

1 **SYSTEMATIC REVIEW PROTOCOL**

2  
3 **WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of**  
4 **exposure to long working hours and of the effect of exposure to long working hours on ischaemic**  
5 **heart disease**

6  
7 Jian Li, MD, PhD <sup>1\*</sup>

8 Chantal Brisson, PhD <sup>2</sup>

9 Els Clays, PhD <sup>3</sup>

10 Marco M. Ferrario, MD, PhD <sup>4</sup>

11 Ivan D. Ivanov, MD, PhD <sup>5</sup>

12 Paul Landsbergis, PhD, EdD, MPH <sup>6</sup>

13 Nancy Leppink, JD <sup>7</sup>

14 Frank Pega, PhD <sup>8</sup>

15 Hynek Pikhart, PhD <sup>9</sup>

16 Annette Prüss-Üstün, PhD <sup>10</sup>

17 Reiner Rugulies, PhD <sup>11</sup>

18 Peter L. Schnall, MD, MPH <sup>12</sup>

19 Gretchen Stevens, PhD <sup>13</sup>

20 Akizumi Tsutsumi, MD, DMSc <sup>14</sup>

21 Yuka Ujita, MD, PhD <sup>15</sup>

22 Johannes Siegrist, PhD <sup>16\*</sup>

23  
24 \* These authors have contributed equally to this work.

25  
26 <sup>1</sup> Institute of Occupational, Social and Environmental Medicine, Centre for Health and Society, Faculty  
27 of Medicine, University of Düsseldorf, Universitätsstraße 1, Düsseldorf 40225, Germany,  
28 [jian.li@uni-duesseldorf.de](mailto:jian.li@uni-duesseldorf.de)

29 <sup>2</sup> Centre de Recherche du CHU de Québec- Université Laval, 1050 Chemin Ste-Foy, Quebec City, G1S  
30 4L8, Quebec, Canada, [Chantal.Brisson@crchudequebec.ulaval.ca](mailto:Chantal.Brisson@crchudequebec.ulaval.ca)

31 <sup>3</sup> Department of Public Health, Ghent University, Campus University Hospital, 4K3, De Pintelaan 185,  
32 B-9000 Ghent, Belgium, [els.clays@UGent.be](mailto:els.clays@UGent.be)

33 <sup>4</sup> Research Centre EPIMED, University of Insubria, Via O Rossi 9, 21100 Varese, Italy,  
34 [marco.ferrario@uninsubria.it](mailto:marco.ferrario@uninsubria.it)

35 <sup>5</sup> Department of Public Health, Environmental and Social Determinants of Health, World Health  
36 Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, [ivanovi@who.int](mailto:ivanovi@who.int)

37 <sup>6</sup> State University of New York-Downstate School of Public Health, 450 Clarkson Ave., Brooklyn, NY  
38 11238, United States of America, [paul.landsbergis@downstate.edu](mailto:paul.landsbergis@downstate.edu)

39 <sup>7</sup> Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International  
40 Labour Organization, Route des Morillons 4, 1211 Geneva, Switzerland, [leppink@ilo.org](mailto:leppink@ilo.org)

41 <sup>8</sup> Department of Public Health, Environmental and Social Determinants of Health, World Health  
42 Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, [pegaf@who.int](mailto:pegaf@who.int)

43 <sup>9</sup> Institute of Epidemiology and Health Care, University College London, 1-19 Torrington Place,  
44 London WC1E 6BT, United Kingdom, [h.pikhart@ucl.ac.uk](mailto:h.pikhart@ucl.ac.uk)

45 <sup>10</sup> Department of Public Health, Environmental and Social Determinants of Health, World Health  
46 Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland, [pruesssa@who.int](mailto:pruesssa@who.int)

47 <sup>11</sup> National Research Centre for the Working Environment, Lersø Parkallé 105; DK-2100 Copenhagen,  
48 Denmark, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, DK-1014

49 Copenhagen, Denmark; and Department of Psychology, University of Copenhagen, Øster  
50 Farimagsgade 2A, DK-1353 Copenhagen, Denmark, [rer@nrcwe.dk](mailto:rer@nrcwe.dk)

51 <sup>12</sup> Center for Occupational and Environmental Health, University of California-Irvine, 100 Theory Way,  
52 Irvine, California, United States of America, [pschnall@workhealth.org](mailto:pschnall@workhealth.org)

53 <sup>13</sup> Department of Information, Evidence and Research, World Health Organization, 20 Avenue Appia,  
54 1211 Geneva 27, Switzerland, [stevensg@who.int](mailto:stevensg@who.int)

55 <sup>14</sup> Department of Public Health, School of Medicine, Kitasato University, 1-15-1 Kitasato, Minami,  
56 Sagamihara, 252-0374, Japan, [akizumi@kitasato-u.ac.jp](mailto:akizumi@kitasato-u.ac.jp)

57 <sup>15</sup> Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International  
58 Labour Organization, Route des Morillons 4, 1211 Geneva, Switzerland, [ujita@ilo.org](mailto:ujita@ilo.org)

59 <sup>16</sup> Life Science Centre, University of Düsseldorf, Merowingerplatz 1a, Düsseldorf 40225, Germany,  
60 [johannes.siegrist@med.uni-duesseldorf.de](mailto:johannes.siegrist@med.uni-duesseldorf.de)

61

62 Corresponding author: Dr. Jian Li; Universitätsstraße 1, Düsseldorf 40225, Germany, Phone: +49-211-  
63 81-08161.

64

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

66

67 Abstract: 459                      Main text: 8,794                      References: 80                      Figures: 1                      Tables: 2  
68 Appendices: 10

69

70 **Abstract**

71

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

78

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

85

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

91

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

101

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

110

111 *PROSPERO registration number:* CRD42017084243

112

113

114 **BACKGROUND**

115

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

128

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

139

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

148

149 **Rationale**

150

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

163

164 Our research will substantially extend the current body of systematic review evidence. For instance, a  
165 2012 systematic review and meta-analysis on the effect of exposure with long working hours on  
166 cardiovascular disease, which included five cohort studies and six case-control studies published up to

167 September 2011, reported a pooled odds ratio of 1.37, with a 95% confidence interval (CI) of 1.11-1.70  
168 <sup>16</sup>. A second systematic review on the effect of long working hours on ischaemic heart disease published  
169 in 2012 included four prospective studies and seven case-control studies published between 1966 and  
170 19 January 2011. For the prospective studies, the authors reported a pooled relative risk of 1.39 (95%  
171 CI: 1.12-1.72), and for the case-control studies a pooled relative risk of 2.43 (95% CI: 1.81-3.26) <sup>17</sup>.  
172 Finally, a third systematic review and meta-analysis published in 2015 of 24 cohort studies (including  
173 20 unpublished studies) in Europe, the USA and Australia up to August 20th 2014 found a relative risk  
174 of 1.13 (95% CI: 1.02-1.26) for the effect of long working hours ( $\geq 55$  hours/week) on ischaemic heart  
175 disease <sup>18</sup>. However, our Systematic Review 1 will be the – to the best of our knowledge – first  
176 systematic review of prevalence studies of exposure to long working hours, and Systematic Review 2  
177 will expand the scope of the existing systematic review evidence by covering evidence from studies  
178 published up to 31 May 2018.

179

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

190

## 191 **Description of the risk factor**

192

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

203

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

217

218 If several studies report exposure levels differing from the standard levels we define here, then, if  
219 possible, we will convert the reported levels to the standard levels and, if not possible, we will report

220 analyses on these alternate exposure levels as supplementary information in the systematic reviews.. In  
 221 the latter case, our protocol will be updated to reflect our new analyses.

222  
 223  
 224

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

	<b>Definition</b>
Risk factor	Long working hours (including those spent in secondary jobs), defined as working hours >40 hours/week, i.e. working hours exceeding standard working hours (35-40 hours/week).
Risk factor levels	Preferable exposure categories are 35-40, 41-48, 49-54 and ≥55 hours/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 hours/week.
Theoretical minimum risk exposure level	Standard working hours defined as working hours of 35-40 hours/week.

225  
 226  
 227

**Description of the outcome**

228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237

The WHO *Global Health Estimates* group outcomes into standard burden of disease categories,<sup>5</sup> based on standard codes from the *International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10)* <sup>22</sup>. The relevant WHO *Global Health Estimates* category for this systematic review is “*II.H.2 Ischaemic heart disease*”.<sup>5</sup> In line with the WHO *Global Health Estimates*, we define the health outcome covered in Systematic Review 2 as ischaemic heart disease, defined as conditions with ICD-10 codes I120 to I125 (Table 2). We will consider prevalence of, incidence of and mortality from ischaemic heart disease. Table 2 presents for each disease or health problem included in the WHO *Global Health Estimates* category the inclusion in this review. This review covers all the relevant WHO *Global Health Estimates* categories.

238 **Table 2: ICD-10 codes and disease and health problems covered by the WHO burden of disease**  
 239 **category *II.H.2 Ischaemic heart disease* and their inclusion in this review**  
 240

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

241  
 242  
 243

**How the risk factor may impact the outcome**

244  
 245  
 246

Figure 1 presents the logic model for our systematic review of the causal relationship between exposure to long working hours and ischaemic heart diseases. This logic model is an *a priori*, process-orientated one <sup>23</sup> that seeks to capture the complexity of the risk factor–outcome causal relationship.<sup>24</sup>

