Environment International 142 (2020) 105746

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

The effect of exposure to long working hours on stroke: A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury

Alexis Descatha[a,b,c,d,*,1], Grace Sembajwe[e,1], Frank Pega[f], Yuka Ujita[g], Michael Baer[h], Fabio Boccuni[i], Cristina Di Tecco[i], Clement Duret[b], Bradley A. Evanoff[j], Diana Gagliardi[i], Lode Godderis[k,l], Seong-Kyu Kang[m], Beon Joon Kim[n], Jian Li[o], Linda L. Magnusson Hanson[p], Alessandro Marinaccio[i], Anna Ozguler[h,q], Daniela Pachito[r], John Pell[s], Fernando Pico[t], Matteo Ronchetti[i], Yves Roquelaure[a], Reiner Rugulies[u,v,w], Martijn Schouteden[l], Johannes Siegrist[x], Akizumi Tsutsumi[y], Sergio Iavicoli[i,1]

[a] UNIV Angers, CHU Angers, Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-49000 Angers, France
[b] AP-HP (Paris Hospital), Occupational Health Unit, Poincaré University Hospital, Garches, France
[c] Versailles St-Quentin Univ–Paris Saclay Univ (UVSQ), UMS 011, UMR-S 1168, France
[d] Inserm, U1168 UMS 011, Villejuif, France
[e] Department of Occupational Medicine Epidemiology and Prevention, Zucker School of Medicine at Hofstra University, Feinstein Institutes for Medical Research, Northwell Health, NY, USA
[f] Environment, Climate Change and Health Department, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland
[g] Labour Administration, Labour Inspection and Occupational Safety and Health Branch, International Labour Organization, Route des Morillons 4, 1211 Geneva, Switzerland
[h] AP-HP (Paris Hospital), SAMU92, Poincaré University Hospital, Garches, France
[i] Inail, Department of Occupational and Environmental Medicine, Epidemiology and Hygiene, Via Fontana Candida 1, 00078 Monte Porzio Catone (Rome), Italy
[j] Division of General Medical Sciences, Washington University School of Medicine, Campus Box 8005, 660 South Euclid Ave, St. Louis, MO 63110, United States
[k] Environment and Health, Kapucijnenvoer 35 blok d – box 7001, 3000 Leuven, Belgium
[l] IDEWE, External Service for Prevention and Protection at Work, Interleuvenlaan 58, 3001 Leuven, Belgium
[m] Department of Occupational and Environmental Medicine, Gachon University Gil Medical Center, Incheon, Republic of Korea
[n] Seoul National University Bundang Hospital, Bundang-gu, Republic of Korea
[o] Department of Environmental Health Sciences, Fielding School of Public Health, School of Nursing, University of California, Los Angeles, United States
[p] Stress Research Institute, Stockholm University, Stockholm, Sweden
[q] Inserm UMS 011, Villejuif, France
[r] Núcleo de Avaliação de Tecnologias em Saúde, Hospital Sírio-Libanês, 142 Barata Ribeiro, Sao Paulo, Brazil
[s] Hunter College Libraries, Social Work and Public Health Library, 2180 3rd Avenue, 110D, New York, NY 10035, United States
[t] Neurology and Stroke Unit, Versailles Hospital, Le Chesnay, France
[u] National Research Centre for the Working Environment, Lersø Parkallé 105, DK-2100 Copenhagen, Denmark
[v] Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, DK-1014 Copenhagen, Denmark
[w] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen, Denmark
[x] Life Science Centre, University of Düsseldorf, Merowingerplatz 1a, Düsseldorf 40225, Germany
[y] Kitasato University School of Medicine, 1-15-1 Kitasato, Minami, Sagamihara 252-0374, Japan

* Corresponding author at: UNIV Angers, CHU Angers, Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-49000 Angers, France.
E-mail addresses: alexis.descatha@inserm.fr (A. Descatha), GSembajwe@northwell.edu (G. Sembajwe), pegaf@who.int (F. Pega), ujita@ilo.org (Y. Ujita), michel.baer@aphp.fr (M. Baer), f.boccuni@inail.it (F. Boccuni), c.ditecco@inail.it (C. Di Tecco), clement.duret@aphp.fr (C. Duret), bevanoff@wustl.edu (B.A. Evanoff), d.gagliardi@inail.it (D. Gagliardi), lode.godderis@kuleuven.be (L. Godderis), sk.kang@gachon.ac.kr (S.-K. Kang), Kim.BJ.Stroke@gmail.com (B.J. Kim), jianli2019@ucla.edu (J. Li), linda.hanson@su.se (L.L. Magnusson Hanson), a.marinaccio@inail.it (A. Marinaccio), anna.ozguler@inserm.fr (A. Ozguler), pachito@uol.com.br (D. Pachito), jpell@hunter.cuny.edu (J. Pell), fpico@ch-versailles.fr (F. Pico), m.ronchetti@inail.it (M. Ronchetti), YvRoquelaure@chu-angers.fr (Y. Roquelaure), RER@nfa.dk (R. Rugulies), Martijn.Schouteden@idewe.be (M. Schouteden), Johannes.Siegrist@med.uni-duesseldorf.de (J. Siegrist), akizumi@kitasato-u.ac.jp (A. Tsutsumi), s.iavicoli@inail.it (S. Iavicoli).
[1] Contributed to this work equally.

https://doi.org/10.1016/j.envint.2020.105746
Received 4 September 2019; Received in revised form 14 April 2020; Accepted 14 April 2020
Available online 03 June 2020

ARTICLE INFO

Handling Editor: Paul Whaley

Keywords:
Global Burden of Disease
Long working hours
Stroke
Occupational
Systematic review
Meta-analysis

ABSTRACT

Background: The World Health Organization (WHO) and the International Labour Organization (ILO) are developing joint estimates of the work-related burden of disease and injury (WHO/ILO Joint Estimates), with contributions from a large network of individual experts. Evidence from mechanistic data and prior studies suggests that exposure to long working hours may cause stroke. In this paper, we present a systematic review and meta-analysis of parameters for estimating the number of deaths and disability-adjusted life years from stroke that are attributable to exposure to long working hours, for the development of the WHO/ILO Joint Estimates.
Objectives: We aimed to systematically review and meta-analyse estimates of the effect of exposure to long working hours (three categories: 41–48, 49–54 and ≥55 h/week), compared with exposure to standard working hours (35–40 h/week), on stroke (three outcomes: prevalence, incidence, and mortality).
Data sources: A protocol was developed and published, applying the Navigation Guide to systematic reviews as an organizing framework where feasible. We searched electronic databases for potentially relevant records from published and unpublished studies, including Ovid MEDLINE, PubMed, EMBASE, Scopus, Web of Science, CISDOC, PsycINFO, and WHO ICTRP. We also searched grey literature databases, Internet search engines, and organizational websites; hand-searched reference lists of previous systematic reviews; and consulted additional experts.
Study eligibility and criteria: We included working-age (≥15 years) individuals in the formal and informal economy in any WHO and/or ILO Member State but excluded children (aged < 15 years) and unpaid domestic workers. We included randomized controlled trials, cohort studies, case-control studies and other non-randomized intervention studies with an estimate of the effect of exposure to long working hours (41–48, 49–54 and ≥55 h/week), compared with exposure to standard working hours (35–40 h/week), on stroke (prevalence, incidence or mortality).
Study appraisal and synthesis methods: At least two review authors independently screened titles and abstracts against the eligibility criteria at a first review stage and full texts of potentially eligible records at a second stage, followed by extraction of data from qualifying studies. Missing data were requested from principal study authors. We combined relative risks using random-effects meta-analysis. Two or more review authors assessed the risk of bias, quality of evidence and strength of evidence, using the Navigation Guide and GRADE tools and approaches adapted to this project.
Results: Twenty-two studies (20 cohort studies, 2 case-control studies) met the inclusion criteria, comprising a total of 839,680 participants (364,616 females) in eight countries from three WHO regions (Americas, Europe, and Western Pacific). The exposure was measured using self-reports in all studies, and the outcome was assessed with administrative health records (13 studies), self-reported physician diagnosis (7 studies), direct diagnosis by a physician (1 study) or during a medical interview (1 study). The outcome was defined as an incident non-fatal stroke event in nine studies (7 cohort studies, 2 case-control studies), incident fatal stroke event in one cohort study and incident non-fatal or fatal ("mixed") event in 12 studies (all cohort studies). Cohort studies were judged to have a relatively low risk of bias; therefore, we prioritized evidence from these studies, but synthesised evidence from case-control studies as supporting evidence. For the bodies of evidence for both outcomes with any eligible studies (i.e. stroke incidence and mortality), we did not have serious concerns for risk of bias (at least for the cohort studies).

Eligible studies were found on the effects of long working hours on stroke incidence and mortality, but not prevalence. Compared with working 35–40 h/week, we were uncertain about the effect on incidence of stroke due to working 41–48 h/week (relative risk (RR) 1.04, 95% confidence interval (CI) 0.94–1.14, 18 studies, 277,202 participants, I^2 0%, low quality of evidence). There may have been an increased risk for acquiring stroke when working 49–54 h/week compared with 35–40 h/week (RR 1.13, 95% CI 1.00–1.28, 17 studies, 275,181participants, I^2 0%, p 0.04, moderate quality of evidence). Compared with working 35–40 h/week, working ≥55 h/week may have led to a moderate, clinically meaningful increase in the risk of acquiring stroke, when followed up between one year and 20 years (RR 1.35, 95% CI 1.13 to 1.61, 7 studies, 162,644 participants, I^2 3%, moderate quality of evidence).

Compared with working 35–40 h/week, we were very uncertain about the effect on dying (mortality) of stroke due to working 41–48 h/week (RR 1.01, 95% CI 0.91–1.12, 12 studies, 265,937 participants, I^2 0%, low quality of evidence), 49–54 h/week (RR 1.13, 95% CI 0.99–1.29, 11 studies, 256,129 participants, I^2 0%, low quality of evidence) and 55 h/week (RR 1.08, 95% CI 0.89–1.31, 10 studies, 664,647 participants, I^2 20%, low quality of evidence).

Subgroup analyses found no evidence for differences by WHO region, age, sex, socioeconomic status and type of stroke. Sensitivity analyses found no differences by outcome definition (exclusively non-fatal or fatal versus "mixed") except for the comparison working ≥55 h/week versus 35–40 h/week for stroke incidence (p for subgroup differences: 0.05), risk of bias ("high"/"probably high" ratings in any domain versus "low"/"probably low" in all domains), effect estimate measures (risk versus hazard versus odds ratios) and comparator (exact versus approximate definition).
Conclusions: We judged the existing bodies of evidence for human evidence as "inadequate evidence for harmfulness" for all exposure categories for stroke prevalence and mortality and for exposure to 41–48 h/week for stroke incidence. Evidence on exposure to 48–54 h/week and ≥55 h/week was judged as "limited evidence for harmfulness" and "sufficient evidence for harmfulness" for stroke incidence, respectively. Producing estimates for the burden of stroke attributable to exposures to working 48–54 and ≥55 h/week appears evidence-based, and the pooled effect estimates presented in this systematic review could be used as input data for the WHO/ILO Joint Estimates.
Protocol identifier: https://doi.org/10.1016/j.envint.2018.06.016.
PROSPERO registration number: CRD42017060124.

1. Background

The World Health Organization (WHO) and the International Labour Organization (ILO) are finalizing their first joint estimates of the work-related burden of disease and injury (WHO/ILO Joint Estimates) (Ryder, 2017). The organizations are estimating the numbers of deaths and disability-adjusted life years (DALYs) that are attributable to exposure to selected occupational risk factors. The WHO/ILO Joint Estimates are based on existing WHO and ILO methodologies for estimating the disease burdens for selected occupational risk factors (Ezzati et al., 2004; International Labour Organization, 1999; 2014; Pruss-Ustun et al., 2017). They will expand existing estimates with those for prioritized additional pairs of occupational risk factors and health outcomes. For this purpose, population attributable fractions (Murray et al., 2004) – the proportional reduction in burden from the health outcome achieved by a reduction of exposure to the risk factor to zero – are being calculated for each additional risk factor-outcome pair. These fractions are being applied to the total disease burden envelopes for health outcomes from the WHO Global Health Estimates for the years 2000–2016 (World Health Organization, 2020).

The WHO/ILO Joint Estimates may include estimates of the burden of stroke attributable to exposure to long working hours, if feasible, as one additional risk factor-outcome pair for which global disease burden has not previously been estimated. To select parameters with the best and least biased evidence for their estimation models, WHO and ILO, supported by a large network of individual experts, have conducted a systematic review and meta-analysis of studies on the relationship between exposure to long working hours and stroke according to protocol (Descatha et al., 2018), and we present these analyses in this paper. WHO and ILO are in parallel also producing a systematic review of studies estimating the prevalence of exposure to long working hours (forthcoming), applying their novel systematic review methods (Pega et al., 2020). The organizations are also conducting or have completed several other systematic reviews and meta-analyses on other additional risk factor-outcome pairs (Godderis et al., 2018; Hulshof et al., 2019; Li et al., 2018; Li et al., 2020; Mandrioli et al., 2018; Paulo et al., 2019; Rugulies et al., 2019; Teixeira et al., 2019; Tenkate et al., 2019). To our knowledge, these are the first systematic reviews and meta-analyses, with a pre-published protocol, conducted specifically for an occupational burden of disease study. The WHO's and ILO's joint estimation methodology and the WHO/ILO Joint Estimates are separate from these systematic reviews, and they will be described in more detail and reported elsewhere.

1.1. Rationale

As the world's population number grows and ages, the global burden of stroke is increasing dramatically (Mukherjee and Patil, 2011), with 16.9 million people suffering a stroke each year and a global incidence of 258/100,000/year (Bejot et al., 2016). To consider the feasibility of estimating the burden of stroke attributable to exposure to long working hours, and to ensure that potential estimates of burden of disease are reported in adherence with the guidelines for accurate and transparent health estimates reporting (GATHER) (Stevens et al., 2016), WHO and ILO require a systematic review of studies on the prevalence of relevant levels of exposure to long working hours (forthcoming), as well as a systematic review and meta-analysis with estimates of the relative effect of exposure to long work hours on stroke prevalence, incidence, and mortality compared with the theoretical minimum risk exposure level. The theoretical minimum risk exposure level is the exposure level that would result in the lowest possible population risk, even if it is not feasible to attain this exposure level in practice (Murray et al., 2004).

