heart disease risk. While empirical tests of this assumption are not available, these contexts can exert a direct effect on working hours. Evidence suggests that economic globalization drives people around the world to work longer.\textsuperscript{25}

Based on knowledge of previous research on long working hours and ischaemic heart disease,\textsuperscript{18,20,21} we assume that the effect of exposure to long working hours on ischaemic heart disease could be modified by country (or WHO region), sex, age, industrial sector, occupation and formality of the economy. Confounding should be considered by, at least, age, sex, and an indicator of socioeconomic position (e.g. income, education or occupational grade). Exceptions are accepted for studies whose study samples were homogenous (such as men only) or that conducted sensitivity analyses to test the presence of confounding (such as sex-disaggregated analyses that can help identify confounding by sex).

Several variables may mediate the effects of this exposure on disease risk through two major pathways. The first one concerns behavioural responses that result in an increase in health-adverse behaviours, such as cigarette smoking, high alcohol consumption, unhealthy diet, and physical inactivity. These behaviours are established risk factors of ischaemic heart disease.\textsuperscript{21,26} Moreover, impaired sleep and poor recovery resulting from this exposure increase the risk of ischaemic heart disease.\textsuperscript{27,28} Chronic psychosocial stress responses define a second pathway mediating the effects of exposure on ischaemic heart disease. According to established physiological evidence recurrent high effort (exposure) results in continued activation of the autonomic nervous/immune systems and associated stress axes, the sympatho-adrenal medullary and the hypothalamic-pituitary adrenal axes, with excessive release of respective stress hormones (adrenalin, noradrenalin and cortisol).\textsuperscript{29-31} In the longer run, this recurrent activation exceeds the regulatory capacity of the cardiovascular system, thus triggering functional dysregulations (e.g. sustained high blood pressure) and structural lesions (e.g. atherogenesis in coronary vessels).\textsuperscript{32}

In addition to epidemiological, clinical and experimental evidence suggesting that chronic psychosocial stress (including that from working long hours) presents a risk factor of ischaemic heart disease, there is indirect evidence on its causal role from animal studies. In classical experiments with cynomolgus macaques a direct effect of exposure to a chronic psychosocial stressor on growth of atherosclerotic plaques in coronary vessels was demonstrated, and this process was prevented by administration of beta-adrenergic blocking agents.\textsuperscript{33}
OBJECTIVES

1. Systematic Review 1: To systematically review quantitative studies of any design on the prevalence of relevant levels of exposure to long working hours in the years 2005 to 2018 among the working-age population, disaggregated by country, sex, age and industrial sector or occupation. Systematic Review 1 will be conducted in a coordinated fashion across all four review groups that examine long working hours with regard to health endpoints (ischaemic heart disease, stroke, depression and alcohol use), led by GS and with JL being the focal point from the working group on long working hours and ischaemic heart disease.

2. Systematic Review 2: To systematically review and meta-analyse randomized controlled studies, cohort studies, case-control studies, and other non-randomized intervention studies including working-age workers (Population) exposed to long working hours (Exposure), compared with workers with the minimum theoretical risk exposure level of 35-40 hours/week (Comparator), in order to estimate the relative effect on ischaemic heart disease (Outcome).
We will apply the Navigation Guide methodology for systematic reviews in environmental and occupational health as our guiding methodological framework, wherever feasible. The guide applies established systematic review methods from clinical medicine, including standard Cochrane Collaboration methods for systematic reviews of interventions, to the field of environmental and occupational health to ensure systematic and rigorous evidence synthesis on environmental and occupational risk factors that reduces bias and maximizes transparency. The need for further methodological development and refinement of the relatively novel Navigation Guide has been acknowledged.

Systematic Review 1 may not map well to the Navigation Guide framework (Figure 1 on page 1009), which is tailored to hazard identification and risk assessment. Nevertheless, steps 1–6 for the stream on human data can be applied to systematically review exposure to risk factors. Systematic Review 2 maps more closely to the Navigation Guide framework, and we will conduct steps 1–6 for the stream on human data, but not conduct any steps for the stream on non-human data, although we will briefly summarize narratively the evidence from non-human data that we are aware of.

We have registered the protocol in PROSPERO under CRD42018084131. This protocol adheres with the preferred reporting items for systematic review and meta-analysis protocols statement (PRISMA-P)\(^\text{36,37}\), with the abstract adhering with the reporting items for systematic reviews in journal and conference abstracts (PRISMA-A)\(^\text{38}\). Any modification of the methods stated in the present protocol will be registered in PROSPERO and reported in the systematic review itself. Systematic Review 1 will be reported according to the GATHER guidelines\(^\text{15}\), and Systematic Review 2 will be reported according to the preferred reporting items for systematic review and meta-analysis statement (PRISMA)\(^\text{39}\). Our reporting of the parameters for estimating the burden of ischaemic heart disease to long working hours in the systematic review will adhere with the requirements of the GATHER guidelines, because the WHO/ILO burden of disease estimates that may be produced consecutive to the systematic review must also adhere to these reporting guidelines.

**Systematic Review 1**

**Eligibility criteria**

The population, exposure, comparator and outcome (PECO) criteria\(^\text{39}\) are described below.

**Types of populations (P)**

We will include studies of the working-age population (≥ 15 years) in the formal and informal economy. Studies of children (aged < 15 years) and unpaid domestic workers will be excluded. Participants residing in any WHO and/or ILO Member State and any industrial setting or occupation will be included. We note that occupational exposure to long working hours may potentially have further population reach (e.g. across generations for workers of reproductive age) and acknowledge that the scope of our systematic reviews will not be able capture these populations and impacts on them. Appendix A provides a complete, but briefer overview of the PECO criteria.

**Types of exposures (E)**
We will include studies that define long working hours in accordance with our standard definition (Table 1). We will prioritize measures of the total number of hours worked, including in both of: main and secondary jobs, self-employment and salaried employment and informal and formal jobs. Cumulative exposure may be the most relevant exposure metric in theory, but we will here also prioritize a non-cumulative exposure metric in practice, because we believe that global exposure data on agreed cumulative exposure measures do not currently exist. We will include all studies where long working hours were measured, whether objectively (e.g. by means of time recording technology), or subjectively, including studies that used measurements by experts (e.g. scientists with subject matter expertise) and self-reports by the worker or workplace administrator or manager. If a study presents both objective and subjective measurements, then we will prioritize objective measurements. We will include studies with measures from any data source, including registry data, in the same analyses and description.

We will include studies on the prevalence of occupational exposure to the risk factor, if it is disaggregated by country, sex (two categories: female, male), age group (ideally in 5-year age bands, such as 20–24 years) and industrial sector (e.g. International Standard Industrial Classification of All Economic Activities, Revision 4 [ISIC Rev. 4]) or occupation (as defined, for example, by the International Standard Classification of Occupations 1988 [ISCO-88] or 2008 [ISCO-08]). We will also extract data on the context of risk factor exposure. Criteria may be revised in order to identify optimal data disaggregation to enable subsequent estimation of the burden of disease.

We shall include studies with exposure data for the years 2005 to 31 May 2018. For optimal modelling of exposure, WHO and ILO require exposure data up to 2018, because recent data points help better estimate time trends, especially where data points may be sparse. The additional rationale for this data collection window is that the WHO and ILO aim to estimate burden of disease in the year 2015, and we believe that the lag time from exposure to outcome will not exceed 10 years; so in their models, the organizations can use the exposure data from as early as 2005 to determine the burden of ischaemic heart disease 10 years later in 2015. To make a conclusive judgment on the best lag time to apply in the model, we will summarize the existing body of evidence on the lag time between exposure to long working hours and ischaemic heart disease in the review.

Both objective and subjective measures will be included. If both subjective and objective measures are presented, then we will prioritize objective ones. Studies with measures from any data source, including registries, will be eligible. The exposure parameter should match the one used in Systematic Review 2 or can be converted to match it.

Types of comparators

There will be no comparator, because we will review risk factor prevalence only.

Types of outcomes

Exposure to the occupational risk factor (i.e. long working hours).

Types of studies

This Systematic Review will include quantitative studies of any design, including cross-sectional studies. These studies must be representative of the relevant industrial sector, relevant occupational
group or the national population. We will exclude qualitative, modelling, and case studies, as well as non-original studies without quantitative data (e.g. letters, commentaries and perspectives).

Study records written in any language will be included. If a study record is written in a language other than those spoken by the authors of this review or those of other reviews in the series (i.e. Arabic, Bulgarian, Chinese, Danish, Dutch, English, French, Finnish, German, Hungarian, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish and Swedish), it will be translated into English. Published and unpublished studies will be included.

Studies conducted using unethical practices will be excluded from the review (e.g. studies that deliberately exposed humans to a known risk factor to human health).

**Types of effect measures**

We will include studies with a measure of the prevalence of a relevant level of exposure to long working hours.

**Information sources and search**

**Electronic academic databases**

We (DG and DP) will at a minimum search the following seven electronic academic databases:


The Ovid Medline search strategy for Systematic Review 1 is presented in Appendix B. We will perform searches in electronic databases operated in the English language using a search strategy in the English language. Consequently, study records that do not report essential information (i.e. title and abstract) in English will not be captured. We will adapt the search syntax to suit the other electronic academic and grey literature databases. When we are nearing completion of the review, we will search the PubMed database for the most recent publications (e.g., e-publications ahead of print) over the last six months. Any deviation from the proposed search strategy in the actual search strategy will be documented.

**Electronic grey literature databases**

GS and AT will at a minimum search the two following electronic academic databases:

1. OpenGrey (http://www.opengrey.eu/)
2. Grey Literature Report (http://greylit.org/)

**Internet search engines**
We (GS and MMF) will also search the Google (www.google.com/) and GoogleScholar (www.google.com/scholar/) Internet search engines and screen the first 100 hits for potentially relevant records.

**Organizational websites**

The websites of the following six international organizations and national government departments will be searched by AD, DG, JP and GS:

1. International Labour Organization (www.ilo.org/).
2. World Health Organization (www.who.int).
5. China National Knowledge Infrastructure (http://www.cnki.net/).

**Hand-searching and expert consultation**

AD, DG, JP, and GS will hand-search for potentially eligible studies in:
- Reference list of previous systematic reviews.
- Reference list of all study records of all included studies.
- Study records published over the past 24 months in the three peer-reviewed academic journals from which we obtain the largest number of included studies.
- Study records that have cited an included study record (identified in Web of Science citation database).
- Collections of the review authors.

Additional experts will be contacted with a list of included studies and study records, with the request to identify potentially eligible additional ones.

**Study selection**

Study selection will be carried out with Covidence \(^{43,44}\) and/or the Rayyan Systematic Reviews Web App \(^{45}\). All study records identified in the search will be downloaded and duplicates will be identified and deleted. Afterwards, at least two review authors (AD and KS), working in pairs, will independently screen against eligibility criteria titles and abstracts (step 1) and then full texts of potentially relevant records (step 2). A third review author (GS) will resolve any disagreements between the pairs of study selectors. If a study record identified in the literature search was authored by a review author assigned to study selection or if an assigned review author was involved in the study, then the record will be re-assigned to another review author for study selection. In the systematic review, we will document the study selection in a flow chart, as per GATHER guidelines \(^ {15} \).

**Data extraction and data items**
A data extraction form will be developed and piloted until there is convergence and agreement among data extractors. At a minimum, two review authors (out of: BAE, ES and LMH) will independently extract the data on exposure to long working hours, disaggregated by country, sex, age and industrial sector or occupation. A third review author (GS) will resolve conflicting extractions. At a minimum, we will extract data on study characteristics (including study authors, study year, study country, participants, exposure and outcome), study design (including study type and measurements of the risk factor and outcome, and response rate), risk of bias (including missing data, as indicated by response rate and other measures) and study context. The estimates of the proportion of the population exposed to the occupational risk factor from included studies will be entered into and managed with, the Review Manager, Version 5.3 (RevMan 5.3) or DistillerSR software.