247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281

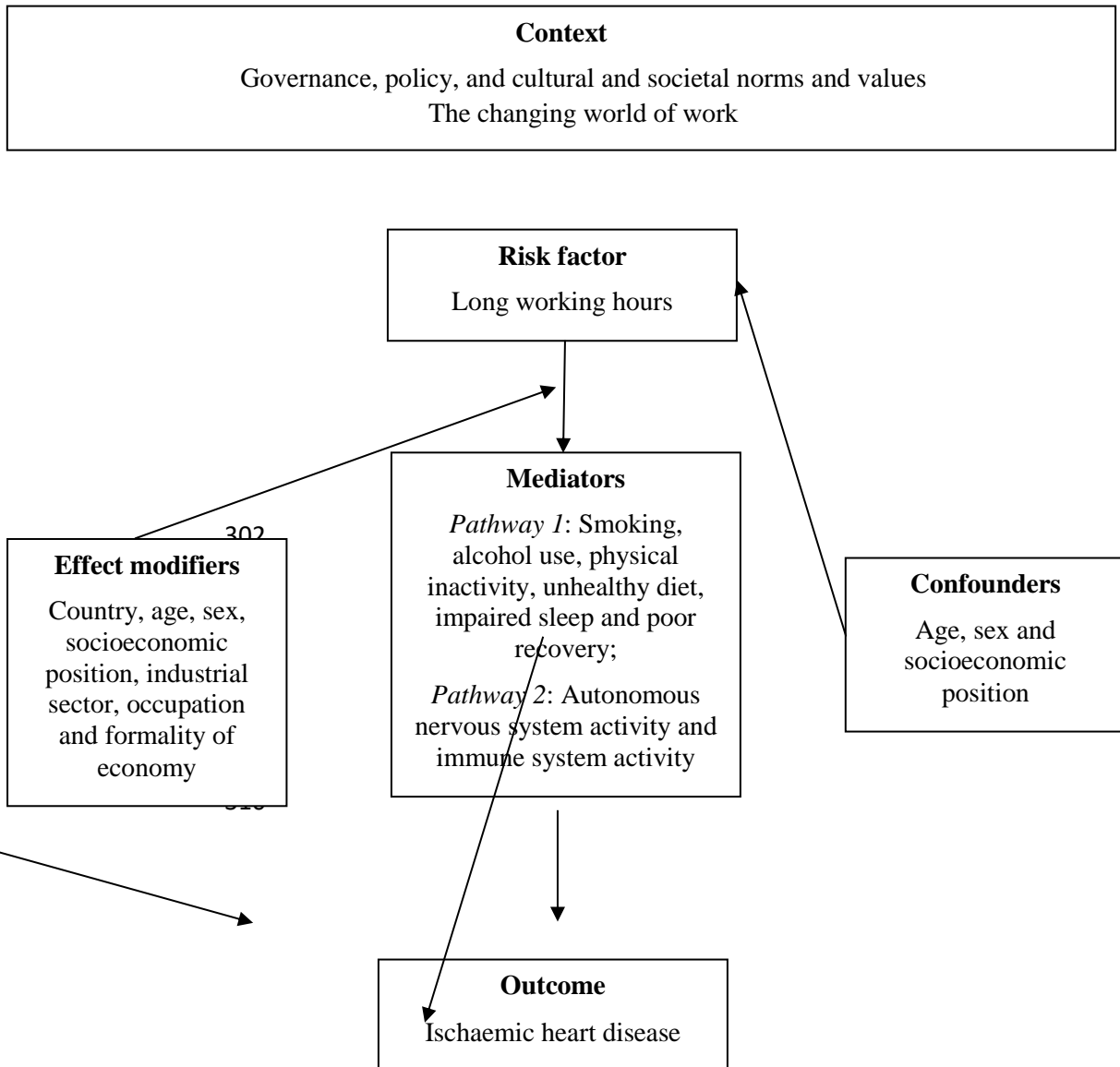
Theoretically, distinct social contexts in labour market are likely to exacerbate or mitigate the effect of exposure to long working hours on ischaemic heart disease risk. While empirical tests of this assumption are not available, these contexts can exert a direct effect on working hours. Evidence suggests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

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

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

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

282 **Figure 1: Logic model of the causal relationship between long working hours and ischaemic heart**  
 283 **disease**



302

311

320 **OBJECTIVES**

- 322 1. Systematic Review 1: To systematically review quantitative studies of any design on the  
 323 prevalence of relevant levels of exposure to long working hours in the years 2005 to 2018  
 324 among the working-age population, disaggregated by country, sex, age and industrial sector or  
 325 occupation. Systematic Review 1 will be conducted in a coordinated fashion across all four  
 326 review groups that examine long working hours with regard to health endpoints (ischaemic  
 327 heart disease, stroke <sup>6</sup>, depression <sup>11</sup> and alcohol use <sup>10</sup>), led by GS and with JL being the focal  
 328 point from the working group on long working hours and ischaemic heart disease.
- 329 2. Systematic Review 2: To systematically review and meta-analyse randomized controlled  
 330 studies, cohort studies, case-control studies, and other non-randomized intervention studies  
 331 including working-age workers (Population) exposed to long working hours (Exposure),  
 332 compared with workers with the minimum theoretical risk exposure level of 35-40 hours/week  
 333 (Comparator), in order to estimate the relative effect on ischaemic heart disease (Outcome).



334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382

## METHODS

We will apply the *Navigation Guide*<sup>34</sup> 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<sup>34</sup>. The need for further methodological development and refinement of the relatively novel *Navigation Guide* has been acknowledged<sup>34</sup>.

Systematic Review 1 may not map well to the *Navigation Guide* framework (Figure 1 on page 1009 in<sup>35</sup>), 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.

<sup>35</sup>  
We have registered the protocol in PROSPERO under CRD42018084131. This protocol adheres with the preferred reporting items for systematic review and meta-analysis protocols statement (PRISMA-P)<sup>36,37</sup>, with the abstract adhering with the reporting items for systematic reviews in journal and conference abstracts (PRISMA-A)<sup>38</sup>. 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<sup>15</sup>, and Systematic Review 2 will be reported according to the preferred reporting items for systematic review and meta-analysis statement (PRISMA)<sup>39</sup>. Our reporting of the parameters for estimating the burden of ischaemic heart disease to long working hours in the systematic review will adhere with the requirements of the GATHER guidelines<sup>15</sup>, because the WHO/ILO burden of disease estimates that may be produced consecutive to the systematic review must also adhere to these reporting guidelines.

### Systematic Review 1

#### *Eligibility criteria*

The population, exposure, comparator and outcome (PECO) criteria<sup>39</sup> are described below.

#### *Types of populations (P)*

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

#### *Types of exposures (E)*

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

395  
396 We will include studies on the prevalence of occupational exposure to the risk factor, if it is  
397 disaggregated by country, sex (two categories: female, male), age group (ideally in 5-year age bands,  
398 such as 20–24 years) and industrial sector (e.g. *International Standard Industrial Classification of All  
399 Economic Activities, Revision 4* [ISIC Rev. 4])<sup>40</sup> or occupation (as defined, for example, by the  
400 *International Standard Classification of Occupations 1988* [ISCO-88]<sup>41</sup> or 2008 [ISCO-08]<sup>42</sup>). We will  
401 also extract data on the context of risk factor exposure. Criteria may be revised in order to identify  
402 optimal data disaggregation to enable subsequent estimation of the burden of disease.

403  
404 We shall include studies with exposure data for the years 2005 to 31 May 2018. For optimal modelling  
405 of exposure, WHO and ILO require exposure data up to 2018, because recent data points help better  
406 estimate time trends, especially where data points may be sparse. The additional rationale for this data  
407 collection window is that the WHO and ILO aim to estimate burden of disease in the year 2015, and we  
408 believe that the lag time from exposure to outcome will not exceed 10 years; so in their models, the  
409 organizations can use the exposure data from as early as 2005 to determine the burden of ischaemic  
410 heart disease 10 years later in 2015. To make a conclusive judgment on the best lag time to apply in the  
411 model, we will summarize the existing body of evidence on the lag time between exposure to long  
412 working hours and ischaemic heart disease in the review.

413  
414 Both objective and subjective measures will be included. If both subjective and objective measures are  
415 presented, then we will prioritize objective ones. Studies with measures from any data source, including  
416 registries, will be eligible. The exposure parameter should match the one used in Systematic Review 2  
417 or can be converted to match it.

418  
419 *Types of comparators*

420  
421 There will be no comparator, because we will review risk factor prevalence only.

422  
423 *Types of outcomes*

424  
425 Exposure to the occupational risk factor (i.e. long working hours).

426  
427 *Types of studies*

428  
429 This Systematic Review will include quantitative studies of any design, including cross-sectional  
430 studies. These studies must be representative of the relevant industrial sector, relevant occupational

431 group or the national population. We will exclude qualitative, modelling, and case studies, as well as  
432 non-original studies without quantitative data (e.g. letters, commentaries and perspectives).

433

434 Study records written in any language will be included. If a study record is written in a language other  
435 than those spoken by the authors of this review or those of other reviews<sup>6-13</sup> in the series (i.e. Arabic,  
436 Bulgarian, Chinese, Danish, Dutch, English, French, Finnish, German, Hungarian, Italian, Japanese,  
437 Norwegian, Portuguese, Russian, Spanish and Swedish), it will be translated into English. Published  
438 and unpublished studies will be included.

439

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

442

#### 443 *Types of effect measures*

444

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

447

#### 448 ***Information sources and search***

449

##### 450 *Electronic academic databases*

451

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

- 453 1. Ovid Medline with Daily Update (2005 to 31 May 2018).
- 454 2. PubMed (2005 to 31 May 2018).
- 455 3. EMBASE (2005 to 31 May 2018).
- 456 4. Scopus (2005 to 31 May 2018).
- 457 5. Web of Science (2005 to 31 May 2018).
- 458 6. CISDOC (2005 to 31 May 2012).
- 459 7. PsychInfo (2005 to 31 May 2018).

460

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

468

##### 469 *Electronic grey literature databases*

470

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

- 472 1. OpenGrey (<http://www.opengrey.eu/>)
- 473 2. Grey Literature Report (<http://greylit.org/>)

474

##### 475 *Internet search engines*

476

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

480

#### 481 *Organizational websites*

482

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

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

494

#### 495 *Hand-searching and expert consultation*

496

497 AD, DG, JP, and GS will hand-search for potentially eligible studies in:

- 498 • Reference list of previous systematic reviews.
- 499 • Reference list of all study records of all included studies.
- 500 • Study records published over the past 24 months in the three peer-reviewed academic journals  
501 from which we obtain the largest number of included studies.
- 502 • Study records that have cited an included study record (identified in Web of Science citation  
503 database).
- 504 • Collections of the review authors.

505

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

508

#### 509 *Study selection*

510

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

520

#### 521 *Data extraction and data items*

522

523 A data extraction form will be developed and piloted until there is convergence and agreement among  
524 data extractors. At a minimum, two review authors (out of: BAE, ES and LMH) will independently  
525 extract the data on exposure to long working hours, disaggregated by country, sex, age and industrial  
526 sector or occupation. A third review author (GS) will resolve conflicting extractions. At a minimum,  
527 we will extract data on study characteristics (including study authors, study year, study country,  
528 participants, exposure and outcome), study design (including study type and measurements of the risk  
529 factor and outcome, and response rate), risk of bias (including missing data, as indicated by response  
530 rate and other measures) and study context. The estimates of the proportion of the population exposed  
531 to the occupational risk factor from included studies will be entered into and managed with, the Review  
532 Manager, Version 5.3 (RevMan 5.3)<sup>46</sup> or DistillerSR<sup>47</sup> softwares.

533

534 We will also extract data on potential conflict of interest in included studies, including the financial  
535 disclosures and funding sources of each author and their affiliated organization. We will use a  
536 modification of a previous method to identify and assess undisclosed financial interests<sup>48</sup>. Where no  
537 financial disclosure/conflict of interest is provided, we will search declarations of interest both in other  
538 records from this study published in the 36 months prior to the included study record and in other  
539 publicly available repositories<sup>49,50</sup>.

540

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

544

#### 545 ***Risk of bias assessment***

546

547 Generally agreed methods (i.e. framework plus tool) for assessing risk of bias do not exist for systematic  
548 reviews of input data for health estimates<sup>51</sup>, for burden of disease studies, of prevalence studies in  
549 general<sup>52</sup>, and those of prevalence studies of occupational and/or environmental risk factors specifically  
550<sup>53-55</sup>. None of the five standard risk of bias assessment methods in systematic reviews<sup>56</sup> are applicable  
551 to assessing prevalence studies. The *Navigation Guide* does not support checklist approaches, such as  
552<sup>52,57</sup>, for assessing risk of bias in prevalence studies.

553

554 We will use a modified version of the *Navigation Guide* risk of bias tool<sup>35</sup> that we developed  
555 specifically for Systematic Review 1 (Appendix C). We will assess risk of bias on the levels of the  
556 individual study and the entire body of evidence. As per our preliminary tool, we will assess risk of bias  
557 along five domains: (i) selection bias; (ii) performance bias; (iii) misclassification bias; (iv) conflict of  
558 interest; and (v) other biases. Risk of bias will be: “low”; “probably low”; “probably high”; “high” or  
559 “not applicable”. To judge the risk of bias in each domain, we will apply our *a priori* instructions  
560 (Appendix C).

