UNIVERSITY OF COPENHAGEN

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 depression

Rugulies, Reiner; Ando, Emiko; Ayuso-Mateos, Jose Luis; Bonafede, Michela; Cabello, Maria; Di Tecco, Cristina; Dragano, Nico; Durand-Moreau, Quentin; Eguchi, Hisashi; Gao, Junling; Garde, Anne H.; Iavicoli, Sergio; Ivanov, Ivan D.; Leppink, Nancy; Madsen, Ida E. H.; Pega, Frank; Pruss-Ustun, Annette M.; Rondinone, Bruna M.; Sørensen, Kathrine; Tsuno, Kanami; Ujita, Yuka; Zadow, Amy

Review article

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 depression

Reiner Rugulies[a,b,c,*], Emiko Ando[d], Jose Luis Ayuso-Mateos[e,f,g], Michela Bonafede[h], Maria Cabello[e,f], Cristina Di Tecco[h], Nico Dragano[i], Quentin Durand-Moreau[j,k], Hisashi Eguchi[l], Junling Gao[m], Anne H. Garde[a,b], Sergio Iavicoli[h], Ivan D. Ivanov[n], Nancy Leppink[o], Ida E.H. Madsen[a], Frank Pega[n], Annette M. Prüss-Üstün[n], Bruna M. Rondinone[h], Kathrine Sørensen[a], Kanami Tsuno[p], Yuka Ujita[o], Amy Zadow[q]

[a] National Research Centre for the Working Environment, Copenhagen, Denmark
[b] Department of Public Health, University of Copenhagen, Copenhagen, Denmark
[c] Department of Psychology, University of Copenhagen, Copenhagen, Denmark
[d] Osaka University, Suita, Osaka, Japan
[e] Department of Psychiatry, Universidad Autonoma de Madrid, Madrid, Spain
[f] Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain
[g] Instituto de Investigación Sanitaria Princesa (IIS-Princesa), Madrid, Spain
[h] Department of Environmental and Occupational Medicine, Epidemiology and Hygiene, INAIL, Monte Porzio Catone, Rome, Italy
[i] Universitätsklinikum Düsseldorf, Düsseldorf, Germany
[j] Occupational Diseases Center, University Hospital of Brest, Brest, France
[k] LABERS EA 3149, University of Brest, Brest, France
[l] Kitasato University School of Medicine, Sagamihara, Kanagawa, Japan
[m] School of Public Health, Fudan University, Shanghai, China
[n] Department of Public Health, Environmental and Social Determinants of Health, World Health Organization, Geneva, Switzerland
[o] Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International Labour Organization, Geneva, Switzerland
[p] Wakayama Medical University, Wakayama-shi, Wakayama, Japan
[q] University of South Australia, Adelaide, Australia

ARTICLE INFO

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 from depression attributable to 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 depression (Systematic Review 2), applying the Navigation Guide systematic review methodology as an organizing framework, conducting both systematic reviews in tandem and in a harmonized way.

Data sources: Separately for Systematic Reviews 1 and 2, we will search electronic academic databases for potentially relevant records from published and unpublished studies, including Medline, EMBASE, Web of Science, CISDOC and PsycINFO. 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) participants in the formal and informal economy in any WHO and/or ILO Member State, but exclude child workers (< 15 years) and unpaid domestic workers. For Systematic Review 1, we will include quantitative prevalence studies of relevant levels of occupational exposure to long working hours (i.e. 35–40, 41–48, 49–54 and ≥55 h/week) stratified by country, sex, age and industrial sector or occupation, in the years 2005–2018. 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 incidence of or mortality due to depression, compared with the theoretical minimum risk exposure level (i.e. 35–40 h/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: CRD42018085729

1. 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) (Ryder, 2017). The organizations plan to estimate the numbers of deaths and disability-adjusted life years (DALYs) that are attributable to selected occupational risk factors 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 (International Labour Organization, 2014; Prüss-Üstün et al., 2017). It will expand these 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 (Murray et al., 2004) – the proportional reduction in burden from the health outcome achieved by a reduction of exposure to the risk factor to zero – 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* (World Health Organization, 2017b).

The WHO/ILO joint methodology will include a methodology for estimating the burden of depression from occupational exposure to long working 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 working hours on depression ('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 (Descatha et al., 2018; Godderis et al., 2018; Hulshof et al., in press; Li et al., 2018; Mandrioli et al., 2018; Paulo et al., Accepted; Teixeira et al., Accepted; Tenkate et al., Accepted). 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, one Systematic Review 1 will be published.

1.1. Rationale

WHO ranks depression as the single largest contributor to non-fatal health loss worldwide, with 7.5% of all years lived with disability attributed to depression in 2015 (World Health Organization, 2017a). To consider the feasibility of estimating the burden of depression due to exposure to long working hours, and to ensure that potential estimates of burden of disease are reported in adherence with the guidelines for accurate and transparent health estimates reporting (GATHER) (Stevens et al., 2016), 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 of studies with estimates of the relative effect of exposure to long working hours on the incidence of and mortality from depression, 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 attain this exposure level in practice (Murray et al., 2004). These data and effect estimates should be tailored to serve as parameters for estimating the burden of depression from exposure to long working hours in the WHO/ILO joint methodology.

To our knowledge, this is the first systematic review that will provide this evidence base for burden of depression attributable to long working hours. Three previous reviews have estimated the association of long working hours with risk of depressive symptoms and depression (Theorell et al., 2015; Virtanen et al., 2018; Watanabe et al., 2016). Theorell et al. reported, based on six cohort studies of high or moderate quality that there was a prospective association of long working weeks with risk of onset of depressive symptoms (Theorell et al., 2015). Using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system, they assessed the evidence as "limited" for women and "very limited" for men. The authors refrained from upgrading the evidence level for long working weeks, because they found the estimates of the association of long working weeks and depression neither consistent, nor large enough for qualifying for an upgrade, and they also did not conduct a meta-analysis of the included effect estimates. In another systematic review, Watanabe et al. examined overtime work and risk of onset of depressive disorders and identified seven cohort studies (Watanabe et al., 2016). The meta-analysis conducted in this systematic review showed an increased, but not statistically significant association of overtime work with risk of depressive disorders (relative risk 1.24; 95% CI 0.88 to 1.75). Virtanen et al. included in

their meta-analysis 10 published cohort studies and 18 prospective cohort studies with individual-participant data, yielding 31 study-specific estimates (as 3 studies of the published studies had provided estimates stratified by sex) (Virtanen et al., 2018). The outcome was named "depressive symptoms" and included both measures of clinical depression and depressive symptoms and of psychological distress. The overall pooled estimate (odds ratio, OR) for the association of long working hours with risk of onset of depressive symptoms was 1.14 (95% CI 1.03 to 1.25). The association was stronger in studies from Asian countries (OR = 1.50, 95% CI 1.13 to 2.01), weaker in European studies (OR = 1.11, 95% CI 1.00 to 1.22) and absent in North American studies (OR = 0.95, 95% CI 0.70 to 1.29). When stratified by clinical depression/depressive symptoms versus psychological distress, the pooled ORs were 1.09 (0.94 to 1.26) and 1.18 (1.06 to 1.32) for clinical depression/depressive symptoms and psychological distress, respectively. Meta-regressions did not show any statistically significant differences in the estimates for clinical depression, depressive symptoms and psychological distress.

The review by Virtanen et al. was published recently (Online First, 8 February 2018) and two authors of the Virtanen et al. article are also authors of this protocol (RR and IEHM). Therefore, we want to briefly delineate the main differences between the Virtanen et al. article and this protocol. First, our search is broader, Virtanen et al. searched two academic databases whereas we will search seven academic databases and two grey literature databases. Second, Virtanen et al. searched studies published until January 2017, whereas we will search for studies published until 30 June 2018. Third, the endpoint in the Virtanen et al. review was depressive symptoms, including but not limited to measures of clinical depression, and psychological stress, whereas our endpoint is restricted to clinical depression. Fourth, Virtanen et al. used the Cochrane's "Tool to Assess Risk of Bias in Cohort Studies" whereas our risk of bias assessment will be derived from the Navigation Guide (Woodruff and Sutton, 2014). Fifth, we aim to do a subgroup analysis stratified by industrial sector or occupation, if data allow this, an analysis not conducted by Virtanen et al. Sixth, Virtanen et al. did not assess the quality of evidence of the summarized results, whereas we aim to asses quality of evidence using the most suitable tools currently available (Guyatt et al., 2011; Higgins and Green, 2011; Morgan et al., 2016). We are not aware of a previous review of prevalence of exposure to long working hours. To the best of our knowledge, this is the first systematic review of parameters required for estimating the global and national burden of depression attributable to long working hours.