Over the last two decades, several reviews have been conducted on the relationship between exposure to psychosocial factors and cardiovascular diseases in general (Huang and Zhang 2006; Huang et al., 2013; Kang et al., 2012a; Kang et al., 2012b; Kivimaki and Kawachi, 2015; Sparks et al., 1997). Kang et al in 2012 systematically reviewed 341 published studies on associations between long working hours and cardiovascular diseases, finding five cohort studies and six case–control studies of which none studied stroke outcomes alone and two studied cardiovascular diseases that included stroke. We are aware of only one previous systematic review and meta-analysis (and individual-participant-data analysis) specific to the effect of exposure to long working hours on stroke (Kivimaki et al., 2015a). This systematic review included one published study and 14 unpublished studies, with evidence up to August of 2014. It found a dose–response association, with relative risk (RR) estimates for stroke of 1.10 (95% CI 0.94–1.28) for study participants working 41–48 h/week; 1.27 (1.03–1.56) for those working 49–54 h/week; and 1.33 (1.11–1.61) for those working ≥55 h/week, compared with participants working standard hours (p for trend <0.0001). A 2018 update of the Kivimaki et al., 2015 systematic review added one additional cohort study (the Danish Labour Force Survey) (Virtanen and Kivimaki, 2018) and found that exposure to working ≥55 h/week led to an increase in risk of stroke by an estimated 21% (95% CI 1.01–1.45; 16 studies). Both the Kivimaki et al. (2015) systematic review and its 2018 update combined studies with non-fatal, fatal, and fatal or non-fatal ("mixed") stroke events in their meta-analyses. However, burden of disease estimation requires separate evidence on stroke incidence (ideally non-fatal events only, but mixed data might also be included) and stroke mortality (ideally fatal events only, but mixed data might also be included). Another systematic review and meta-analysis on the effects of exposure to long working hours on occupational health outcomes (Wong et al. (2019) included stroke as one (of many) physiological health outcomes, making it not directly comparable with the current systematic review. To our knowledge, none of these prior systematic reviews had a pre-published protocol and/or missed other essential aspects of a systematic review. Our systematic review is fully compliant with the latest systematic review methods (including use of a protocol), expands the scope of the existing systematic review evidence by covering evidence from studies published up to 31 August 2019, and includes workers in the formal and the informal economy (104th International Labour Conference, 2015).

1.2. Description of the risk factor

Burden of disease estimation requires unambiguous definitions of the risk factor, risk factor levels, and the theoretical minimum risk exposure level. Long working hours were defined as working hours exceeding standard working hours, i.e. working for ≥41 h/week

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

	Definition
Risk factor	Long working hours (including those spent in secondary jobs), defined as working hours >40 h/week, i.e. working hours exceeding standard working hours (35–40 h/week).
Risk factor levels	Four levels: 1. 35–40 h/week. 2. 41–48 h/week. 3. 49–54 h/week. 5. ≥55 h/week.
Theoretical minimum risk exposure level	Standard working hours, defined as working hours of 35–40 h/week (level 1).

Source: (Descatha et al., 2018).

(Table 1). Based on results from earlier studies on long working hours and health endpoints (Kivimaki et al., 2015a; Kivimaki and Kawachi 2015; Kivimaki et al., 2015b; Virtanen et al., 2012), the pre-defined exposure level categories for our systematic review were 35–40, 41–48, 49–54 and ≥55 h/week (Table 1).

The theoretical minimum risk exposure was standard working hours defined as 35–40 h/week (Table 1). The theoretical minimum risk exposure might be lower than standard working hours, but we excluded working hours <35 h/week because studies indicate that a proportion of individuals working less than those standard hours do so because of existing health problems (Kivimaki et al., 2015c; Virtanen et al., 2012). Consequently, studies using individuals working less than standard hours as the reference group were excluded from the review and meta-analysis. The category 35–40 h/week has also been used as the reference group in many large studies and previous systematic reviews (Bejot et al., 2016; Kivimaki et al., 2015a; Stevens et al., 2016; Virtanen et al., 2012).

1.3. Definition of the outcome

The WHO Global Health Estimates group outcomes into standard burden of disease categories (World Health Organization 2018), based on standard codes from the *International Statistical Classification of Diseases and Related Health Problems 10th Revision* (ICD-10) (World Health Organization, 2015). The relevant WHO Global Health Estimates category for this systematic review was "*II.H.4 Stroke*" (World Health Organization, 2018). In line with the WHO Global Health Estimates, we defined the health outcome covered in this Systematic Review as stroke, determined as conditions with ICD-10 codes I60 to I69 (Table 2). We considered prevalence of, incidence of, and mortality from stroke. Table 2 presents, for each disease or health problem included in the WHO Global Health Estimates category, the inclusion criteria for this review. This review covers all the relevant WHO Global Health Estimates categories.

1.4. How the risk factor may impact the outcome

Fig. 1 presents the logic model for our systematic review of the causal relationship between exposure to long working hours and stroke, taken from our protocol (Descatha et al., 2018). This logic model was an a priori, process-oriented one (Rehfuess et al., 2018) that sought to capture the complexity of the risk factor–outcome causal relationship (Anderson et al., 2011).

Based on knowledge of previous research on long working hours and stroke, we assumed that the effect of long working hours on stroke could be modified by country (or WHO region), sex, age, industrial sector, occupation, and formality of the economy. Confounding was considered by, at least, age, sex, and an indicator of socioeconomic status (e.g. income, education or occupational grade). Exceptions were accepted for studies whose samples were homogeneous (such as men only) or who conducted sensitivity analyses (such as sex-specific analyses).

Several variables may mediate the effects of long working hours on disease risk through two major pathways. The first one highlights behavioural responses that result in an increase in health-adverse behaviours, such as cigarette smoking, high alcohol consumption, unhealthy diet, and physical inactivity (Taris et al., 2011; Virtanen et al., 2015). Impaired sleep and poor recovery due to long working hours has also been shown to increase the risk of stroke (Sonnentag et al., 2017; Virtanen et al., 2009). Chronic psychosocial stress responses define a second pathway that would mediate the effects of long working hours on stroke. According to established physiological evidence, recurrent high physical and psychological effort (seen in long working hours) results in biological mechanisms that lead to the excessive release of stress hormones adrenalin, noradrenalin, and cortisol (Chandola et al., 2010; Jarczok et al., 2013; Nakata 2012). Over time, 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) (Kivimaki and Steptoe, 2018).

Working long hours may have a direct influence on stroke through a physiological response. In fact, chronic psychosocial stress has been shown to activate structures in the prefrontal cortex and limbic system stimulating abnormal levels of stress hormones, as well as arousing the sympathetic and vagal tone via the hypothalamic–pituitaryadrenal and sympatho-adrenal medullary axes (Steptoe and Kivimaki, 2012, 2013). These reactions may alter a range of endocrine, immune and inflammatory biomarkers with adverse effects on the cardiovascular system, such as high blood pressure (Hayashi et al., 1996), other cardiometabolic risk factors (McEwen 1998a, b), and growth of carotid intima-media thickness (Krause et al., 2009).

2. Objectives

To systematically review and meta-analyse evidence on the effect of exposure to long working hours (three categories: 41–48, 49–54, and ≥55 h/week) on stroke prevalence, incidence and mortality among workers of working age, compared with the minimum risk exposure level (standard working hours: 35–40 h/week).

3. Methods

3.1. Developed protocol

The Navigation Guide (Woodruff and Sutton 2014) for systematic reviews in environmental and occupational health was used as our guiding methodological framework, and applied wherever feasible. The Navigation Guide applies established systematic review methods from clinical medicine, including standard Cochrane methods for systematic reviews of interventions, to the field of environmental and occupational health to ensure systematic and rigorous evidence synthesis that reduces bias and maximizes transparency (Woodruff and Sutton, 2014).

Table 2
ICD-10 codes and disease and health problems covered by the WHO Global Health Estimates categories "II.H.4 Stroke" and their inclusion in the systematic review.

ICD-10 code	Disease or health problem	Included in this review
I60	Subarachnoid haemorrhage	Yes
I61	Intracerebral haemorrhage	Yes
I62	Other nontraumatic intracranial haemorrhage	Yes
I63	Cerebral infarction	Yes
I64	Stroke, not specified as haemorrhage or infarction	Yes
I65	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction	Yes
I66	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction	Yes
I67	Other cerebrovascular diseases	Yes
I68	Cerebrovascular disorders in diseases classified elsewhere	Yes
I69	Sequelae of cerebrovascular disease	Yes

Source: (Descatha et al., 2018).

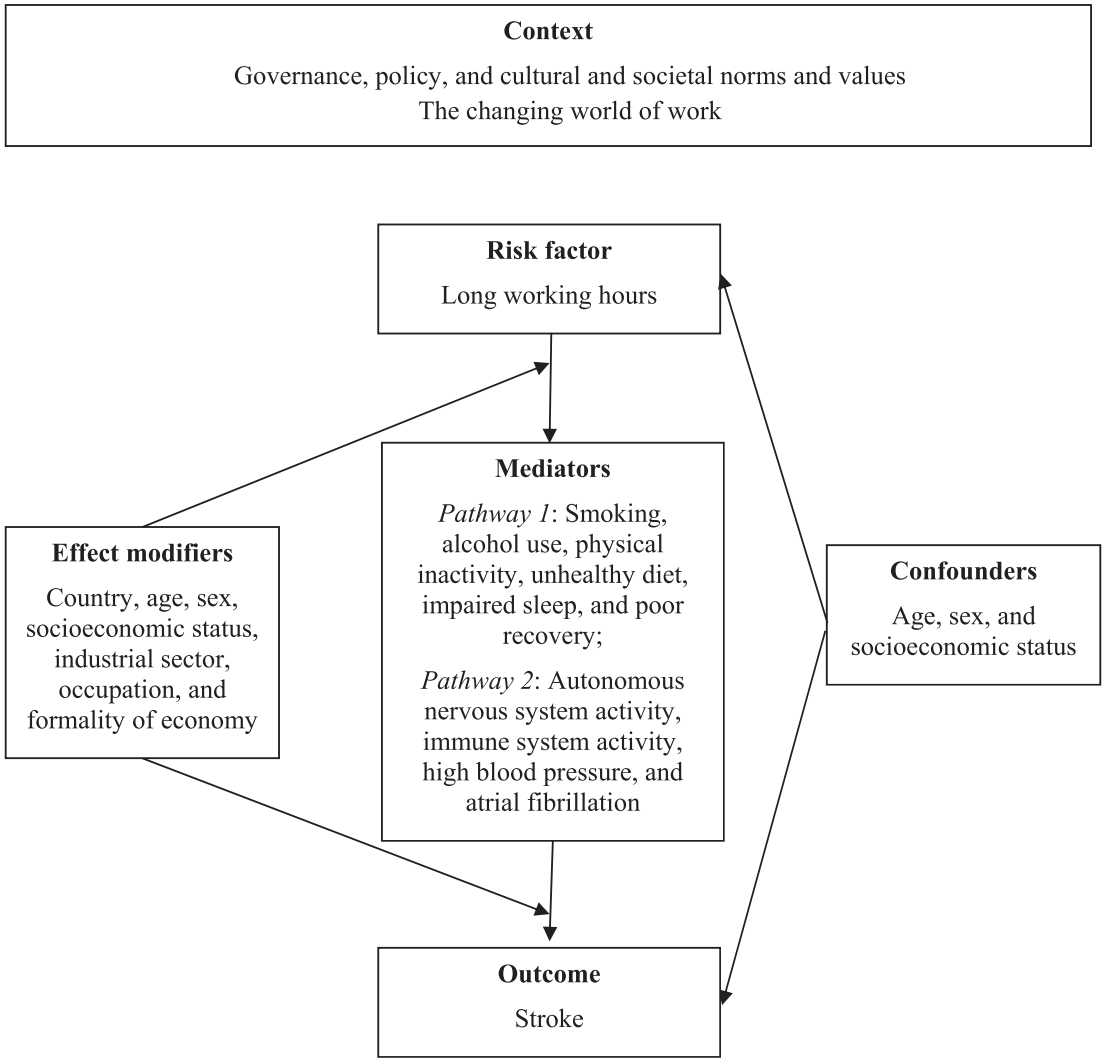

Fig. 1. Logic model of the possible causal relationship between exposure to long working hours and stroke. Source: (Descatha et al., 2018). The modifier arcs are shown here as connecting arrows to make the chart simpler to read. The hypothesized modifying effect illustrates the potential relationship between the risk factor and/or the mediators and between the mediators and the outcome.

The need for further methodological development and refinement of the relatively novel Navigation Guide has been acknowledged (Woodruff and Sutton, 2014). Our systematic review used most of the Navigation Guide framework, and steps 1–6 for the stream on human data were conducted; we left out steps for the stream on non-human data, opting instead for a brief narrative of that evidence.

We registered the protocol in PROSPERO under CRD42017060124. This registered protocol adheres with the preferred reporting items for systematic review and meta-analysis protocols statement (PRISMA-P) (Moher et al., 2015; Shamseer et al., 2015), with the abstract adhering with the reporting items for systematic reviews in journal and conference abstracts (PRISMA-A) (Beller et al., 2013). Any modification of the methods stated in the protocol was registered in PROSPERO and reported here, in the systematic review itself. Our review has been presented in concordance with the preferred reporting items for systematic review and meta-analysis statement (PRISMA) (Liberati et al., 2009). The reporting of the parameters for estimating the burden of stroke that is attributable to exposure to long working hours in the systematic review adheres to the requirements of the GATHER guidelines (Stevens et al., 2016) because the WHO/ILO Joint Estimates that may be produced consecutive to this systematic review must also adhere to these reporting guidelines.

All methods and reporting guidelines were standardised across all systematic reviews conducted for the WHO/ILO Joint Estimates (Descatha et al., 2018; Godderis et al., 2018; Hulshof et al., 2019; Li et al., 2018; Li et al., 2020; MandrioLi et al., 2018; Paulo et al., 2019; Rugulies et al., 2019; Teixeira et al., 2019; Tenkate et al., 2019).