561

562 All risk of bias assessors (BAE, DG, LMH and GS) will trial the tool until they synchronize their  
563 understanding and application of each risk of bias domain, considerations and criteria for ratings. At  
564 least two study authors (out of: BAE, DG and LMH) will then independently judge the risk of bias for  
565 each study by outcome, and a third author (GS) will resolve any conflicting judgments. We will present  
566 the findings of our risk of bias assessment for each eligible study in a standard ‘Risk of bias’ table<sup>58</sup>.  
567 Our risk of bias assessment for the entire body of evidence will be presented in a standard ‘Risk of bias  
568 summary’ figure<sup>58</sup>.

569

#### 570 ***Synthesis of results***

571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617

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

### ***Quality of evidence assessment***

There is no agreed method for assessing quality of evidence in systematic reviews of the prevalence of occupational and/or environmental risk factors. We will adopt/adapt from the latest *Navigation Guide* instructions for grading<sup>35</sup>, 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<sup>59</sup>. We will grade the evidence, using the three *Navigation Guide* quality of evidence ratings: “high”, “moderate” and “low”<sup>35</sup>. Within each of the relevant reasons for downgrading, we will rate any concern per reason as “none”, “serious” or “very serious”. We will start at “high” for non-randomized studies and will downgrade for no concern by nil, for a serious concern by one grade (-1), and for a very serious concern by two grades (-2). We will not up-grade or down-grade the quality of evidence for the three other reasons normally considered in GRADE assessments (i.e. large effect, dose-response and plausible residual confounding and bias), because we consider them irrelevant for prevalence estimates.

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

### ***Strength of evidence assessment***

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

## **Systematic Review 2**

### ***Eligibility criteria***

The PECO<sup>39</sup> criteria are described below.

### ***Types of populations***

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

626

#### 627 *Types of exposures*

628

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

638

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

642

#### 643 *Types of comparators*

644

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

647

#### 648 *Types of outcomes*

649

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

655

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

657

i) Diagnosis by a physician with imaging.

658

ii) Hospital discharge records.

659

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

660

iv) Medically certified cause of death.

661

All other measures will be excluded from this systematic review.

662

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

665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712

*Types of studies*

We will include studies that investigate the effect of long working hours on ischaemic heart disease for any years. Eligible study designs will be randomized controlled trials (including parallel-group, cluster, cross-over and factorial trials), cohort studies (both prospective and retrospective), case-control studies, and other non-randomized intervention studies (including quasi-randomized controlled trials, controlled before-after studies and interrupted time series studies). We included a broader set of observational study designs than is commonly included, because a recent augmented Cochrane Review of complex interventions identified valuable additional studies using such a broader set of study designs<sup>60</sup>. As we have an interest in quantifying risk and not in qualitative assessment of hazard<sup>61</sup>, we will exclude all other study designs (e.g. uncontrolled before-and-after, cross-sectional, qualitative, modelling, case and non-original studies).

Records published in any year and any language will be included. Again, the search will be conducted using English language terms, so that records published in any language that present essential information (i.e. title and abstract) in English will be included. If a record is written in a language other than those spoken by the authors of this review or those of other reviews in the series<sup>6-13</sup>, then the record will be translated into English. Published and unpublished studies will be included.

Studies conducted using unethical practices will be excluded.

*Types of effect measures*

We will include measures of the relative effect of a relevant level of long working hours on the risk of having, developing or dying from ischaemic heart disease, compared with the theoretical minimum risk exposure level. Included relative effect measures are risk ratios and odds ratios for prevalence and mortality measures and hazard ratios for incidence measures (e.g., developed or died from ischaemic heart disease). Measures of absolute effects will be excluded (e.g. mean differences in risks or odds). Measures of absolute effects (e.g. mean differences in risks or odds) will be converted into relative effect measures, but if conversion is impossible, they will be excluded. To ensure comparability of effect estimates and facilitate meta-analysis, if a study presents an odds ratio, then we will convert it into a risk ratio, if possible, using the guidance provided in the Cochrane Collaboration's handbook for systematic reviews of interventions<sup>62</sup>.

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

If a study presents estimates for the effect from two or more alternative models that have been adjusted for different variables, then we will systematically prioritize the estimate from the model that we consider best adjusted, applying the lists of confounders and mediators identified in our logic model (Figure 1). We will prioritize estimates from models adjusted for more potential confounders over those from models adjusted for fewer. For example, if a study presents estimates from a crude, unadjusted model (Model A), a model adjusted for one potential confounder (Model B) and a model adjusted for



713 two potential confounders (Model C), then we will prioritize the estimate from Model C. We will  
714 prioritize estimates from models unadjusted for mediators over those from models that adjusted for  
715 mediators, because adjustment for mediators can introduce bias. For example, if Model A has been  
716 adjusted for two confounders, and Model B has been adjusted for the same two confounders and a  
717 potential mediator, then we will choose the estimate from Model A over that from Model B. We  
718 prioritize estimates from models that can adjust for time-varying confounders that are at the same time  
719 also mediators, such as marginal structural models<sup>63</sup> over estimates from models that can only adjust  
720 for time-varying confounders, such as fixed-effects models<sup>64</sup>, over estimates from models that cannot  
721 adjust for time-varying confounding. If a study presents effect estimates from two or more potentially  
722 eligible models, then we will explain specifically why we prioritized the selected model.

723

## 724 *Information sources and search*

725

### 726 *Electronic academic databases*

727

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

- 729 1. International Clinical Trials Register Platform (to May 31<sup>st</sup> 2018).
- 730 2. Ovid MEDLINE with Daily Update (1946 to May 31<sup>st</sup> 2018).
- 731 3. PubMed (1946 to May 31<sup>st</sup> 2018).
- 732 4. EMBASE (1947 to May 31<sup>st</sup> 2018).
- 733 5. Scopus (1788 to May 31<sup>st</sup> 2018).
- 734 6. Web of Science (1945 to May 31<sup>st</sup> 2018).
- 735 7. CISDOC (1901 to 2012).
- 736 8. PsychInfo (1880 to May 31<sup>st</sup> 2018).

737

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

744

### 745 *Electronic grey literature databases*

746

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

- 748 1. OpenGrey (<http://www.opengrey.eu/>)
- 749 2. Grey Literature Report (<http://greyLit.org/>)

750

### 751 *Internet search engines*

752

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

756

### 757 *Organizational websites*

758

759 The websites of the seven following international organizations and national government departments  
760 will be searched for both systematic reviews by GS and HP:

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

770

771 *Hand-searching and expert consultation*

772

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

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

781

782 Additional experts will be contacted with a list of included studies, with the request to identify

783 potentially eligible additional studies.

784

785 *Study selection*

786

787 Study selection will be carried out with the Rayyan Systematic Reviews Web App<sup>45</sup>. All study records

788 identified in the search will be downloaded and duplicates will be identified and deleted. Afterwards,

789 at least two review authors (PLS and JL), working in pairs, will independently screen titles and abstracts

790 (step 1) and then full texts (step 2) of potentially relevant records. A third review author (JS) will resolve

791 any disagreements between the two review authors. If a study record identified in the literature search

792 was authored by a review author assigned to study selection or if an assigned review author was

793 involved in the study, then the record will be re-assigned to another review author for study selection. The

794 study selection will be documented in a flow chart in the systematic review, as per PRISMA guidelines

795<sup>39</sup>.

796

797 *Data extraction and data items*

798

799 A data extraction form will be developed and trialled until data extractors reach convergence and

800 agreement. At a minimum, two review authors (RR and JL) will extract data on study characteristics

801 (including study authors, study year, study country, participants, exposure and outcome), study design

802 (including summary of study design, comparator, epidemiological models used and effect estimate

803 measure), risk of bias (including selection bias, reporting bias, confounding, and reverse causation) and

804 study context (e.g. data on contemporaneous exposure to other occupational risk factors potentially

805 relevant for deaths or other health loss from ischaemic heart disease). A third review author (JS) will

806 resolve conflicts in data extraction. Data will be entered into and managed with the Review Manager,

807 Version 5.3 (RevMan 5.3)<sup>46</sup> or DistillerSR<sup>47</sup> softwares, but the Health Assessment Workspace

808 Collaborative (HAWC, <http://hawc.readthedocs.io/en/latest/>)<sup>65</sup> may also be used in parallel or to  
809 prepare data for entry into RevMan 5.3.

810

811 We will also extract data on potential conflict of interest in included studies. For each author and  
812 affiliated organization of each included study record, we will extract their financial disclosures and  
813 funding sources. We will use a modification of a previous method to identify and assess undisclosed  
814 financial interest of authors<sup>48</sup>. Where no financial disclosure or conflict of interest statements are  
815 available, we will search the name of all authors in other study records gathered for this study and  
816 published in the prior 36 months and in other publicly available declarations of interests<sup>49,50</sup>.

817

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

821

### 822 ***Risk of bias assessment***

823

824 Standard risk of bias tools do not exist for systematic reviews for hazard identification in occupational  
825 and environmental health, nor for risk assessment. The five methods specifically developed for  
826 occupational and environmental health are for either or both hazard identification and risk assessment,  
827 and they differ substantially in the types of studies (randomized, observational and/or simulation  
828 studies) and data (e.g. human, animal and/or *in vitro*) they seek to assess<sup>56</sup>. However, all five methods,  
829 including the *Navigation Guide*<sup>35</sup>, assess risk of bias in human studies similarly<sup>56</sup>.

830

831 The *Navigation Guide* was specifically developed to translate the rigor and transparency of systematic  
832 review methods applied in the clinical sciences to the evidence stream and decision context of  
833 environmental health<sup>34</sup>, which includes workplace environment exposures and associated health  
834 outcomes. The guide is our overall organizing framework, and we will also apply its risk of bias  
835 assessment method in Systematic Review 2. The *Navigation Guide* risk of bias assessment method  
836 builds on the standard risk of bias assessment methods of the Cochrane Collaboration<sup>58</sup> and the US  
837 Agency for Healthcare Research and Quality<sup>66</sup>. Some further refinements of the *Navigation Guide*  
838 method may be warranted<sup>67</sup>, but it has been successfully applied in several completed and ongoing  
839 systematic reviews<sup>68-75</sup>. In our application of the *Navigation Guide* method, we will draw heavily on  
840 one of its latest versions, as presented in the protocol for an ongoing systematic review<sup>35</sup>. Should a  
841 more suitable method become available, we may switch to it.

842

843 We will assess risk of bias on the individual study level and on the body of evidence overall. The nine  
844 risk of bias domains included in the *Navigation Guide* method for human studies are: (i) source  
845 population representation; (ii) blinding; (iii) exposure assessment; (iv) outcome assessment; (v)  
846 confounding; (vi) incomplete outcome data; (vii) selective outcome reporting; (viii) conflict of interest;  
847 and (ix) other sources of bias. While two of the earlier case studies of the *Navigation Guide* did not  
848 utilize outcome assessment as a risk of bias domain for studies of human data<sup>68-71</sup>, all of the subsequent  
849 reviews have included this domain<sup>35,72-75</sup>. Risk of bias or confounding ratings will be: “low”; “probably  
850 low”; “probably high”; “high” or “not applicable”<sup>35</sup>. To judge the risk of bias in each domain, we will  
851 apply *a priori* instructions (Appendix H), which we have adopted or adapted from an ongoing  
852 *Navigation Guide* systematic review<sup>35</sup>). For example, a study will be assessed as carrying “low” risk  
853 of bias from source population representation, if we judge the source population to be described in  
854 sufficient detail (including eligibility criteria, recruitment, enrollment, participation and loss to follow  
855 up) and the distribution and characteristics of the study sample to indicate minimal or no risk of

856 selection effects. The risk of bias at study level will be determined by the worst rating in any bias  
857 domain for any outcome. For example, if a study is rated as “probably high” risk of bias in one domain  
858 for one outcome and “low” risk of bias in all other domains for the outcome and in all domains for all  
859 other outcomes, the study will be rated as having a “probably high” risk of bias overall.