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" (p. 4) (104th International Labour Conference, 2015). We consider the formality of the economy studied in studies included in both Systematic Reviews.

1.2. 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 (both in main and secondary jobs) exceeding standard working hours, i.e. working hours of ≥ 41 h/week. Based on results from earlier studies on long working hours and health endpoints (Kivimäki et al., 2015a; Kivimäki et al., 2015b; Virtanen et al., 2015), the preferred four exposure level categories for our review are 35–40, 41–48, 49–54 and ≥ 55 h/week, allowing calculations of potential dose-response associations. If the studies provide the preferred exposure categories, we will use the preferred exposure categories, if they provide other exposure categories, we will use the other exposure categories, as long as exposure exceeds 40 h/week.

The theoretical minimum risk exposure is standard working hours defined as 35–40 h/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 h/week, because studies indicate that a proportion of individuals working less than standard hours do so because of existing health problems (Kivimäki et al., 2015a; Virtanen et al., 2012). In other words, poor health might have selected a certain proportion of individuals into working fewer than standard working hours and therefore a group working fewer than standard working hours cannot serve as a comparator. Consequently, if a study uses as the reference group individuals working less than standard hours, or combines individuals working standard hours and individuals working less than standard hours as the reference group, then these studies will be excluded from the review and meta-analysis. 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.

1.3. Description of the outcome

The WHO *Global Health Estimates* group outcomes into standard burden of disease categories (World Health Organization, 2017b), based on standard codes from the *International Statistical Classification of Diseases and Related Health Problems 10th Revision* (ICD-10) (World Health Organization, 2015). The relevant *WHO Global Health Estimates* category for this systematic review is "*II.E.1 Major depressive disorders*" (World Health Organization, 2017b). In line with the WHO *Global Health Estimates*, we define the health outcome covered in Systematic Review 2 as depression, corresponding with the ICD-10 codes F32 (depressive episode), F33 (recurrent depressive disorder) and F34.1 (dysthymia). We will consider prevalence of, incidence of and mortality from depression. Table 2 presents for each disease or health problem included in the WHO *Global Health Estimates* category the inclusion

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

	Definition
Risk factor	Long working hours (including those spent in secondary jobs), defined as working hours > 40/week hours, i.e. working hours exceeding standard working hours (35–40 h/week).
Risk factor levels	Preferable exposure categories are 35–40, 41–48, 49–54 and ≥ 55 h/week. However, whether we can use these categories will depend on the information provided in the studies. If the preferable exposure categories are not available, we will use the exposure categories provided by the studies as long as these exposure categories exceed 40 h/week.
Theoretical minimum risk exposure level	Standard working hours defined as working hours of 35–40 h/week.

Table 2
ICD-10 codes and disease and health problems covered by the WHO burden of disease category *II.E.1 Depressive disorders* and their inclusion in this review.

ICD-10 code	Disease or health problem	Included in this review
F32	Depressive episode	Yes
F33	Recurrent depressive disorder	Yes
F34.1	Dysthymia	Yes

criteria for this review. This review covers all the relevant WHO *Global Health Estimates* categories.

1.4. How the risk factor may impact the outcome

Fig. 1 presents the logic model for our systematic review of the causal relationship between exposure to long working hours and depression. This logic model is an a priori, process-orientated one (Rehfuess et al., 2018) that seeks to capture complexity of the risk factor-outcome causal relationship (Anderson et al., 2011).

Based on knowledge of previous research on long working hours and depression we assume that the effect of long working hours on risk of depression may be mediated via (a) disturbance of work/life balance, (b) exhaustion, (c) emotional distress, (d) health-related behaviors, such as lack of physical activity, high alcohol consumption and reduced sleeping hours, and (e) psycho-physiological changes, such as activation of the hypothalamic-pituitary-adrenal (HPA) axis, inflammation processes, circadian disruptions, and sleep impairment (Baglioni et al., 2011; Bannai and Tamakoshi, 2014; Bergs et al., 2018; Boden and Fergusson, 2011; Fujimura et al., 2014; Gold, 2015; Kronfeld-Schor and Einat, 2012; McEwen, 2004, 2012; Pariante and Lightman, 2008; Pittenger and Duman, 2008; Virtanen et al., 2009; Virtanen et al., 2015).

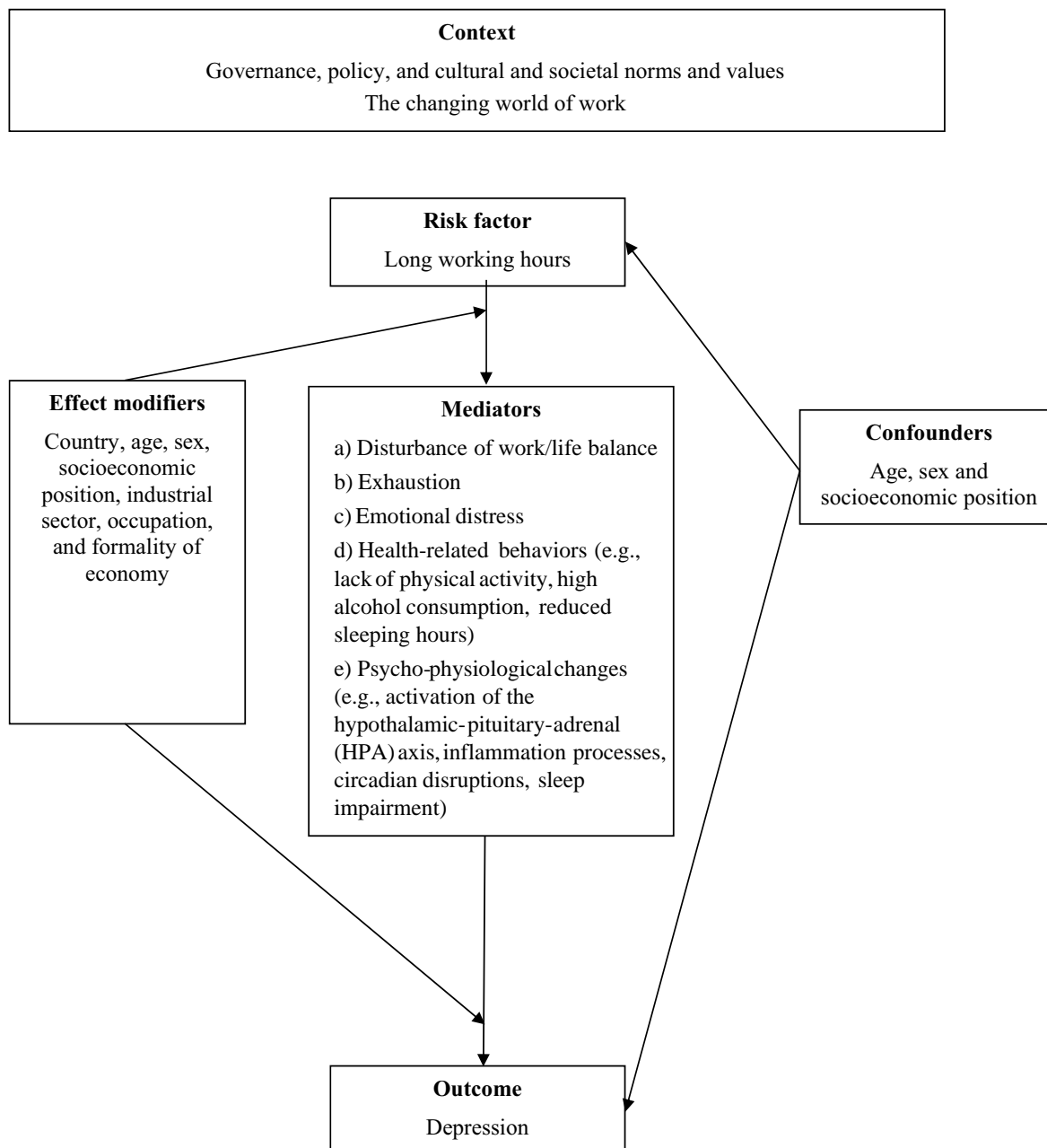As possible confounders we included age, sex and socioeconomic