3.2. Searched literature

3.2.1. Electronic academic databases

We searched the eight following electronic academic databases to the specified date:

1. International Clinical Trials Register Platform (to 31 May 2018).
2. Ovid MEDLINE with Daily Update (1 January 1946 to 31 May 2018, and updated on 3 April 2020).
3. PubMed (1 January 1946 to 31 May 2018, and updated on to 3 April 2020).
4. EMBASE (1 January 1947 to 31 May 2018).
5. Scopus (1 January 1788 to 31 May 2018).
6. Web of Science (1 January 1945 to 31 May 2018).
7. CISDOC (1 January 1901 to 31 December 2012, searched on 31 May 2018).
8. PsychInfo (1880 to 31 May 2018).

The Ovid MEDLINE search strategy was presented in the protocol (Descatha et al., 2018). To identify studies on stroke, we adopted or adapted several search terms or strings used in a recent Cochrane Review on Cerebrolysin for acute ischaemic stroke (Ziganshina et al., 2017). The full search strategies for all databases were revised by an information scientist and are presented in Appendix 3. Deviations from the planned search strategy are documented in *Section 8*. We performed searches in electronic databases operated in the English language using a search strategy in the English language in June and July 2018. When we neared completion of the review, we conducted a search of the Ovid MEDLINE and PubMed databases on 3 April 2020 to capture the most recent publications (e.g., publications ahead of print).

3.2.2. Electronic grey literature databases

We searched the following two electronic academic databases in May 2018:

OpenGrey (http://www.opengrey.eu/).
Grey Literature Report (https://www.nyam.org/library/collections-and-resources/grey-literature-report/).

3.2.3. Internet search engines

We also searched the internet search engines Google (www.google.com/) and GoogleScholar (www.google.com/scholar/), and screened the first 100 hits for potentially relevant records, as has been done in Cochrane Reviews (Pega et al., 2015; Pega et al., 2017).

3.2.4. Organizational websites

The websites of the seven following international organizations and national government departments were searched in May 2018:

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

3.2.5. Hand-searching and expert consultation

We hand-searched for potentially eligible studies in:

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

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

3.3. Selected studies

Study selection was carried out using the Covidence systematic review software (Covidence systematic review software). All study records identified in the search were downloaded and duplicates were identified and deleted. Afterwards, at least two review authors, working in pairs, independently screened titles and abstracts (step 1) and then full texts (step 2) of potentially relevant records. A third review author resolved any disagreements between the first two review authors. Records were not assigned to reviewers who had been authors of a given study. For example, since some authors of this systematic review are also authors of the Fadel 2019 study (Fadel et al., 2019), we ensured that study selection and all steps of the systematic review for this study were conducted exclusively by reviewers who were *not* authors of the Fadel 2019 study. The study selection was documented in a flow chart in the systematic review, as per PRISMA guidelines (Liberati et al., 2009), Appendix 8.

3.4. Eligibility criteria

The PECO criteria (Morgan et al., 2018) are described below.

3.4.1. Types of populations

We included studies of the working-age population (≥15 years) in the formal and informal economy. Studies of children (aged <15 years) and unpaid domestic workers were excluded. Participants residing in any Member (or member) State of WHO and/or ILO and any industrial sector or occupation were included. Appendix F of our protocol provides a complete, but briefer overview of the PECO criteria (Descatha et al., 2018).

3.4.2. Types of exposures

We included studies that defined long working hours in accordance with our standard definition (Table 1). We prioritized measures of the total number of hours worked, including in both main and secondary jobs, self-employment, and salaried employment, whether in the informal or the formal economy.

We included studies with objective (e.g. by means of time recording technology) and subjective measurements of long working hours, including studies that used estimates by experts (e.g. scientists with subject matter expertise) and self-reports by workers, workplace administrators, and managers. If a study presented both objective and subjective measurements, we prioritized objective ones. Studies with measures from any data source, including registry data, were included.

For studies that reported exposure levels differing from our standard levels (Table 1), we converted the reported levels to the standard levels if possible and reported analyses on these alternate exposure levels if impossible. Exposures collected and/or reported in a non-preferred unit (e.g. hours/day) were converted to the preferred unit (i.e. hours/week).

3.4.3. Types of comparators

The comparator was participants exposed to the theoretical minimum risk exposure level of 35–40 h worked per week (Table 1).

3.4.4. Types of outcomes

This systematic review included three outcomes:

1. Has had a stroke (stroke prevalence).
2. Acquired stroke (stroke incidence).
3. Died due to stroke (stroke mortality).

We included studies that defined stroke in accordance with our standard definition (Table 2). We expected that most studies examining exposure to long working hours and its effect on stroke have documented ICD-10 diagnostic codes. In the remaining cases, methods that approximate ICD-10 criteria ascertained the outcome, such as physician-obtained self-reports (see also Appendix 4 in the supplementary data and Section 5.3).

The following measurements of stroke are regarded as eligible:

i. Diagnosis by a physician with imaging.
ii. Hospital discharge records.
iii. Other relevant administrative data (e.g., records of sickness absence or disability).
iv. Medically certified cause of death.

All other measures were excluded from this systematic review.

Objective (e.g., health records) and subjective (e.g., self-reports) measures of the outcome are eligible. If a study presents both objective and subjective measurements, we prioritized the objective ones.

Studies with "mixed" outcome definitions (i.e., including both fatal stroke events and non-fatal stroke events) provide evidence on both the outcome stroke incidence and the outcome stroke mortality. These studies were consequently included in analyses on both of these outcomes, as long as they were sufficiently homogeneous statistically with studies of non-fatal events only and fatal events only, respectively (as determined by sensitivity analyses; *Section 3.9*).

3.4.5. Types of studies

We included studies that investigated the effect of long working hours on stroke for any years. Eligible study designs were 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 designs (Arditi et al., 2016). Since our objective was to quantify risk rather than simply assess hazard (Barroga and Kojima, 2013), we excluded 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 were included. Because the search was conducted using English language terms, records published in any language that presented essential information (i.e. title and abstract) in English were included. If a record was written in a language other than those spoken by the authors of this review or other reviews (Godderis et al., 2018; Hulshof et al., 2019; Li et al., 2018; Mandrioli et al., 2018; Paulo et al., 2019; Rugulies et al., 2019; Teixeira et al., 2019; Tenkate et al., 2019) in the series (i.e. Arabic, Bulgarian, Chinese, Danish, Dutch, English, French, Finnish, German, Hungarian, Italian, Japanese, Norwegian, Portuguese, Russian, Spanish, Swedish and Thai), the record was translated into English. Published and unpublished studies were included. Studies conducted using unethical practices were excluded (e.g., studies that deliberately exposed humans to a known risk factor to human health).

3.4.6. Types of effect measures

We included measures of the effect of a relevant level of long working hours on the risk of stroke (prevalence, incidence and mortality), compared with the theoretical minimum risk exposure level. We included relative effect measures such as RRs and ORs for mortality measures and hazard rate ratios for incidence measures (e.g., developed or died from stroke). Measures of absolute effects (e.g., mean differences in risks or odds) were converted into relative effect measures, but if conversion was impossible, they were excluded. To ensure comparability of effect estimates and facilitate meta-analysis, if a study presented an OR, then we converted it into a RR, if possible, using the guidance provided in the Cochrane handbook for systematic reviews of interventions (Higgins and Green, 2011).

If a study presented estimates for the effect from two or more alternative models that had been adjusted for different variables, then we prioritized the estimate from the model that provided information on the relevant confounders or mediators (at least the core variables defined in Fig. 1: age, sex, and socioeconomic status). We prioritized estimates from models adjusted for more potential confounders over those from models adjusted for fewer. We prioritized estimates from models unadjusted for mediators over those from models adjusted for mediators, because adjustment for mediators can introduce bias. We prioritized estimates from models that can adjust for time-varying confounders that are at the same time also mediators, such as marginal structural models (Pega et al., 2016), over estimates from models that can only adjust for time-varying confounders, such as fixed-effects models (Gunasekara et al., 2014), over estimates from models that cannot adjust for time-varying confounding. If a study presents effect estimates from two or more potentially eligible models, then we documented why we prioritized the selected model.

3.5. Extracted data

A standard data extraction form was developed and trialled until data extractors reached convergence and agreement. At least two review authors independently extracted data on study characteristics (including study authors, study year, study country, participants, exposure, and outcome), study design (including study type, comparator, epidemiological model(s) used, and effect estimate measure) and risk of bias (including source population representation, blinding, exposure assessment, outcome assessment, confounding, incomplete outcome data, selective outcome reporting, conflict of interest, and other sources of bias). A third review author resolved conflicts in data extraction. Data were entered into and managed with Excel.

We also extracted data on potential conflict of interest in included studies. For each author and affiliated organization of each included study record, we extracted their financial disclosures and funding sources. We used a modification of a previous method to identify and assess undisclosed financial interest of authors (Forsyth et al., 2014). Where no financial disclosure or conflict of interest statements were available, we searched the name of all authors in other study records gathered for this study and published in the prior 36 months and in other publicly available declarations of interests (Drazen et al., 2010a, 2010b).

3.6. Requested missing data

We requested missing data from the principal study author by email or phone, using the contact details provided in the principal study record. If we did not receive a positive response from the principal study author, we sent follow-up emails twice, at two and four weeks. We present a description of missing data, the study author from whom the data were requested, the date of requests sent, the date on which data were received (if any), and a summary of the responses provided by the study authors (Appendix 1 in the Supplementary data). If we did not receive some or all the requested missing data, we nevertheless retained the study in the systematic review as long as it fulfilled our eligibility criteria.

3.7. Assessed risk of bias

Standard risk of bias tools do not exist for systematic reviews for hazard identification or those for risk assessment in occupational and environmental health. The five such tools developed specifically for occupational and environmental health are for either or both hazard identification and risk assessment, and they differ substantially in the types of studies (randomized, observational and/or simulation studies) and data (e.g. human, animal and/or *in vitro*) they seek to assess (Rooney et al., 2016). However, all five tools, including the Navigation Guide, assess risk of bias in human studies similarly (Rooney et al., 2016).

Consistent with using the Navigation Guide as our organizing framework, we used its risk of bias tool, which builds on the standard risk of bias assessment methods of Cochrane (Higgins et al., 2011) and the US Agency for Healthcare Research and Quality (Viswanathan et al., 2008), and has been successfully applied in several completed and ongoing systematic reviews (Johnson et al., 2016, 2014; Koustas et al., 2014; Lam et al., 2014, 2017, 2016a; Vesterinen et al., 2014). To adhere with methods in the Navigation Guide, we used updates from its latest version published in the protocol for an ongoing systematic review

(Lam et al., 2016a).

We assessed risk of bias on the individual study level and across the body of evidence for each outcome. To judge the risk of bias in each domain, we applied a priori instructions (Li et al., 2018b), which were adapted from an ongoing Navigation Guide systematic review (Lam et al., 2016a), and further described in our protocol (Descatha et al., 2018).

All risk of bias assessors jointly trialled the application of the risk of bias criteria until they had synchronized their understanding and application of these criteria. Two or more study authors independently assessed the risk of bias for each study by outcome. Where individual assessments differed, a third author resolved the conflict. For each included study, we reported the risk of bias assessment by domains in a standard 'Risk of bias' table (Higgins et al., 2011). For the entire body of evidence, we presented the study-level risk of bias ratings by domains in a 'Risk of bias summary' figure (Higgins et al., 2011).

3.8. Synthesised evidence (including conducted meta-analysis)

We conducted meta-analyses separately for estimates of the effect on stroke prevalence, incidence, and mortality. If we found two or more studies with an eligible effect estimate, two or more review authors independently investigated the clinical heterogeneity (Deeks et al., 2011) of the studies in terms of participants (including country, sex, age and industrial sector or occupation), level of risk factor exposure, comparator, and outcomes following our protocol (Descatha et al., 2018). If the effect estimates differed considerably by WHO region, sex, and/or age, or a combination of these, then we synthesised evidence for the relevant populations defined by these variables, or combination thereof. If we found effect estimates to be clinically homogeneous across WHO regions, sex, and/or age groups, then we combined studies from all these populations into one pooled effect estimate that would be applied across all combinations of WHO regions, sexes, and age groups in the WHO/ILO Joint Estimate.

If we judged two or more studies for the relevant combination of WHO region, sex, and age group, or combination thereof, to be sufficiently clinically homogeneous to potentially be combined using quantitative meta-analysis, then we tested the statistical heterogeneity of the studies using the I^2 statistic (Figueroa, 2014). If two or more clinically homogeneous studies were found to be sufficiently homogeneous statistically to be combined in a meta-analysis, we pooled the RRs of the studies in a quantitative meta-analysis, using the inverse variance method with a random effects model to account for cross-study heterogeneity (Figueroa, 2014). Meta-analyses were conducted in RevMan 5.3.

We did not quantitatively combine data from studies with different designs (e.g. we did not combine cohort studies with case-controls studies) or levels of adjustment (e.g., we did not combine unadjusted models with adjusted models). We only pooled studies that we judged to have a minimum acceptable level of adjustment for the three core confounders identified (Fig. 1, *Section 3.4.5*).

If we found studies with "pure" outcome definitions (i.e. capturing exclusively either non-fatal or fatal stroke events) and "mixed" outcome definitions (i.e. capturing any stroke events, whether non-fatal *or* fatal), then we conducted "exploratory subgroup analyses" in which we subgrouped studies by "pure" versus "mixed" outcome definitions. Before conducting these analyses, we formulated the following rules for determining inclusion of these studies in quantitative meta-analyses:

- If there was no evidence for (meaningful) subgroup differences, then we would pool studies with "mixed" and "pure" outcome definitions.
- If there was evidence for (meaningful) subgroup differences, then we would not pool studies with "mixed" and "pure" outcome definitions.

If quantitative synthesis was not feasible (for instance, due to different exposure levels than defined above), we synthesised the study findings narratively and identified the estimates that we judged to be the highest quality evidence available.

3.9. Conducted subgroup and sensitivity analyses

We conducted subgroup analyses for the main meta-analysis and comparison of interest (i.e., the meta-analysis of cohort studies for the comparison of worked ≥55 h/week versus worked 35–40 h/week). We conducted subgroup analyses by:

- WHO region.
- Sex.
- Age group.
- Socio-economic status (SES).
- Type of stroke (ischaemic versus haemorrhagic).

There were insufficient data to conduct subgroup analyses by occupation, industrial sector, and formality of economy.