We will also extract data on potential conflict of interest in included studies, including the financial disclosures and funding sources of each author and their affiliated organization. We will use a modification of a previous method to identify and assess undisclosed financial interests. Where no financial disclosure/conflict of interest is provided, we will search declarations of interest both in other records from this study published in the 36 months prior to the included study record and in other publicly available repositories.

We will request missing data from the principal study author by email or phone, using the contact details provided in the principal study record. If no response is received, we will follow up twice via email, at two and four weeks.

### Risk of bias assessment

Generally agreed methods (i.e. framework plus tool) for assessing risk of bias do not exist for systematic reviews of input data for health estimates, for burden of disease studies, of prevalence studies in general, and those of prevalence studies of occupational and/or environmental risk factors specifically. None of the five standard risk of bias assessment methods in systematic reviews are applicable to assessing prevalence studies. The Navigation Guide does not support checklist approaches, such as, for assessing risk of bias in prevalence studies.

We will use a modified version of the Navigation Guide risk of bias tool that we developed specifically for Systematic Review 1 (Appendix C). We will assess risk of bias on the levels of the individual study and the entire body of evidence. As per our preliminary tool, we will assess risk of bias along five domains: (i) selection bias; (ii) performance bias; (iii) misclassification bias; (iv) conflict of interest; and (v) other biases. Risk of bias will be: “low”; “probably low”; “probably high”; “high” or “not applicable”. To judge the risk of bias in each domain, we will apply our a priori instructions (Appendix C).

All risk of bias assessors (BAE, DG, LMH and GS) will trial the tool until they synchronize their understanding and application of each risk of bias domain, considerations and criteria for ratings. At least two study authors (out of: BAE, DG and LMH) will then independently judge the risk of bias for each study by outcome, and a third author (GS) will resolve any conflicting judgments. We will present the findings of our risk of bias assessment for each eligible study in a standard ‘Risk of bias’ table. Our risk of bias assessment for the entire body of evidence will be presented in a standard ‘Risk of bias summary’ figure.

### Synthesis of results
We will neither produce any summary measures, nor synthesise the evidence quantitatively. The included evidence will be presented in what could be described as an ‘evidence map’. All included data points from included studies will be presented, together with meta-data on the study design, number of participants, characteristics of population, setting, and exposure measurement of the data point.

Quality of evidence assessment

There is no agreed method for assessing quality of evidence in systematic reviews of the prevalence of occupational and/or environmental risk factors. We will adopt/adapt from the latest Navigation Guide instructions for grading, including criteria (Appendix D). We will downgrade for the following five reasons from the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach: (i) risk of bias; (ii) inconsistency; (iii) indirectness; (iv) imprecision; and (v) publication bias. We will grade the evidence, using the three Navigation Guide quality of evidence ratings: “high”, “moderate” and “low”. Within each of the relevant reasons for downgrading, we will rate any concern per reason as “none”, “serious” or “very serious”. We will start at “high” for non-randomized studies and will downgrade for no concern by nil, for a serious concern by one grade (-1), and for a very serious concern by two grades (-2). We will not up-grade or down-grade the quality of evidence for the three other reasons normally considered in GRADE assessments (i.e. large effect, dose-response and plausible residual confounding and bias), because we consider them irrelevant for prevalence estimates.

All quality of evidence assessors (BAE, LMH and DG) will trial the application of our instructions and criteria for quality of evidence assessment until their understanding and application is synchronized. At least two review authors (LMH and DG) will independently judge the quality of evidence for the entire body of evidence by outcome. A third review author (GS) will resolve any conflicting judgments. In the systematic review, for each outcome, we will present our assessments of the risk for each GRADE domain, as well as an overall GRADE rating.

Strength of evidence assessment

To our knowledge, no agreed method exists for rating strength of evidence in systematic reviews of prevalence studies. We (AD and GS) will rate the strength of the evidence for use as input data for estimating national-level exposure to the risk factor. Our rating will be based on a combination of the following four criteria: (i) quality of the entire body of evidence; (ii) population coverage of evidence (WHO regions and countries); (iii) confidence in the entire body of evidence; and (iv) other compelling attributes of the evidence that may influence certainty. We will rate the strength of the evidence as either “potentially sufficient” or “potentially inadequate” for use as input data (Appendix E).

Systematic Review 2

Eligibility criteria

The PECO criteria are described below.

Types of populations
We will include studies of the working-age population (≥ 15 years) in the formal and informal economy. Studies of children (aged < 15 years) and unpaid domestic workers will be excluded. Data on the formal and informal economy that the workers work in will be extracted. Participants residing in any WHO and/or ILO Member State and any industrial setting or occupational group will be included. We note that occupational exposure to long working hours may potentially have further population reach (e.g. across generations for workers of reproductive age) and acknowledge that the scope of our systematic reviews will not be able capture these populations and impacts on them. Appendix F provides a complete, but briefer overview of the PECO criteria.

**Types of exposures**

We will include studies that define long working hours in accordance with our standard definition (Table 1). We will again prioritize measures of the total number of hours worked, including in both of: main and secondary jobs, self-employment and salaried employment and informal and formal jobs. We will include all studies where long working hours were measured, whether objectively (e.g. by means of time recording technology), or subjectively, including studies that used measurements by experts (e.g. scientists with subject matter expertise) and self-reports by the worker or workplace administrator or manager. If a study presents both objective and subjective measurements, then we will prioritize objective measurements. We will include studies with measures from any data source, including registry data, in the same analyses and description.

Regarding years of data coverage in our systematic review, we will include studies that define exposure to long working hours in accordance with our standard definition (Table 1). Studies from any year will be included.

**Types of comparators**

The included comparator will be participants exposed to the theoretical minimum risk exposure level (Table 1). We will exclude all other comparators.

**Types of outcomes**

We will include studies that define ischaemic heart disease in accordance with our standard definition of this outcome (Table 2). Other coronary-related unspecific symptoms (e.g. chest pain) will be excluded. We expect that most studies examining exposure to long working hours and its effect on ischaemic heart disease have documented ICD-10 diagnostic codes. In the remaining cases, methods that approximate ICD-10 criteria will ascertain ischaemic heart disease.

The following measurements of ischaemic heart disease will be regarded as eligible:

i) Diagnosis by a physician with imaging.

ii) Hospital discharge records.

iii) Other relevant administrative data (e.g. records of sickness absence or disability).

iv) Medically certified cause of death.

All other measures will be excluded from this systematic review.

Objective and subjective measures of the outcome will be eligible. If a study presents both objective and subjective measurements, then we will prioritize the objective ones.
Types of studies

We will include studies that investigate the effect of long working hours on ischaemic heart disease for any years. Eligible study designs will be randomized controlled trials (including parallel-group, cluster, cross-over and factorial trials), cohort studies (both prospective and retrospective), case-control studies, and other non-randomized intervention studies (including quasi-randomized controlled trials, controlled before-after studies and interrupted time series studies). We included a broader set of observational study designs than is commonly included, because a recent augmented Cochrane Review of complex interventions identified valuable additional studies using such a broader set of study designs. As we have an interest in quantifying risk and not in qualitative assessment of hazard, we will exclude all other study designs (e.g. uncontrolled before-and-after, cross-sectional, qualitative, modelling, case and non-original studies).

Records published in any year and any language will be included. Again, the search will be conducted using English language terms, so that records published in any language that present essential information (i.e. title and abstract) in English will be included. If a record is written in a language other than those spoken by the authors of this review or those of other reviews in the series, then the record will be translated into English. Published and unpublished studies will be included. Studies conducted using unethical practices will be excluded.

Types of effect measures

We will include measures of the relative effect of a relevant level of long working hours on the risk of having, developing or dying from ischaemic heart disease, compared with the theoretical minimum risk exposure level. Included relative effect measures are risk ratios and odds ratios for prevalence and mortality measures and hazard ratios for incidence measures (e.g., developed or died from ischaemic heart disease). Measures of absolute effects will be excluded (e.g. mean differences in risks or odds). Measures of absolute effects (e.g. mean differences in risks or odds) will be converted into relative effect measures, but if conversion is impossible, they will be excluded. To ensure comparability of effect estimates and facilitate meta-analysis, if a study presents an odds ratio, then we will convert it into a risk ratio, if possible, using the guidance provided in the Cochrane Collaboration’s handbook for systematic reviews of interventions.

As shown in our logic framework (Figure 1), we a priori consider the following variables to be potential effect modifiers of the effect of long working hours on ischaemic heart disease: country, age, sex, industrial sector, occupation and formality of employment. We consider age, sex and socio-economic position to be potential confounders. Potential mediators are: smoking, alcohol use, physical inactivity, unhealthy diet, impaired sleep, poor recovery, autonomous nervous system activity and immune system activity.

If a study presents estimates for the effect from two or more alternative models that have been adjusted for different variables, then we will systematically prioritize the estimate from the model that we consider best adjusted, applying the lists of confounders and mediators identified in our logic model (Figure 1). We will prioritize estimates from models adjusted for more potential confounders over those from models adjusted for fewer. For example, if a study presents estimates from a crude, unadjusted model (Model A), a model adjusted for one potential confounder (Model B) and a model adjusted for
two potential confounders (Model C), then we will prioritize the estimate from Model C. We will prioritize estimates from models unadjusted for mediators over those from models that adjusted for mediators, because adjustment for mediators can introduce bias. For example, if Model A has been adjusted for two confounders, and Model B has been adjusted for the same two confounders and a potential mediator, then we will choose the estimate from Model A over that from Model B. We prioritize estimates from models that can adjust for time-varying confounders that are at the same time also mediators, such as marginal structural models over estimates from models that can only adjust for time-varying confounders, such as fixed-effects models, over estimates from models that cannot adjust for time-varying confounding. If a study presents effect estimates from two or more potentially eligible models, then we will explain specifically why we prioritized the selected model.

Information sources and search

Electronic academic databases

At a minimum, we (CB, EC and PL) will search the eight following electronic academic databases:

1. International Clinical Trials Register Platform (to May 31st 2018).
7. CISDOC (1901 to 2012).

The Ovid Medline search strategy for Systematic Review 2 is presented in Appendix G. We will perform searches in electronic databases operated in the English language using a search strategy in the English language. We will adapt the search syntax to suit the other electronic academic and grey literature databases. When we are nearing completion of the review, we will search the PubMed database for the most recent publications (e.g., e-publications ahead of print) over the last six months. Any deviation from the proposed search strategy in the actual search strategy will be documented.

Electronic grey literature databases

At a minimum, we (GS and AT) will search the two following two electronic academic databases:

1. OpenGrey (http://www.opengrey.eu/)
2. Grey Literature Report (http://greylit.org/)

Internet search engines

We (GS and MMF) will also search the Google (www.google.com/) and GoogleScholar (www.google.com/scholar/) Internet search engines and screen the first 100 hits for potentially relevant records.

Organizational websites

The websites of the seven following international organizations and national government departments will be searched for both systematic reviews by GS and HP:
Hand-searching and expert consultation

We (GS and JL) will hand-search for potentially eligible studies in:

- Reference list of previous systematic reviews.
- Reference list of all included study records.
- Study records published over the past 24 months in the three peer-reviewed academic journals with the largest number of included studies.
- Study records that have cited the included studies (identified in Web of Science citation database).
- Collections of the review authors.

Additional experts will be contacted with a list of included studies, with the request to identify potentially eligible additional studies.