Fig. 1. Logic model of the possible causal relationship between long working hours and depression.

position, i.e. we assume that these variables may impact both long working hours and risk of depression. It is well established that women and individuals of low socioeconomic position have a higher risk of depression than men and individuals of high socioeconomic position (Kessler et al., 2003; Lorant et al., 2003; Wittchen and Jacobi, 2005). With regard to age, some studies indicate that 12-month prevalence of depression is modestly higher in young adulthood than middle adulthood (Kessler et al., 2003; Wittchen and Jacobi, 2005), although birth cohort effects may also play a role, with a higher prevalence of depression in more recent birth cohorts (Kessler et al., 2003). Age, sex and socioeconomic position may also be related to lengths of working hours, although the direction of the relations may be dependent on other variables and contextual factors (Bannai et al., 2016; Larsen et al., 2017; Lee et al., 2016; O'Reilly and Rosato, 2013; Organisation for Economic Co-operation and Development (OECD), 2018; Wirtz et al., 2012), thus, it appears reasonable to regard these three variables as potential confounders for the association of long working hours with depression. We will address this possible confounding in Systematic Review 2 by including only studies in the meta-analysis that have adjusted or stratified for age, sex and socioeconomic position.

It is possible that age, sex and socioeconomic position are not only confounders, but also effect modifiers for the association of long working hours and depression. We will address this by conducting meta-analyses stratified by age, sex and socioeconomic position, if the data allow this. We further consider as effect modifiers country, industrial sector, occupation and formality of economy and will also conduct meta-analyses stratified by these variables, if data allow this.

Fig. 1 also considers macro and meso-level context that may impact either the prevalence of long working hours or the effect of long working hours on depression, or both (Commission of Social Determinants of Health, 2008; Dahlgreen and Whitehead, 2006; Martikainen et al., 2002; Rugulies et al., 2004).

2. 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 2005–2018 among the working-age population, disaggregated by country, sex, age and industrial sector or occupation. Systematic Review 1 will be conducted in a co-ordinated fashion across all four review groups that examine long working hours with regard to health endpoints (i.e. ischaemic heart disease (Li et al., 2018), stroke (Descatha et al., 2018), alcohol use (Godderis et al., 2018) and depression (this review)) led by Grace Sembajwe from the stroke review group.

2. Systematic Review 2: To systematically review and meta-analyse randomized control trials, cohort studies, case-control studies and other non-randomized intervention studies including estimates of the relative effect of a relevant level of occupational exposure to long working hours on depression in any year among the working-age population, compared with the minimum risk exposure level of 35–40 h/week.

3. Methods

We will apply the *Navigation Guide* (Woodruff and Sutton, 2014) 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 (Woodruff and Sutton, 2014). The need for further methodological development and refinement of the relatively novel *Navigation Guide* has been acknowledged (Woodruff and Sutton, 2014).

Systematic Review 1 may not map well to the *Navigation Guide* framework (see Fig. 1 on page 1009 in Woodruff and Sutton, 2014), 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 CRD42018085729. This protocol adheres with the preferred reporting items for systematic review and meta-analysis protocols statement (PRISMA-P) (Moher et al., 2015; Shamseer et al., 2015), with the abstract adhering with the reporting items for systematic reviews in journal and conference abstracts (PRISMA-A) (Beller et al., 2013). 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 (Stevens et al., 2016), and Systematic Review 2 will be reported according to the preferred reporting items for systematic review and meta-analysis statement (PRISMA) (Liberati et al., 2009). Our reporting of the parameters for estimating the burden of depression from occupational exposure to long working hours in the systematic review will adhere with the requirements of the GATHER guidelines (Stevens et al., 2016), because the WHO/ILO burden of disease estimates that may be produced consecutive to the systematic review must also adhere to these reporting guidelines.

3.1. Systematic Review 1

3.1.1. Eligibility criteria

The population, exposure, comparator and outcome (PECO) criteria (Liberati et al., 2009) are described below.

3.1.1.1. Types of populations. We will include studies of working-age (≤15 years) workers 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.

3.1.1.2. Types of exposures. 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]) (United Nations, 2008) or occupation (as defined, for example, by the *International Standard Classification of Occupations 1988* [ISCO-88] (International Labour Organization, 1987) or *2008* [ISCO-08] (International Labour Organization, 2012)). 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 depression 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 depression 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.

3.1.1.3. Types of comparators.
There will be no comparator, because we will review risk factor prevalence only.

3.1.1.4. Types of outcomes.
Exposure to the occupational risk factor (i.e. long working hours).

3.1.1.5. 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 (Descatha et al., 2018; Godderis et al., 2018; Hulshof et al., in press; Li et al., 2018; Mandrioli et al., 2018; Paulo et al., Accepted; Teixeira et al., Accepted; Tenkate et al., Accepted) in the series (i.e. Arabic, Bulgarian, Chinese, Danish, Dutch, English, French, Finnish, German, Hungarian, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish, Swedish and Thai), it will be translated into English. Published and unpublished studies will be included.

Studies conducted using unethical practices will be excluded from the review.

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

3.1.2. Information sources and search
3.1.2.1. Electronic academic databases.
We (that is a research team formed from researchers across the four long working hour review groups, including JLAM, MB, MC, CDT, BMR, KS and KT from this review group) will, at a minimum, search the following seven electronic academic databases:

1. Ovid MEDLINE with Daily Update (1 January 2005 to 31 May 2018)
2. PubMed (1 January 2005 to 31 May 2018)
3. EMBASE (1 January 2005 to 31 May 2018)
4. Scopus (1 January 2005 to 31 May 2018)
5. Web of Science (1 January 2005 to 31 May 2018)
6. CISDOC (1 January 2005 to 31 July 2018)
7. PsycINFO (1 January 2005 to 31 July 2018)

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.

3.1.2.2. Electronic grey literature databases.
We will, at a minimum, search the two following electronic grey literature databases:

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

3.1.2.3. Internet search engines.
We will 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, as has been done previously in Cochrane Reviews (Pega et al., 2015; Pega et al., 2017).

3.1.2.4. Organizational websites.
We will search, at a minimum, the websites of the following seven international organizations and national government departments:

1. International Labour Organization (www.ilo.org/).
2. World Health Organization (www.who.int).
3. European Agency for Safety and Health at Work (https://osha.europa.eu/en).
4. Eurostat (www.ec.europa.eu/eurostat/web/main/home).
5. China National Knowledge Infrastructure (http://www.cnki.net/).
6. Finnish Institute of Occupational Health (https://www.ttl.fi/en/).
7. National Institute of Occupational Safety and Health (NIOSH) of the United States of America, using the NIOSH data and statistics gateway (https://www.cdc.gov/niosh/data/).

3.1.2.5. Hand searching and expert consultation.
We will hand search for potentially eligible studies in.

- Reference lists of previous systematic reviews.
- Reference lists 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.

3.1.3. Study selection
Study selection will be carried out with Covidence (Babineau, 2014; Covidence systematic review software) and/or the Rayyan Systematic

Reviews Web App (Ouzzani et al., 2016). All study records identified in the search will be downloaded and duplicates will be identified and deleted. Afterwards, at least two review authors (from researchers across the four long working hour review groups, including JLAM, MB, MC, CDT, BMR, KS and KT from this review group), 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 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 (Stevens et al., 2016).

3.1.4. 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 (from researchers across the four long working hour review groups, including JLAM, MB, MC, CDT, BMR, KS and KT from this review group), 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 will resolve conflicting extractions. At a minimum, we will extract data on study characteristics (including study authors, study year, study country, participants and exposure), study design (including study type), 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) (2014) or DistillerSR (EvidencePartner, 2017) softwares.

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 (Forsyth et al., 2014). 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 (Drazen et al., 2010a; Drazen et al., 2010b).

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.