We conducted the following sensitivity analyses:

- Studies with exclusively non-fatal or fatal stroke events, compared with studies with "mixed" (non-fatal and/or fatal) stroke events.
- Studies judged to be of "high"/"probably high" risk of bias in any domain, compared with "low"/"probably low" risk of bias in all domains.
- Studies with documented or approximated ICD-10 diagnostic codes (e.g., as recorded in administrative health records), compared with studies without ICD-10 diagnostic codes (e.g., self-reports).
- Studies with different effect estimates (e.g., risk versus odds versus hazard rate ratios)
- Studies with approximate comparator definition (e.g. 7–8 h/day), compared with strict comparator definition (i.e., 35–40 h/ week).

We did not conduct sensitivity analysis on risk of bias from conflict of interest, as no included study was rated as "high"/"probably high" on this domain.

3.10. Assessed quality of evidence

We assessed quality of evidence using a modified version of the Navigation Guide quality of evidence assessment approach (Lam et al., 2016a). The approach is based on the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach (Schünemann et al., 2011) adapted specifically to systematic reviews in occupational and environmental health (Morgan et al., 2016).

At least two review authors assessed quality of evidence for the entire body of evidence by outcome, with any disagreements resolved by a third review author. We adapted the latest Navigation Guide instructions (Lam et al., 2016c) for grading the quality of evidence and presented the adapted instructions in our protocol (Descatha et al., 2018). We downgraded the quality of evidence for the following five reasons: (i) risk of bias; (ii) inconsistency; (iii) indirectness; (iv) imprecision; and (v) publication bias (Balshem et al., 2011). These items were considered downgrades if they could not be explained. When our systematic review had included ten or more studies, we aimed to generate an Egger's funnel plot to judge concerns on publication bias. If it included nine or fewer studies, we judged the risk of publication bias qualitatively.

We graded the quality of the entire body of evidence by outcome, using the three Navigation Guide standard quality of evidence ratings: "high", "moderate" and "low" (Lam et al., 2016a). Within each of the relevant domains, we rated the concern for the quality of evidence, using the ratings "none", "serious" and "very serious". As per Navigation Guide, we started at "high" for randomized studies and "moderate"

for observational studies. Quality was downgraded for a serious concern by one grade (−1) and for a very serious concern by two grades (−2). We upgraded the quality of evidence for large effect, dose–response, and residual confounding and bias not possibly explaining the effect. There had to be compelling reasons to upgrade or downgrade. If we had a serious concern for risk of bias in a body of evidence consisting of observational studies (−1), but had no other concerns, and had no reasons for upgrading, then we downgraded the quality of evidence by one grade from "moderate" to "low".

3.11. Assessed strength of evidence

Our systematic review included observational epidemiologic studies of human data only, and no other streams of evidence (e.g. no studies of non-human data). The standard Navigation Guide methodology (Lam et al., 2016a) allows for rating human and non-human animal studies separately, and then combining the strength of evidence for each stream for an overall strength of evidence rating. However, the Navigation Guide also allows for rating one stream of evidence based on the factors described above (i.e., risk of bias, indirectness, inconsistency, imprecisions, publication bias, large magnitude of effect, dose–response, and residual confounding) to arrive at an overall rating of the quality of evidence as 'high', 'moderate' or 'low' (see above and the protocol). The approach of evaluating only the human evidence stream is consistent with the GRADE methodology that has adopted the Bradford Hill considerations (Schunemann et al., 2011). So, using the method above based on the Navigation Guide incorporates the considerations of Bradford Hill (Table 3).

An additional step described in the protocol integrates the quality of the evidence (as described above) with other elements including direction of effect, confidence in the effect, and other compelling attributes of the data that may influence our certainty to allow for an overall rating that consists of "sufficient evidence of toxicity/harmfulness", "limited evidence of toxicity/harmfulness", "inadequate evidence of toxicity/harmfulness" and "evidence of lack of toxicity/harmfulness" based on human evidence. This approach to evaluate only the human evidence has been applied in previous systematic reviews (Lam et al., 2017; Lam et al., 2016a, 2016b) and verified by the US National Academy of Sciences (National Academies of Sciences 2017). It also provides two steps that integrate Bradford Hill criteria (evaluating the quality of the evidence and then evaluating the overall strength of evidence). Finally, the GRADE quality of evidence ratings (which are the same as for Navigation Guide) are analogous to the final ratings from Bradford Hill for causality (Schunemann et al., 2011) (Table 4).

4. Results

4.1. Study selection

Of the total of 7522 individual study records identified in our searches, 6 records reporting results from 22 studies fulfilled the eligibility criteria and were included in the systematic review (Fig. 2). For the 30 excluded studies that most closely resembled inclusion criteria, the reasons for exclusion are listed in Appendix 2 in the Supplementary data. Of the 22 included studies, all were included in the quantitative meta-analysis (Fig. 2).

4.2. Characteristics of included studies

The characteristics of the included studies are summarized in Table 5.

4.2.1. Study type

Twenty (20) cohort studies and two case-control studies were included. The type of effect estimates most commonly reported were hazard and odds ratios (10 studies each), and rate/risk ratios (2 studies). All included studies adjusted for the minimum set of pre-specified confounders. Some retrospective cohort and case-controls studies also adjusted for further potential confounders (Table 5).

4.2.2. Population studied

The included studies captured a total of 839,680 workers (364,616 females and 475,064 males). All but one of the studies (21) examined both female and male workers; one study captured males only (Hayashi et al., 2019). The studied age groups were working age (18–74 years at maximum range, mostly 20–65 years), with mean age varying between 39 and 54 years.

Study populations represented the WHO regions of Europe (13 studies from five countries), the Americas (six studies from one country), and the Western Pacific (three studies from two countries). The most commonly studied countries were the USA (six studies), Denmark (four studies), and Finland, Sweden, and the Republic of Korea (2 studies each).

Most studies did not provide quantitative breakdowns of participants by industrial sector and occupation. However, all included studies appeared to cover several industrial sectors and occupations. Only one study (Fadel et al., 2019) provided breakdowns by industrial sector and on occupations at the one digit level of the 2008 *International Standard Classification of Occupations* (ISCO08) (International Labour Organization 1987).

Table 3
Bradford Hill considerations and their relationship to GRADE and the Navigation Guide for evaluating the overall quality of the evidence for human observational studies.

Bradford Hill	GRADE	Navigation Guide
Strength	Strength of association and imprecision in effect estimate	Strength of association and imprecision in effect estimate
Consistency	Consistency across studies, i.e., across different situations (different researchers)	Consistency across studies, i.e., across different situations (different researchers)
Temporality	Study design, properly designed and conducted observational studies	Study design, properly designed and conducted observational studies
Biological Gradient	Dose response gradient	Dose response gradient
Specificity	Indirectness	Indirectness
Coherence	Indirectness	Indirectness
Experiment	Study design, properly designed and conducted observational studies	Study design, properly designed and conducted observational studies
Analogy	Existing association for critical outcomes leads to not downgrading the quality, indirectness	Existing association for critical outcomes leads to not downgrading the quality, indirectness. Evaluating the overall strength of body of human evidence allows consideration of other compelling attributes of the data that may influence certainty.

Footnotes: Adapted from (Schünemann et al., 2011).

Table 4
Interpretation of the GRADE ratings of the overall quality of evidence and the Navigation Guide ratings for strength of evidence evaluation.

GRADE rating for quality of evidence	Interpretation of GRADE rating	Navigation Guide rating for strength of evidence for human evidence	Interpretation of Navigation Guide rating
High	There is high confidence that the true effect lies close to that of the estimate of the effect.	**Sufficient evidence of harmfulness**	A positive relationship is observed between exposure and outcome where chance, bias, and confounding can be ruled out with reasonable confidence. The available evidence includes results from one or more well-designed, well conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies.
Moderate	There is moderate confidence in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.	**Limited evidence of harmfulness**	A positive relationship is observed between exposure and outcome where chance, bias, and confounding cannot be ruled out with reasonable confidence. Confidence in the relationship is constrained by such factors as: the number, size, or quality of individual studies, 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.
Low	The panel's confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect	**Inadequate evidence of harmfulness**	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 assessment of effects.
Very Low	There is little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.		

Footnotes: Adapted from (Schunemann et al., 2011) and (Lam et al., 2016a).

4.2.3. Exposure studied

All studies measured exposure to long working hours with either self-reported questionnaires or face-to-face interviews. Other measures such as official or company records were not used.

4.2.4. Comparator studied

The comparator for almost all studies (19) was the minimum risk exposure level: 35–40 h/week (standard working hours). Three studies reported the comparator in a non-preferred unit (i.e. hours/day), which we converted to the preferred unit of hours/week:

- The Kim case-control study (Kim et al., 2013) reported the comparator as 5–8 h/day (converted to 25–39.9 h/week); we judged the impact on results of this approximate definition as likely to be small (if any) because very few workers reported working <7 h per day (<35 h/week) in this study.
- The Japan public health center-based prospective (JPHC) study cohort II used in the Hayashi et al. study (Hayashi et al., 2019) used a comparator of 7–9 h/day (converted to 35–45 h/week), with most participants reporting working 44 h/week, with no information reported on part-time work.
- The Constances study reported by Fadel et al. (2019) used a Yes/No question on working greater than 10 h/day for at least 50 days/year (part time jobs were excluded).

We considered that these studies provided reasonable approximations of the hours/week definitions used in the study protocol, and performed additional sensitivity analyses to check for heterogeneity by approximate versus exact comparator definition (see *Section 3.9*).

4.2.5. Outcomes studied

No studies reported any evidence on the outcome of stroke prevalence.

Twenty-one studies (19 cohort studies and two case-control studies) reported evidence on the outcome "Acquired stroke" (or stroke incidence). Of these, nine studies (seven cohort studies and two case-control studies) defined the outcome as incidence of a non-fatal stroke event, and 12 studies (all cohort studies) as an incident event that was either non-fatal or fatal ("mixed"). Incident stroke was defined as first time event in 16 studies and as retrospective lifetime new episode of stroke in one study (Fadel et al., 2019). Most of these studies used administrative data (8) and one study used physician interviews. The time between the exposure and the outcome studied varied from 8 years to 20 years (and working time in a respective study) for incidence of stoke, and 8 to 20 years for mortality from stroke.

Thirteen studies (all cohort studies) reported evidence on the outcome "died due to stroke" (stroke mortality). One of these studies defined the outcome as a fatal stroke event, and 12 studies used a "mixed" outcome definition including both fatal and/or non-fatal stroke events.

4.3. Risk of bias at individual study level

4.3.1. Acquired stroke (stroke incidence)

The risk of bias ratings for each domain for all 21 included studies for this outcome are presented in Fig. 3, and the justification for each rating for each domain by included study is presented in Appendix 4 in the Supplementary data. In assessing the quality of evidence for the incidence of stroke, we prioritized the evidence from the 19 cohort studies as the main evidence for the outcome due to less risk of bias (Fig. 3) and we deprioritized the evidence from case-control studies as supporting evidence only.

4.3.1.1. Selection bias. Selection bias was assessed by determining whether the groups being compared were the same in all relevant ways (or as close to this as possible) apart from the exposure. For the 19 cohort studies included for this outcome, selection bias was rated high for one study (Hayashi et al., 2019) because data were collected from only one prefecture (out of six possible regions) and because female workers were excluded (Tsugane et al., 2001, Li et al., 2019, see specific comment in Appendix 4). Selection bias was rated as probably high for seven studies (Kivimaki 2015 - ACL 1986, Kivimaki 2015 – COPSOQ I 1997, Kivimaki 2015 – COPSOQ II 2004, Kivimaki 2015 – FPS 2000, Kivimaki 2015 – HESSUP 1998, Kivimaki 2015 – IPAW 1996, Kivimaki 2015 – Whitehall II 1991) based on only a specific population or industry being included and a very low study participation rate; probably low for seven studies (Hannerz et al., 2018, Kivimaki 2015 - MIDUS 1995, Kivimaki 2015 – PUMA 1999, Kivimaki 2015 – WLSG 1992, Kivimaki 2015 – WLSS 1993, Kivimaki 2015 – WOLF N 1996, Kivimaki 2015 – WOLF S 1992) because of either low study participation rate (<70%) with possible difference between responders and non-responders, or only indirect evidence being available on inclusion criteria, recruitment and enrolment procedures;

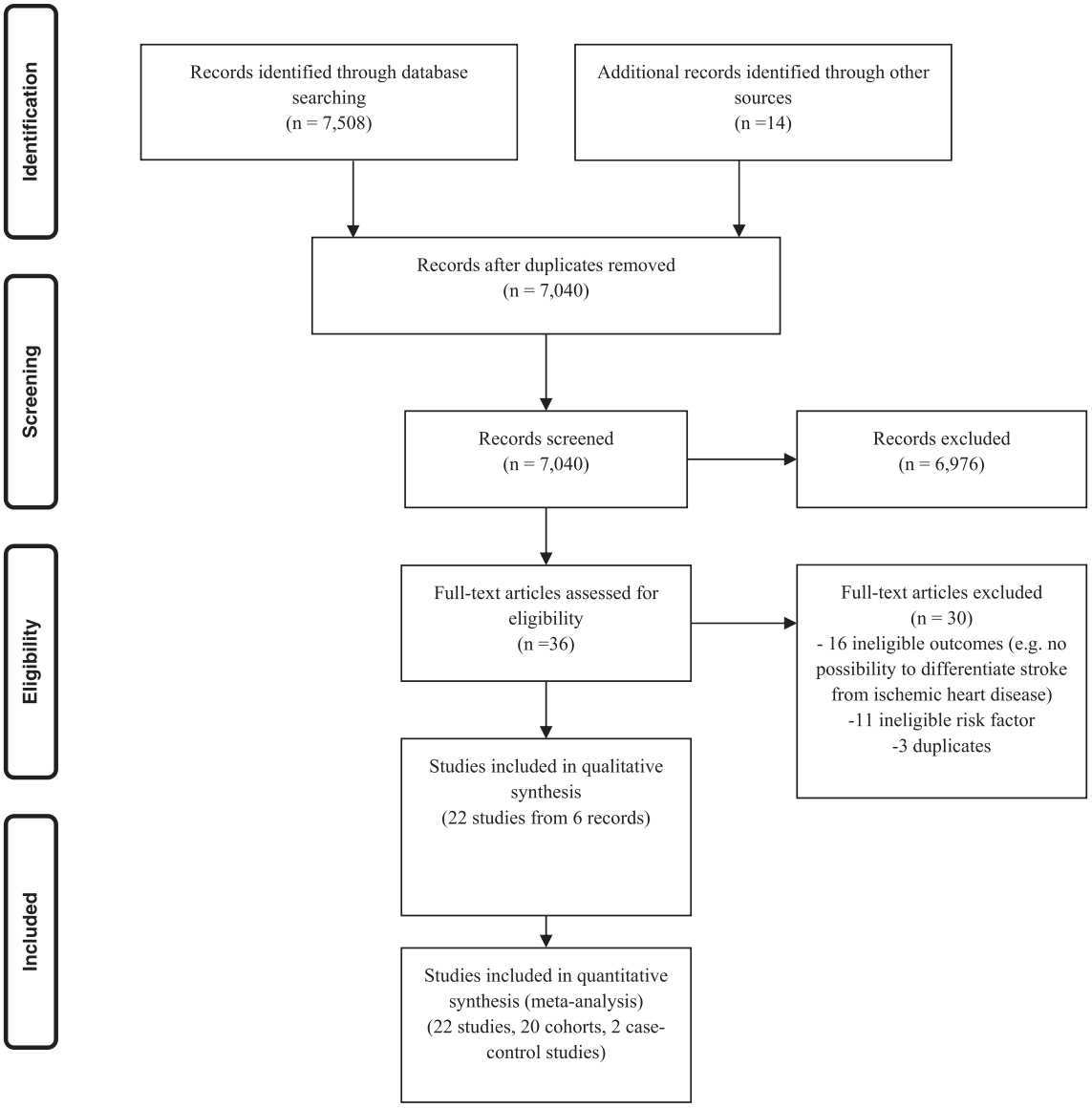

Fig. 2. Flow diagram of study selection.

and low for four studies for which we had no or only minor concerns for risk of bias from selection.