Study selection

Study selection will be carried out with the Rayyan Systematic Reviews Web App. All study records identified in the search will be downloaded and duplicates will be identified and deleted. Afterwards, at least two review authors (PLS and JL), working in pairs, will independently screen titles and abstracts (step 1) and then full texts (step 2) of potentially relevant records. A third review author (JS) will resolve any disagreements between the two review authors. If a study record identified in the literature search was authored by a review author assigned to study selection or if an assigned review author was involved the study, then the record will be re-assigned to another review author for study selection. The study selection will be documented in a flow chart in the systematic review, as per PRISMA guidelines.

Data extraction and data items

A data extraction form will be developed and trialled until data extractors reach convergence and agreement. At a minimum, two review authors (RR and JL) will extract data on study characteristics (including study authors, study year, study country, participants, exposure and outcome), study design (including summary of study design, comparator, epidemiological models used and effect estimate measure), risk of bias (including selection bias, reporting bias, confounding, and reverse causation) and study context (e.g. data on contemporaneous exposure to other occupational risk factors potentially relevant for deaths or other health loss from ischaemic heart disease). A third review author (JS) will resolve conflicts in data extraction. Data will be entered into and managed with the Review Manager, Version 5.3 (RevMan 5.3) or DistillerSR softwares, but the Health Assessment Workspace.
Collaborative (HAWC, http://hawc.readthedocs.io/en/latest/) may also be used in parallel or to prepare data for entry into RevMan 5.3.

We will also extract data on potential conflict of interest in included studies. For each author and affiliated organization of each included study record, we will extract their financial disclosures and funding sources. We will use a modification of a previous method to identify and assess undisclosed financial interest of authors. Where no financial disclosure or conflict of interest statements are available, we will search the name of all authors in other study records gathered for this study and published in the prior 36 months and in other publicly available declarations of interests.

We will request missing data from the principal study author by email or phone, using the contact details provided in the principal study record. If we do not receive a positive response from the study author, we will send follow-up emails twice, at two and four weeks.

Risk of bias assessment

Standard risk of bias tools do not exist for systematic reviews for hazard identification in occupational and environmental health, nor for risk assessment. The five methods specifically developed for occupational and environmental health are for either or both hazard identification and risk assessment, and they differ substantially in the types of studies (randomized, observational and/or simulation studies) and data (e.g. human, animal and/or in vitro) they seek to assess. However, all five methods, including the Navigation Guide, assess risk of bias in human studies similarly.

The Navigation Guide was specifically developed to translate the rigor and transparency of systematic review methods applied in the clinical sciences to the evidence stream and decision context of environmental health, which includes workplace environment exposures and associated health outcomes. The guide is our overall organizing framework, and we will also apply its risk of bias assessment method in Systematic Review 2. The Navigation Guide risk of bias assessment method builds on the standard risk of bias assessment methods of the Cochrane Collaboration and the US Agency for Healthcare Research and Quality. Some further refinements of the Navigation Guide method may be warranted, but it has been successfully applied in several completed and ongoing systematic reviews. In our application of the Navigation Guide method, we will draw heavily on one of its latest versions, as presented in the protocol for an ongoing systematic review. Should a more suitable method become available, we may switch to it.

We will assess risk of bias on the individual study level and on the body of evidence overall. The nine risk of bias domains included in the Navigation Guide method for human studies are: (i) source population representation; (ii) blinding; (iii) exposure assessment; (iv) outcome assessment; (v) confounding; (vi) incomplete outcome data; (vii) selective outcome reporting; (viii) conflict of interest; and (ix) other sources of bias. While two of the earlier case studies of the Navigation Guide did not utilize outcome assessment as a risk of bias domain for studies of human data, all of the subsequent reviews have included this domain. Risk of bias or confounding ratings will be: “low”; “probably low”; “probably high”; “high” or “not applicable”. To judge the risk of bias in each domain, we will apply a priori instructions (Appendix H), which we have adopted or adapted from an ongoing Navigation Guide systematic review. For example, a study will be assessed as carrying “low” risk of bias from source population representation, if we judge the source population to be described in sufficient detail (including eligibility criteria, recruitment, enrollment, participation and loss to follow up) and the distribution and characteristics of the study sample to indicate minimal or no risk of...
selection effects. The risk of bias at study level will be determined by the worst rating in any bias
domain for any outcome. For example, if a study is rated as “probably high” risk of bias in one domain
for one outcome and “low” risk of bias in all other domains for the outcome and in all domains for all
other outcomes, the study will be rated as having a “probably high” risk of bias overall.

All risk of bias assessors (EC, AT and PL) will jointly trial the application of the risk of bias criteria
until they have synchronized their understanding and application of these criteria. At least two study
authors (EC and AT) will independently judge the risk of bias for each study by outcome. Where
individual assessments differ, a third author (PL) will resolve the conflict. In the systematic review, for
each included study, we will report our study-level risk of bias assessment by domain in a standard
‘Risk of bias’ table. For the entire body of evidence, we will present the study-level risk of bias
assessments in a ‘Risk of bias summary’ figure.

Synthesis of results

We will conduct meta-analyses separately for estimates of the effect on incidence and mortality. Studies
of different designs will not be combined quantitatively. If we find two or more studies with an eligible
effect estimate, two or more review authors (JS and JL) will independently investigate the clinical
heterogeneity of the studies in terms of participants (including country, sex, age and industrial sector or
occupation), level of risk factor exposure, comparator and outcomes. If we find that effect estimates
differ considerably by country, sex and/or age, or a combination of these, then we will synthesise
evidence for the relevant populations defined by country, sex and/or age, or combination thereof.
Differences by country could include or be expanded to include differences by country group (e.g.
WHO region or World Bank income group). If we find that effect estimates are clinically homogenous
across countries, sexes and age groups, then we will combine studies from all of these populations into
one pooled effect estimate that could be applied across all combinations of countries, sexes and age
groups in the WHO/ILO joint methodology.

If we judge two or more studies for the relevant combination of country, sex and age group, or
combination thereof, to be sufficiently clinically homogenous to potentially be combined quantitatively
using quantitative meta-analysis, then we will test the statistical heterogeneity of the studies using the
$\hat{I}^2$ statistic. If two or more clinically homogenous studies are found to be sufficiently homogenous
statistically to be combined in a meta-analysis, we will pool the risk ratios of the studies in a quantitative
meta-analysis, using the inverse variance method with a random effects model to account for cross-
study heterogeneity. The meta-analysis will be conducted in RevMan 5.3, but the data for entry into
these programmes may be prepared using another recognized statistical analysis programme, such as
Stata. We will neither quantitatively combine data from studies with different designs (e.g. combining
cohort studies with case-controls studies), nor unadjusted and adjusted models. We will only combine
studies that we judge to have a minimum acceptable level of adjustment for confounders. If quantitative
synthesis is not feasible, then we will synthesise the study findings narratively and identify the estimates
that we judged to be the highest quality evidence available.

Additional analyses

If we source micro-data on exposure, outcome and potential confounding variables, we may conduct
meta-regressions to adjust optimally for potential confounders.
If there is evidence for differences in effect estimates by country, sex, age, industrial sector and/or occupation, or by a combination of these variables, then we will conduct subgroup analyses by the relevant variable or combination of variables, as feasible. Where both studies on workers in the informal economy and in the formal economy are included, then we will conduct sub-group analyses by formality of economy. Findings of these subgroup analyses, if any, will be used as parameters for estimating burden of disease specifically for relevant populations defined by these variables. We will also conduct subgroup analyses by study design (e.g. randomized controlled trials versus cohort studies versus case-control studies).

We will perform a sensitivity analyses that will include only studies judged to be of “low” or “probably low” risk of bias from conflict of interest; judged to be of “low” or “probably low” risk of bias; and with documented or approximated ICD-10 diagnostic codes. We may also conduct a sensitivity analysis using an alternative meta-analytic model, namely the inverse variance heterogeneity (IVhet) model.

Quality of evidence assessment

We will assess quality of evidence using a modified version of the Navigation Guide quality of evidence assessment tool. The tool is based on the GRADE approach adapted specifically to systematic reviews in occupational and environmental health. Should a more suitable method become available, we may switch to it.

At least two review authors (JS and JL) will assess quality of evidence for the entire body of evidence by outcome, with any disagreements resolved by a third review author. We will adopt or adapt the latest Navigation Guide instructions (Appendix D) for grading the quality of evidence. We will downgrade the quality of evidence for the following five GRADE reasons: (i) risk of bias; (ii) inconsistency; (iii) indirectness; (iv) imprecision; and (v) publication bias. If our systematic review includes ten or more studies, we will generate a funnel plot to judge concerns on publication bias. If it includes nine or fewer studies, we will judge the risk of publication bias qualitatively. To assess risk of bias from selective reporting, protocols of included studies, if any, will be screened to identify instances of selective reporting.

We will grade the evidence, using the three Navigation Guide standard quality of evidence ratings: “high”, “moderate” and “low”. Within each of the relevant domains, we will rate the concern for the quality of evidence, using the ratings “none”, “serious” and “very serious”. As per Navigation Guide, we will start at “high” for randomized studies and “moderate” for observational studies. Quality will be downgraded for no concern by nil grades (0), for a serious concern by one grade (-1) and for a very serious concern by two grades (-2). We will upgrade the quality of evidence for the following other reasons: large effect, dose-response and plausible residual confounding and bias. For example, if we have a serious concern for risk of bias in a body of evidence consisting of observational studies (-1), but no other concerns, and there are no reasons for upgrading, then we will downgrade its quality of evidence by one grade from “moderate” to “low”.

Strength of evidence assessment

We will apply the standard Navigation Guide methodology to rate the strength of the evidence. The rating will be based on a combination of the following four criteria: (i) quality of the body of evidence; (ii) direction of the effect; (iii) confidence in the effect; and (iv) other compelling attributes of the data that may influence our certainty. The ratings for strength of evidence for the
effect of long working hours on ischaemic heart disease will be "sufficient evidence of toxicity/harmfulness", "limited of toxicity/harmfulness", "inadequate of toxicity/harmfulness" and "evidence of lack of toxicity/harmfulness" (Appendix I).
FINANCIAL SUPPORT

All authors are salaried staff members of their respective institutions. The publication was prepared with financial support from the WHO cooperative agreement with the Centres for Disease Control and Prevention National Institute for Occupational Safety and Health of the United States of America on implementing Resolution WHA 60.26 "Workers' Health: Global Plan of Action" (Grant 1 E11 OH0010676-02).

SPONSORS

The sponsors of this systematic review are the World Health Organization and the International Labour Organization.

AUTHOR CONTRIBUTIONS

IDI, NL, FP and APU had the idea for the systematic review. IDI, NL, FP, and YU gathered the review team. FP led and all authors contributed to the development of the standard methodology for all systematic reviews in the series. FP led and all authors contributed to the development and writing of the standard template for all protocols in the series. JL and JS are the lead reviewers of this systematic review. JL and JS wrote the first draft of this protocol, using the protocol template prepared by FP; and CB, EC, MMF, PL, JL, FP, HP, RR, JS, PLS, AT and YU made substantial contributions to the revisions of the manuscript. The search strategy was mainly developed and piloted by DP, GS, and JL. JL, FP and JS are experts in epidemiology, JL and JS are experts in occupational psychosocial risk factors and cardiovascular diseases, and FP is an expert in systematic review methodology. FP coordinated all inputs from WHO, ILO and external experts and ensured consistency across the systematic reviews of the series. JL and JS are the guarantors of the systematic review.