3.1.5. 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 (The GATHER Working Group, 2016), for burden of disease studies, of prevalence studies in general (Munn et al., 2014) and those of prevalence studies of occupational and/or environmental risk factors specifically (Krauth et al., 2013; Mandrioli and Silbergeld, 2016; Vandenberg et al., 2016). None of the five standard risk of bias assessment methods in systematic reviews (Rooney et al., 2016) are applicable to assessing prevalence studies. The *Navigation Guide* does not support checklist approaches, such as Hoy et al. (2012) and Munn et al. (2014), for assessing risk of bias in prevalence studies.

We will use a modified version of the *Navigation Guide* risk of bias tool (Lam et al., 2016c) 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 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 will then independently judge the risk of bias for each study by outcome, and a third author 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 (Higgins et al., 2011). Our risk of bias assessment for the entire body of evidence will be presented in a standard 'Risk of bias summary' figure (Higgins et al., 2011).

3.1.6. 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.

3.1.7. 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 (Lam et al., 2016c), 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 (Schünemann et al., 2011). We will grade the evidence, using the three *Navigation Guide* quality of evidence ratings: "high", "moderate" and "low" (Lam et al., 2016c). 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 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 will independently judge the quality of evidence for the entire body of evidence by outcome. A third review author 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.

3.1.8. Strength of evidence assessment

To our knowledge, no agreed method exists for rating strength of evidence in systematic reviews of prevalence studies. We 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).

3.2. Systematic Review 2

3.2.1. Eligibility criteria

The PECO (Liberati et al., 2009) criteria are described below.

3.2.1.1. Types of populations.
We will include studies of working-age (≤15 years) workers 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, if feasible. 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 F provides a complete, but briefer overview of the PECO criteria.

3.2.1.2. 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, studies from any year will be included.

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

3.2.1.4. Types of outcomes.
We will include studies that define depression in accordance with our standard definition of this outcome (Table 2) that is depressive episode (ICD-10, F32), recurrent depressive disorder (F33) and dysthymia (F34.1). Other affective disorders, e.g., bipolar disorders, will be excluded. We expect that most studies examining long working hours and depression will not have documented ICD-10 diagnostic codes, but will have ascertained depression with methods that approximate ICD-10 criteria (e.g., a validated depression rating scale filled in by the worker). We will include both self-reported and non-self-reported measurements of the outcome, but will prioritize non-self-reported over self-reported ones.

The following measurements of depression are eligible:

i. Psychiatric diagnostic interview.
ii. Diagnosis by a physician, psychologist or other qualified health professional.
iii. Hospital admission or discharge record.
iv. Administrative data (e.g., disability pensioning with the diagnosis of depression).
v. Register data of treatment for depression, with antidepressant medication, psychotherapy or both; will only be included if there is documentation that the treatment was for depression and not for other types of disorders.
vi. Self-administered rating scale for depression that was previously validated against a clinical measure of depression and that dichotomized respondents into cases versus non-cases (e.g., Center of Epidemiological Studies Depression Scale (CES-D) (Radloff, 1977) or Major Depression Inventory (MDI) (Bech et al., 2001)) or other

validated self-administered rating scales.
vii. Medically certified cause of death.

Because the endpoint of our study is binary, studies exclusively reporting depression as a continuous variable (e.g., level of depressive symptoms) will be excluded, as will be all other measurements.

3.2.1.5. Types of studies.
We will include studies that investigated the effect of long working hours on depression for any years. Eligible study designs will be randomized controlled trials (including parallel-group, cluster, cross-over and factorial trials) and 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 include 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 (Arditi et al., 2016). All other study designs, such as uncontrolled before-and-after, cross-sectional, qualitative, modelling, case and non-original studies will be excluded.

With regard to cohort studies, we will include only studies that have excluded individuals with depression at baseline (to reduce the risk of reverse causation). However, it is possible that some studies have additionally measured levels of non-clinical depressive symptoms at baseline and have included this measure as a covariate.

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 (Descatha et al., 2018; Godderis et al., 2018; Hulshof et al., in press; Li et al., 2018; Mandrioli et al., 2018; Paulo et al., Accepted; Teixeira et al., Accepted; Tenkate et al., Accepted), then the record will be translated into English. Published and unpublished studies will be included.

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

3.2.1.6. Types of effect measures.
We will include measures of the relative effect of a relevant level of long working hours on the risk of developing or dying from depression, compared with the theoretical minimum risk exposure level. Effect estimates of prevalence measures only will be excluded. We will include relative effect and incidence measures such as risk ratios, odds ratios and hazard ratios. 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 (Higgins and Green, 2011).

As shown in our logic framework (Fig. 1), we a priori consider the following variables to be potential effect modifiers of the effect of long working hours on depression: country, age, sex, industrial sector, occupation and formality of employment. We consider age, sex and socioeconomic position to be potential confounders. Potential mediators are: disturbance of work/life balance; exhaustion; emotional distress; health-related behaviors; and psychophysiological changes.

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 (Fig. 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 (Pega et al., 2016) over estimates from models that can only adjust for time-varying confounders, such as fixed-effects models (Gunasekara et al., 2014), 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.

3.2.2. Information sources and search

3.2.2.1. Electronic academic databases. We (MB, CDT, BMR and KS) will, at a minimum, search the seven following electronic academic databases:

1. International Clinical Trials Register Platform (to 30 June 2018)
2. Ovid MEDLINE with Daily Update (1946 to 30 June 2018)
3. PubMed (1946 to 30 June 2018)
4. EMBASE (1947 to 30 June 2018)
5. Web of Science (1945 to 30 June 2018)
6. CISDOC (1901 to 30 June 2018)
7. PsycINFO (1880 to 30 June 2018)

The Ovid Medline search strategy for Systematic Review 2 is presented in Appendix G. We will perform searchers in the electronic databases operated in the English literature using a search strategy in the English language. We will perform searches in electronic databases operated in the English language using a search strategy in the English language. We (CDT, KS and RR) 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.

3.2.2.2. Electronic grey literature databases. We (MB, CDT, BMR and KS) will, at a minimum, search the two following electronic grey literature databases:

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

3.2.2.3. Internet search engines. We (MB, CDT, BMR and KS) will 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.

3.2.2.4. Organizational websites. We (MB, CDT, BMR and KS) will search the websites of the six following international organizations and national government departments:

1. International Labour Organization (www.ilo.org/)
2. World Health Organization (www.who.int)
3. EUROSTAT (www.ec.europa.eu/eurostat/web/main/home)
4. National Institute of Occupational Safety and Health (NIOSH) of the United States of America, using the NIOSH data and statistics gateway (https://www.cdc.gov/niosh/data/)
5. Finnish Institute of Occupational Health (https://www.ttl.fi/en/)

6. International Commission of Occupational Health (ICOH) Scientific Committee on Work Organization and Psychosocial Factors (ICOH-WOPS) http://www.icohweb.org/site/scientific-committee-detail.asp?sc=33

3.2.2.5. Hand searching and expert consultation. We (MB, CDT, BMR, KS, SI and RR) will hand search for potentially eligible studies in.

- Reference lists of previous systematic reviews.
- Reference lists 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.

3.2.3. Study selection

Study selection will be carried out in a reference manager database, such as Covidence (Babineau, 2014; Covidence systematic review software) or the Rayyan Systematic Reviews Web App (Ouzzani et al., 2016). All study records identified in the search will be downloaded and duplicates will be identified and deleted. Afterwards, at least two review authors (out of: RR, EA, JLAM, MB, MC, CDT, ND, QDM, HE, JG, AHG, SI, IEHM, BMR, KS, KT and AZ), working in pairs, will independently screen titles and abstracts (step 1) and then full texts (step 2) of potentially relevant records. Any disagreements between the two review authors will be resolved by discussion and the involvement of a third review author (MB, CDT, BMR or KS). 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 (Liberati et al., 2009).

3.2.4. Data extraction and data items

A data extraction form will be developed and trialed until data extractors reach convergence and agreement. At a minimum, two review authors (out of: MB, CDT, BMR and KS) 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 risk of depression). A third review author (out of: MB, CDT, BMR, KS, RR) will resolve conflicts in data extraction. Data will be entered into and managed with the RevMan 5.3 software (2014).

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 (Forsyth et al., 2014). Where no financial disclosure or conflict of interest statements are 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 (Drazen et al., 2010a; Drazen et al., 2010b).

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.