For case-control studies, the risk of selection bias was rated probably high in both studies. Indeed, unlike cohort studies, most case-control studies due to their high selectivity in from where they drew the cases or controls did not adequately represent their target population.

4.3.1.2. Performance bias. For the included cohort studies, blinding of study participants and study personnel to exposure assignment and to study participants' characteristics was usually not mentioned. Because data were extracted from multipurpose studies that include many other exposures and outcomes we rated all cohort studies as probably low risk of performance bias.

For the two case-control studies, the risk of performance bias was rated as probably low in one study (Jeong, 2013) because there was only indirect evidence that the study personnel were blinded to exposure assignment or participants' characteristics; there was no concern about blinding of study personnel in the other case-control study (Kim, 2013), which we rated as low risk of performance bias.

4.3.1.3. Detection bias (exposure assessment). For cohort studies, although an objective assessment of the exposure would have reduced the risk of detection bias by self-reported exposures, we judged that standard self-reported assessment of exposure did not introduce a noteworthy risk of detection bias. Self-reports of (long) working hours have been validated against objective measures of long working hours (Imai et al., 2016). Other data coming from physician working hours also showed good validity of such self-reported work-exposures (Todd et al, 2010; Saunders et al., 2005). In addition, self-report might be more representative of the total work exposure (mail, phone, social network), including any secondary jobs (Aziz et al., 2019). In our opinion it is unlikely that use of self-reported exposures in the included cohort studies introduced any substantial detection bias. We consequently rated all studies as probably carrying a low risk of detection bias in the exposure assessment.

For case-control studies, we rated risk of detection bias as probably low in all studies. We again judged that self-report of (long) working hours was unlikely to have introduced any substantial risk of detection bias, since such subjective measures have been validated against objective measures of long working hours (e.g. Imai et al., 2016).

4.3.1.4. Detection bias (outcome assessment). For the included cohort

Table 5
Characteristics of included studies in the systematic review of long working hours and stroke.

(Part I: study population and study type)

Study	Study population							Study type				
Study ID	Total number of study participants	Number of female study participants	Country of study population	Geographic location (specify as 'national' or list regions or sites)	Industrial sector (specify ISIC-4 code provided in worksheet "Industrial sector codes")	Occupation (specify ISCO-08 code provided in worksheet "Occupation code")	Age	Study design	Study period (month of first collection of any data and month of last collection of any data)	Follow-up period (period in months between exposure and outcome)	Response at baseline (%)	Loss to follow-up (%)
Fadel 2019	143,592	71,041	France	National	All (except farmers)	Not reported	18–69 years	Retrospective cohort study	2012 Jan-2018 Dec	Working time (retrospective)	96.7	12.5
Hannerz 2018	149,811	77,902	Denmark	National	Not reported	Not reported	15–75 years at baseline	Prospective cohort study	1999 (baseline) to 2013	Average 7.7 years (based on Ischemic heart disease study data)	Between 53 and 70	0.5
Hayashi 2019	15,277	0	Japan	region	Not reported	All (homemaker excluded)	40–59 years at baseline (means 48.1 to 51.2)	Prospective cohort study	1993 (baseline) to2012	20 years	82.7	17.1
Jeong 2013	1018	149 (male 82.8%)	Korea, Republic of	National	Not reported	Not reported	Mean age: cases 49.5 years, controls 49.8 years	Case-control study	November 2009 to October 2011	Not applicable (retrospective)	80.9 (among controls)	15.1
Kim 2013	2820	1413	Korea, Republic of	National	Not reported	Not reported	Mean age: cases 54.1 years, controls 53.6 years	Case-control study	October 2002 - March 2004	Not applicable (retrospective)	53.7	11.6
Kivimaki 2015 - ACL 1986	1502	802	United States	National	Not reported	Not reported	Mean age 44.5 years at baseline	Prospective cohort study	1986 (baseline) to 2002	16 years	68	19
Kivimaki 2015 - Alameda 1973	1585	666	United States	Region	Not reported	Not reported	Mean age 44.4 years at baseline	Prospective cohort study	1973 (baseline) to 1994	20 years	86	40
Kivimaki 2015 - COPSOQ-I 1997	1803	876	Denmark	National	Not reported	Not reported	Mean age 40.6 years at baseline	Prospective cohort study	1997 (baseline)	Not reported	61	<10
Kivimaki 2015 - COPSOQ-II 2004	3389	1785	Denmark	National	Not reported	Not reported	Mean age 42.7 years at baseline	Prospective cohort study	2004 (baseline)	Not reported	59	<10
Kivimaki 2015 - DWECS 2000	5535	2590	Denmark	National	Not reported	Not reported	Mean age 41.8 years at baseline	Prospective cohort study	2000 (baseline)	Not reported	75	<10
Kivimaki 2015 - FPS 2000	44,565	35,840	Finland	National	Not reported	Not reported	Mean age 44.6 years at baseline	Prospective cohort study	2000 (baseline)	Not reported	68	<10
Kivimaki 2015 - HeSSup 1998	16,150	8971	Finland	National	Not reported	Not reported	Mean age 39.6 years at baseline	Prospective cohort study	1998 (baseline)	Not reported	40	<10
Kivimaki 2015 - MIDUS 1995	3303	1637	United States	National	Not reported	Not reported	Mean age 44.2 years at baseline	Prospective cohort study	1995 (baseline) to 2005	10 years	61	32
Kivimaki 2015–IPAW 1996	2021	1360	Denmark	Regional	84, 86, 87	all	Mean age 41.2 years at baseline	Prospective cohort study	1996–1997 (baseline)	Not reported	76	<10

(*continued on next page*)

Table 5 (*continued*)

(Part I: study population and study type)

Study	Study population							Study type				
Study ID	Total number of study participants	Number of female study participants	Country of study population	Geographic location (specify as 'national' or list regions or sites)	Industrial sector (specify ISIC-4 code provided in worksheet "Industrial sector codes")	Occupation (specify ISCO-08 code provided in worksheet "Occupation code")	Age	Study design	Study period (month of first collection of any data and month of last collection of any data)	Follow-up period (period in months between exposure and outcome)	Response at baseline (%)	Loss to follow-up (%)
Kivimaki 2015 - NHANES I 1982	4875	2800	United States	National	Not reported	Not reported	Mean age 48.8 years at baseline	Prospective cohort study	1982 (baseline) to 1992	10 years	93	<10
Kivimaki 2015 - O'Reilly 2013	414,949	144,938	United Kingdom (Northern Ireland)	Region	Not reported	Not reported	Mean age 39.0 years at baseline	Prospective cohort study	2001 (baseline) to 2009	8.7 years (linkage)	88	<10
Kivimaki 2015–PUMA 1999	1783	1473	Denmark	Regional	82 (Human service)	all	Mean age 39.6 years at baseline	Prospective cohort study	1999 (baseline)	Not reported	80	<10
Kivimaki 2015 -Whitehall II 1991	7614	2327	United Kingdom	Regional	84	all	Mean age 49.1 at baseline	Prospective cohort study	1991–1994 (baseline)	Not reported	74	<10
Kivimaki 2015 - WLSG 1992	5421	2883	United States	Region	Not reported	Not reported	Mean age 54.1 years at baseline	Prospective cohort study	1992 (baseline) to 2003–2005	10 years	82	17
Kivimaki 2015 - WLSS 1993	2366	1299	United States	Region	Not reported	Not reported	Mean age 52.4 years at baseline	Prospective cohort study	1993 (baseline) to 2004–2007	12 years	56	26
Kivimaki 2015 - WOLF-N 1996	4648	772	Sweden	Region	Not reported	Not reported	Mean age 44.0 years at baseline	Prospective cohort study	1996–1998 (baseline)	Not reported	76	<10
Kivimaki 2015 - WOLF-S 1992	5554	2395	Sweden	Region	Not reported	Not reported	Mean age 41.5 years at baseline	Prospective cohort study	1992–1995 (baseline)	Not reported	82	<10

(Part II: exposure assessment and comparator)

Study	Exposure assessment							Comparator
Study ID	Exposure definition (i.e. how was the exposure defined?)	Unit for which exposure was assessed	Mode of exposure data collection	Exposure assessment methods	Levels/intensity of exposure (specify unit)	Number of study participants in exposed group (highest exposed group)	Number of study participants in unexposed group	Definition of comparator (define comparator group, including specific level of exposure)
Fadel 2019	Work more than 10 h, more than 50 days year	Individual level	Pen-and-paper survey	Self-report	Yes versus No (plus number of years of exposure)	42,542	95,391	No work more than 10 h, more than 50 days a year
Hannerz 2018	Weekly working hours	Individual level	Telephone interview	Self-report	35–40 h/week, 41–48 h/ week, 49–54 h/week, ≥55 h/week	5099	124,944	35–40 h/week
Hayashi 2019	Working hours by day	Individual level	Survey	Self-report	<7 h/ day ; 7–<9 h/day; 9–<11 h/day; ≥11 h/day	2042	6596	7–9 h /day
Jeong 2013	Working hours per week (short term)	Individual level	Face-to-face survey	Self-report	≤ 40 h/week, 40–50 h/ week, 50–55 h/week, ≥55 h/week	Not reported	Not reported	≤ 40 h/week

(*continued on next page*)

Table 5 (*continued*)

(Part II: exposure assessment and comparator)

Study	Exposure assessment							Comparator
Study ID	Exposure definition (i.e. how was the exposure defined?)	Unit for which exposure was assessed	Mode of exposure data collection	Exposure assessment methods	Levels/intensity of exposure (specify unit)	Number of study participants in exposed group (highest exposed group)	Number of study participants in unexposed group	Definition of comparator (define comparator group, including specific level of exposure)
Kim 2013	Hours worked/day	Individual level	Face-to-face survey	Face-to-face interviews administered by trained nurses	4–8 h/ day 5–8 h/day; 9–12 h/day; ≥13 h/day	1330	1473	4–8 h/day
Kivimaki 2015 - ACL 1986	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	181	1321	35–40 h/week
Kivimaki 2015 - Alameda 1973	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	152	1433	35–40 h/week
Kivimaki 2015 - COPSOQ-I 1997	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	109	1694	35–40 h/week
Kivimaki 2015 - COPSOQ-II 2004	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	177	3212	35–40 h/week
Kivimaki 2015 - DWECS 2000	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	440	5095	35–40 h/week
Kivimaki 2015 - FPS 2000	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	1414	43,151	35–40 h/week
Kivimaki 2015 - HeSSup 1998	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	1417	14,733	35–40 h/week
Kivimaki 2015 - MIDUS 1995	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	464	2893	35–40 h/week
Kivimaki 2015–IPAW 1996	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week,	Not reported	Not reported	35–40 h/week
Kivimaki 2015 - NHANES I 1982	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	477	4398	35–40 h/week
Kivimaki 2015 - O'Reilly 2013	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, ≥55 h/ week	39,069	375,880	35–40 h/week
Kivimaki 2015–PUMA 1999	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks	Not reported	Not reported	35–40 h/week
Kivimaki 2015 -Whitehall II 1991	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	745	6869	35–40 h/week
Kivimaki 2015 - WLSG 1992	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥5 h/week	724	4697	35–40 h/week
Kivimaki 2015 - WLSS 1993	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks, ≥55 h/week	324	2042	35–40 h/week
Kivimaki 2015 - WOLF-N 1996	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/ week, 49/54 h/weeks	Not reported	Not reported	35–40 h/week

(*continued on next page*)

Table 5 (*continued*)

(Part II: exposure assessment and comparator)

Study	Exposure assessment							Comparator
Study ID	Exposure definition (i.e. how was the exposure defined?)	Unit for which exposure was assessed	Mode of exposure data collection	Exposure assessment methods	Levels/intensity of exposure (specify unit)	Number of study participants in exposed group (highest exposed group)	Number of study participants in unexposed group	Definition of comparator (define comparator group, including specific level of exposure)
Kivimaki 2015 - WOLF-S 1992	Weekly working hours	Individual level	Survey	Self-report	35–40 h/week, 41/48 h/week, 49/54 h/weeks, ≥55 h/week	232	5322	35–40 h/week

(Part III: outcome assessment and statistical modelling)