ACKNOWLEDGMENTS

We thank Alexis Descatha, Cristina Di Tecco, Diana Gagliardi, Sergio Iavicoli, Daniela V. Pachito, and Cristina Di Tecco for their feedback on an earlier version of this protocol. We thank Frida Fischer, Anders Knutsson and Mikael Sallinen for their feedback on the search strategy. We are grateful to Lisa Bero, Rebecca Morgan, Susan Norris, Holger J. Schünemann, Patrice Sutton, and Tracey Woodruff for their feedback on the methods for this protocol. We thank Paul Whaley and Tim Driscoll for their editorial guidance. The authors alone are responsible for the views expressed in this article and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

CONFLICT OF INTEREST

None declared.
REFERENCES

1. Ryder G. Welcome address from the Director General of the International Labour Organization. XXI World Congress on Safety and Health at Work; 2017; Sands Expo and Convention Centre, Singapore; 2017.


44. Covidence systematic review software VHI, Melbourne, Australia. Available at www.covidence.org.


79. Rooney AA. Protocol to evaluate the evidence for an association between perfluorooctanoic acid or perfluorooctane sulfonate exposure and immunotoxicity. In: Translation OoHAa, editor. Durham, NC: National Toxicology Program Division, National Institute of Environmental Health Sciences; 2015.

Appendix A: Overview of inclusion and exclusion criteria, Systematic Review 1

Include

1. Quantitative studies of any design, including cross-sectional studies, on the prevalence of relevant levels of long working hours for the years 2005 to 2018
2. Studies of the working-age population (≥ 15 years), including workers in the informal economy
3. Studies on populations residing in any Member States of WHO and/or ILO and any industrial sector or occupational group setting
4. Studies on the prevalence of relevant levels of long working hours, if the prevalence is disaggregated by WHO region, sex (two categories: female, male), age group and industrial sector or occupation
5. Studies published between 2005-2018 and any language with essential information (title and abstract) in English

Exclude

1. Studies of unpaid domestic workers
2. Studies of children (aged < 15 years)
3. Studies with data only on years other than between 2005 and 2018
4. Qualitative, modelling, and case studies, as well as non-original studies without quantitative data (e.g. letters, commentaries and perspectives)
Appendix B: Ovid Medline search strategy, Systematic Review 1

Population

No search string

Exposure

1. exp "personnel staffing and scheduling"/
2. "personnel staffing and scheduling".ti,ab,kw.
3. shift work schedule.ti,ab,kw.
4. work schedule tolerance.ti,ab,kw.
5. workload.kw.
6. workday shifts.ti,ab,kw.
7. overwork*.ti,ab,kw.
8. overtime.ti,ab,kw.
9. workweek*.ti,ab,kw.
10. (work* adj3 hour*).ti,ab,kw.
11. (work* adj3 schedul*).ti,ab,kw.
12. work* adj3 roster.ti,ab,kw.
13. (work* adj3 organi#ation).ti,ab,kw.
14. (work* adj3 time*).ti,ab,kw.
15. (work* adj3 overload*).ti,ab,kw.
16. (work* adj3 extend*).ti,ab,kw.
17. (work* adj3 compress*).ti,ab,kw.
18. (work* adj3 week*).ti,ab,kw.
19. (work* adj3 day?).ti,ab,kw.
20. (job? adj3 hour*).ti,ab,kw.
21. (job? adj3 schedul*).ti,ab,kw.
23. (job? adj3 organi#ation).ti,ab,kw.
24. (job? adj3 time*).ti,ab,kw.
25. (job? adj3 overload*).ti,ab,kw.
26. (job? adj3 extend*).ti,ab,kw.
27. (job? adj3 compress*).ti,ab,kw.
28. (job? adj3 week*).ti,ab,kw.
29. (job? adj3 day?).ti,ab,kw.
30. (shift? adj3 hour*).ti,ab,kw.
31. (shift? adj3 schedul*).ti,ab,kw.
32. (shift? adj3 roster).ti,ab,kw.
33. (shift? adj3 organi#ation).ti,ab,kw.
34. (shift? adj3 time*).ti,ab,kw.
35. (shift? adj3 overload*).ti,ab,kw.
36. (shift? adj3 extend*).ti,ab,kw.
37. (shift? adj3 compress*).ti,ab,kw.
38. (shift? adj3 week*).ti,ab,kw.
39. (shift? adj3 day?).ti,ab,kw.
40. (work* and (life* or live*) and (balances* or imbalances* or unbalances* or interference*)).ti,ab,kw.
41. (work* and famil* and conflict*).ti,ab,kw.
42. or/1-41

Study design

43. prevalence.tw.
44. incidence.tw.
45. epidemiol*.tw.
46. survey.tw.
47. rapid assessment.tw.
48. situation assessment.tw.
49. situational assessment.tw.
50. rar.tw.
51. cohort.tw.
52. surveillance.tw.
53. seroprevalence.tw.
54. seroincidence.tw.
55. seroepidemiol*.tw.
56. screening.mp.
57. exp epidemiologic methods/
58. exp epidemiologic studies/
59. exp sentinel surveillance/
60. exp seroepidemiologic studies/
61. exp cohort studies/
62. exp cross-sectional studies/
63. exp longitudinal studies/
64. exp follow-up studies/
65. exp prospective studies/
66. or/43-65

67. 42 and 66
Appendix C: Rate risk of bias, Systematic Review 1

Tool development

Departing from the Navigation Guide risk of bias tool, we first selected the five relevant risk of bias domains (i.e., selection bias, performance bias, exposure measurement, conflict of interest, and other bias), and we adopted or adapted the relevant considerations and criteria for ratings. Second, we integrated key considerations from the GATHER Guidelines and existing checklists for prevalence studies. Third, we integrated relevant criteria for ratings for the exposure measurement domain from the US Office of Health Assessment and Translation / National Toxicology Program risk of bias tool. Finally, systematic review methodologists from all systematic reviews in the series jointly improved the preliminary tool, and they will further refine and test it over the course of the systematic reviews. If a more suitable method becomes available over the course of the systematic review, then we may switch to it.

Instructions

- Please evaluate each individual study for the following five risk of bias domains, indicated by one key question each.
- For each risk of bias domain, please rate the risk as “low”; “probably low”; “probably high”; “high”; or “not applicable”.
- For each risk of bias domain, please provide a justification for your rating.
- If there is empirical evidence or other knowledge that informs the direction of bias, please include this in your answer. However, if there is not enough information to robustly indicate direction of a potential bias, please do not guess the direction of the bias.
- Some internal validity issues could potentially be appropriately captured in considerations for several different risk of bias domains. In this situation, please select the single most appropriate domain to evaluate this potential bias, to avoid double-counting the same internal validity concern.

1. Is there a risk that exposures captured in the study sample do not represent exposures in the target population in a manner that might introduce selection bias?

The target population is defined as the population, for which study investigators aim to assess exposures.

Examples of considerations for this risk of bias domain include:

1. the study sample is an adequate representation of the target population
2. the study sample and the target population have similar characteristics (e.g., sociodemographic characteristics, occupation and disease status)
3. participant inclusion and exclusion were appropriately defined
4. the sampling frame is representative of the target population
5. either the study sample was drawn at random, or a census was undertaken
6. the proportion of persons invited to participate in the study who did participate in the study was acceptable
7. the proportion of persons invited to participate in the study who did participate in the study was comparable across exposure levels
8. the reasons for non-participation in the study were acceptable
9. the proportion of study participant who participated in the exposure assessment was acceptable
10. the proportion of study participant who participated in the exposure assessment was comparable across exposure levels
11. the reasons for non-participation in exposure assessment were acceptable

If feasible, please also access and consider information reported in other study records from the study.

Criteria for a rating of LOW risk of bias (i.e., the answer to Question 1 above is: “No”):

a) The descriptions of the target population, inclusion/exclusion criteria, recruitment and enrollment procedures (including sampling frame), participation/response rates are sufficiently detailed, and adequate data were supplied on the distribution of relevant study sample and population characteristics, so that you judge the risk of selection effects to be low.

OR

b) The descriptions and/or data as indicated in “a)” above do suggest the potential for selection effects. However, adequate information was given that you can judge any potential selection effects as not differential across sub-groups defined by exposure levels.

OR

c) The descriptions and/or data as indicated in “a)” above do suggest the potential for selection effects, and there was no information suggesting that potential selection effects were not differential across sub-groups defined by exposure levels. However, drivers of selection effects were well-understood, these drivers of selection effects were measured in the data set, and appropriate post-hoc statistical methods were used to control for potential selection bias.

Criteria for the rating of PROBABLY LOW risk of bias (i.e., the answer to Question 1 above is: “Probably No”):

There is insufficient information about participant selection to permit you to judge the risk of bias to be low. However, indirect evidence suggests that inclusion/exclusion criteria, recruitment and enrollment procedures, and participation/response rates were consistent across groups as described by the criteria for a rating of low risk of bias.

Criteria for the rating of PROBABLY HIGH risk of bias (i.e., the answer to Question 1 above is: “Probably Yes”):

There is insufficient information about participant selection to permit you to judge the risk of bias to be high. However, indirect evidence suggests that inclusion/exclusion criteria, recruitment and enrollment procedures, and participation/response rates were inconsistent across groups, as described by the criteria for a rating of high risk of bias.

Criteria for the rating of HIGH risk of bias (i.e., i.e., the answer to Question 1 above is: “Yes”):
a) The descriptions of the target population, inclusion/exclusion criteria, recruitment and enrollment procedures, participation/response rates and/or data on the distribution of relevant study sample and population characteristics suggest that the risk of selection effects was substantial;

AND

b) No information was provided to indicate that potential selection effects were *not* differential across sub-groups defined by exposure level;

AND

c) No appropriate post-hoc techniques were used to control for potential selection bias.

Criteria for the rating of NOT APPLICABLE (Question 1 is not applicable to the study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

2. **Were exposure assessors and other study personnel blinded to relevant participant characteristics (e.g. occupation and/or disease status) or could exposure assessors’ or other study personnel’s prior knowledge of participant characteristics have influenced the exposure assessment in a manner that might introduce performance bias?**

Criteria for a rating of LOW risk of bias (i.e., the answer to Question 2 above is: “No”):

a) Exposure assessors and study personnel were blinded to relevant participant characteristics (e.g. occupation and/or disease status), and the blinding was probably not broken.

OR

b) Exposure assessors and study personnel were not blinded to relevant participant characteristics. However, this lack of blinding is unlikely to have influenced the exposure measurement (e.g., it was likely that exposure was systematically assessed similarly across sub-groups defined by participant characteristic);

Criteria for the rating of PROBABLY LOW risk of bias (i.e., the answer to Question 2 above is: “Probably No”):

The information on blinding is insufficient to permit a rating of low risk of bias. However, indirect evidence suggests that the exposure assessors and study personnel were adequately blinded, as described by the criteria for a rating of low risk of bias.

Criteria for the rating of PROBABLY HIGH risk of bias (i.e., the answer to Question 2 above is: “Probably Yes”):

The information on blinding is insufficient to permit a rating of high risk of bias. However,
indirect evidence suggests that the exposure assessors and study personnel were not adequately blinded, as described by the criteria for a rating of high risk of bias.

Criteria for the rating of HIGH risk of bias (i.e., the answer to Question 2 above is: “Yes”):

- Exposure assessors and study personnel were not at all blinded or incompletely blinded, and the exposure measures are likely to be influenced by the lack of blinding (e.g., exposure was systematically assessed differentially for sub-groups defined by participant characteristics).

OR

- Blinding of exposure assessors and study personnel was attempted, but may have been broken, and the breaking of the blinding may have introduced bias.

Criteria for the rating of NOT APPLICABLE (Question 2 is not applicable to the study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

3. Is there a risk that the methods for assessing exposure might consistently over- or under-estimate exposure in a manner that might introduce misclassification bias?