3.2.5. 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 (Rooney et al., 2016). However, all five methods, including the *Navigation Guide* (Lam et al., 2016c), assess risk of bias in human studies similarly (Rooney et al., 2016).

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 (Woodruff and Sutton, 2014), 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 (Higgins et al., 2011) and the US Agency for Healthcare Research and Quality (Viswanathan et al., 2008). Some further refinements of the *Navigation Guide* method may be warranted (Goodman et al., 2017), but it has been successfully applied in several completed and ongoing systematic reviews (Johnson et al., 2016; Johnson et al., 2014; Koustas et al., 2014; Lam et al., 2016a; Lam et al., 2014; Lam et al., 2016b; Vesterinen et al., 2014). 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 (Lam et al., 2016c). 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 (Johnson et al., 2014; Koustas et al., 2014; Lam et al., 2014; Vesterinen et al., 2014) all of the subsequent reviews have included this domain (Johnson et al., 2016; Lam et al., 2016a; Lam et al., 2017; Lam et al., 2016b; Lam et al., 2016c). Risk of bias or confounding ratings will be: "low"; "probably low"; "probably high"; "high" or "not applicable" (Lam et al., 2016c). 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 (Lam et al., 2016c). 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 (MB, CDT, BMR, KS, RR and SI) 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 (out of: MB, CDT, BMR and KS) will independently judge the risk of bias for each study by outcome. Where individual assessments differ, a third author (MB, CDT, MB, BMR, RR or SI) 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 (Higgins et al., 2011). For the entire body of evidence, we will present the study-level risk of bias assessments in a 'Risk of bias summary' figure (Higgins et al., 2011).

3.2.6. 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 (out of: CDT, KS, RR and SI) will independently investigate the clinical heterogeneity of the studies in terms of participants (including country, sex, age, socioeconomic position and industrial sector or occupation), level of risk factor exposure, comparator and outcomes. If we find that effect estimates differ considerably by country, sex, socioeconomic position and industrial sector or occupation, or a combination of these, then we will synthesise evidence for the relevant populations defined by country, sex, age, socioeconomic position and industrial sector or occupation, 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, age, socioeconomic position occupation and industrial sector, 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 I^2 statistic (Higgins et al., 2003). 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 (Higgins and Green, 2011). The meta-analysis will be conducted in RevMan 5.3 (2014), but the data for entry into these programmes may be prepared using another recognized statistical analysis programme, such as Stata (Stata Cooperation, 2017). 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. More specifically, the analyses have to be adjusted or stratified for (i) sex, (ii) age and (iii) a measure of socioeconomic position (e.g., education, income or occupational grade) to be included in the meta-analysis. 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.

3.2.7. Additional analyses

If there is evidence for differences in effect estimates by country, sex, age, socioeconomic position and industrial sector or occupation, or a combination of these variables, then we will conduct subgroup analyses by these variables. If studies on workers in the informal economy and in the formal economy are included, then we will conduct subgroup analysis by formality of economy studied. 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 examine the potential of these variables to be effect modification in a meta-regression, if feasible. In addition, we may conduct meta-regressions or stratified analyses for other potential effect modifiers, if allowed by the data.

If feasible, sensitivity analyses will be conducted that will include only studies judged to be of "low" or "probably low" risk of bias. If feasible, we will conduct sensitivity analyses that are stratified by whether the estimate was based on a documented ICD-10 diagnostic

code or was based on an approximation of an ICD-10 diagnostic code. We may also conduct a sensitivity analysis using an alternative meta-analytic model, namely the inverse variance heterogeneity (IVhet) model.

A recent systematic review and meta-analysis on job strain and risk of hospital treatment for depression showed that depressive symptoms are likely partly an intermediate step in the pathway linking occupational exposure and risk of depression (Madsen et al., 2017). Consequently, we regard depressive symptoms as a mediator and do not include in the main meta-analysis estimates that are adjusted for depressive symptoms, unless the analysis used a model that can adjust for this mediation (e.g. an appropriately specified marginal structural model). However, because baseline depressive symptoms may also be a confounder if they have caused both reporting of long working hours at baseline and incidence of depression at follow-up (Madsen et al., 2017), we will conduct an additional analysis with estimates that are adjusted for baseline depressive symptoms, if studies have provided such estimates.

3.2.8. Quality of evidence assessment

We will assess quality of evidence using a modified version of the *Navigation Guide* quality of evidence assessment tool (Lam et al., 2016c). The tool is based on the GRADE approach (Schünemann et al., 2011) adapted specifically to systematic reviews in occupational and environmental health (Morgan et al., 2016). Should a more suitable method become available, we may switch to it.

Working in pairs, we (MB, CDT, BMR, KS, RR and SI) will assess quality of evidence for the entire body of evidence by outcome, with any disagreements resolved by a third review author (RR or SI). We will adopt or adapt the latest *Navigation Guide* instructions (Appendix D) for grading the quality of evidence (Lam et al., 2016c). 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" (Lam et al., 2016c). 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 downgrade 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 up-grade 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".

3.2.9. Strength of evidence assessment

We will apply the standard *Navigation Guide* methodology (Lam et al., 2016c) 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 depression will be "sufficient evidence of toxicity/harmfulness", "limited of toxicity/harmfulness", "inadequate of toxicity/harmfulness" and "evidence of lack of toxicity/harmfulness" (Appendix I).

Author contributions

IDI, NL, FP and APÜ 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. RR is the lead reviewer of Systematic Review 2. RR wrote the first draft of this protocol, using the protocol template prepared by FP, and all authors made substantial contributions to the revisions of the manuscript. The search strategy was developed and piloted by KS in collaboration with a research librarian. FP coordinated all inputs from the World Health Organization, International Labour Organization and external experts and ensured consistency across the systematic reviews of this series. RR is the guarantor of Systematic Review 2.

Conflict of interest

CDT and SI report participation in projects granted from the Italian Ministry of Health. All other authors declare no conflict of interest.

Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2018.11.011.

References

104th International Labour Conference, 2015. Transition From the Informal to the Formal Economy (Recommendation No. 204). International Labour Organization, Geneva.

Anderson, L.M., Petticrew, M., Rehfuess, E., Armstrong, R., Ueffing, E., Baker, P., Francis, D., Tugwell, P., 2011. Using logic models to capture complexity in systematic reviews. Res. Synth. Methods 2, 33–42.

Arditi, C., Burnand, B., Peytremann-Bridevaux, I., 2016. Adding non-randomised studies to a Cochrane review brings complementary information for healthcare stakeholders:

an augmented systematic review and meta-analysis. BMC Health Serv. Res. 16, 598.

Babineau, J., 2014. Product review: covidence (systematic review software). J. Can. Health Libr. Assoc. (JCHLA) 32, 68–71.

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., Lombardo, C., Riemann, D., 2011. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. J. Affect. Disord. 135, 10–19.

Bannai, A., Tamakoshi, A., 2014. The association between long working hours and health: a systematic review of epidemiological evidence. Scand. J. Work Environ. Health 40, 5–18.

Bannai, A., Yoshioka, E., Saijo, Y., Sasaki, S., Kishi, R., Tamakoshi, A., 2016. The risk of developing diabetes in association with long working hours differs by shift work schedules. J. Epidemiol. 26, 481–487.

Bech, P., Rasmussen, N.A., Olsen, L.R., Noerholm, V., Abildgaard, W., 2001. The sensitivity and specificity of the major depression inventory, using the present state examination as the index of diagnostic validity. J. Affect. Disord. 66, 159–164.

Beller, E.M., Glasziou, P.P., Altman, D.G., Hopewell, S., Bastian, H., Chalmers, I., Gotzsche, P.C., Lasserson, T., Tovey, D., Group, P.f.A, 2013. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. PLoS Med. 10, e1001419.

Bergs, Y., Hofs, H., Kant, I., Slangen, J., Jansen, N.W.H., 2018. Work–family conflict and depressive complaints among Dutch employees: examining reciprocal associations in a longitudinal study. Scand. J. Work Environ. Health 44, 69–79.

Boden, J.M., Fergusson, D.M., 2011. Alcohol and depression. Addiction 106, 906–914.

Commission on Social Determinants of Health, 2008. Closing the gap in a generation: health equity through action on the social determinants of health. In: Final Report of the Commission on Social Determinants of Health. World Health Organization, Geneva.