860

861 All risk of bias assessors (EC, AT and PL) will jointly trial the application of the risk of bias criteria  
862 until they have synchronized their understanding and application of these criteria. At least two study  
863 authors (EC and AT) will independently judge the risk of bias for each study by outcome. Where  
864 individual assessments differ, a third author (PL) will resolve the conflict. In the systematic review, for  
865 each included study, we will report our study-level risk of bias assessment by domain in a standard  
866 ‘Risk of bias’ table <sup>58</sup>. For the entire body of evidence, we will present the study-level risk of bias  
867 assessments in a ‘Risk of bias summary’ figure <sup>58</sup>.

868

### 869 *Synthesis of results*

870

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

883

884 If we judge two or more studies for the relevant combination of country, sex and age group, or  
885 combination thereof, to be sufficiently clinically homogenous to potentially be combined quantitatively  
886 using quantitative meta-analysis, then we will test the statistical heterogeneity of the studies using the  
887  $I^2$  statistic <sup>76</sup>. If two or more clinically homogenous studies are found to be sufficiently homogenous  
888 statistically to be combined in a meta-analysis, we will pool the risk ratios of the studies in a quantitative  
889 meta-analysis, using the inverse variance method with a random effects model to account for cross-  
890 study heterogeneity <sup>76</sup>. The meta-analysis will be conducted in RevMan 5.3, but the data for entry into  
891 these programmes may be prepared using another recognized statistical analysis programme, such as  
892 Stata. We will neither quantitatively combine data from studies with different designs (e.g. combining  
893 cohort studies with case-controls studies), nor unadjusted and adjusted models. We will only combine  
894 studies that we judge to have a minimum acceptable level of adjustment for confounders. If quantitative  
895 synthesis is not feasible, then we will synthesise the study findings narratively and identify the estimates  
896 that we judged to be the highest quality evidence available.

897

### 898 *Additional analyses*

899

900 If we source micro-data on exposure, outcome and potential confounding variables, we may conduct  
901 meta-regressions to adjust optimally for potential confounders.

902

903 If there is evidence for differences in effect estimates by country, sex, age, industrial sector and/or  
904 occupation, or by a combination of these variables, then we will conduct subgroup analyses by the  
905 relevant variable or combination of variables, as feasible. Where both studies on workers in the informal  
906 economy and in the formal economy are included, then we will conduct sub-group analyses by formality  
907 of economy. Findings of these subgroup analyses, if any, will be used as parameters for estimating  
908 burden of disease specifically for relevant populations defined by these variables. We will also conduct  
909 subgroup analyses by study design (e.g. randomized controlled trials versus cohort studies versus case-  
910 control studies).

911  
912 We will perform a sensitivity analyses that will include only studies judged to be of “low” or “probably  
913 low” risk of bias from conflict of interest; judged to be of “low” or “probably low” risk of bias; and  
914 with documented or approximated ICD-10 diagnostic codes. We may also conduct a sensitivity analysis  
915 using an alternative meta-analytic model, namely the inverse variance heterogeneity (IVhet) model <sup>77</sup>.

916

### 917 ***Quality of evidence assessment***

918

919 We will assess quality of evidence using a modified version of the *Navigation Guide* quality of evidence  
920 assessment tool <sup>35</sup>. The tool is based on the GRADE approach <sup>59</sup> adapted specifically to systematic  
921 reviews in occupational and environmental health <sup>78</sup>. Should a more suitable method become available,  
922 we may switch to it.

923

924 At least two review authors (JS and JL) will assess quality of evidence for the entire body of evidence  
925 by outcome, with any disagreements resolved by a third review author. We will adopt or adapt the latest  
926 *Navigation Guide* instructions (Appendix D) for grading the quality of evidence <sup>35</sup>. We will downgrade  
927 the quality of evidence for the following five GRADE reasons: (i) risk of bias; (ii) inconsistency; (iii)  
928 indirectness; (iv) imprecision; and (v) publication bias. If our systematic review includes ten or more  
929 studies, we will generate a funnel plot to judge concerns on publication bias. If it includes nine or fewer  
930 studies, we will judge the risk of publication bias qualitatively. To assess risk of bias from selective  
931 reporting, protocols of included studies, if any, will be screened to identify instances of selective  
932 reporting.

933

934 We will grade the evidence, using the three *Navigation Guide* standard quality of evidence ratings:  
935 “high”, “moderate” and “low” <sup>35</sup>. Within each of the relevant domains, we will rate the concern for the  
936 quality of evidence, using the ratings “none”, “serious” and “very serious” . As per *Navigation Guide*,  
937 we will start at “high” for randomized studies and “moderate” for observational studies. Quality will be  
938 downgrade for no concern by nil grades (0), for a serious concern by one grade (-1) and for a very  
939 serious concern by two grades (-2). We will up-grade the quality of evidence for the following other  
940 reasons: large effect, dose-response and plausible residual confounding and bias. For example, if we  
941 have a serious concern for risk of bias in a body of evidence consisting of observational studies (-1),  
942 but no other concerns, and there are no reasons for upgrading, then we will downgrade its quality of  
943 evidence by one grade from “moderate” to “low”.

944

### 945 ***Strength of evidence assessment***

946

947 We will apply the standard *Navigation Guide* methodology <sup>35</sup> to rate the strength of the evidence.  
948 The rating will be based on a combination of the following four criteria: (i) quality of the body of  
949 evidence; (ii) direction of the effect; (iii) confidence in the effect; and (iv) other compelling  
950 attributes of the data that may influence our certainty. The ratings for strength of evidence for the

951 effect of long working hours on ischaemic heart disease will be “sufficient evidence of  
952 toxicity/harmfulness”, “limited of toxicity/harmfulness”, “inadequate of toxicity/harmfulness” and  
953 “evidence of lack of toxicity/harmfulness” (Appendix I).

954

955

956

957 **FINANCIAL SUPPORT**

958

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

964

965 **SPONSORS**

966

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

969

970 **AUTHOR CONTRIBUTIONS**

971

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

983

984 **ACKNOWLEDGMENTS**

985

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

993

994 **CONFLICT OF INTEREST**

995

996 None declared.

997

998

999 **REFERENCES**

1000

1001 1. Ryder G. Welcome address from the Director General of the International Labour Organization. XXI  
1002 World Congress on Safety and Health at Work; 2017; Sands Expo and Convention Centre, Singapore; 2017.

1003 2. Pruss-Ustun A, Wolf J, Corvalan C, Bos R, Neira M. Preventing disease through healthy environments: a  
1004 global assessment of the burden of disease from environmental risks In: Department of Public Health EasDoH,  
1005 editor. Geneva: World Health Organization; 2017.

1006 3. International Labour Organization. Safety and health at work : a vision for sustainable prevention: XX  
1007 World Congress on Safety and Health at Work 2014: Global Forum for Prevention, 24-27 August 2014, Frankfurt,  
1008 Germany. Geneva: International Labour Organization; 2014.

1009 4. Murray CJL, Ezzati M, Lopez AD, Rodgers A, Vander Hoorn S. Comparative Quantification of Health Risks:  
1010 Conceptual Framework and Methodological Issues. In: Ezzati M, Lopez AD, Rodgers A, Murray CJL, eds.  
1011 Comparative Quantification of Health Risks: Global and Regional Burden of Disease Attributable to Selected  
1012 Major Risk Factors. Geneva: World Health Organization; 2004: 1-39.

1013 5. World Health Organization. WHO methods and data sources for global burden of disease estimates  
1014 2000-2015. Global Health Estimates Technical Paper WHO/HIS/IER/GHE/2017.1. In: Department of Information  
1015 EaR, editor. Geneva: World Health Organization; 2017.

1016 6. Descatha A, Sembajwe G, Baer M, et al. WHO/ILO work-related burden of disease and injury: Protocol  
1017 for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on  
1018 stroke. *Environment International* In press.

1019 7. Hulshof C, Colosio C, Ivaonv ID, et al. WHO/ILO work-related burden of disease and injury: Protocol for  
1020 systematic reviews of occupational exposure to ergonomic risk factors and of the effect of occupational  
1021 exposure to ergonomic risk factors on osteoarthritis and other musculoskeletal diseases. *Environment*  
1022 *International* Submitted.

1023 8. John SM, Akagwu OC, Akparibo IY, et al. WHO/ILO work-related burden of disease and injury: Protocol  
1024 for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational  
1025 exposure to solar ultraviolet radiation on melanoma and non-melanoma skin cancer. *Environment International*  
1026 Submitted.

1027 9. Mandrioli D, Schlünssen V, Adam B, et al. WHO/ILO work-related burden of disease and injury: Protocols  
1028 for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational  
1029 exposure to dusts and/or fibres on pneumoconiosis. *Environment International* In press.

1030 10. Pachito DV, Bakusic J, Boonen E, et al. WHO/ILO work-related burden of disease and injury: Protocol for  
1031 systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on  
1032 alcohol use and alcohol use disorder. *Environment International* Under review.

1033 11. Rugulies RF, Ando E, Ayuso Mateos JL, et al. WHO/ILO work-related burden of disease and injury:  
1034 Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working  
1035 hours on depression. *Environment International* Under review.

1036 12. Teixeira LR, Azevedo TM, Bortkiewicz AT, et al. WHO/ILO work-related burden of disease and injury:  
1037 Protocol for systematic reviews of occupational exposure to noise and of the effect of occupational exposure to  
1038 noise on cardiovascular disease. *Environment International* Submitted.

1039 13. Tenkate TD, Paulo M, Adam B, et al. WHO/ILO work-related burden of disease and injury: Protocol for  
1040 systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational  
1041 exposure to solar ultraviolet radiation on cataract. *Environment International* Submitted.

1042 14. Li J, Brisson C, Clays E, et al. WHO/ILO work-related burden of disease and injury: Protocol for systematic  
1043 reviews of exposure to long working hours and of the effect of exposure to long working hours on ischaemic  
1044 heart disease. *Environment International* Accepted.

1045 15. Stevens GA, Alkema L, Black RE, et al. Guidelines for Accurate and Transparent Health Estimates  
1046 Reporting: the GATHER statement. *Lancet* 2016; **388**(10062): e19-e23.