Study	Outcome assessment							Statistical modelling						
Study ID	Definition of outcome	Which International Classification of Diseases (ICD) code was reported for the outcome (if any)?	Diagnostic assessment method	Number of cases with outcome of interest in exposed group	Number of non-cases (i.e. without outcome of interest) in exposed group	Number of cases with outcome of interest in unexposed group	Number of non-cases (i.e. without outcome of interest) in unexposed group	Adjusted for confounding by: age	Adjusted for confounding by: sex	Adjusted for confounding by: Socioeconomic status (please specify indicator, e.g. level of education)	Other potential confounders adjusted for (please specify)	Adjusted for mediation by (please specify)	Other potential mediators adjusted for	Treatment effect measure type
Fadel 2019	Stroke (acquired)	ICD10: I60/I61/I63/I64	Medical interview	364	Not reported	763	Not reported	Yes	Yes	Yes, PCS (French socioeconomic status)	No (but available in not prioritized model, cerebrovascular risk factors)	No (but available in not prioritized model, cerebrovascular risk factors)	No	Hazard rate ratio
Hannerz 2018	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	Clinically registered hospital treatment or death	59	Not reported	1418	Not reported	Yes	Yes	Yes ESeC (European socioeconomic classification)	Yes (calendar time, time passed since start of follow-up)	No	No	Hazard rate ratio
Hayashi 2019	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	National stroke criteria	59	Not reported	283	Not reported	Yes	No (but only male)	No	Yes (cerebrovascular risk factors)	No	No	Hazard rate ratio
Jeong 2013	Stroke (acquired)	ICD10: I60/I61/I63	Diagnosis based on registration	Not applicable because of study design	Not applicable because of study design	Not applicable because of study design	Not applicable because of study design	Yes	Yes	Yes (level education)	Yes (cerebrovascular risk factor and white or blue collar)	No	No	Odds ratio
Kim 2013	Stroke (acquired)	ICD10: I60/I61/I63	Physician diagnostic record	Not applicable because of study design	Not applicable because of study design	Not applicable because of study design	Not applicable because of study design	Yes	No (paired)	No	Yes (Phenylpropanolamine)	No	No	Odds ratio
Kivimaki 2015 - ACL 1986	Stroke (acquired)	ICD10: I60/I61/I63/I64	Self-report (diagnosed by a physician)	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Odds ratio
	Stroke (acquired)	ICD10: I60/I61/I63/I64	Self-report (diagnosed	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/	No	No	No	Odds ratio

(*continued on next page*)

Table 5 (*continued*)

(Part III: outcome assessment and statistical modelling)

Study	Outcome assessment							Statistical modelling						
Study ID	Definition of outcome	Which International Classification of Diseases (ICD) code was reported for the outcome (if any)?	Diagnostic assessment method	Number of cases with outcome of interest in exposed group	Number of non-cases (i.e. without outcome of interest) in exposed group	Number of cases with outcome of interest in unexposed group	Number of non-cases (i.e. without outcome of interest) in unexposed group	Adjusted for confounding by: age	Adjusted for confounding by: sex	Adjusted for confounding by: Socioeconomic status (please specify indicator, e.g. level of education)	Other potential confounders adjusted for (please specify)	Adjusted for mediation by (please specify)	Other potential mediators adjusted for	Treatment effect measure type
Kivimaki 2015 - Alameda 1973			by a physician)							intermediate/ high classes)				
Kivimaki 2015 - COPSOQ-I 1997	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - COPSOQ-II 2004	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - DWECS 2000	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - FPS 2000	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - HeSSup 1998	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015–IPAW 1996	Stroke (acquired + fatal)	ICD10: I60/ I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - MIDUS 1995	Stroke (acquired)	ICD10: I60/ I61/I63/I64	Self-report (diagnosed by a physician)	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Odds ratio
Kivimaki 2015 - NHANES I 1982	Stroke (acquired)	ICD10: I60/ I61/I63/I64	Self-report (diagnosed by a physician)	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/ intermediate/ high classes)	No	No	No	Odds ratio
Kivimaki 2015 - O'Reilly 2013	Stroke (fatal)	ICD-10 I60-69	Death certificate	Not reported	Not reported	Not reported	Not reported	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Hazard rate ratio

(*continued on next page*)

Table 5 (*continued*)

(Part III: outcome assessment and statistical modelling)

Study	Outcome assessment							Statistical modelling						
Study ID	Definition of outcome	Which International Classification of Diseases (ICD) code was reported for the outcome (if any)?	Diagnostic assessment method	Number of cases with outcome of interest in exposed group	Number of non-cases (i.e. without outcome of interest) in exposed group	Number of cases with outcome of interest in unexposed group	Number of non-cases (i.e. without outcome of interest) in unexposed group	Adjusted for confounding by: age	Adjusted for confounding by: sex	Adjusted for confounding by: Socioeconomic status (please specify indicator, e.g. level of education)	Other potential confounders adjusted for (please specify)	Adjusted for mediation by (please specify)	Other potential mediators adjusted for	Treatment effect measure type
Kivimaki 2015–PUMA 1999	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 -Whitehall II 1991	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Hazard rate ratio
Kivimaki 2015 - WLSG 1992	Stroke (acquired)	ICD10: I60/I61/I63/I64	Self-report (diagnosed by a physician)	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Odds ratio
Kivimaki 2015 - WLSS 1993	Stroke (acquired)	ICD10: I60/I61/I63/I64	Self-report (diagnosed by a physician)	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Odds ratio
Kivimaki 2015 - WOLF-N 1996	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Odds ratio
Kivimaki 2015 - WOLF-S 1992	Stroke (acquired + fatal)	ICD10: I60/I61/I63/I64	Hospital and mortality record	Not reported	Not reported	Not reported	Not reported	Yes	Yes	Yes (not specified, low/intermediate/high classes)	No	No	No	Hazard rate ratio

Footnotes: N/A because of study design.

	Fadel 2019	Hannerz 2018	Hayashi 2019	Jeong 2013	Kim 2013	Kivimaki 2015 - ACL 1986	Kivimaki 2015 - Alameda 1973	Kivimaki 2015 - COPSOQ I 1997	Kivimaki 2015 - COPSOQ II 2004	Kivimaki 2015 - DWECS 2000	Kivimaki 2015 - FPS 2000	Kivimaki 2015 - HeSSup 1998	Kivimaki 2015 - IPAW 1996	Kivimaki 2015 - MIDUS 1995	Kivimaki 2015 - NHANES 1982	Kivimaki 2015 - PUMA 1999	Kivimaki 2015 - Whitehall II 1991	Kivimaki 2015 - WLSG 1992	Kivimaki 2015 - WLSS 1993	Kivimaki 2015 - WOLF N 1996	Kivimaki 2015 - WOLF S 1992
1. Are the study groups at risk of not representing their source populations in a manner that might introduce selection bias?	Low	Probably low	High	Probably high	Probably high	Probably high	Low	Probably high	Probably high	Low	Probably high	Probably high	Probably high	Probably low	Low	Probably low	Probably high	Probably low	Probably low	Probably low	Probably low
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Probably low	Probably low	Probably low	Low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
3. Were exposure assessment methods lacking accuracy?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
4. Were outcome assessment methods lacking accuracy?	Probably low	Low	Low	Low	Low	Probably low	Probably low	Low	Low	Low	Low	Low	Low	Probably low	Probably low	Low	Low	Probably low	Probably low	Low	Low
5. Was potential confounding inadequately incorporated?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
6. Were incomplete outcome data inadequately addressed?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably high	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably high	Probably low	Probably low	Probably low	Probably low	Probably high	Probably low	Probably low
7. Does the study report appear to have selective outcome reporting?	Probably low	Probably low	Probably low	Probably low	Probably low	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	Low	Low	Low	Low	Low	Probably low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low

Fig. 3. Summary of risk of bias, Acquired stroke (non-fatal and mixed non-fatal/fatal stroke). Footnotes: N/A = not applicable.

studies, the 10 studies with mixed outcome definition that comprised both fatal and non-fatal stroke events used administrative data, either physician-based clinical diagnoses or ICD-coded records. Therefore, we rated risk of detection bias for these studies as low or probably low. For the incidence studies based on self-reported outcome data, our rating for detection bias was probably low. In the previous meta-analysis by Kivimaki et al. no difference was observed in subgroup analyses with and without self-report, an observation repeated by Wong et al more recently (Kivimaki et al., 2015a; Wong et al., 2019). Fair validity of self-reported stroke for clinical practice is established (Woodfield et al., 2015a; Woodfield et al., 2015b). The clinical experts in our group (neurologists, emergency physicians) indicated that self-report of a previous stroke is generally considered as a stroke in clinical practice. Though appropriate QA/QC for methods were not clearly specified for all included studies, we judged risk of detection bias in the outcome assessment to be probably low or low.

For case-control studies, all the included studies were rated as low risk of detection bias, because the definition of cases was based on diagnostic procedures including interviews conducted by trained interviewers (Jeong et al., 2013) and imaging and clinical examinations (Kim et al., 2013).

4.3.1.5. Confounding. For cohort studies, 18 included studies calculated effect estimates using appropriate statistical models that adjusted for confounding by the minimum set of confounders (age, sex, and SES) that we had pre-specified in our protocol. The only exception was Hayashi et al. (2019), which was restricted to male participants and did not adjust for confounding by sex (Hayashi et al., 2019). Some studies also controlled for potential other confounders, mediators, and/or moderators and/or reported that additional factors were controlled for but did not affect the results (Fadel et al., 2019; Hayashi et al., 2019; Jeong et al., 2013; Kim et al., 2013). All studies were rated as probably low risk of confounding.

The two included case-control studies calculated effect estimates using appropriate statistical models that adjusted for confounding by the minimum set of confounders pre-specified in our protocol. These studies were considered probably low risk for confounding.

4.3.1.6. Selection bias (incomplete outcome data). In all 19 included cohort studies, we judged that participants were followed up for a sufficient length of exposure to long working hours (at least 5 years) for them to reasonably have acquired the outcome due to this exposure. The proportion of outcome data missing at follow-up over time, as documented at study terminus, was acceptably low (i.e. $<20\%$). There was balance across exposure groups in the survey non-response at baseline, item non-response at baseline, missing participants at final follow up and missing outcome data at final follow-up, with similar reasons for missing study participants and/or outcome data across groups (if reported), and/or the missing outcome data were imputed using appropriate statistical methods.

Based on these considerations, we judged most of the cohort studies to have probably low risk of selection bias due to incomplete outcome data, except three studies with an attrition over 20% that might lead to bias (Kivimaki 2015 – Alameida 1973, Kivimaki 2015 - MIDUS 1995, Kivimaki 2015 – WLSS 1993)

For case-control studies, outcome data were complete, with no outcome data missing from any study participant by definition. Therefore, all studies were rated as probably low risk.

4.3.1.7. Reporting bias. In all cohort studies with pre-published protocols, the outcomes were reported in the included study record as they had been pre-specified in the protocol. In the cohort studies without a pre-published protocol, the outcomes were reported in the results sections of the study records as they had been anticipated in the abstracts and methods sections in the study record, and we also did not find any other evidence that reporting may have been biased. We consequently judged risk of reporting bias as probably low in all published cohort studies, while we considered it not applicable for the included unpublished studies because none of these were originally designed to study the effect of exposure to long working hours on stroke.

For case-control studies, there was no concern for selective reporting bias, as the outcomes were reported in the results sections of the study records as they had been anticipated in the abstracts and methods sections in the study record. However, as we had to request the authors to provide additional data and to re-run some analysis, we rated all studies as having probably low risk of reporting bias.

4.3.1.8. Conflict of interest. None of the included studies received financial support from a company or other entity with a financial interest in the study findings. All were funded by public research agencies or related organizations that were free from commercial interests in the study findings; were authored only by persons who were not affiliated with companies or other entities with vested interests; and/or no conflict of interest was declared by study authors.

Therefore, we rated 18 studies as having low risk of bias from conflict of interest. The only exception was one cohort study that was rated as probably low risk of bias from conflict of interest, because the study authors did not explicitly report the lack of conflict of interest, although there was no evidence of commercial interests influencing the

study (Kivimaki 2015 – Alameda 1973).

Similarly, we judged all case-control studies to have low risk of bias in this domain, because these studies were also conducted exclusively by researchers who were publicly funded, and we also found no evidence of commercial interests influencing these studies.

4.3.1.9. Other risk of bias. We also considered other types of bias for cohort studies. In Fadel, the retrospective design might have led to a bias in interpretation (Fadel et al., 2019). However, the analyses with exposure lag time led us to rate the risk of this potential bias as probably low. We also considered short periods of follow-up time along with the rarity of the event, and possible healthy workers effects in our rating of probably low risk of bias for the cohort studies.

For the case-control studies, we considered whether inclusion of recurrent stroke might have produced a bias, but did not find a study reporting data on first and recurrent stroke. We also considered potential bias due to missing information on stroke diagnosis. However, given the selection criteria, such bias seemed unlikely. Therefore, we rated all studies as probably low risk.

4.3.2. Died from stroke

The risk of bias ratings for each domain for all included studies for this outcome are presented in Fig. 4, and the justification for each rating for each domain by included study is presented in Appendix 5 in the Supplementary data. We judged the risk of bias to be probably low in all included studies (Fig. 4). No case-control studies were included.

4.3.2.1. Selection bias. The only mortality study (O'Reilly and Rosato 2013, cited in Kivimaki et al., 2015a) was considered low risk of bias. From mixed fatal/non-fatal studies (already detailed, see also Fig. 4, Appendix 5), two were classified as low risk and another four were probably low risk of selection bias; one study was rated at high risk, and six at probably high risk. So, we rated selection bias between probably high and low.

4.3.2.2. Performance bias. Blinding of group assignment is usually not mentioned in the study records. However, there is a very low probability of risk of bias due to potential impact of knowledge of group assignment on exposure or outcome, particularly as outcome is mostly assessed by administrative data or death register data. Therefore, all studies are rated as probably low risk.

4.3.2.3. Detection bias (exposure assessment). Although an objective assessment of the exposure would have strengthened the quality, the uniform standard self-report assessment of exposure was considered to reduce this detection bias. The validity of this information was shown to be quite high (Imai et al., 2016), and it is unlikely that reduced precision of measurement exerted a systematic bias. Therefore, all studies were rated as probably low risk.

4.3.2.4. Detection bias (outcome assessment). All studies with mixed fatal and non-fatal stroke outcomes used administrative data, either physician-based clinical diagnoses or stroke-classified records. Studies used death registry data with specific ICD codes were very accurate. For each included study, we consequently rated the risk of bias to be low.