Covidence Systematic Review Software. Melbourne, Australia. Available from. http://www.covidence.org.

Dahlgreen, G., Whitehead, M., 2006. European Strategies for Tackling Social Inequalities in Health. Levelling Up: Part 2. WHO Regional Office for Europe, Copenhagen.

Descatha, A., Sembajwe, G., Baer, M., Boccuni, F., Di Tecco, C., Duret, C., Evanoff, B.A., Gagliardi, D., Ivanov, I.D., Leppink, N., Magnusson Hanson, L.L., Marinaccio, A., Ozguler, A., Pega, F., Pico, F., Prüss-Üstün, A.M., Ronchetti, M., Roquelaure, Y., Sabbath, E., Stevens, G.A., Tsutsumi, A., Ujita, Y., Iavicoli, S., 2018. 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 stroke. Environ. Int. 19, 366–378.

Drazen, J.M., de Leeuw, P.W., Laine, C., Mulrow, C., DeAngelis, C.D., Frizelle, F.A., Godlee, F., Haug, C., Hebert, P.C., James, A., Kotzin, S., Marusic, A., Reyes, H., Rosenberg, J., Sahni, P., Van der Weyden, M.B., Zhaori, G., 2010a. Toward more uniform conflict disclosures: the updated ICMJE conflict of interest reporting form. JAMA 304, 212–213.

Drazen, J.M., Van der Weyden, M.B., Sahni, P., Rosenberg, J., Marusic, A., Laine, C., Kotzin, S., Horton, R., Hebert, P.C., Haug, C., Godlee, F., Frizelle, F.A., de Leeuw, P.W., DeAngelis, C.D., 2010b. Uniform format for disclosure of competing interests in ICMJE journals. JAMA 303, 75–76.

EvidencePartner, 2017. DistillerSR. Available from. https://www.evidencepartners.com/products/distillersr-systematic-review-software, Accessed date: 13 December 2017.

Forsyth, S.R., Odierna, D.H., Krauth, D., Bero, L.A., 2014. Conflicts of interest and critiques of the use of systematic reviews in policymaking: an analysis of opinion articles. Syst. Rev. 3, 122.

Fujimura, Y., Sekine, M., Tatsuse, T., 2014. Sex differences in factors contributing to family-to-work and work-to-family conflict in Japanese civil servants. J. Occup. Health 56, 485–497.

Godderis, L., Bakusic, J., Boonen, E., Delvaux, E., Ivanov, I.D., Lambrechts, M.-C., Latorraca, C.O., Leppink, N., Martimbianco, A.L., Pega, F., Prüss-Üstün, A.M., Riera, R., Ujita, Y., Pachito, D.V., 2018. 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 alcohol use and alcohol use disorder. Environ. Int. 120, 22–33.

Gold, P.W., 2015. The organization of the stress system and its dysregulation in depressive illness. Mol. Psychiatry 20, 32–47.

Goodman, J.E., Lynch, H.N., Beck, N.B., 2017. More clarity needed in the navigation guide systematic review framework. Environ. Int. 102, 74–75.

Gunasekara, F.I., Richardson, K., Carter, K., Blakely, T., 2014. Fixed effects analysis of repeated measures data. Int. J. Epidemiol. 43, 264–269.

Guyatt, G.H., Oxman, A.D., Schunemann, H.J., Tugwell, P., Knottnerus, A., 2011. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. J. Clin. Epidemiol. 64, 380–382.

Higgins, J., Green, S. (Eds.), 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]. The Cochrane Collaboration Available from. http://handbook.cochrane.org.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. Br. Med. J. 327, 557–560.

Higgins, J., Altman, D., Sterne, J., 2011. Chapter 8: Assessing risk of bias in included studies. In: Higgins, J., Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 510 [Updated March 2011]. The Cochrane Collaboration Available from. http://handbookcochraneorg.

Hoy, D., Brooks, P., Woolf, A., Blyth, F., March, L., Bain, C., Baker, P., Smith, E., Buchbinder, R., 2012. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J. Clin. Epidemiol. 65, 934–939.

Hulshof, C., Colosio, C., Ivaovn, I.D., Kuijer, P., Leppink, N., Mandic-Rajcevic, S., Masci, F., Neupane, S., Nygård, C.-H., Oakman, J., Pega, F., Prakash, K., Proper, K., Prüss-Üstün, A.M., Ujita, Y., Van der Molen, H., Frings-Dresen, M., 2018. WHO/ILO work-

related burden of disease and injury: protocol for systematic reviews of occupational exposure to ergonomic risk factors and of the effect of occupational exposure to ergonomic risk factors on osteoarthritis and other musculoskeletal diseases. Environ. Int. https://doi.org/10.1016/j.envint.2018.09.053. (in press).

International Labour Organization, 1987. ISCO–88: International Standard Classification of Occupations. International Labour Organization, Geneva.

International Labour Organization, 2012. ISCO–08: International Standard Classification of Occupations. International Labour Organization, Geneva.

International Labour Organization, 2014. Safety and Health at Work: A Vision for Sustainable Prevention: XX World Congress on Safety and Health at Work 2014: Global Forum for Prevention, 24–27 August 2014, Frankfurt, Germany. International Labour Organization, Geneva.

Johnson, P.I., Sutton, P., Atchley, D.S., Koustas, E., Lam, J., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The navigation guide-evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1028–1039.

Johnson, P.I., Koustas, E., Vesterinen, H.M., Sutton, P., Atchley, D.S., Kim, A.N., Campbell, M., Donald, J.M., Sen, S., Bero, L., Zeise, L., Woodruff, T.J., 2016. Application of the navigation guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan. Environ. Int. 92–93, 716–728.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 289, 3095–3105.

Kivimäki, M., Jokela, M., Nyberg, S.T., Singh-Manoux, A., Fransson, E.I., Alfredsson, L., Bjorner, J.B., Borritz, M., Burr, H., Casini, A., Clays, E., De Bacquer, D., Dragano, N., Erbel, R., Geuskens, G.A., Hamer, M., Hooftman, W.E., Houtman, I.L., Jockel, K.H., Kittel, F., Knutsson, A., Koskenvuo, M., Lunau, T., Madsen, I.E.H., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Rugulies, R., Salo, P., Shipley, M.J., Siegrist, J., Steptoe, A., Suominen, S.B., Theorell, T., Vahtera, J., Westerholm, P.J., Westerlund, H., O'Reilly, D., Kumari, M., Batty, G.D., Ferrie, J.E., Virtanen, M., for the IPD-Work Consortium, 2015a. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. Lancet 386, 1739–1746.

Kivimäki, M., Virtanen, M., Kawachi, I., Nyberg, S.T., Alfredsson, L., Batty, G.D., Bjorner, J.B., Borritz, M., Brunner, E.J., Burr, H., Dragano, N., Ferrie, J.E., Fransson, E.I., Hamer, M., Heikkilä, K., Knutsson, A., Koskenvuo, M., Madsen, I.E.H., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Rugulies, R., Salo, P., Siegrist, J., Steptoe, A., Suominen, S., Theorell, T., Vahtera, J., Westerholm, P.J., Westerlund, H., Singh-Manoux, A., Jokela, M., for the IPD-Work Consortium, 2015b. Long working hours, socioeconomic status, and the risk of incident type 2 diabetes: a meta-analysis of published and unpublished data from 222,120 individuals. Lancet Diabetes Endocrinol. 3, 27–34.

Koustas, E., Lam, J., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The navigation guide-evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1015–1027.

Krauth, D., Woodruff, T.J., Bero, L., 2013. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. Environ. Health Perspect. 121, 985–992.

Kronfeld-Schor, N., Einat, H., 2012. Circadian rhythms and depression: human psychopathology and animal models. Neuropharmacology 62, 101–114.

Lam, J., Koustas, E., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The navigation guide-evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1040–1051.

Lam, J., Koustas, E., Sutton, P., Cabana, M., Whitaker, E., Padula, A., Vesterinen, H., Daniels, N., Woodruff, T.J., 2016a. Applying the Navigation Guide: Case Study #6. Association Between Formaldehyde Exposures and Asthma. (In preparation).