- 1047 16. Kang MY PH, Seo JC, Kim D, Lim YH, Lim S, Cho SH, Hong YC. Long working hours and cardiovascular  
1048 disease: a meta-analysis of epidemiologic studies. *Journal of Occupational & Environmental Medicine* 2012;  
1049 **54**(5): 532–7.
- 1050 17. Virtanen M HK, Jokela M, Ferrie JE, Batty GD, Vahtera J, Kivimäki M. Long working hours and coronary  
1051 heart disease: a systematic review and meta-analysis. *American Journal of Epidemiology* 2012; **176**(7): 586–96.
- 1052 18. Kivimaki M, Jokela M, Nyberg ST, et al. Long working hours and risk of coronary heart disease and  
1053 stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. *Lancet*  
1054 2015; **386**(10005): 1739-46.
- 1055 19. 104th International Labour Conference. Transition from the Informal to the Formal Economy  
1056 (Recommendation No. 204). Geneva: International Labour Organization 2015.
- 1057 20. Kivimaki M, Virtanen M, Kawachi I, et al. Long working hours, socioeconomic status, and the risk of  
1058 incident type 2 diabetes: a meta-analysis of published and unpublished data from 222 120 individuals. *Lancet*  
1059 *Diabetes Endocrinol* 2015; **3**(1): 27-34.
- 1060 21. Virtanen M JM, Nyberg ST, Madsen IE, Lallukka T, Ahola K, Alfredsson L, Batty GD, Bjorner JB, Borritz M,  
1061 Burr H, Casini A, Clays E, De Bacquer D, Dragano N, Erbel R, Ferrie JE, Fransson EI, Hamer M, Heikkilä K, Jöckel  
1062 KH, Kittel F, Knutsson A, Koskenvuo M, Ladwig KH, Lunau T, Nielsen ML, Nordin M, Oksanen T, Pejtersen JH,  
1063 Pentti J, Rugulies R, Salo P, Schupp J, Siegrist J, Singh-Manoux A, Steptoe A, Suominen SB, Theorell T, Vahtera J,  
1064 Wagner GG, Westerholm PJ, Westerlund H, Kivimäki M. Long working hours and alcohol use: systematic review  
1065 and meta-analysis of published studies and unpublished individual participant data. *BMJ* 2015; **13**.
- 1066 22. World Health Organization. ICD-10: International Statistical Classification of Diseases and Related Health  
1067 Problems: 10th Revision. Geneva: World Health Organization; 2015.
- 1068 23. Rehfuss EA, Booth A, Brereton L, et al. Towards a taxonomy of logic models in systematic reviews and  
1069 health technology assessments: A priori, staged, and iterative approaches. *Res Synth Methods* 2017.
- 1070 24. Anderson LM, Petticrew M, Rehfuss E, et al. Using logic models to capture complexity in systematic  
1071 reviews. *Res Synth Methods* 2011; **2**(1): 33-42.
- 1072 25. Lee SM, D.; Messenger J. C. Working time around the world: trends in working hours, laws and policies  
1073 in a global comparative perspective. Geneva: International Labour Office; 2007.
- 1074 26. Taris TW YJ, Beckers DG, Verheijden MW, Geurts SA, Kompier MA. Investigating the associations among  
1075 overtime work, health behaviors, and health: a longitudinal study among full-time employees. *International*  
1076 *Journal of Behavioral Medicine* 2011 **18**(4): 352–60.
- 1077 27. Virtanen M, Ferrie JE, Gimeno D, et al. Long working hours and sleep disturbances: the Whitehall II  
1078 prospective cohort study. *Sleep* 2009; **32**(6): 737-45.
- 1079 28. Sonnentag S, Venz L, Casper A. Advances in recovery research: What have we learned? What should be  
1080 done next? *J Occup Health Psychol* 2017; **22**(3): 365-80.
- 1081 29. Chandola T, Heraclides A, Kumari M. Psychophysiological biomarkers of workplace stressors. *Neurosci*  
1082 *Biobehav Rev* 2010; **35**(1): 51-7.
- 1083 30. Jarczok MN, Jarczok M, Mauss D, et al. Autonomic nervous system activity and workplace stressors--a  
1084 systematic review. *Neurosci Biobehav Rev* 2013; **37**(8): 1810-23.
- 1085 31. Nakata A. Psychosocial job stress and immunity: a systematic review. *Methods Mol Biol* 2012; **934**: 39-  
1086 75.
- 1087 32. Kivimaki M, Steptoe A. Effects of stress on the development and progression of cardiovascular disease.  
1088 *Nat Rev Cardiol* 2017.
- 1089 33. Kaplan JR, Manuck SB. Antiatherogenic effects of beta-adrenergic blocking agents: theoretical,  
1090 experimental, and epidemiologic considerations. *Am Heart J* 1994; **128**(6 Pt 2): 1316-28.
- 1091 34. Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and  
1092 transparent method for translating environmental health science into better health outcomes. *Environ Health*  
1093 *Perspect* 2014; **122**(10): 1007-14.

1094 35. Lam J, Sutton P, Padula AM, et al. Applying the Navigation Guide Systematic Review Methodology Case  
1095 Study #6: Association between Formaldehyde Exposure and Asthma: A Systematic Review of the Evidence:  
1096 Protocol. San Francisco, CA: University of California at San Francisco, 2016.

1097 36. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis  
1098 protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015; **4**: 1.

1099 37. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis  
1100 protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; **349**: g7647.

1101 38. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts: reporting systematic reviews in journal  
1102 and conference abstracts. *PLoS medicine* 2013; **10**(4): e1001419.

1103 39. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and  
1104 meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine*  
1105 2009; **6**(7): e1000100.

1106 40. United Nations. ISIC Rev. 4: International Standard Industrial Classification of All Economic Activities,  
1107 Revision 4. Statistical Papers Series M No. 4/Rev.4. In: Affairs DoEaS, editor. New York, NY: United Nations;  
1108 2008.

1109 41. International Labour Organization. ISCO–88: International Standard Classification of Occupations.  
1110 Geneva: International Labour Organization; 1987.

1111 42. International Labour Organization. ISCO–08: International Standard Classification of Occupations.  
1112 Geneva: International Labour Organization; 2012.

1113 43. Babineau J. Product Review: Covidence (Systematic Review Software). *Journal of the Canadian Health*  
1114 *Libraries Association (JCHLA)* 2014; **32**(2): 68-71.

1115 44. Covidence systematic review software VHI, Melbourne, Australia. Available at [www.covidence.org](http://www.covidence.org).

1116 45. Ouzzani M, et al. Rayyan - a web and mobile app for systematic reviews. *Systematic Reviewss* 2016; **5**(1):  
1117 210-.

1118 46. Review Manager (RevMan). Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane  
1119 Collaboration; 2014.

1120 47. EvidencePartner. DistillerSR. Accessed from: [https://www.evidencepartners.com/products/distillersr-  
1121 systematic-review-software/](https://www.evidencepartners.com/products/distillersr-systematic-review-software/); EvidencePartner; 2017.

1122 48. Forsyth SR, Odierna DH, Krauth D, Bero LA. Conflicts of interest and critiques of the use of systematic  
1123 reviews in policymaking: an analysis of opinion articles. *Syst Rev* 2014; **3**: 122.

1124 49. Drazen JM, de Leeuw PW, Laine C, et al. Toward more uniform conflict disclosures: the updated ICMJE  
1125 conflict of interest reporting form. *JAMA* 2010; **304**(2): 212-3.

1126 50. Drazen JM, Van der Weyden MB, Sahni P, et al. Uniform format for disclosure of competing interests in  
1127 ICMJE journals. *JAMA* 2010; **303**(1): 75-6.

1128 51. The GATHER Working Group. The GATHER Statement: Explanation and Elaboration. Geneva: World  
1129 Health Organization, 2016.

1130 52. Munn Z, Moola S, Riitano D, Lisy K. The development of a critical appraisal tool for use in systematic  
1131 reviews addressing questions of prevalence. *Int J Health Policy Manag* 2014; **3**(3): 123-8.

1132 53. Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of  
1133 published animal studies: a systematic review. *Environ Health Perspect* 2013; **121**(9): 985-92.

1134 54. Mandrioli D, Silbergeld EK. Evidence from Toxicology: The Most Essential Science for Prevention. *Environ*  
1135 *Health Perspect* 2016; **124**(1): 6-11.

1136 55. Vandenberg LN, Agerstrand M, Beronius A, et al. A proposed framework for the systematic review and  
1137 integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environ Health* 2016; **15**(1): 74.

1138 56. Rooney AA, Cooper GS, Jahnke GD, et al. How credible are the study results? Evaluating and applying  
1139 internal validity tools to literature-based assessments of environmental health hazards. *Environ Int* 2016; **92-93**:  
1140 617-29.

1141 57. Hoy D, Brooks P, Woolf A, et al. Assessing risk of bias in prevalence studies: modification of an existing  
1142 tool and evidence of interrater agreement. *J Clin Epidemiol* 2012; **65**(9): 934-9.

- 1143 58. Higgins J, Altman D, Sterne J. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S,  
1144 eds. Cochrane Handbook for Systematic Reviews of Interventions Version 510 [updated March 2011] The  
1145 Cochrane Collaboration, 2011 Available from <http://handbookcochraneorg>; 2011.
- 1146 59. Schünemann H, Oxman A, Vist G, et al. Chapter 12: Interpreting results and drawing conclusions. In:  
1147 Higgins J, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 510 [updated  
1148 March 2011]. Available from [www.handbook.cochrane.org](http://www.handbook.cochrane.org); The Cochrane Collaboration; 2011.
- 1149 60. Arditi C, Burnand B, Peytremann-Bridevaux I. Adding non-randomised studies to a Cochrane review  
1150 brings complementary information for healthcare stakeholders: an augmented systematic review and meta-  
1151 analysis. *BMC Health Serv Res* 2016; **16**(1): 598.
- 1152 61. Barroga EF, Kojima T. Research study designs: an appraisal for peer reviewers and science editors. *Eur*  
1153 *Sci Ed* 2013; **(2013): 44-45**.
- 1154 62. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated  
1155 March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org>; 2011.
- 1156 63. Pega F, Blakely T, Glymour MM, Carter KN, Kawachi I. Using Marginal Structural Modeling to Estimate  
1157 the Cumulative Impact of an Unconditional Tax Credit on Self-Rated Health. *Am J Epidemiol* 2016; **183**(4): 315-  
1158 24.
- 1159 64. Gunasekara FI, Richardson K, Carter K, Blakely T. Fixed effects analysis of repeated measures data. *Int J*  
1160 *Epidemiol* 2014; **43**(1): 264-9.
- 1161 65. Shapiro A. Health Assessment Collaborative (HAWC). 2013. . 2013. <https://hawcproject.org> (accessed  
1162 January 7 2017).
- 1163 66. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in  
1164 Systematic Reviews of Health Care Interventions. *Methods Guide for Effectiveness and Comparative*  
1165 *Effectiveness Reviews*. Rockville (MD); 2008.
- 1166 67. Goodman JE, Lynch HN, Beck NB. More clarity needed in the Navigation Guide systematic review  
1167 framework. *Environ Int* 2017; **102**: 74-5.
- 1168 68. Johnson PI, Sutton P, Atchley DS, et al. The Navigation Guide - evidence-based medicine meets  
1169 environmental health: systematic review of human evidence for PFOA effects on fetal growth. *Environ Health*  
1170 *Perspect* 2014; **122**(10): 1028-39.
- 1171 69. Koustas E, Lam J, Sutton P, et al. The Navigation Guide - evidence-based medicine meets environmental  
1172 health: systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environ Health Perspect*  
1173 2014; **122**(10): 1015-27.
- 1174 70. Lam J, Koustas E, Sutton P, et al. The Navigation Guide - evidence-based medicine meets environmental  
1175 health: integration of animal and human evidence for PFOA effects on fetal growth. *Environ Health Perspect*  
1176 2014; **122**(10): 1040-51.
- 1177 71. Vesterinen H, Johnson P, Atchley D, et al. The relationship between fetal growth and maternal  
1178 glomerular filtration rate: a systematic review. *J Maternal Fetal Neonatal Med* 2014: 1-6.
- 1179 72. Johnson PI, Koustas E, Vesterinen HM, et al. Application of the Navigation Guide systematic review  
1180 methodology to the evidence for developmental and reproductive toxicity of triclosan. *Environ Int* 2016; **92-93**:  
1181 716-28.
- 1182 73. Lam J, Sutton P, Halladay A, et al. Applying the Navigation Guide Systematic Review Methodology Case  
1183 Study #4: Association between Developmental Exposures to Ambient Air Pollution and Autism. *PLoS One* 2016;  
1184 **21**(11(9)).
- 1185 74. Lam J, Lanphear B, Bellinger D, et al. Developmental PBDE exposure and IQ/ADHD in childhood: A  
1186 systematic review and meta-analysis. *Environ Health Perspectives* 2017; **125**(8).
- 1187 75. Lam J, Koustas E, Sutton P, et al. Applying the Navigation Guide: Case Study #6. Association Between  
1188 Formaldehyde Exposures and Asthma. *In preparation* 2016.
- 1189 76. Figueroa JL. Distributional effects of Oportunidades on early child development. *Social science &*  
1190 *medicine* 2014; **113**: 42-9.