4.3.2.5. Confounding. Studies appropriately accounted for most, but not all of the important confounders. Included confounders were age, sex, and socioeconomic status. Some studies additionally controlled for further factors that could act as mediators or moderators, or reported that additional factors were controlled for but did not affect the results. All reports used appropriate statistical techniques of confounder control. As these procedures are not expected to introduce substantial bias, all studies were rated as probably low risk at the individual study level.

4.3.2.6. Selection bias (incomplete outcome data). Participants were followed long enough to obtain outcome measurements. Any one of the following criteria may have been additionally applied: (1) Attrition or missing data were balanced in numbers across exposure groups, with similar reasons for missing data across groups; (2) Missing data were imputed using appropriate methods; (3) more than 50% of the baseline population was followed-up with outcome data. Based on these considerations, all studies were rated as probably low risk of bias.

4.3.2.7. Reporting bias. All the studies' pre-specified outcomes were outlined either in the pre-published protocol or in the published manuscript (Abstract and/or Methods sections). A respective bias is therefore unlikely, and all studies were rated as probably low risk.

4.3.2.8. Conflict of interest. All cohort studies did not receive support from a company or other entity having a financial interest in the outcome of the study; were funded by public research agencies or related organizations that were free from commercial interests in study findings; had authors unaffiliated with companies or other entities with vested interests, but instead were mostly affiliated with public institutions or organizations; and/or had study authors declaring no conflict of interest. Therefore, all studies were rated low risk of bias.

	Hannerz 2018	Hayashi 2019	Kivimaki 2015 - COPSOQ I 1997	Kivimaki 2015 - COPSOQ II 2004	Kivimaki 2015 - DWES 2000	Kivimaki 2015 - Hessup 1998	Kivimaki 2015 - FPS 2000	Kivimaki 2015 - IPAW 1996	Kivimaki 2015 O'Reilly 2013	Kivimaki 2015 - PUMA 1999	Kivimaki 2015 - Whitehall II 1991	Kivimaki 2015 - WOLF N 1996	Kivimaki 2015 - WOLF S 1992
1. Are the study groups at risk of not representing their source	Probably low	High	Probably high	Probably high	Low	Probably high	Probably high	Probably high	Low	Probably low	Probably high	Probably low	Probably low
2. Was knowledge of the group assignments inadequately	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
3. Were exposure assessment methods lacking accuracy?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
4. Were outcome assessment methods lacking accuracy?	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
5. Was potential confounding inadequately incorporated?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
6. Were incomplete outcome data inadequately addressed?	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low
7. Does the study report appear to have selective outcome	Probably low	Probably low	N/A	N/A	N/A	N/A	N/A	N/A	Probably low	N/A	N/A	N/A	N/A
8. Did the study receive any support from a company, study	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
9. Did the study appear to have other problems that could put it	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low	Probably low

Fig. 4. Summary of risk of bias, Died from stroke. Footnotes: N/A = Not applicable.

4.3.2.9. Other risk of bias. We did not find any evidence to assume high or probably high risk of other types of risk of bias in these studies. Similarly to acquired stroke, this category of bias was classified as probably low taking into account similar confounders.

4.4. Synthesis of results

4.4.1. Outcome: Has stroke (stroke prevalence)

No eligible study was found on the effect of long working hours on stroke prevalence.

4.4.2. Outcome: acquired stroke (stroke incidence)

4.4.2.1. Comparison: worked 41–48 h/week, compared with worked 35–40 h/week. A total of 20 studies (18 cohort studies and two case-control studies) with a total of 281,139 participants reported data on this comparison for this outcome. We analysed evidence from cohort studies separately from case-control studies. Because we judged cohort studies to carry a lower risk of bias than case-control studies (*Section 4.3.1*), our main meta-analysis for this comparison is of the cohort studies.

Eighteen cohort studies with a total of 277,202 participants from three WHO regions reported estimates of the effect of exposure to long working hours on the risk of acquiring stroke when working 41–48 h/week, compared with 35–40 h/week. These studies were somewhat heterogeneous in that six studies defined the outcome as a non-fatal stroke event (Kivimaki 2015 -ACL, Kivimaki 2015 -Alameda, Kivimaki 2015 –MIDUS, Kivimaki 2015 -NHANES I, Kivimaki 2015 -WLSG, Kivimaki 2015 -WLSS), whereas 12 of the studies defined the outcome as a non-fatal or fatal (or "mixed") stroke event (Hannerz et al., 2018, Hayashi et al., 2019, Kivimaki 2015 -COPSOQ-I, Kivimaki 2015 -COPSOQ-II, Kivimaki 2015 –DWECS, Kivimaki 2015 -FPS, Kivimaki 2015 -HeSSup, Kivimaki 2015 -IPAW, Kivimaki 2015 -PUMA, Kivimaki 2015 –Whiteall II Kivimaki 2015 - WOLF-N, Kivimaki 2015 -WOLF-S). Because fatal and non-fatal stroke events share an identical pathophysiological basis we considered studies with "pure" non-fatal events and studies with both fatal and non-fatal ("mixed") events to be sufficiently homogeneous clinically be included in the same meta-analysis (Kivimaki et al., 2015a; Li et al., 2020). Subgrouping "pure" non-fatal and "mixed" event studies demonstrated no evidence for subgroup differences (Appendix 4), suggesting that these studies are sufficiently homogeneous statistically to be combined. There did not appear to be an increased risk of acquiring stroke in people working 41–48 h/week compared with people working 35–40 h/week (relative risk (RR) 1.04, 95% CI 0.94–1.14, 18 studies, 277,202 participants, I^2 0%; Fig. 5).

Two case-control studies included a total of 3,937 participants. The combined analysis had a large confidence interval indicating a non-significant protective effect of acquiring stroke (OR 0.68, 95% CI 0.13–3.43, 2 studies, 3,937 participants, I^2 89%, Fig. 6).

4.4.2.2. Comparison: worked 49–54 h/week, compared with worked 35–40 h/week. A total of 19 studies (17 cohort studies and two case-control studies) with a total of 279,118 participants reported data on this comparison for this outcome. We again meta-analysed evidence from cohort studies separately from that from case-control studies and prioritized evidence from cohort studies over that from case-control studies, as outlined above (*Section 4.4.2.1*). Our main meta-analysis for this comparison for this outcome again is also that of the eligible cohort studies.

Seventeen cohort studies with a total of 275,181 participants from three WHO regions reported estimates of the effect of exposure to long working hours on the risk of acquiring stroke when working 41–48 h/week, compared with 35–40 h/week. These studies were somewhat heterogeneous in that six studies defined the outcome as a non-fatal stroke event (Kivimaki 2015 -ACL, Kivimaki 2015 -Alameda, Kivimaki 2015 –MIDUS, Kivimaki 2015 -NHANES I, Kivimaki 2015 -WLSG, Kivimaki 2015 -WLSS), whereas 11 of the studies defined the outcome as a non-fatal or fatal (or "mixed") stroke event (Hannerz et al., 2018, Hayashi et al., 2019, Kivimaki 2015 -COPSOQ-I, Kivimaki 2015 -COPSOQ-II, Kivimaki 2015 –DWECS, Kivimaki 2015 -FPS, Kivimaki 2015 -HeSSup, Kivimaki 2015 -PUMA, Kivimaki 2015 –Whiteall II Kivimaki 2015 - WOLF-N, Kivimaki 2015 -WOLF-S). As with the previous comparison for the same outcome, and as done previously (Li et al., 2020), we again judged these studies to be sufficiently homogeneous clinically to potentially be combined, again found no evidence for subgroup differences between studies defined by these outcome definitions (Appendix 4), and therefore again decided to combine these studies in one meta-analysis, as has also been done previously (Kivimaki et al., 2015a; Virtanen and Kivimaki, 2018). There appear to be a small increased risk of acquiring stroke in people working 49–54 h/week compared with people working 35–40 h/week (RR 1.13, 95% CI 1.001–1.275, 17 studies, 275,181participants, I^2 0%, p 0.048, Fig. 7).

The two case-control studies with eligible evidence including a total

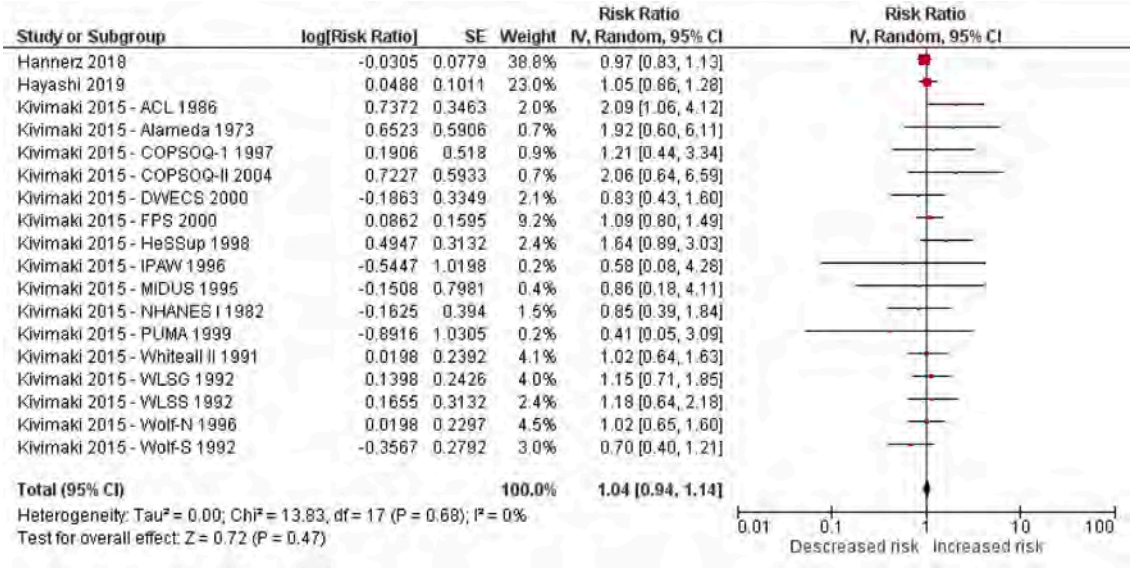

Fig. 5. Main meta-analysis of prioritized evidence (cohort studies), Outcome: acquired stroke. Comparison: worked 41–48 h/week compared with worked 35–40 h/week. Footnote: The approximate comparator for Hayashi 2019 was 35–45 h/day (vs. 45–55 h/week).

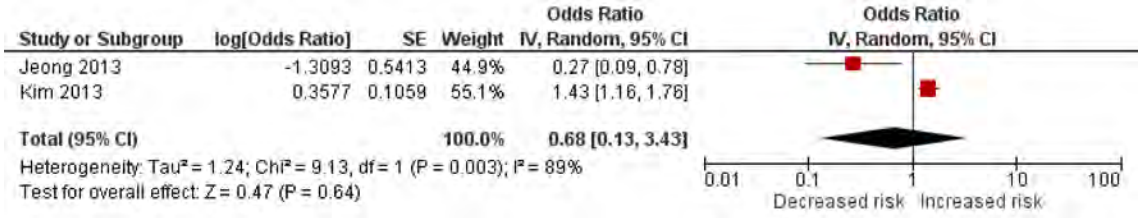

Fig. 6. Supporting meta-analysis of deprioritized evidence (case-control studies), Outcome: acquired stroke. Comparison: worked 41–48 h/week compared with worked 35–40 h/week or a similar comparison (40–50 h/week for Jeong 2013 and 45–55 h/week for Kim 2013). Footnote: 40–50 h/week for Jeong 2013 and 45–55 h/week for Kim 2013 vs. 35–45 h/week (approximated).

of 3,937 participants, with two approximated categories (50–54 and 45–60 h/week) and found opposite results, with a combined OR showing reduced risk of acquiring stroke, with upper CI above 1 (OR 0.91, 95% CI 0.35–2.37, 2 studies, 3,937 participants, I^2 89% Fig. 8).

4.4.2.3. Comparison: worked ≥55 h/week, compared with worked 35–40 h/week. A total of 18 studies (16 cohort studies and 2 case-control studies) with a total of 416,279 participants reported data on this comparison for this outcome. We again meta-analysed evidence from cohort studies separately from that from case-control studies, and prioritized evidence from cohort studies over that from case-control studies as described above.

Sixteen cohort studies with a total of 412,742 participants from three WHO regions reported estimates of the effect of exposure to long working hours on the risk of acquiring stroke when working ≥55 h/week, compared with working 35–40 h/week. The included studies were again somewhat heterogeneous in outcome definition, with seven studies defining the outcome as a non-fatal stroke event (Fadel 2019, Kivimaki 2015 -ACL, Kivimaki 2015 -Alameda, Kivimaki 2015 –MIDUS, Kivimaki 2015 -NHANES I, Kivimaki 2015 -WLSG, Kivimaki 2015 -WLSS) and 9 studies defining the outcome as a non-fatal or fatal (or "mixed") stroke event (Hannerz et al., 2018, Hayashi et al., 2019, Kivimaki 2015 -COPSOQ-I, Kivimaki 2015 -COPSOQ-II, Kivimaki 2015 –DWECS, Kivimaki 2015 -FPS, Kivimaki 2015 -HeSSup, Kivimaki 2015 –Whiteall II, Kivimaki 2015 -WOLF-S). Contrary to the previous section, though these studies might have been sufficiently homogeneous clinically to potentially be combined, there was a significant group difference with large heterogeneity I^2 = 74% p = 0.05) precluding consideration as a mixed group. Only the seven studies with the outcome as a non-fatal stroke event were included here (Appendix 4). There appeared to be an increased risk of acquiring stroke in people working ≥55 h/week compared with people working 35–40 h/week (RR 1.35, 95% CI 1.13 to 1.61, 7 studies, 162.644 participants, I^2 3%, Fig. 9).

The two case-control studies with eligible evidence included a total of 3,937 participants, and indicated a significant effect of acquiring stroke (OR 1.96, 95% CI 1.50 to 2.55, 2 studies, 3,937 participants, Fig. 10).

4.4.3. Outcome: died from stroke

4.4.3.1. Comparison: worked 41–48 h/week, compared with worked 35–40 h/week. A total of 12 cohort studies with a total of 265,937 participants from one WHO region reported estimates of the effect of exposure to long working hours on the risk of dying from stroke when working 41–48 h/week, compared with 35–40 h/week. All these studies defined the outcome as a non-fatal or fatal (or "mixed") stroke event. All of these studies could be included in a quantitative meta-analysis. There did not appear to be an increased risk of dying from stroke in people working 41–48 h/week compared with people working 35–40 h/week (RR (1.01, 95% CI 0.91 to 1.12, 12 studies, 265,937 participants, I^2 0%, Fig. 11).