Lam, J., Sutton, P., Halladay, A., Davidson, L.I., Lawler, C., Newschaffer, C.J., Kalkbrenner, A., Joseph, J., Zilber School of Public Health, Windham, G.C., Daniels, N., Sen, S., Woodruff, T.J., 2016b. Applying the Navigation Guide Systematic Review Methodology Case Study #4: Association between Developmental Exposures to Ambient Air Pollution and Autism. PLoS One 21.

Lam, J., Sutton, P., Padula, A.M., Cabana, M.D., Koustas, E., Vesterinen, H.M., Whitaker, E., Skalla, L., Daniels, N., Woodruff, T.J., 2016c. Applying the Navigation Guide Systematic Review Methodology Case Study #6: Association between Formaldehyde Exposure and Asthma: A Systematic Review of the Evidence: Protocol. University of California at San Francisco, San Francisco, CA.

Lam, J., Lanphear, B., Bellinger, D., Axelrad, D., McPartland, J., Sutton, P., Davidson, L.I., Daniels, N., Sen, S., Woodruff, T.J., 2017. Developmental PBDE exposure and IQ/ADHD in childhood: a systematic review and meta-analysis. Environ. Health Perspect. 125.

Larsen, A.D., Hannerz, H., Moller, S.V., Dyreborg, J., Bonde, J.P., Hansen, J., Kolstad, H.A., Hansen, A.M., Garde, A.H., 2017. Night work, long work weeks, and risk of accidental injuries. A register-based study. Scand. J. Work Environ. Health 43, 578–586.

Lee, D.W., Hong, Y.C., Min, K.B., Kim, T.S., Kim, M.S., Kang, M.Y., 2016. The effect of long working hours on 10-year risk of coronary heart disease and stroke in the Korean population: the Korea National Health and Nutrition Examination Survey (KNHANES), 2007 to 2013. Ann. Occup. Environ. Med. 28, 64.

Li, J., Brisson, C., Clays, E., Ferrario, M.M., Ivanov, I.D., Landsbergis, P., Leppink, N., Pega, F., Pikhart, H., Prüss-Üstün, A.M., Rugulies, R., Schnall, P.L., Stevens, G.A., Tsutsumi, A., Ujita, Y., Siegrist, J.W.H.O., 2018. 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. Environ. Int. 119, 558–569.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6, e1000100.

Lorant, V., Deliege, D., Eaton, W., Robert, A., Philippot, P., Ansseau, M., 2003. Socioeconomic inequalities in depression: a meta-analysis. Am. J. Epidemiol. 157, 98–112.

Madsen, I.E.H., Nyberg, S.T., Magnusson Hanson, L.L., Ferrie, J.E., Ahola, K., Alfredsson, L., Batty, G.D., Bjorner, J.B., Borritz, M., Burr, H., Chastang, J.F., de Graaf, R., Dragano, N., Hamer, M., Jokela, M., Knutsson, A., Koskenvuo, M., Koskinen, A., Leineweber, C., Niedhammer, I., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Plaisier, I., Salo, P., Singh-Manoux, A., Suominen, S., Ten Have, M., Theorell, T., Toppinen-Tanner, S., Vahtera, J., Väänänen, A., Westerholm, P.J.M., Westerlund, H., Fransson, E.I., Heikkilä, K., Virtanen, M., Rugulies, R., Kivimäki, M., for the IPD-Work Consortium, 2017. Job strain as a risk factor for clinical depression: systematic review and meta-analysis with additional individual participant data. Psychol. Med. 47, 1342–1356.

Mandrioli, D., Silbergeld, E.K., 2016. Evidence from toxicology: the most essential science for prevention. Environ. Health Perspect. 124, 6–11.

Mandrioli, D., Schlünssen, V., Adam, B., Cohen, A.A., Chen, W., Colosio, C., Fischer, A., Godderis, L., Göen, T., Ivanov, I.D., Leppink, N., Mandic-Rajcevic, S., Masci, F., Nemery, B., Pega, F., Prüss-Üstün, A.M., Sgargi, D., Ujita, Y., Van der Mierden, S., Zungu, M., Scheepers, P., 2018. WHO/ILO work-related burden of disease and injury: protocols for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. Environ. Int. 119, 174–185.

Martikainen, P., Bartley, M., Lahelma, E., 2002. Psychosocial determinants of health in social epidemiology. Int. J. Epidemiol. 31, 1091–1093.

McEwen, B.S., 2004. Protection and damage from acute and chronic stress: allostasis and allostatic overload and relevance to the pathophysiology of psychiatric disorders. Ann. N. Y. Acad. Sci. 1032, 1–7.

McEwen, B.S., 2012. Brain on stress: how the social environment gets under the skin. Proc. Natl. Acad. Sci. U. S. A. 109 (Suppl. 2), 17180–17185.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., Group, P.-P, 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4 (1).

Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., Guyatt, G., Hooijmans, C., Langendam, M., Mandrioli, D., Mustafa, R.A., Rehfuess, E.A., Rooney, A.A., Shea, B., Silbergeld, E.K., Sutton, P., Wolfe, M.S., Woodruff, T.J., Verbeek, J.H., Holloway, A.C., Santesso, N., Schunemann, H.J., 2016. GRADE: assessing the quality of evidence in environmental and occupational health. Environ. Int. 92–93, 611–616.

Munn, Z., Moola, S., Riitano, D., Lisy, K., 2014. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. Int. J. Health Policy Manag. 3, 123–128.

Murray, C.J.L., Ezzati, M., Lopez, A.D., Rodgers, A., Vander Hoorn, S., 2004. Comparative quantification of health risks: conceptual framework and methodological issues. In: Ezzati, M., Lopez, A.D., Rodgers, A., Murray, C.J.L. (Eds.), Comparative Quantification of Health Risks: Global and Regional Burdedn of Disease Attributable to Selected Major Risk Factors. World Health Organization, Geneva.

O'Reilly, D., Rosato, M., 2013. Worked to death? A census-based longitudinal study of the relationship between the numbers of hours spent working and mortality risk. Int. J. Epidemiol. 42, 1820–1830.

Organisation for Economic Co-operation and Development (OECD), 2018. OECD.Stat: Average Usual Weekly Hours Worked on the Main Job. Available from. http://stats.oecd.org/Index.aspx?DatasetCode=AVE_HRS, Accessed date: 21 March 2018.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan-a web and mobile app for systematic reviews. Syst. Rev. 5, 210.

Pariante, C.M., Lightman, S.L., 2008. The HPA axis in major depression: classical theories and new developments. Trends Neurosci. 31, 464–468.

Paulo, M.S., Akagwu, O.C., Akparibo, I.Y., Al-Rifai, R.H., Balazs, A., Bazrafshan, S.K., Gobba, F., Ivanov, I.D., Kezic, S., Leppink, N., Loney, T., Modenese, A., Pega, F., Peters, C., Tenkate, T.D., Ujita, Y., Wittlich, M., John, S.M., 2018. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on melanoma and non-melanoma skin cancer. Environ. Int. https://doi.org/10.1016/j.envint.2018.09.039. (Accepted for publication).

Pega, F., Liu, S.Y., Walter, S., Lhachimi, S.K., 2015. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst. Rev. 9, CD011247.

Pega, F., Blakely, T., Glymour, M.M., Carter, K.N., Kawachi, I., 2016. Using marginal structural modeling to estimate the cumulative impact of an unconditional tax credit on self-rated health. Am. J. Epidemiol. 183, 315–324.

Pega, F., Liu, S.Y., Walter, S., Pabayo, R., Saith, R., Lhachimi, S.K., 2017. Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst. Rev. 11, CD011135.

Pittenger, C., Duman, R.S., 2008. Stress, depression, and neuroplasticity: a convergence of mechanisms. Neuropsychopharmacology 33, 88–109.

Prüss-Üstün, A.M., Wolf, J., Corvalan, C., Bos, R., Neira, M., 2017. Preventing Disease Through Healthy Environments: A Global Assessment of the Burden of Disease From Environmental Risks. World Health Organization, Geneva.

Radloff, L.S., 1977. The CES-D Scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1, 385–401.

Rehfuess, E.A., Booth, A., Brereton, L., Burns, J., Gerhardus, A., Mozygemba, K., Oortwijn, W., Pfadenhauer, L.M., Tummers, M., van der Wilt, G.J., Rohwer, A., 2018. Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged, and iterative approaches. Res. Synth. Methods 9, 13–24.