- 1191 77. Doi SAR, Furuya-Kanamori L, Thalib L, Barendregt JJ. Meta-analysis in evidence-based healthcare: a  
1192 paradigm shift away from random effects is overdue. *Int J Evid Based Healthc* 2017; **15**(4): 152-60.
- 1193 78. Morgan RL, Thayer KA, Bero L, et al. GRADE: Assessing the quality of evidence in environmental and  
1194 occupational health. *Environ Int* 2016; **92-93**: 611-6.
- 1195 79. Rooney AA. Protocol to evaluate the evidence for an association between perfluorooctanoic acid or  
1196 perfluorooctane sulfonate exposure and immunotoxicity. In: Translation OoHAa, editor. Durham, NC: National  
1197 Toxicology Program Division, National Institute of Environmental Health Sciences; 2015.
- 1198 80. Anderson LM, Petticrew M, Rehfues E, et al. Using logic models to capture complexity in systematic  
1199 reviews. *Research synthesis methods* 2011; **2**(1): 33-42.

1200

1201 **Appendix A: Overview of inclusion and exclusion criteria, Systematic Review 1**

1202

1203 ***Include***

1204

- 1205 1. Quantitative studies of any design, including cross-sectional studies, on the prevalence of relevant levels  
1206 of long working hours for the years 2005 to 2018
- 1207 2. Studies of the working-age population ( $\geq 15$  year), including workers in the informal economy
- 1208 3. Studies on populations residing in any Member States of WHO and/or ILO and any industrial sector or  
1209 occupational group setting
- 1210 4. Studies on the prevalence of relevant levels of long working hours, if the prevalence is disaggregated by  
1211 WHO region, sex (two categories: female, male), age group and industrial sector or occupation
- 1212 5. Studies published between 2005-2018 and any language with essential information (title and abstract) in  
1213 English

1214

1215 ***Exclude***

1216

- 1217 1. Studies of unpaid domestic workers
- 1218 2. Studies of children (aged  $< 15$  years)
- 1219 3. Studies with data only on years other than between 2005 and 2018
- 1220 4. Qualitative, modelling, and case studies, as well as non-original studies without quantitative data (e.g.  
1221 letters, commentaries and perspectives)
- 1222
- 1223

## Appendix B: Ovid Medline search strategy, Systematic Review 1

### *Population*

No search string

### *Exposure*

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

37. (shift? adj3 compress\*).ti,ab,kw.
38. (shift? adj3 week\*).ti,ab,kw.
39. (shift? adj3 day?).ti,ab,kw.
40. (work\* and (life\* or live\*) and (balances\* or imbalances\* or unbalances\* or interference\*)).ti,ab,kw.
41. (work\* and famil\* and conflict\*).ti,ab,kw.
42. or/1-41

*Study design*

43. prevalence.tw.
  44. incidence.tw.
  45. epidemiol\*.tw.
  46. survey.tw.
  47. rapid assessment.tw.
  48. situation assessment.tw.
  49. situational assessment.tw.
  50. rar.tw.
  51. cohort.tw.
  52. surveillance.tw.
  53. seroprevalence.tw.
  54. seroincidence.tw.
  55. seroepidemiol\*.tw.
  56. screening.mp.
  57. exp epidemiologic methods/
  58. exp epidemiologic studies/
  59. exp sentinel surveillance/
  60. exp seroepidemiologic studies/
  61. exp cohort studies/
  62. exp cross-sectional studies/
  63. exp longitudinal studies/
  64. exp follow-up studies/
  65. exp prospective studies/
  66. or/43-65
67. 42 and 66

## Appendix C: Rate risk of bias, Systematic Review 1

### Tool development

Departing from the Navigation Guide risk of bias tool <sup>35</sup>, we first selected the five relevant risk of bias domains (i.e., selection bias, performance bias, exposure measurement, conflict of interest, and other bias), and we adopted or adapted the relevant considerations and criteria for ratings. Second, we integrated key considerations from the GATHER Guidelines <sup>51</sup> and existing checklists for prevalence studies <sup>52,57</sup>. Third, we integrated relevant criteria for ratings for the exposure measurement domain from the US Office of Health Assessment and Translation / National Toxicology Program risk of bias tool <sup>79</sup>. Finally, systematic review methodologists from all systematic reviews in the series <sup>6-8,10-14</sup> jointly improved the preliminary tool, and they will further refine and test it over the course of the systematic reviews. If a more suitable method becomes available over the course of the systematic review, then we may switch to it.

### Instructions

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

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

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

Examples of considerations for this risk of bias domain include:

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



10. the proportion of study participant who participated in the exposure assessment was comparable across exposure levels
11. the reasons for non-participation in exposure assessment were acceptable

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

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

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

OR

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

OR

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

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

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

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

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

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

- a) The descriptions of the target population, inclusion/exclusion criteria, recruitment and enrollment procedures, participation/response rates and/or data on the distribution of relevant study sample and population characteristics suggest that the risk of selection effects was substantial;

AND

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

AND

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

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

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

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

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

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

OR

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

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

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

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

The information on blinding is insufficient to permit a rating of high risk of bias. However,

indirect evidence suggests that the exposure assessors and study personnel were not adequately blinded, as described by the criteria for a rating of high risk of bias.

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

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

OR

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

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

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

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

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

***List of potential considerations:***

Quality of exposure measurement:

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

Data collection:

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

Sample size

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

Validity of prevalence calculation:

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

Missing data:

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

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

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

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

OR

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

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

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

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

OR

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

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

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

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

OR

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

OR

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

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

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

OR

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

OR

- It is unclear how exposure measurement was obtained.

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

There is evidence that exposure measurement methods are not capable of introducing risk of bias

in the study.

**4. Did the study receive any support from a company, study author or other entity with a potential financial interest in the exposures assessed?**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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





## Appendix D: Instructions for grading the quality of evidence, Systematic Reviews 1 and 2

Most of the text from these instructions and criteria for judging risk of bias has been adopted verbatim or adapted from one of the latest Navigation Guide systematic reviews<sup>35</sup>.

### A. Grading Quality

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

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

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

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

1. In deciding on the overall quality of evidence, one does not average across studies (for instance if some studies have no serious limitations, some serious limitations, and some very serious limitations, one does not automatically rate quality down by one level because of an average rating of serious limitations). Rather, judicious consideration of the contribution of each study, with a general guide to focus on the high-quality studies is warranted.<sup>1</sup>
2. This judicious consideration requires evaluating the extent to which each study contributes toward the estimate of magnitude of effect. The contribution that each study makes will usually reflect study sample size and number of outcome events. Larger studies with many events will contribute more, much larger studies with many more events will contribute much more.
3. One should be conservative in the judgment of rating down. That is, one should be confident that there is substantial risk of bias across most of the body of available evidence before one rates down for risk of bias.

---

<sup>1</sup> a Note: Limitations to GRADE's risk of bias assessments as stated by GRADE: "First, empirical evidence supporting the criteria is limited. Attempts to show systematic difference between studies that meet and do not meet specific criteria have shown inconsistent results. Second, the relative weight one should put on the criteria remains uncertain. The GRADE approach is less comprehensive than many systems, emphasizing simplicity and parsimony over completeness. GRADE's approach does not provide a quantitative rating of risk of bias. Although such a rating has advantages, we share with the Cochrane Collaboration methodologists a reluctance to provide a risk of bias score that, by its nature, must make questionable assumptions about the relative extent of bias associated with individual items and fails to consider the context of the individual items."

4. The risk of bias should be considered in the context of other limitations. If, for instance, reviewers find themselves in a close-call situation with respect to two quality issues (risk of bias and, say, precision), GRADE suggests rating down for at least one of the two.

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

Rating for Risk of Bias (Study Limitations)	Rationale for your judgment
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels	
Human	

## Category 2. Indirectness of Evidence

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

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

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

Based on GRADE (Guyatt et al. 2011), evidence can be indirect in one of three ways.<sup>2</sup>

1. The population studied differs from the population of interest (the term applicability is often used for this form of indirectness). GRADE states that in general, one should not rate down for population differences unless one has compelling reason to think that the biology in the population of interest is so different than the population tested that the magnitude of effect will differ substantially. According to GRADE, most often, this will not be the case.
2. The intervention (exposure) tested may differ from the exposure of interest, i.e., a difference in the chemical, route and/or dose. Decisions regarding indirectness of populations and exposure depend on an understanding of whether biological or social factors are sufficiently different that one might expect substantial differences in the magnitude of effect. GRADE also states, “As with all other aspects of rating quality of evidence, there is a continuum of similarity of the intervention that will require judgment. It is rare, and usually unnecessary, for the intended populations and interventions to be identical to those in the studies, and we should

---

<sup>2</sup> GRADE includes a fourth type of indirectness that occurs when there are no direct (i.e., head-to-head) comparisons between two or more interventions of interest. This criterion is not relevant to our study question; it could be relevant to future case studies.

only rate down if the differences are considered sufficient to make a difference in outcome likely.”

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

Rating for Indirectness		Rationale for your judgment
0 no change		
-1 decrease quality 1 level		
-2 decrease quality 2 levels		
Human		

### Category 3. Inconsistency of Evidence

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

According to Cochrane, “when studies yield widely differing estimates of effect (heterogeneity or variability in results) investigators should look for robust explanations for that heterogeneity.

...When heterogeneity exists and affects the interpretation of results, but authors fail to identify a plausible explanation, the quality of the evidence decreases.”

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

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

Based on GRADE, reviewers should consider rating down for inconsistency when:

1. Point estimates vary widely across studies;
2. Confidence intervals <sup>80</sup> show minimal or no overlap;
3. The statistical test for heterogeneity-which tests the null hypothesis that all studies in a meta-analysis have the same underlying magnitude of effect- shows a low P-value;
4. The  $I^2$  -which quantifies the proportion of the variation in point estimates due to among-study differences-is large. (I.e., the  $I^2$  index quantifies the degree of heterogeneity in a meta-analysis).