The following list of potential considerations is a collection of factors that may potentially systematically influence the internal validity of the exposure assessment (not those that may randomly affect overall study results). These considerations should be interpreted only as suggested ones and not as a scoring or a checklist.

**List of potential considerations:**

Quality of exposure measurement:

1. Was the case definition used in the exposure measurement appropriate and consistently applied?
2. Was the exposure self-reported by the participant for whom the exposure was measured?
3. In self-reported exposure measurement, were study participants aware of their own health status?
4. Was the exposure measure in other ways dependent on study participants’ characteristics (e.g. occupation and/or disease status)?
5. Was the exposure assessment based on individual exposure measurements (e.g. personal dosimetry) or aggregate environmental/ambient/area exposure assessment?
6. Was the exposure assessment influenced by non-occupational exposure (e.g., the environmental concentration of an occupational risk factor)?
7. Did the exposure assessment appropriately take into account the adoption of protective gear and other preventive strategies?
8. If technical devices were used for exposure assessment, did the technical requirements of the devices used to perform the exposure measurement meet specific and established standards (e.g. devices were properly calibrated)?
9. Has the exposure measurement been validated for relevant populations and settings?

Data collection:
10. Were data collected directly from the participants (as opposed to by proxy)?
11. Was the same data collection mode used for measuring exposure among all study participants?

Sample size

12. Was the size of the total study sample appropriate?
13. Was the size of the study sub-samples for each exposure level was appropriate

Validity of prevalence calculation:

14. Was the length of the shortest prevalence period for the assessed exposure appropriate (e.g. the shortest prevalence period covered was at least one full working day)?
15. Were the numerator and denominator for the prevalence estimate appropriate?

Missing data:

16. Were missing data appropriately accounted for (e.g., missing data were appropriately imputed)?

Criteria for a rating of LOW risk of bias (i.e., the answer to Question 3 above is: “No”):

The reviewers judge that there is low risk of exposure misclassification, i.e.:

   a) There is high confidence in the accuracy of the exposure measurement method (e.g., the methods have been tested for validity and reliability in measuring the targeted exposure). In other words, *direct* evidence suggests that exposure was consistently assessed (i.e., under the same method and time-frame) using well-established methods that directly measure exposure; the exposure measurement assessed relevant levels of the exposure over relevant reporting periods.

   OR

   b) Less-established or less direct exposure measurements are validated against well-established or direct methods. For example, the exposure was assessed using less-established methods that directly measure exposure, but the methods are validated against well-established methods; exposure was assessed for a relevant reporting period; the exposure measurement assessed relevant levels of the exposure over relevant reporting periods.

Criteria for the rating of PROBABLY LOW risk of bias (i.e., the answer to Question 3 above is: “ Probably No”):

There is insufficient information about the exposure measurement methods to permit a rating of low risk of bias. However, there is *indirect* evidence that exposure measurement methods were accurate, as described by the criteria for a rating of low risk of bias:

   a) *Indirect* evidence suggests that the exposure was consistently assessed using well-established methods that directly measure exposure

   OR
b) Exposure was assessed using indirect measures (e.g., questionnaire or occupational exposure measurement by a certified occupational health and safety professional) that have been validated or empirically shown to be consistent with methods that directly measure exposure (i.e., inter-methods validation: one method vs. another). And the exposure measurement assessed relevant levels of the exposure over relevant reporting periods.

Criteria for the rating of PROBABLY HIGH risk of bias (i.e., the answer to Question 3 above is: “Probably Yes”):

There is insufficient information about the exposure measurement methods to permit a rating of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a rating of high risk of bias:

a) Indirect evidence that the exposure was assessed using poorly validated methods that directly measure exposure.

OR

b) There is evidence that the exposure was assessed using indirect measures that have not been validated or empirically shown to be consistent with methods that directly measure exposure (e.g., questionnaire, job-exposure matrix or self-report without validation).

OR

c) Insufficient information is provided about the exposure measurement method, including validity and reliability, but there is no evidence for concern about the exposure measurement method used.

Criteria for the rating of HIGH risk of bias (i.e., the answer to Question 3 above is: “Yes”):

- Direct evidence suggests that the exposure was assessed using methods with poor validity.

OR

- Evidence suggests exposure misclassification (e.g., differential recall of self-reported exposure).

OR

- It is unclear how exposure measurement was obtained.

Criteria for the rating of NOT APPLICABLE (Question 3 is not applicable to the study):

There is evidence that exposure measurement methods are not capable of introducing risk of bias
4. Did the study receive any support from a company, study author or other entity with a potential financial interest in the exposures assessed?

Criteria for a judgment of LOW risk of bias: Financial conflicts of interest are defined per the July 2010 version of the International Committee of Medical Journal Editors uniform disclosure form for potential conflicts of interest and included: concurrent or former board membership, concurrent or former consultancy work, concurrent or former industry employment, expert testimony, industry grants (issued or pending), payment for lectures including service on speakers bureaus, payment for manuscript preparation, patents (planned, pending, or issued), royalties, payment for development of educational presentations, stock or stock options, and travel reimbursement, or other relations with relevant industries (ICMJE 2010).

Criteria for a rating of LOW risk of bias (i.e., the answer to Question 4 above is: “No”):

The study did not receive support from a company, study author or other entity that had a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations or academic grants funded by government, foundations and/or non-profit organizations;
- Treatment used in the study (if any) was purchased from a supplier;
- Company-affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects, for which there is a potential conflict of interest).

Criteria for the rating of PROBABLY LOW risk of bias (i.e., the answer to Question 4 above is: “Probably No”):

There is insufficient information to permit a rating of low risk of bias, but there is indirect evidence which suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a rating of low risk of bias.

Criteria for the rating of PROBABLY HIGH risk of bias (i.e., the answer to Question 4 above is: “Probably Yes”):

There is insufficient information to permit a rating of high risk of bias, but there is indirect evidence which suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a rating of high risk of bias.
Criteria for the rating of HIGH risk of bias (i.e., the answer to Question 4 above is: “Yes”):

The study received support from a company, study author or other entity that had a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals, equipment or testing provided at no cost;
- Writing services;
- Author/staff from the study was an employee or otherwise affiliated with a company with a financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the rating of NOT APPLICABLE (Question 4 is not applicable to the study):

There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

5. Did the study appear to have other problems that could put it at a risk of bias?

Criteria for a rating of LOW risk of bias (i.e. the answer to Question 5 above is: “No”):

The study appears to be free of other sources of bias.

Criteria for the rating of PROBABLY LOW risk of bias (i.e., the answer to Question 5 above is: “Probably No”):

There is insufficient information to permit a rating of low risk of bias, but there is indirect evidence which suggests the study was free of other threats to validity.

Criteria for the rating of PROBABLY HIGH risk of bias (i.e., the answer to Question 5 above is: “Probably Yes”):

There is insufficient information to permit a rating of high risk of bias, but there is indirect evidence which suggests the study was not free of other threats to validity, as described by the criteria for a rating of high risk of bias.

Criteria for the rating of HIGH risk of bias (i.e., the answer to Question 5 above is: “Yes”):

There is at least one important other risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used;
- Stopped early due to some data-dependent process (including a formal-stopping rule);
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a sub-group showing greater or lesser effect);
- Has been claimed to have been fraudulent; and/or
- Had some other risk of bias.
Appendix D: Instructions for grading the quality of evidence, Systematic Reviews 1 and 2

Most of the text from these instructions and criteria for judging risk of bias has been adopted verbatim or adapted from one of the latest Navigation Guide systematic reviews.  

A. Grading Quality

Each of the categories to consider in downgrading or upgrading the evidence is described in detail below. Please record your results on the chart at the end of each category, including a brief explanation for your ratings.

Category 1. Quality of Study Limitations (Risk of Bias)

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

The evidence from studies can be rated down if most of the relevant evidence comes from studies that suffer from a high risk of bias. Risk of bias is rated by outcome across studies. Study limitations for each outcome for individual studies and across studies are summarized in the heat maps. GRADE outlines the following principles for moving from risk of bias in individual studies to rating quality of evidence across studies.

1. In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies is warranted.

2. This judicious consideration requires evaluating the extent to which each study contributes toward the estimate of magnitude of effect. The contribution that each study makes will usually reflect study sample size and number of outcome events. Larger studies with many events will contribute more, much larger studies with many more events will contribute much more.

3. One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.

---

1 A Note: Limitations to GRADE's risk of bias assessments as stated by GRADE: "First, empirical evidence supporting the criteria is limited. Attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain. The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE’s approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items.”
4. The risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), GRADE suggests rating down for at least one of the two.

5. Notwithstanding the first four principles, reviewers will face close-call situations. You should acknowledge that you are in such a situation, make it explicit why you think this is the case, and make the reasons for your ultimate judgment apparent.

<table>
<thead>
<tr>
<th>Rating for Risk of Bias (Study Limitations)</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>-1 decrease quality 1 level</td>
<td></td>
</tr>
<tr>
<td>-2 decrease quality 2 levels</td>
<td></td>
</tr>
</tbody>
</table>

**Category 2. Indirectness of Evidence**

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Quality of evidence (your confidence in estimates of effect) may decrease when substantial differences exist between the population, exposure, or outcomes measured in the research studies under consideration in the review.

Evidence is direct when it directly compares the exposures in which we are interested in the populations in which we are interested and measures outcomes important to the study question (in GRADE the outcomes must be important to patients).

Based on GRADE (Guyatt et al. 2011), evidence can be indirect in one of three ways.

1. The population studied differs from the population of interest (the term applicability is often used for this form of indirectness). GRADE states that in general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different than the population tested that the magnitude of effect will differ substantially. According to GRADE, most often, this will not be the case.

2. The intervention (exposure) tested may differ from the exposure of interest, i.e., a difference in the chemical, route and/or dose. Decisions regarding indirectness of populations and exposure depend on an understanding of whether biological or social factors are sufficiently different that one might expect substantial differences in the magnitude of effect. GRADE also states, “As with all other aspects of rating quality of evidence, there is a continuum of similarity of the intervention that will require judgment. It is rare, and usually unnecessary, for the intended populations and interventions to be identical to those in the studies, and we should not rate down for this reason.”

---

2 GRADE includes a fourth type of indirectness that occurs when there are no direct (i.e., head-to-head) comparisons between two or more interventions of interest. This criterion is not relevant to our study question; it could be relevant to future case studies.
only rate down if the differences are considered sufficient to make a difference in outcome likely.”

3. Outcomes may differ from those of primary interest; for instance, surrogate outcomes that are not themselves important, but measured in the presumption that changes in the surrogate reflect changes in an important outcome. The difference between desired and measured outcomes may relate to time frame. When there is a discrepancy between the time frame of measurement and that of interest, whether to rate down by one or two levels will depend on the magnitude of the discrepancy. Another source of indirectness related to measurement of outcomes is the use of substitute or surrogate endpoints in place of the exposed population’s important outcome of interest. In general, the use of a surrogate outcome requires rating down the quality of evidence by one, or even two, levels. Consideration of the biology, mechanism, and natural history of the disease can be helpful in making a decision about indirectness. Surrogates that are closer in the putative causal pathway to the adverse outcomes warrant rating down by only one level for indirectness. GRADE states that rarely, surrogates are sufficiently well established that one should choose not to rate down quality of evidence for indirectness. In general, evidence based on surrogate outcomes should usually trigger rating down, whereas the other types of indirectness will require a more considered judgment.

<table>
<thead>
<tr>
<th>Rating for Indirectness</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>-1 decrease quality 1 level</td>
<td></td>
</tr>
<tr>
<td>-2 decrease quality 2 levels</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
</tbody>
</table>

**Category 3. Inconsistency of Evidence**

Possible ratings: 0 = no change; -1 or -2 downgrade 1 or 2 levels

According to Cochrane, “when studies yield widely differing estimates of effect (heterogeneity or variability in results) investigators should look for robust explanations for that heterogeneity. …When heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of the evidence decreases.”