Review Manager (RevMan). Version 5.3. The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen.

Rooney, A.A., Cooper, G.S., Jahnke, G.D., Lam, J., Morgan, R.L., Boyles, A.L., Ratcliffe, J.M., Kraft, A.D., Schunemann, H.J., Schwingl, P., Walker, T.D., Thayer, K.A., Lunn, R.M., 2016. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. Environ. Int. 92-93, 617–629.

Rugulies, R., Aust, B., Syme, S.L., 2004. Epidemiology of health and illness. A socio-psycho-physiological perspective. In: Sutton, S., Baum, A., Johnston, M. (Eds.), The Sage Handbook of Health Psychology. Sage, London.

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

Schünemann, H., Oxman, A., Vist, G., Higgins, J., Deeks, J., Glasziou, P., Guyatt, G., 2011. Chapter 12: interpreting results and drawing conclusions. In: Higgins, J., Green, S. (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 510 [Updated March 2011]. The Cochrane Collaboration Available from. http://www.handbook.cochrane.org.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., Group, P.-P, 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br. Med. J. 349, g7647.

Stata Cooperation, 2017. Stata 15. Stata Cooperation, College Station, TX.

Stevens, G.A., Alkema, L., Black, R.E., Boerma, J.T., Collins, G.S., Ezzati, M., Grove, J.T., Hogan, D.R., Hogan, M.C., Horton, R., Lawn, J.E., Marusic, A., Mathers, C.D., Murray, C.J., Rudan, I., Salomon, J.A., Simpson, P.J., Vos, T., Welch, V., 2016. Guidelines for accurate and transparent health estimates reporting: the GATHER statement. Lancet 388, e19–e23.

Teixeira, L.R., Azevedo, T.M., Bortkiewicz, A.T., Corrêa da Silva, D.T., De Abreu, W., De Almeida, M.S., De Araújo, M.A.N., Gadzicka, E.H., Ivanov, I.D., Leppink, N., Macedo, M.R.V., Maciel, E.M.G.S., Pawlaczyk-Luszczynska, M.S., Pega, F., Prüss-Üstün, A.M., Siedlecka, J.M., Ujita, Y., Braga, J.U., 2018. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of occupational exposure to noise and of the effect of occupational exposure to noise on cardiovascular disease. Environ. Int. https://doi.org/10.1016/j.envint.2018.09.040. (Accepted for publication).

Tenkate, T.D., Paulo, M., Adam, B., Al-Rifai, R.H., Chou, B.R., Gobba, F., Ivanov, I.D., Leppink, N., Loney, T., Pega, F., Peters, C., Prüss-Üstün, A.M., Ujita, Y., Wittlich, M., Modenese, A., 2018. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on cataract. Environ. Int. https://doi.org/10.1016/j.envint.2018.10.001. (Accepted for publication).

The GATHER Working Group, 2016. The GATHER Statement: Explanation and Elaboration. World Health Organization, Geneva.

Theorell, T., Hammarström, A., Aronsson, G., Träskman Bendz, L., Grape, T., Hogstedt, C., Marteinsdottir, I., Skoog, I., Hall, C., 2015. A systematic review including meta-analysis of work environment and depressive symptoms. BMC Public Health 15, 738.

United Nations, 2008. In: Affairs D.o.E.a.S (Ed.), ISIC Rev. 4: International Standard Industrial Classification of All Economic Activities, Revision 4. Statistical Papers Series M No. 4/Rev.4. United Nations, New York, NY.

Vandenberg, L.N., Agerstrand, M., Beronius, A., Beausoleil, C., Bergman, A., Bero, L.A., Bornehag, C.G., Boyer, C.S., Cooper, G.S., Cotgreave, I., Gee, D., Grandjean, P., Guyton, K.Z., Hass, U., Heindel, J.J., Jobling, S., Kidd, K.A., Kortenkamp, A., Macleod, M.R., Martin, O.V., Norinder, U., Scheringer, M., Thayer, K.A., Toppari, J., Whaley, P., Woodruff, T.J., Ruden, C., 2016. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. Environ. Health 15 (74).

Vesterinen, H., Johnson, P., Atchley, D., Sutton, P., Lam, J., Zlatnik, M., Sen, S., Woodruff, T., 2014. The relationship between fetal growth and maternal glomerular filtration rate: a systematic review. J. Matern. Fetal Neonatal Med. 1–6.

Virtanen, M., Ferrie, J.E., Gimeno, D., Vahtera, J., Elovainio, M., Singh-Manoux, A., Marmot, M.G., Kivimäki, M., 2009. Long working hours and sleep disturbances: the Whitehall II prospective cohort study. Sleep 32, 737–745.

Virtanen, M., Heikkila, K., Jokela, M., Ferrie, J.E., Batty, G.D., Vahtera, J., Kivimäki, M., 2012. Long working hours and coronary heart disease: a systematic review and meta-analysis. Am. J. Epidemiol. 7, 586–596.

Virtanen, M., Jokela, M., Nyberg, S.T., Madsen, I.E.H., Lallukka, T., Ahola, K., Alfredsson, L., Batty, G.D., Bjorner, J.B., Borritz, M., Burr, H., Casini, A., Clays, E., De Bacquer, D., Dragano, N., Erbel, R., Ferrie, J.E., Fransson, E.I., Hamer, M., Heikkilä, K., Jockel, K.H., Kittel, F., Knutsson, A., Koskenvuo, M., Ladwig, K.H., Lunau, T., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Rugulies, R., Salo, P., Shupp, J., Siegrist, J., Singh-Manoux, A., Steptoe, A., Suominen, S.B., Theorell, T., Vahtera, J., Wagner, G.G., Westerholm, P.J., Westerlund, H., Kivimäki, M., for the IPD-Work Consortium, 2015. Long working hours and alcohol use: systematic review and meta-analysis of published studies and unpublished individual participant data. Br. Med. J. 350, g7772.

Virtanen, M., Jokela, M., Madsen, I.E.H., Magnusson Hanson, L.L., Lallukka, T., Nyberg, S.T., Alfredsson, L., Batty, G.D., Bjorner, J.B., Borritz, M., Burr, H., Dragano, N., Erbel, R., Ferrie, J.E., Heikkilä, K., Knutsson, A., Koskenvuo, M., Nielsen, M.L., Oksanen, T., Pejtersen, J.H., Pentti, J., Rahkonen, O., Rugulies, R., Salo, P., Schupp, J., Shipley, M.J., Siegrist, J., Singh-Manoux, A., Suominen, S.B., Theorell, T., Vahtera, J., Wagner, G.G., Wang, J.L., Yiengprugsawan, V., Westerlund, H., Kivimäki,

M., 2018. Long working hours and depressive symptoms: systematic review and meta-analysis of published studies and unpublished individual participant data. Scand. J. Work Environ. Health 44, 239–250.

Viswanathan, M., Ansari, M.T., Berkman, N.D., Chang, S., Hartling, L., McPheeters, M., Santaguida, P.L., Shamliyan, T., Singh, K., Tsertsvadze, A., Treadwell, J.R., 2008. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews, Rockville (MD).

Watanabe, K., Imamura, K., Kawakami, N., 2016. Working hours and the onset of depressive disorder: a systematic review and meta-analysis. Occup. Environ. Med. 73, 877–884.

Wirtz, A., Lombardi, D.A., Willetts, J.L., Folkard, S., Christiani, D.C., 2012. Gender differences in the effect of weekly working hours on occupational injury risk in the United States working population. Scand. J. Work Environ. Health 38, 349–357.

Wittchen, H.U., Jacobi, F., 2005. Size and burden of mental disorders in Europe-a critical review and appraisal of 27 studies. Eur. Neuropsychopharmacol. 15, 357–376.

Woodruff, T.J., Sutton, P., 2014. The navigation guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ. Health Perspect. 122, 1007–1014.

World Health Organization, 2015. ICD-10: International Statistical Classification of Diseases and Related Health Problems: 10th Revision. World Health Organization, Geneva.

World Health Organization, 2017a. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, Geneva.

World Health Organization, 2017b. WHO Methods and Data Sources for Global Burden of Disease Estimates 2000–2015. World Health Organization, Geneva.