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

Rating for Inconsistency		Rationale for your judgment
0 no change		
-1 decrease quality 1 level		
-2 decrease quality 2 levels		
Human		

#### Category 4. Imprecision of Evidence

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

Cochrane states that when studies have few participants and few events, and thus have wide confidence intervals <sup>80</sup>, authors can lower their rating of the quality of evidence. These ratings of precision are made as judgments by review authors. The ratings are made by looking across studies, or, if available, on the results of a meta-analysis.

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

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

from high to moderate). If the CI is very wide, GRADE might rate down by two levels.

<b>Rating for Imprecision</b> 0 no change -1 decrease quality 1 level -2 decrease quality 2 levels	<b>Rationale for your judgment</b>
Human	

### Category 5. Publication Bias

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

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

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

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

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

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

substantial potential biases.”

<b>Rating for Publication Bias</b>	<b>Rationale for your judgment</b>
0 no change -1 decrease quality 1 level -2 decrease quality 2 levels	
Human	

## Upgrade Categories

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

### Category 6. Large Magnitude of Effect

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

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

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

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

<b>Rating for Large Magnitude of Effect</b>	<b>Rationale for your judgment</b>
0 no change +1 increase quality 1 level +2 increase quality 2 levels	

Human		
-------	--	--

**Category 7. Dose-response**

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

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

<b>Rating for Dose-Response</b>	<b>Rationale for your judgment</b>	
0 no change		
+1 increase quality 1 level		
+2 increase quality 2 levels		
Human		

**Category 8. Residual Confounding Increases Confidence**

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

Upgrade if consideration of all plausible residual confounders, biases, or effect modification would underestimate the effect or suggest a spurious effect when results show no effect. If a study reports an association despite the presence of residual confounding, biases or effect modification that would diminish the association, confidence in the association is increased (National Toxicology Program 2015). GRADE provides an illustrative example related to bias: rating up observational evidence finding lack of association between vaccination and autism, which occurred despite empirically confirmed bias that parents of autistic children may be more likely to remember their vaccine experience. The negative findings despite this form of recall bias suggest rating up the quality of evidence (Guyatt et al. 2011).<sup>88</sup>

<b>Rating for Residual Confounding Increases Confidence</b>	<b>Rationale for your judgment</b>	
0 no change		
+1 increase quality 1 level		
+2 increase quality 2 levels		
Human		

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

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

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

---- High

---- Moderate

---- Low



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

2  
3  
4  
5  
6  
7  
8

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

Potentially sufficient evidence for use as input data	The body of evidence was rated as being of high or moderate quality; it covers at least one country each from at least two WHO regions; the review authors had confidence in the feasibility of using the evidence as input data; and there are no compelling attributes of the evidence that may reduce certainty in it.
Potentially insufficient or inadequate evidence for use as input data	The body of evidence was rated as of low quality; it covered fewer than one country each from at least two WHO regions; the review authors were very uncertain about the feasibility of using the evidence as input data or had no confidence in it; and/or there are compelling attributes of the evidence that may reduce certainty in it.

9

10 **Appendix F: Overview of inclusion and exclusion criteria, Systematic Review 2**

11

12 ***Include***

13

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

28

29 ***Exclude***

30

- 31 1. Studies of unpaid domestic workers
- 32 2. Studies of children ( $< 15$  years)
- 33 3. Studies reporting on measures of absolute effects (e.g. mean differences in risks or odds), if they cannot
- 34 be converted into eligible relative measures
- 35 4. Cross-sectional studies, qualitative, modelling, and case studies, as well as non-original studies without
- 36 quantitative data (e.g. letters, commentaries and perspectives)

37

38 **Appendix G: Ovid Medline search strategy, Systematic Review 2**

39

40 *Population*

41

42 No search string

43

44 *Exposure*

45

46 1. exp "personnel staffing and scheduling"/

47 2. "personnel staffing and scheduling".ti,ab,kw.

48 3. shift work schedule.ti,ab,kw.

49 4. work schedule tolerance.ti,ab,kw.

50 5. workload.kw.

51 6. workday shifts.ti,ab,kw.

52 7. overwork\*.ti,ab,kw.

53 8. overtime.ti,ab,kw.

54 9. workweek\*.ti,ab,kw.

55 10. (work\* adj3 hour\*).ti,ab,kw.

56 11. (work\* adj3 schedul\*).ti,ab,kw.

57 12. work\* ad3 roster.ti,ab,kw.

58 13. (work\* adj3 organi#ation).ti,ab,kw.

59 14. (work\* adj3 time\*).ti,ab,kw.

60 15. (work\* adj3 overload\*).ti,ab,kw.

61 16. (work\* adj3 extend\*).ti,ab,kw.

62 17. (work\* adj3 compress\*).ti,ab,kw.

63 18. (work\* adj3 week\*).ti,ab,kw.

64 19. (work\* adj3 day?).ti,ab,kw.

65 20. (job? adj3 hour\*).ti,ab,kw.

66 21. (job? adj3 schedul\*).ti,ab,kw.

67 22. (job? adj3 roster).ti,ab,kw.

68 23. (job? adj3 organi#ation).ti,ab,kw.

69 24. (job? adj3 time\*).ti,ab,kw.

70 25. (job? adj3 overload\*).ti,ab,kw.

71 26. (job? adj3 extend\*).ti,ab,kw.

72 27. (job? adj3 compress\*).ti,ab,kw.

73 28. (job? adj3 week\*).ti,ab,kw.

74 29. (job? adj3 day?).ti,ab,kw.

75 30. (shift? adj3 hour\*).ti,ab,kw.

76 31. (shift? adj3 schedul\*).ti,ab,kw.

77 32. (shift? adj3 roster).ti,ab,kw.

78 33. (shift? adj3 organi#ation).ti,ab,kw.

79 34. (shift? adj3 time\*).ti,ab,kw.

80 35. (shift? adj3 overload\*).ti,ab,kw.

81 36. (shift? adj3 extend\*).ti,ab,kw.

82 37. (shift? adj3 compress\*).ti,ab,kw.

83 38. (shift? adj3 week\*).ti,ab,kw.

84 39. (shift? adj3 day?).ti,ab,kw.

85 40. (work\* and (life\* or live\*) and (balances\* or imbalances\* or unbalances\* or interference\*)).ti,ab,kw.  
86 41. (work\* and famil\* and conflict\*).ti,ab,kw.  
87 42. or/1-41  
88  
89 *Study design*  
90  
91 43. exp Clinical Trial/  
92 44. trial\$.tw.  
93 45. experiment\$.tw  
94 46. (intervention adj3 (study or studies or analys\$)).tw.  
95 47. Epidemiologic Studies/  
96 48. Observational Study/  
97 49. ((observational or epidemiologic\$) adj (study or studies or analys\$)).tw.  
98 50. exp Cohort Studies/  
99 51. cohort\$.tw.  
100 52. (panel\$ adj3 (study or studies or analys\$ or data)).tw.  
101 53. (follow up adj (study or studies or analys\$)).tw.  
102 54. (repeat\$ adj measure\$).tw.  
103 55. longitudinal\$.tw.  
104 56. retrospective\$.tw.  
105 57. exp Case Control Studies/  
106 58. (case\$ adj3 control\$).tw.  
107 59. (exposure\$ adj4 (study or studies or analys\$)).tw.  
108 60. before-after.tw  
109 61. pre-post.tw  
110 62. nonexperimental.ti,ab.  
111 63. non-experimental.ti,ab.  
112 64. nonrandomized.ti,ab.  
113 65. nonrandomised.ti,ab.  
114 66. non-randomized.ti,ab.  
115 67. non-randomised.ti,ab.  
116 68. or/43-67  
117  
118 *Outcome*  
119  
120 69. exp Myocardial Ischemia/  
121 70. Heart Diseases/  
122 71. angina.tw.  
123 72. (heart adj3 disease\$).tw.  
124 73. (coronary adj3 disease\$).tw.  
125 74. (isch?emic adj disease\*).tw.  
126 75. myocardial infarct\$.tw.  
127 76. (heart adj3 infarct\$).tw.  
128 77. ((artery occlusion\* or artery disease\* or arterioscleros\* or atheroscleros\*) adj2 coronary).ti,ab.  
129 78. (IHD or CHD or CAD or AMI or MI).ti,ab.  
130 79. exp Myocardial Revascularization/  
131 80. (coronary adj3 bypass\$).tw.

132 81. (coronary adj3 angioplast\$.tw.  
133 82. cabg.tw.  
134 83. ptca.tw.  
135 84. or/69-83  
136  
137 85. 42 and 68 and 84  
138  
139  
140

141 **Appendix H: Instructions for making risk of bias determinations, Systematic Reviews 2**

142  
143 Most of the text from these instructions and criteria for judging risk of bias has been adopted verbatim  
144 or adapted from one of the latest Navigation Guide systematic reviews <sup>35</sup>.

145  
146 Instructions:

- 147  
148  *Please evaluate each individual study for the following nine risk of bias domains. Please answer “low risk,” “probably low risk,” “probably high risk,” “high risk,” or “not applicable” and provide details/justification for your rating. If there is empirical evidence or other knowledge that informs the direction of bias, please include this in your answer as well; however, if there is not enough information to do so please do not guess at the direction of bias.*
- 149  
150  
151  
152  
153
- 154  *Additionally, please note that some internal validity issues could potentially be appropriately captured in several different risk of bias considerations. In this situation, please select the single most appropriate domain to evaluate this potential bias, to avoid double-counting the same internal validity concern.*
- 155  
156  
157

158  
159 *2. Are the study groups at risk of not representing their source populations in a manner that might introduce selection bias?*

160  
161  
162 The source population is viewed as the population for which study investigators are targeting their study question of interest. Examples of considerations for this risk of bias domain include: 1) level of detail reported for participant inclusion/exclusion (including details from previously published papers referenced in the article for an existing cohort); 163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191

170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191

answer: “No”): EITHER:

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

OR

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

OR

c) Although the descriptions and/or data as indicated in “a” above suggested the potential for selection effects and there was no support indicating that potential selection effects were *not* differential across both exposure and outcome, selection factors appeared to be well-understood, were measured in the data set, and appropriate adjustment post hoc techniques were used to control for selection bias.

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

193

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

199

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

201

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

207

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

209

210 d) There were indications from descriptions of the source population,  
211 inclusion/exclusion criteria, recruitment and enrollment procedures,  
212 participation and follow-up rates, or data on the distribution of relevant study  
213 sample and population characteristics that risk of selection effects were  
214 substantial; and

215 e) There was no support to indicate that potential selection effects were *not*  
216 differential across both exposure and outcome; and

217 f) Adjustment post hoc techniques were not used to control for selection bias.

218

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

221

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

224

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

228

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

231 • No blinding, but the review authors judge that the outcome measures as well  
232 as the exposure measures are not likely to be influenced by lack of blinding  
233 (such as differential outcome assessment where the outcome is assessed using  
234 different measurement or estimation metrics across the exposure groups, or  
235 differential exposure assessment where exposure is assessed using different  
236 measurement or estimation metrics across the diagnostic or outcome groups);  
237 or

238 • Blinding of key study personnel was ensured, and it is unlikely that the  
239 blinding could have been broken; or

240 • Some key study personnel were not blinded, but exposure and outcome  
241 assessment was blinded and the non-blinding of others is unlikely to  
242 introduce bias.