Based on GRADE (Guyatt et al. 2011), a body of evidence is not rated up in quality if studies yield consistent results, but may be rated down in quality if inconsistent. Their stated reason is that a consistent bias will lead to consistent, spurious findings.

GRADE suggests rating down the quality of evidence if large inconsistency (heterogeneity) in study results remains after exploration of a priori hypotheses that might explain heterogeneity. Judgment of the extent of heterogeneity is based on similarity of point estimates, extent of overlap of confidence intervals, and statistical criteria. GRADE’s recommendations refer to inconsistencies in effect size, specifically to relative measures (risk ratios and hazard ratios or odds ratios), not absolute measures.

Based on GRADE, reviewers should consider rating down for inconsistency when:
1. Point estimates vary widely across studies;

2. Confidence intervals show minimal or no overlap;

3. The statistical test for heterogeneity—which tests the null hypothesis that all studies in a meta-analysis have the same underlying magnitude of effect—shows a low P-value;

4. The $I^2$—which quantifies the proportion of the variation in point estimates due to among-study differences—is large. (I.e., the $I^2$ index quantifies the degree of heterogeneity in a meta-analysis).

GRADE states that inconsistency is important only when it reduces confidence in results in relation to a particular decision. Even when inconsistency is large, it may not reduce confidence in results regarding a particular decision. For example, studies that are inconsistent related to the magnitude of a beneficial or harmful effect (but are in the same direction) would not be rated down; in instances when results are inconsistent as to whether there is a benefit or harm of treatment, GRADE would rate down the quality of evidence as a result of variability in results, because the meaning of the inconsistency is so relevant to the decision to treat or not to treat.

<table>
<thead>
<tr>
<th>Rating for Inconsistency</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>-1 decrease quality 1 level</td>
<td></td>
</tr>
<tr>
<td>-2 decrease quality 2 levels</td>
<td></td>
</tr>
</tbody>
</table>

**Category 4. Imprecision of Evidence**

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

Cochrane states that when studies have few participants and few events, and thus have wide confidence intervals, authors can lower their rating of the quality of evidence. These ratings of precision are made as judgments by review authors. The ratings are made by looking across studies, or, if available, on the results of a meta-analysis.

GRADE defines evidence quality differently for systematic reviews and guidelines. For systematic reviews, quality refers to confidence in the estimates of effect. For guidelines, quality refers to the extent to which confidence in the effect estimate is adequate to support a particular decision (Guyatt et al. 2011). For the purpose of step 3 of Navigation Guide, we will use the systematic review definition, because the decision phase does not occur until step 4 when recommendations for prevention are made. Thus, when reviewing the data for imprecision, evaluate your confidence in the estimate of the effect.

According to GRADE, to a large extent, CIs inform the impact of random error on evidence quality. Thus, when considering imprecision, the issue is whether the CI around the estimate of exposure effect is sufficiently narrow. If it is not, GRADE rates down the evidence quality by one level (for instance,
from high to moderate). If the CI is very wide, GRADE might rate down by two levels.

<table>
<thead>
<tr>
<th>Rating for Imprecision</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>-1 decrease quality 1 level</td>
<td></td>
</tr>
<tr>
<td>-2 decrease quality 2 levels</td>
<td></td>
</tr>
</tbody>
</table>

**Category 5. Publication Bias**

Possible ratings: 0 = no change; -1 or -2 downgrade 1 or 2 levels

GRADE (Guyatt et al. 2011) and Cochrane (Higgins and Green 2011) assess publication bias in a similar manner. Whereas “selective outcome reporting” is assessed for each study included in the review as part of the risk of bias assessment, “publication bias” is assessed on the body of evidence. GRADE states that “when an entire study remains unreported and the results relate to the size of the effect- publication bias- one can assess the likelihood of publication bias only by looking at a group of studies.”

Cochrane’s definition of publication bias is “the publication or non-publication of research findings depending on the nature and direction of the results.” Cochrane and GRADE are primarily concerned with overestimates of true effects of treatments or pharmaceuticals, especially related to “small studies effects”, i.e., the tendency for estimates of an intervention to be more beneficial in smaller studies. There is empirical evidence in the clinical sciences that publication and other reporting biases result in over estimating the effects of interventions (Higgins and Green 2011).

In contrast, in environmental health, we are primarily concerned with underestimating the true effects of a chemical exposure, since in many cases population wide exposure has already occurred. We are also concerned that studies finding no association are less likely to be published because journals are less likely to publish “negative” findings.

Applying this inverted concern to GRADE’s assessment for publication bias, leads to these considerations when rating publication bias:

- Early negative studies, particularly if small in size, are suspect. (GRADE is concerned with early positive studies).
- Authors of systematic reviews should suspect publication bias when studies are uniformly small, particularly when sponsored by the industry. (Same as GRADE)
- Empirical examination of patterns of results (e.g., funnel plots) may suggest publication bias but should be interpreted with caution. (Same as GRADE)
- More compelling than any of these theoretical exercises is authors’ success in obtaining the results of some unpublished studies and demonstrating that the published and unpublished data show different results. (Same as GRADE)
- Comprehensive searches of the literature including unpublished studies, i.e., the grey literature, and a search for research in other languages are important to addressing publication bias. Note that Cochrane also states “comprehensive searching is not sufficient to prevent some
substantial potential biases.”

<table>
<thead>
<tr>
<th>Rating for Publication Bias</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>-1 decrease quality 1 level</td>
<td></td>
</tr>
<tr>
<td>-2 decrease quality 2 levels</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
</tbody>
</table>

**Upgrade Categories**

GRADE states that the circumstances for upgrading likely occur infrequently and are primarily relevant to observational and other non-randomized studies. Although it is possible to rate up results from randomized controlled trials, GRADE has yet to find a compelling circumstance for doing so (Guyatt et al. 2011). GRADE specifies 3 categories for increasing the quality of evidence (Guyatt et al. 2011)

**Category 6. Large Magnitude of Effect**

Possible ratings: 0 = no change; +1 or +2 upgrade 1 or 2 levels

Modeling studies suggests that confounding (from non-random allocation) alone is unlikely to explain associations with a relative risk (RR) greater than 2 (or less than 0.5), and very unlikely to explain associations with an RR greater than 5 (or less than 0.2). Thus, these are the definitions of “large magnitude of effect” used by GRADE to upgrade 1 or 2 levels, respectively. Also, GRADE is more likely to rate up if the effect is rapid and out of keeping with prior trajectory; usually supported by indirect evidence. GRADE presents empirical evidence to support these conclusions, and states that “although further research is warranted, both modeling and empirical work suggest the size of bias from confounding is unpredictable in direction but bounded in size.

Hence, the GRADE group has previously suggested guidelines for rating quality of evidence up by one category (typically from low to moderate) for associations greater than 2, and up by two categories for associations greater than 5.”

Applying the GRADE definitions of large magnitude of effect i.e., RR greater than 2 or 5 is problematic in environmental health because for dichotomous outcomes RR is a function of the exposure comparator; these definitions also are not applicable to results from continuous variables. At present, we do not have an empirically defined “large magnitude of effect.” Therefore, for the purpose of this case study, review authors should assess whether the results indicate a large magnitude of effect using their expert judgment of “large effects” in environmental health and state their definition for discussion by the group.

<table>
<thead>
<tr>
<th>Rating for Large Magnitude of Effect</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>+1 increase quality 1 level</td>
<td></td>
</tr>
<tr>
<td>+2 increase quality 2 levels</td>
<td></td>
</tr>
</tbody>
</table>
Human

**Category 7. Dose-response**

Possible ratings: 0 = no change; +1 or +2 upgrade 1 or 2 levels

Possible considerations include consistent dose response gradients in one or multiple studies, and/or dose response across studies, depending on the overall relevance to the body of evidence.

<table>
<thead>
<tr>
<th>Rating for Dose-Response</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>+1 increase quality 1 level</td>
<td></td>
</tr>
<tr>
<td>+2 increase quality 2 levels</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
</tbody>
</table>

**Category 8. Residual Confounding Increases Confidence**

Possible ratings: 0=no change; +1 or +2 upgrade 1 or 2 levels

Upgrade if consideration of all plausible residual confounders, biases, or effect modification would underestimate the effect or suggest a spurious effect when results show no effect. If a study reports an association despite the presence of residual confounding, biases or effect modification that would diminish the association, confidence in the association is increased (National Toxicology Program 2015). GRADE provides an illustrative example related to bias: rating up observational evidence finding lack of association between vaccination and autism, which occurred despite empirically confirmed bias that parents of autistic children may be more likely to remember their vaccine experience. The negative findings despite this form of recall bias suggest rating up the quality of evidence (Guyatt et al. 2011).  

<table>
<thead>
<tr>
<th>Rating for Residual Confounding Increases Confidence</th>
<th>Rationale for your judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 no change</td>
<td></td>
</tr>
<tr>
<td>+1 increase quality 1 level</td>
<td></td>
</tr>
<tr>
<td>+2 increase quality 2 levels</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
</tr>
</tbody>
</table>

The results of the reviewers’ ratings by population will be compiled and discussed leading to a final decision on overall quality of human evidence. The rationale for the decision will be fully documented.

**Final decision on overall quality of human evidence:**

(Example: Moderate quality is upgraded 1 step to high for XYZ reason(s))

---- High
--- Moderate

--- Low
Appendix E: Rate the strength of evidence, Systematic Review 1

The strength of evidence for use as input data will be rated based on a combination of four criteria: (1) quality of the entire body of evidence; (2) population coverage of evidence (WHO regions and countries); (3) confidence in the entire body of evidence; and (4) other compelling attributes of the evidence that may influence certainty. The strength of evidence ratings are summarized below, where their meaning is further defined.

<table>
<thead>
<tr>
<th>Potentially sufficient evidence for use as input data</th>
<th>The body of evidence was rated as being of high or moderate quality; it covers at least one country each from at least two WHO regions; the review authors had confidence in the feasibility of using the evidence as input data; and there are no compelling attributes of the evidence that may reduce certainty in it.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potentially insufficient or inadequate evidence for use as input data</td>
<td>The body of evidence was rated as of low quality; it covered fewer than one country each from at least two WHO regions; the review authors were very uncertain about the feasibility of using the evidence as input data or had no confidence in it; and/or there are compelling attributes of the evidence that may reduce certainty in it.</td>
</tr>
</tbody>
</table>
Appendix F: Overview of inclusion and exclusion criteria, Systematic Review 2

Include

1. Studies of working-age population (≥ 15 year), including workers in the informal economy
2. Studies on populations residing in any Member States of WHO and/or ILO and any occupational group or industrial setting
3. Studies that defined ischaemic heart disease in accordance with our standard definition
4. Studies with the following designs: randomized controlled trials (including parallel-group, cluster, crossover and factorial ones), prospective and retrospective cohort studies, case-control studies and other non-randomized intervention studies that estimate the effect of long working hours on ischaemic heart disease, for any years
5. Studies with measures of the relative effect of a relevant level of long working hours on the risk of having, getting or dying from ischaemic heart disease, compared with the theoretical minimum risk exposure level of working 35-40 hours/week
6. Included measures are relative effect measures, risk ratios and odds ratios for prevalence and mortality measures and hazard ratios for incidence measures
7. Studies published in any year and any language with essential information (title and abstract) in English