243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292

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

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

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

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

Criteria for the judgment of HIGH risk of bias (i.e. answer:

“Yes”): Any of the following:

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

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

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

#### 4. *Were exposure assessment methods lacking accuracy?*

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

#### **List of Considerations:**

*Possible sources of exposure assessment metrics:*

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

*For each, overall considerations include:*

- 1) *What is the quality of the source of the metric being used?*
- 2) *Is the exposure measured in the study a surrogate for the exposure?*



- 293 3) *What was the temporal coverage (i.e. short or long-term exposure)?*  
294 4) *Did the analysis account for prediction uncertainty?*  
295 5) *How was missing data accounted for, and any data imputations incorporated?*  
296 6) *Were sensitivity analyses performed?*

297

298 *In particular, for exposure assessment models:*

- 299 1) *Were the input data in the study suspected to systematically under- or over-*  
300 *estimate exposure?*  
301 2) *What type of model was used?*  
302 3) *What was geographic/spatial accuracy (county, census tract, organization,*  
303 *individual residence)?*  
304 4) *What was the temporal specificity and variation?*  
305 5) *What was the space-time coverage of the model?*  
306 6) *Were time-activity patterns accounted for?*

307

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

309

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

- 311 • There is high confidence in the accuracy of the exposure assessment  
312 methods, such as methods that have been tested for validity and reliability  
313 in measuring the targeted exposure.

314

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

316

317 There is insufficient information about the exposure assessment methods to permit  
318 a judgment of low risk of bias, but there is indirect evidence that suggests that  
319 methods were robust, as described by the criteria for a judgment of low risk of bias.

320

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

322

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

327

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

329

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

- 332 • There is low confidence in the accuracy of the exposure assessment methods;  
333 or  
334 • Less-established or less direct exposure measurements are not validated  
335 and are suspected to introduce bias that impacts the outcome assessment;  
336 or  
337 • Uncertain how exposure information was obtained.

338

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

341

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

344

345 *5. Were outcome assessment methods lacking accuracy?*

346

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

348

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

350

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

364

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

366

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

373

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

375

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

380

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

382

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

385

- 386 • There is low confidence in the accuracy of the outcome assessment methods;  
387 or
- 388 • Less-established or less direct outcome measurements are not validated  
389 and are suspected to introduce bias that impacts the outcome assessment
- 390 • Uncertain how outcome information was obtained

391

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

394

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

397

398 *6. Was potential confounding inadequately incorporated?*

399

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

403 Tier I: Important confounders

- 404 • Age, sex and socioeconomic position.

405

406 Tier II: Other potentially important confounders:

407 None identified.

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

409

410 The study appropriately assessed and accounted for (i.e. matched, stratified,  
411 excluded certain populations or statistically controlled for) all important  
412 confounders (Tier I) using appropriate statistical techniques, or reported that  
413 important confounders were evaluated and omitted because inclusion did not  
414 substantially affect the results. The determination of specific confounders may also  
415 be informed by, but not limited to, the studies included in the overall review,

416

417 AND the study appropriately assessed and accounted for (i.e. matched, stratified,  
418 or statistically controlled for) other potentially important confounders relevant  
419 (Tier II) using appropriate statistical techniques, or reported that these confounders  
420 were evaluated and omitted because inclusion did not substantially affect the  
421 results,

422 AND the important potential confounders were measured consistently across study  
423 groups using valid and reliable methods, or the influence of covariate measurement  
424 error was determined, through sensitivity analysis, to be minimal.

425

426

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

428

429 The study appropriately accounted for most but not all of the important confounders  
430 (Tier

431 I) or used appropriate statistical techniques;

432

433 AND some of the other potentially important confounders relevant (Tier II) using  
434 appropriate statistical techniques,

435 OR reported that these confounders were evaluated and omitted because inclusion  
436 did not substantially affect the results;

437 AND this is not expected to introduce substantial bias.

438

439 Criteria for the judgment of PROBABLY HIGH risk of bias (i.e. answer: “Probably  
440 Yes”): The study evaluated some but not all of the important confounders (Tier I),

441 AND some but not all of the other potentially important confounders relevant (Tier

442 II), OR used questionable statistical techniques for confounder adjustment;  
443 AND this is expected to introduce substantial bias.

444

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

446

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

452

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

455

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

458

459 *7. Were incomplete outcome data inadequately addressed?*

460

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

462

463 Participants were followed long enough to obtain outcome measurements

464

465 OR any one of the following:

466

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

479

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

481

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

486

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

488

489 There is insufficient information about incomplete outcome data to permit a  
490 judgment of high risk of bias, but there is indirect evidence which suggests

491 incomplete outcome data was not adequately addressed, as described by the criteria  
492 for a judgment of high risk of bias.

493

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

495

496 Participants were not followed long enough to obtain outcome  
497 measurements OR any one of the following:

- 498 • Reason for missing outcome data likely to be related to true outcome, with  
499 either imbalance in numbers or reasons for missing data across exposure  
500 groups; or
- 501 • For dichotomous outcome data, the proportion of missing outcomes  
502 compared with observed event risk enough to induce biologically  
503 relevant bias in intervention effect estimate; or
- 504 • For continuous outcome data, plausible effect size (difference in  
505 means or standardized difference in means) among missing outcomes  
506 enough to induce biologically relevant bias in observed effect size; or
- 507 • Potentially inappropriate application of imputation.

508

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

511

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

514

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

516

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

518

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

523

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

525

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

533

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

535

536 There is insufficient information about selective outcome reporting to permit a  
537 judgment of high risk of bias, but there is indirect evidence which suggests the  
538 study was not free of selective reporting, as described by the criteria for a judgment  
539 of high risk of bias. This includes if a pre-published protocol is not available and  
540 the study’s pre-specified (primary and secondary) outcomes outlined in the

541 published manuscript's methods, abstract, and/or introduction section that are of  
542 interest in the review have not been reported in the pre-specified way.

543

544 Criteria for the judgment of HIGH risk of bias (i.e. answer:

545 "Yes"): Any one of the following:

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

556

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

559

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

562

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

565

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

567

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

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

587

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

589

590 There is insufficient information to permit a judgment of low risk of bias, but there

591 is indirect evidence which suggests the study was free of support from a company,  
592 study author, or other entity having a financial interest in the outcome of the study,  
593 as described by the criteria for a judgment of low risk of bias.

594

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

596

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

601

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

603

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

606

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

607

608

609

610

611

612

613

614

615

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

618

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

621

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

623

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

625

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

626

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

627

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

630

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

632

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

636

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

638

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

639

- Had a potential source of bias related to the specific study design used; or

640  
641  
642  
643  
644  
645  
646  
647  
648

- Stopped early due to some data-dependent process (including a formal-stopping rule); or
- The conduct of the study is affected by interim results (e.g. recruiting additional participants from a subgroup showing greater or lesser effect); or
- Has been claimed to have been fraudulent; or
- Had some other problem.



649 **Appendix I: Rate the strength of evidence, Systematic Review 2**

650

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

655

Sufficient evidence of toxicity/harmfulness	The available evidence usually includes consistent results from well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies. For human evidence a positive relationship is observed between exposure and outcome where chance, bias, and confounding, can be ruled out with reasonable confidence.
Limited evidence of toxicity/harmfulness	The available evidence is sufficient to determine the effects of the exposure, but confidence in the estimate is constrained by such factors as: the number, size, or quality of individual studies, the confidence in the effect, or inconsistency of findings across individual studies. As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion. For human evidence a positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence.
Inadequate evidence of toxicity/harmfulness	Studies permit no conclusion about a toxic effect. The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of: the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an estimation of effects.
Evidence of lack of toxicity/harmfulness	The available evidence includes consistent results from well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies. For human evidence more than one study showed no effect on the outcome of interest at the full range of exposure levels that humans are known to encounter, where bias and confounding can be ruled out with reasonable confidence. The conclusion is limited to the age at exposure and/or other conditions and levels of exposure studied.

656

657 **Appendix J: Parameters for estimating the national and global burden of ischemic heart disease attributable to long-working hours: Protocol for a**  
 658 **systematic review and meta-analysis (Jian Li)**

#	Item	Guidance	On page #	Author Comments
<b>Title</b>				
1	Identification	Identify the report as a systematic review, or systematic review and meta-analysis, as appropriate.	1	
2	Update	If the protocol is for an update of a previous systematic review, identify as such.	NA	Not an update of a previous systematic review
<b>Registration</b>				
3	Registration	If registered, provide the name of the registry (e.g. PROSPERO) and registration number.	4	PROSPERO-ID: CRD42017084243
<b>Authors</b>				
4	Contact	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author.	1-2	
5	Contributions	Describe contributions of protocol authors and identify the guarantor of the review.	24	
<b>Amendments</b>				
6	Amendments	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments.	NA	
<b>Support</b>				

#	Item	Guidance	On page #	Author Comments
7	Sources	Indicate sources of financial or other support for the review.	24	
8	Sponsor	Provide name for the review funder/s and/or sponsor/s	24	
9	Roles	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol.	24	
<b>Introduction</b>				
10	Rationale	Describe the rationale for the review in the context of what is already known	5-6	
11	Objectives	Provide an explicit statement of the question(s) the review will address, with specific reference to: <ul style="list-style-type: none"> <li>• Participants</li> <li>• Interventions / Exposures (as appropriate)</li> <li>• Comparisons</li> <li>• Outcomes</li> <li>• Study design</li> </ul>	9	
<b>Methods</b>				
12	Eligibility criteria	Specify the study characteristics (e.g. PICO/PECO, study design, setting, time frame) and report characteristics (e.g. years considered, language, publication status) to be used as criteria for eligibility for the review.	11-15	
13	Information sources	Describe all intended information sources (e.g. electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage.	15-16	

#	Item	Guidance	On page #	Author Comments
14	Search strategy	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated.	15-16, 27-30	
15	Data management	Describe the mechanism(s) that will be used to manage records and data throughout the review.	15-16	
16	Selection process	State the process that will be used for selecting studies (e.g. two independent reviewers) through each phase of the review (i.e. screening, eligibility, and inclusion in meta-analysis).	16-17	
17	Data collection process	Describe planned method of extracting data from reports (e.g. piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators.	17-18	
18	Data items	List and define all variables for which data will be sought (e.g. PICO items, funding sources), any pre-planned data assumptions and simplifications	11-12	
19	Outcomes and prioritisation	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale.	12	
20	Bias in individual studies	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	18-20	
21	Data synthesis criteria	Describe criteria under which study data will be quantitatively synthesized	21-22	
22	Summary measures	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g. $I^2$ , Kendall's tau).	21-22	

#	Item	Guidance	On page #	Author Comments
23	Additional analyses	Describe any proposed additional analyses (e.g. sensitivity or subgroup analyses, meta-regression)	23	
24	Alternative synthesis	If quantitative synthesis is not appropriate, describe the type of summary planned	NA	
25	Meta-bias	Specify any planned assessment of meta-bias(es) (e.g. publication bias across studies, selective reporting within studies)	20-21	
26	Confidence in cumulative evidence	Describe how the strength of the body of evidence will be assessed (e.g. GRADE)	20-21	

659

660 *Environment International* modified PRISMA-P report adapted from: Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart  
661 LA. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4(1):1. (Changes are minor,  
662 with text edits to accommodate the subject matter of the journal and formatting to fit page.)

663