Exclude

1. Studies of unpaid domestic workers
2. Studies of children (< 15 years)
3. Studies reporting on measures of absolute effects (e.g. mean differences in risks or odds), if they cannot be converted into eligible relative measures
4. Cross-sectional studies, qualitative, modelling, and case studies, as well as non-original studies without quantitative data (e.g. letters, commentaries and perspectives)
Appendix G: Ovid Medline search strategy, Systematic Review 2

Population

No search string

Exposure

1. exp "personnel staffing and scheduling"/
2. "personnel staffing and scheduling".ti,ab,kw.
3. shift work schedule.ti,ab,kw.
4. work schedule tolerance.ti,ab,kw.
5. workload.kw.
6. workday shifts.ti,ab,kw.
7. overwork*.ti,ab,kw.
8. overtime.ti,ab,kw.
9. workweek*.ti,ab,kw.
10. (work* adj3 hour*).ti,ab,kw.
11. (work* adj3 schedul*).ti,ab,kw.
12. work* ad3 roster.ti,ab,kw.
13. (work* adj3 organis#ation).ti,ab,kw.
14. (work* adj3 time*).ti,ab,kw.
15. (work* adj3 overload*).ti,ab,kw.
16. (work* adj3 extend*).ti,ab,kw.
17. (work* adj3 compress*).ti,ab,kw.
18. (work* adj3 week*).ti,ab,kw.
19. (work* adj3 day?).ti,ab,kw.
20. (job? adj3 hour*).ti,ab,kw.
21. (job? adj3 schedul*).ti,ab,kw.
23. (job? adj3 organis#ation).ti,ab,kw.
24. (job? adj3 time*).ti,ab,kw.
25. (job? adj3 overload*).ti,ab,kw.
26. (job? adj3 extend*).ti,ab,kw.
27. (job? adj3 compress*).ti,ab,kw.
28. (job? adj3 week*).ti,ab,kw.
29. (job? adj3 day?).ti,ab,kw.
30. (shift? adj3 hour*).ti,ab,kw.
31. (shift? adj3 schedul*).ti,ab,kw.
32. (shift? adj3 roster).ti,ab,kw.
33. (shift? adj3 organis#ation).ti,ab,kw.
34. (shift? adj3 time*).ti,ab,kw.
35. (shift? adj3 overload*).ti,ab,kw.
36. (shift? adj3 extend*).ti,ab,kw.
37. (shift? adj3 compress*).ti,ab,kw.
40. (work* and (life* or live*) and (balances* or imbalances* or unbalances* or interference*)).ti,ab,kw.
41. (work* and famil* and conflict*).ti,ab,kw.
42. or/1-41

Study design

43. exp Clinical Trial/
44. trial$.tw.
45. experiment$.tw
46. (intervention adj3 (study or studies or analys$)).tw.
47. Epidemiologic Studies/
48. Observational Study/
49. ((observational or epidemiologic$) adj (study or studies or analys$)).tw.
50. exp Cohort Studies/
51. cohort$.tw.
52. (panel$ adj3 (study or studies or analys$ or data)).tw.
53. (follow up adj (study or studies or analys$)).tw.
54. (repeat$ adj measure$).tw.
55. longitudinal$.tw.
56. retrospective$.tw.
57. exp Case Control Studies/
58. (case$ adj3 control$).tw.
59. (exposure$ adj4 (study or studies or analys$)).tw.
60. before-after.tw
61. pre-post.tw
62. nonexperimental.ti,ab.
63. non-experimental.ti,ab.
64. nonrandomized.ti,ab.
65. nonrandomised.ti,ab.
66. non-randomized.ti,ab.
67. non-randomised.ti,ab.
68. or/43-67

Outcome

69. exp Myocardial Ischemia/
70. Heart Diseases/
71. angina.tw.
72. (heart adj3 disease$).tw.
73. (coronary adj3 disease$).tw.
74. (isch?emic adj disease*).tw.
75. myocardial infarct$.tw.
76. (heart adj3 infarct$).tw.
77. ((artery occlusion* or artery disease* or arterioscleros* or atheroscleros*) adj2 coronary).ti,ab.
78. (IHD or CHD or CAD or AMI or MI).ti,ab.
79. exp Myocardial Revascularization/
80. (coronary adj3 bypass$).tw.
81. (coronary adj3 angioplast$).tw.
82. cabg.tw.
83. ptca.tw.
84. or/69-83
85. 42 and 68 and 84
Appendix H: Instructions for making risk of bias determinations, Systematic Reviews 2

Most of the text from these instructions and criteria for judging risk of bias has been adopted verbatim or adapted from one of the latest Navigation Guide systematic reviews.

Instructions:

Please evaluate each individual study for the following nine risk of bias domains. Please answer “low risk,” “probably low risk,” “probably high risk,” “high risk,” or “not applicable” and provide details/justification for your rating. If there is empirical evidence or other knowledge that informs the direction of bias, please include this in your answer as well; however, if there is not enough information to do so please do not guess at the direction of bias.

Additionally, please note that some internal validity issues could potentially be appropriately captured in several different risk of bias considerations. In this situation, please select the single most appropriate domain to evaluate this potential bias, to avoid double-counting the same internal validity concern.

1. Are the study groups at risk of not representing their source populations in a manner that might introduce selection bias?

The source population is viewed as the population for which study investigators are targeting their study question of interest. Examples of considerations for this risk of bias domain include: 1) level of detail reported for participant inclusion/exclusion (including details from previously published papers referenced in the article for an existing cohort); 2) participation rates and whether this differed by exposure or outcome group; 3) attrition rates and reasons; and 4) comparisons of study characteristics between the study population and full cohort.

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”): EITHER:

a) The descriptions of the source population, inclusion/exclusion criteria, recruitment and enrollment procedures, participation and follow-up rates were sufficiently detailed, and adequate data were supplied on the distribution of relevant study sample and population characteristics to support the assertion that risk of selection effects was minimal.

OR

b) Although the descriptions and/or data as indicated in “a” above suggested the potential for selection effects, adequate support was given indicating that potential selection effects were not differential across both exposure and outcome.

OR

c) Although the descriptions and/or data as indicated in “a” above suggested the potential for selection effects and there was no support indicating that potential selection effects were not differential across both exposure and outcome, selection factors appeared to be well-understood, were measured in the data set, and appropriate adjustment post hoc techniques were used to control for selection bias.
Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information about participant selection to permit a judgment of low risk of bias, but there is indirect evidence which suggests that inclusion/exclusion criteria, recruitment and enrollment procedures, and participation and follow-up rates were consistent across groups as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information about participant selection to permit a judgment of high risk of bias, but there is indirect evidence which suggests that inclusion/exclusion criteria, recruitment and enrollment procedures, and participation and follow-up rates were inconsistent across groups, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

d) There were indications from descriptions of the source population, inclusion/exclusion criteria, recruitment and enrollment procedures, participation and follow-up rates, or data on the distribution of relevant study sample and population characteristics that risk of selection effects were substantial; and
e) There was no support to indicate that potential selection effects were not differential across both exposure and outcome; and
f) Adjustment post hoc techniques were not used to control for selection bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that participant selection is not an element of study design capable of introducing risk of bias in the study.

3. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”): Any of the following:

- No blinding, but the review authors judge that the outcome measures as well as the exposure measures are not likely to be influenced by lack of blinding (such as differential outcome assessment where the outcome is assessed using different measurement or estimation metrics across the exposure groups, or differential exposure assessment where exposure is assessed using different measurement or estimation metrics across the diagnostic or outcome groups); or
- Blinding of key study personnel was ensured, and it is unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but exposure and outcome assessment was blinded and the non-blinding of others is unlikely to introduce bias.
Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information about blinding to permit a judgment of low risk of bias, but there is indirect evidence which suggests the study was adequately blinded, as described by the criteria for a judgment of low risk of bias. For example, investigators were effectively blinded to the exposure and/or outcome groups if the exposure was measured by a separate entity and the outcome was obtained from a hospital record.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information about blinding to permit a judgment of high risk of bias, but there is indirect evidence which suggests the study was not adequately blinded, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”): Any of the following:

- No blinding or incomplete blinding, and the outcome measures or exposure measures is likely to be influenced by lack of blinding (i.e. differential outcome or exposure assessment); or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken so as to introduce bias; or
- Some key study personnel were not blinded, and the non-blinding of others was likely to introduce bias.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

4. Were exposure assessment methods lacking accuracy?

The following list of considerations represents a collection of factors proposed by experts in various fields that may potentially influence the internal validity of the exposure assessment in a systematic manner (not those that may randomly affect overall study results). These should be interpreted only as suggested considerations, and should not be viewed as scoring or a checklist.

List of Considerations:

Possible sources of exposure assessment metrics:

1) Official Records (Ministry of Labour or other official sources)
2) Organization
3) Self reported
4) Combination of the above options

For each, overall considerations include:

1) What is the quality of the source of the metric being used?
2) Is the exposure measured in the study a surrogate for the exposure?
3) What was the temporal coverage (i.e. short or long-term exposure)?
4) Did the analysis account for prediction uncertainty?
5) How was missing data accounted for, and any data imputations incorporated?
6) Were sensitivity analyses performed?

In particular, for exposure assessment models:
1) Were the input data in the study suspected to systematically under- or over-
estimate exposure?
2) What type of model was used?
3) What was geographic/spatial accuracy (county, census tract, organization, individual residence)?
4) What was the temporal specificity and variation?
5) What was the space-time coverage of the model?
6) Were time-activity patterns accounted for?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):
The reviewers judge that there is low risk of exposure misclassification, i.e.:
- There is high confidence in the accuracy of the exposure assessment methods, such as methods that have been tested for validity and reliability in measuring the targeted exposure.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):
There is insufficient information about the exposure assessment methods to permit a judgment of low risk of bias, but there is indirect evidence that suggests that methods were robust, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):
There is insufficient information about the exposure assessment methods to permit a judgment of high risk of bias, but there is indirect evidence that suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):
The reviewers judge that there is high risk of exposure misclassification and any one of the following:
- There is low confidence in the accuracy of the exposure assessment methods;
or
- Less-established or less direct exposure measurements are not validated and are suspected to introduce bias that impacts the outcome assessment;
or
- Uncertain how exposure information was obtained.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):
There is evidence that exposure assessment methods are not capable of introducing risk of bias in the study.

5. Were outcome assessment methods lacking accuracy?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

The reviewers judge that there is low risk of outcome misclassification, i.e.:

- Outcomes were assessed and defined consistently across all study participants, using valid and reliable measures (all non-fatal or fatal ischaemic heart disease events (ICD-10 codes I20–I25) with solid medical records). Note that all outcome assessment measures captured in the PECO statement are considered beforehand to be valid and reliable, unless other information provided within the study warrants a consideration otherwise; or
- Less-established or less direct outcome measurements are validated against well-established or direct methods; or
- Appropriate sensitivity analyses were conducted that suggest the influence of outcome misclassification would be minimal
- AND, if applicable, appropriate QA/QC for methods is described and is satisfactory.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information about the outcome assessment methods to permit a judgment of low risk of bias, but there is indirect evidence which suggests that methods were robust, as described by the criteria for a judgment of low risk of bias. Appropriate QA/QC for methods are not described but the review authors judge that the outcome and the outcome assessment are objective and uniform across study groups.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information about the outcome assessment methods to permit a judgment of high risk of bias, but there is indirect evidence which suggests that methods were not robust, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

The reviewers judge that there is high risk of outcome misclassification and any one of the following:

- There is low confidence in the accuracy of the outcome assessment methods; or
- Less-established or less direct outcome measurements are not validated and are suspected to introduce bias that impacts the outcome assessment
- Uncertain how outcome information was obtained
Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that outcome assessment methods are not capable of introducing risk of bias in the study.

6. Was potential confounding inadequately incorporated?

List of important potential confounders, collectively generated by review authors prior to the initiation of screening for studies based on expert opinion and knowledge gathered from the literature:

Tier I: Important confounders
- Age, sex and socioeconomic position.

Tier II: Other potentially important confounders:

None identified.

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

The study appropriately assessed and accounted for (i.e. matched, stratified, excluded certain populations or statistically controlled for) all important confounders (Tier I) using appropriate statistical techniques, or reported that important confounders were evaluated and omitted because inclusion did not substantially affect the results. The determination of specific confounders may also be informed by, but not limited to, the studies included in the overall review, and the study appropriately assessed and accounted for (i.e. matched, stratified, or statistically controlled for) other potentially important confounders relevant (Tier II) using appropriate statistical techniques, or reported that these confounders were evaluated and omitted because inclusion did not substantially affect the results, and the important potential confounders were measured consistently across study groups using valid and reliable methods, or the influence of covariate measurement error was determined, through sensitivity analysis, to be minimal.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

The study appropriately accounted for most but not all of the important confounders (Tier I) or used appropriate statistical techniques; and some of the other potentially important confounders relevant (Tier II) using appropriate statistical techniques, or reported that these confounders were evaluated and omitted because inclusion did not substantially affect the results; and this is not expected to introduce substantial bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”): The study evaluated some but not all of the important confounders (Tier I), and some but not all of the other potentially important confounders relevant (Tier II) using appropriate statistical techniques; or OR reported that these confounders were evaluated and omitted because inclusion did not substantially affect the results; and AND this is not expected to introduce substantial bias.
II), OR used questionable statistical techniques for confounder adjustment; AND this is expected to introduce substantial bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

The study did not account for or evaluate multiple important confounders (Tier I), AND did not account for or evaluate multiple other potentially important confounders relevant (Tier II), OR the important potential confounders were inappropriately measured and/or inappropriately analyzed across study groups.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that outcome assessment methods are not capable of introducing risk of bias in the study.

7. Were incomplete outcome data inadequately addressed?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

Participants were followed long enough to obtain outcome measurements OR any one of the following:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to introduce bias); or
- Attrition or missing outcome data balanced in numbers across exposure groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a relevant impact on the exposure effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a relevant impact on the observed effect size; or
- Missing data have been imputed using appropriate methods

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information about incomplete outcome data to permit a judgment of low risk of bias, but there is indirect evidence which suggests incomplete outcome data was adequately addressed, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information about incomplete outcome data to permit a judgment of high risk of bias, but there is indirect evidence which suggests
incomplete outcome data was not adequately addressed, as described by the criteria
for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

Participants were not followed long enough to obtain outcome
measurements OR any one of the following:
- Reason for missing outcome data likely to be related to true outcome, with
either imbalance in numbers or reasons for missing data across exposure
groups; or
- For dichotomous outcome data, the proportion of missing outcomes
compared with observed event risk enough to induce biologically
relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in
means or standardized difference in means) among missing outcomes
enough to induce biologically relevant bias in observed effect size; or
- Potentially inappropriate application of imputation.

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to
study):

There is evidence that incomplete outcome data is not capable of introducing risk
of bias in the study.

8. Does the study report appear to have selective outcome reporting?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

All of the study’s pre-specified (primary and secondary) outcomes outlined in the pre-
published protocol or the published manuscript’s methods, abstract, and/or
introduction section that are of interest in the review have been reported in the pre-
specified way.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information about selective outcome reporting to permit a
judgment of low risk of bias, but there is indirect evidence which suggests the
study was free of selective reporting, as described by the criteria for a judgment of
low risk of bias. This includes if a pre-published protocol is not available but the
study’s pre-specified (primary and secondary) outcomes outlined in the published
manuscript’s methods, abstract, and/or introduction section that are of interest in
the review have been reported in the pre-specified way.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information about selective outcome reporting to permit a
judgment of high risk of bias, but there is indirect evidence which suggests the
study was not free of selective reporting, as described by the criteria for a judgment
of high risk of bias. This includes if a pre-published protocol is not available and
the study’s pre-specified (primary and secondary) outcomes outlined in the
published manuscript’s methods, abstract, and/or introduction section that are of interest in the review have not been reported in the pre-specified way.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”): Any one of the following:

- Not all of the study’s pre-specified primary outcomes (as outlined in the pre-published protocol or published manuscript’s methods, abstract, and/or introduction) have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified; or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected effect); or
- One or more outcomes of interest are reported incompletely

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

9. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;
- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information to permit a judgment of low risk of bias, but there
is indirect evidence which suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of low risk of bias.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence which suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Chemicals, equipment or testing provided at no cost;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;
- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest

Criteria for the judgment of NOT APPLICABLE (risk of bias domain is not applicable to study):

There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

10. **Did the study appear to have other problems that could put it at a risk of bias?**

Criteria for a judgment of LOW risk of bias (i.e. answer: “No”):

The study appears to be free of other sources of bias.

Criteria for the judgment of PROBABLY LOW risk of bias (i.e. answer: “Probably No”):

There is insufficient information to permit a judgment of low risk of bias, but there is indirect evidence which suggests the study was free of other threats to validity.

Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably Yes”):

There is insufficient information to permit a judgment of high risk of bias, but there is indirect evidence which suggests the study was not free of other threats to validity, as described by the criteria for a judgment of high risk of bias.

Criteria for the judgment of HIGH risk of bias (i.e. answer: “Yes”):

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used; or
• Stopped early due to some data-dependent process (including a formal-stopping rule); or
• The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
• Has been claimed to have been fraudulent; or
• Had some other problem.
Appendix I: Rate the strength of evidence, Systematic Review 2

The strength of evidence will be rated based on a combination of four criteria: (1) Quality of the entire body of evidence; (2) Direction of the effect estimate; (3) Confidence in the effect estimate; and (4) Other compelling attributes of the evidence that may influence certainty. The strength of evidence ratings are summarized below, where their meaning is further defined.

<table>
<thead>
<tr>
<th>Strength of Evidence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficient evidence of toxicity/harmfulness</td>
<td>The available evidence usually includes consistent results from well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies. For human evidence a positive relationship is observed between exposure and outcome where chance, bias, and confounding, can be ruled out with reasonable confidence.</td>
</tr>
<tr>
<td>Limited evidence of toxicity/harmfulness</td>
<td>The available evidence is sufficient to determine the effects of the exposure, but confidence in the estimate is constrained by such factors as: the number, size, or quality of individual studies, the confidence in the effect, or inconsistency of findings across individual studies. As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion. For human evidence a positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence.</td>
</tr>
<tr>
<td>Inadequate evidence of toxicity/harmfulness</td>
<td>Studies permit no conclusion about a toxic effect. The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an estimation of effects.</td>
</tr>
<tr>
<td>Evidence of lack of toxicity/harmfulness</td>
<td>The available evidence includes consistent results from well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies. For human evidence more than one study showed no effect on the outcome of interest at the full range of exposure levels that humans are known to encounter, where bias and confounding can be ruled out with reasonable confidence. The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.</td>
</tr>
</tbody>
</table>
### Parameters for estimating the national and global burden of ischemic heart disease attributable to long-working hours: Protocol for a systematic review and meta-analysis (Jian Li)

<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>Guidance</th>
<th>On page</th>
<th>Author Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identification</td>
<td>Identify the report as a systematic review, or systematic review and meta-analysis, as appropriate.</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Update</td>
<td>If the protocol is for an update of a previous systematic review, identify as such.</td>
<td>NA</td>
<td>Not an update of a previous systematic review</td>
</tr>
<tr>
<td></td>
<td>Registration</td>
<td>If registered, provide the name of the registry (e.g. PROSPERO) and registration number.</td>
<td>4</td>
<td>PROSPERO-ID: CRD42017084243</td>
</tr>
<tr>
<td></td>
<td>Authors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contact</td>
<td>Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author.</td>
<td>1-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contributions</td>
<td>Describe contributions of protocol authors and identify the guarantor of the review.</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Amendments</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>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Item</td>
<td>Guidance</td>
<td>On page #</td>
<td>Author Comments</td>
</tr>
<tr>
<td>----</td>
<td>----------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>7</td>
<td>Sources</td>
<td>Indicate sources of financial or other support for the review.</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Sponsor</td>
<td>Provide name for the review funder/s and/or sponsor/s</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Roles</td>
<td>Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol.</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

**Introduction**

| 10 | Rationale      | Describe the rationale for the review in the context of what is already known                                                             | 5-6       |                 |
|    | Objectives     | Provide an explicit statement of the question(s) the review will address, with specific reference to:                                    | 9         |                 |
|    |                | • Participants                                                                                                                            |           |                 |
|    |                | • Interventions / Exposures (as appropriate)                                                                                              |           |                 |
|    |                | • Comparisons                                                                                                                             |           |                 |
|    |                | • Outcomes                                                                                                                               |           |                 |
|    |                | • Study design                                                                                                                            |           |                 |

**Methods**

<p>| 12 | Eligibility criteria | Specify the study characteristics (e.g. PICO/PECO, study design, setting, time frame) and report characteristics (e.g. years considered, language, publication status) to be used as criteria for eligibility for the review. | 11-15     |                 |
|    | Information sources  | Describe all intended information sources (e.g. electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage.                     | 15-16     |                 |</p>
<table>
<thead>
<tr>
<th>#</th>
<th>Item</th>
<th>Guidance</th>
<th>On page #</th>
<th>Author Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Search strategy</td>
<td>Present draft of search strategy to be used for at least one electronic</td>
<td>15-16, 27-30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>database, including planned limits, such that it could be repeated.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Data management</td>
<td>Describe the mechanism(s) that will be used to manage records and data</td>
<td>15-16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>throughout the review.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Selection process</td>
<td>State the process that will be used for selecting studies (e.g. two</td>
<td>16-17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>independent reviewers) through each phase of the review (i.e. screening,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>eligibility, and inclusion in meta-analysis).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Data collection process</td>
<td>Describe planned method of extracting data from reports (e.g. piloting</td>
<td>17-18</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>forms, done independently, in duplicate), any processes for obtaining</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>and confirming data from investigators.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Data items</td>
<td>List and define all variables for which data will be sought (e.g. PICO</td>
<td>11-12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>items, funding sources), any pre-planned data assumptions and</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>simplifications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Outcomes and prioritisation</td>
<td>List and define all outcomes for which data will be sought, including</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>prioritization of main and additional outcomes, with rationale.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Bias in individual studies</td>
<td>Describe anticipated methods for assessing risk of bias of individual</td>
<td>18-20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>studies, including whether this will be done at the outcome or study</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>level, or both; state how this information will be used in data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>synthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Data synthesis criteria</td>
<td>Describe criteria under which study data will be quantitatively</td>
<td>21-22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>synthesized</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Summary measures</td>
<td>If data are appropriate for quantitative synthesis, describe planned</td>
<td>21-22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>summary measures, methods of handling data, and methods of combining</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>data from studies, including any planned exploration of consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(e.g. $I^2$, Kendall’s tau).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Item</td>
<td>Guidance</td>
<td>On page #</td>
<td>Author Comments</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>23</td>
<td>Additional analyses</td>
<td>Describe any proposed additional analyses (e.g. sensitivity or subgroup analyses, meta-regression)</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Alternative synthesis</td>
<td>If quantitative synthesis is not appropriate, describe the type of summary planned</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Meta-bias</td>
<td>Specify any planned assessment of meta-bias(es) (e.g. publication bias across studies, selective reporting within studies)</td>
<td>20-21</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Confidence in cumulative evidence</td>
<td>Describe how the strength of the body of evidence will be assessed (e.g. GRADE)</td>
<td>20-21</td>
<td></td>
</tr>
</tbody>
</table>