4.4.3.2. Comparison: worked 49–54 h/week, compared with worked 35–40 h/week. A total of 11 cohort studies with a total of 256,129 participants from one WHO region reported estimates of the effect of exposure to long working hours on the risk of dying from stroke when working 49–54 h/week, compared with 35–40 h/week. Again, all these included studies defined the outcome as a non-fatal or fatal (or "mixed") stroke event, and we again judged these studies to be sufficiently homogeneous clinically to be combined in a meta-analysis. There did not appear to be an increased risk of dying from

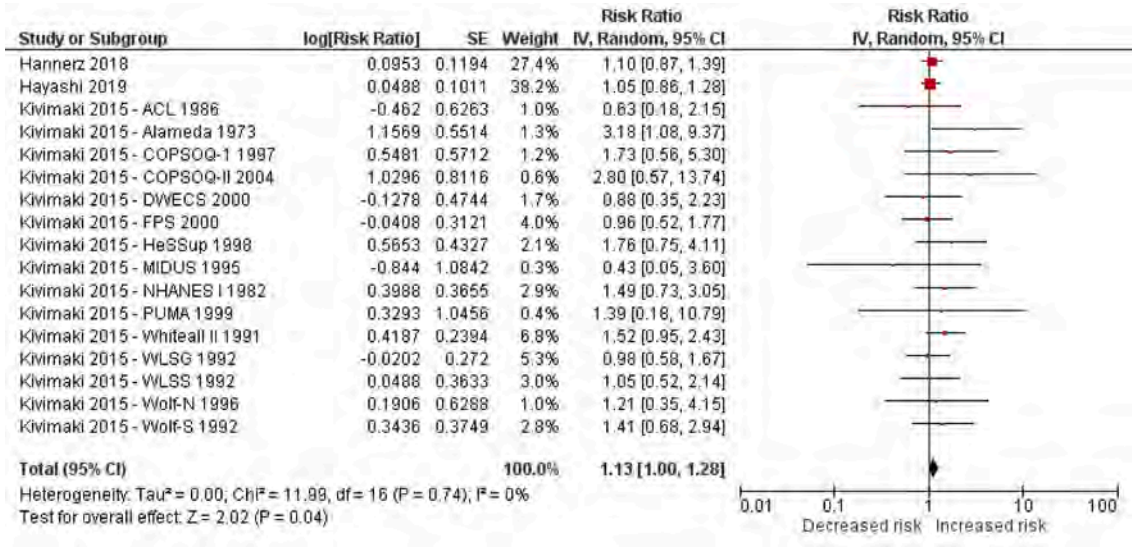

Fig. 7. Main meta-analysis of prioritized evidence (cohort studies), Outcome: Acquired stroke. Comparison; Worked 49–54 h/week compared with worked 35–40 h/week. Footnote: The approximate comparator for Hayashi 2019 was 35–45 h/day (vs. 45–55 h/week).

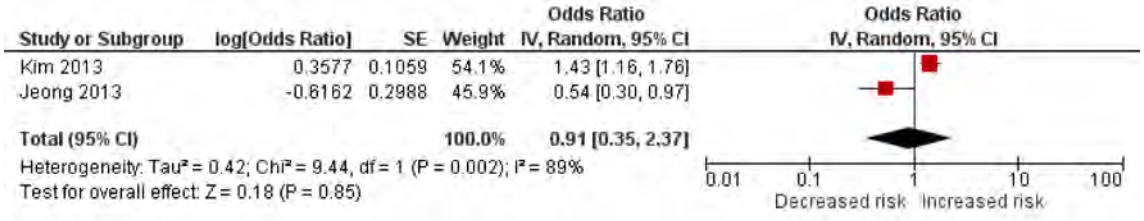

Fig. 8. Supporting meta-analysis of deprioritized evidence (case-control studies), Outcome: Acquired stroke. Comparison: Worked 49–54 h/week compared with worked 35–40 h/week (or a similar comparison). Footnote: 50–55 h/week for Jeong 2013, and 45–55 h/week for Kim 2013 vs. 35–45 h/week (approximated).

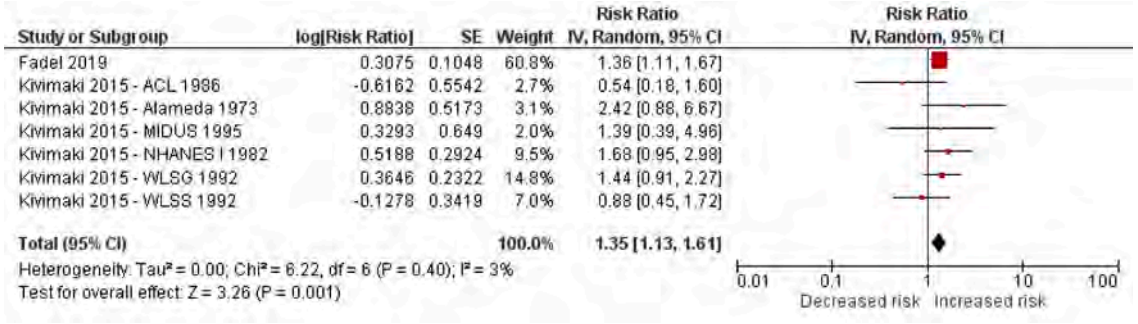

Fig. 9. Main meta-analysis of prioritized evidence (cohort studies), Outcome: Acquired stroke (incidence only). Comparison: worked ≥55 h/week compared with worked 35–40 h/week. Footnote: The approximated comparator for Fadel 2019 was less than 10 h/day for more than 50 days a year (versus more than 10 h/day for more than 50 days a year).

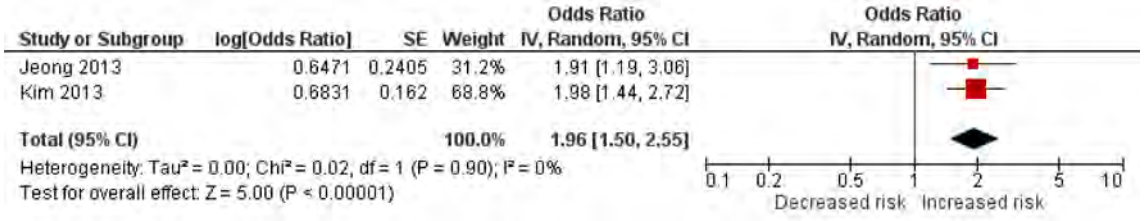

Fig. 10. Supporting meta-analysis of deprioritized evidence (case-control studies), Outcome: Acquired stroke. Comparison; Worked ≥55 h/week compared with worked 35–40 h/week (or a similar comparison). Footnote: ≥55 h/week for Kim 2013 vs. 35 to 45 h/week (approximated).

stroke in people working 48–54 h/week compared with people working 35–40 h/week (RR 1.13, 95% CI 0.99 to 1.29, 11 studies, 256,129 participants, I^2 0%; Fig. 12).

4.4.3.3. Comparison; worked ≥55 h/week, compared with worked 35–40 h/week. A total of 10 cohort studies with a total of 664,647 participants from two WHO regions reported estimates of the effect of exposure to long working hours on the risk of dying from stroke when working ≥55 h/week, compared with 35–40 h/week. These studies that we pooled in our meta-analysis were somewhat heterogeneous in that one study defined the outcome as a fatal stroke event (Kivimaki 2015 –O Reilly 2013), whereas nine of the studies defined the outcome as a non-fatal or fatal (or "mixed") stroke event (Hannerz et al., 2018, Hayashi et al., 2019, Kivimaki 2015 -COPSOQ-I, Kivimaki 2015

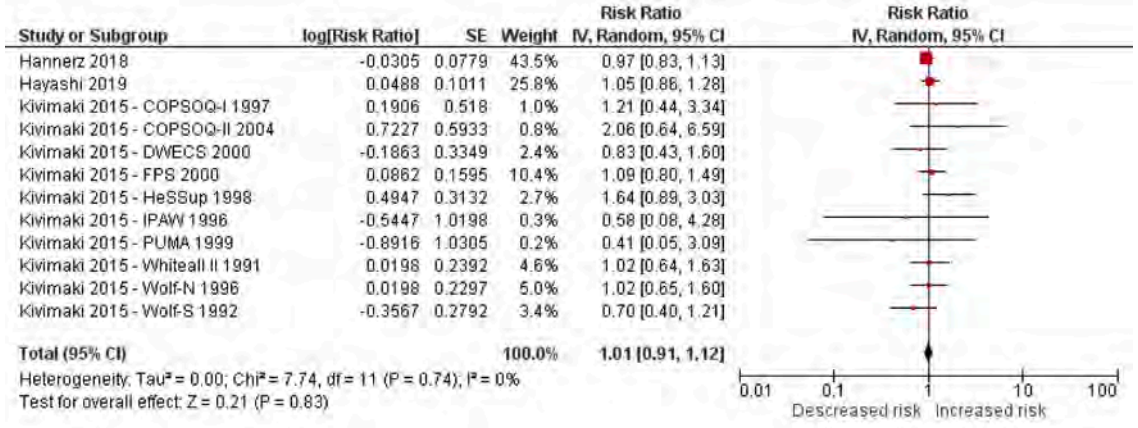

Fig. 11. Main meta-analysis (cohort studies). Outcome: Died from stroke. Comparison: Worked 41–48 h/week compared with worked 35–40 h/week. Footnote: The approximate comparator for Hayashi 2019 was 35–45 h/day (vs. 45–55 h/week).

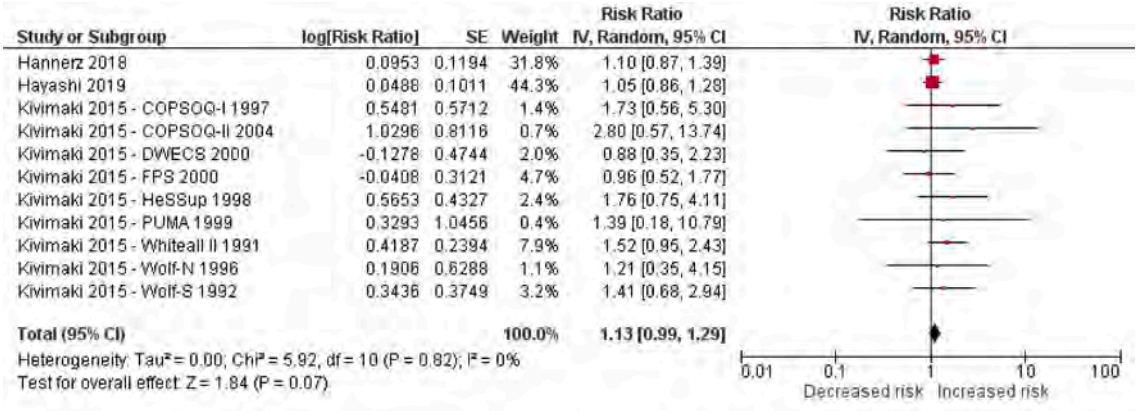

Fig. 12. Main meta-analysis (cohort studies). Outcome: Died from stroke. Comparison: Worked 49–54 h/week compared with worked 35–40 h/week. Footnote: The approximate comparator for Hayashi 2019 was 35–45 h/day (vs. 45–55 h/week).

-COPSOQ-II, Kivimaki 2015 –DWECS, Kivimaki 2015 -FPS, Kivimaki 2015 -HeSSup, Kivimaki 2015 –Whiteall II, Kivimaki 2015 -WOLF-S). Applying the same criteria as in case of acquired stroke (*Section 4.4.2.3*), the heterogeneity of included studies was judged to be low. All these studies could consequently be included in a quantitative meta-analysis. There did not appear to be an increased risk of dying from stroke in people working ≥55 h/week compared with people working 35–40 h/week (RR 1.08, 95% CI 0.89–1.31, 10 studies, 664,647 participants, I^2 = 20%, Fig. 13).

4.5. Additional analyses

4.5.1. Subgroup analyses

Subgroup analyses were performed for data from the main meta-analysis (cohort studies) with comparison between the group worked ≥55 h/week and the group worked 35–40 h/week. These analyses include subgrouping by WHO region, sex, age, industry, occupation, SES, and type of stroke (Table 6). Because no cohort study reported effect estimates by subtype of stroke, subgroup analyses by type of stroke were exceptionally conducted using data from the two included case-control studies (deprioritized evidence). These subgroup analyses found no evidence for meaningful subgroup differences by WHO region, sex, age, SES, and type of stroke. The forest plots are presented in Appendix 7 in the Supplementary data.

4.5.2. Sensitivity analyses

Sensitivity analyses were also performed for data from the main meta-analysis (cohort studies) with comparison between the group that worked ≥55 h/week and the group that worked 35–40 h/week. There were no meaningful differences by risk of bias, effect estimate measurement and definition of the comparator (Table 7; Appendix 6).

4.6. Quality of evidence

4.6.1. Outcome: acquired stroke (stroke incidence)

4.6.1.1. Comparison: worked 41–48 h/week, compared with worked 35–40 h/week. We did not have any serious concerns regarding risk of bias in the body of evidence on this comparison for this outcome, because we judged the risk of bias to be probably low. We also had no serious concerns regarding inconsistency, specifically regarding the cohort studies that were judged to be of higher quality. We did not have serious concerns for indirectness, regarding the combination of the outcome definition including "mixed" (fatal or non-fatal) events and non-fatal events. Our exploratory subgroup analyses did not indicate any difference between mixed events and non-fatal events (Appendix 6 in the Supplementary data). We therefore did not downgrade the quality of evidence for risk of bias, inconsistency, or indirectness (±0 levels). We had serious concerns for imprecision, given large CIs in the pooled effect estimates, and we therefore downgraded by one level (−1). We did not have any serious concerns for publication bias. We upgraded neither for a large effect estimate, nor for evidence for a dose–response. In conclusion, we started at "moderate" for observational studies and downgraded by one level (−1) for imprecision to a final rating of "low".

4.6.1.2. Comparison: worked 49–54 h/week, compared with worked 35–40 h/week. As above (4.6.1.1) we had no serious concerns regarding risk of bias, inconsistency, indirectness, or publication bias.

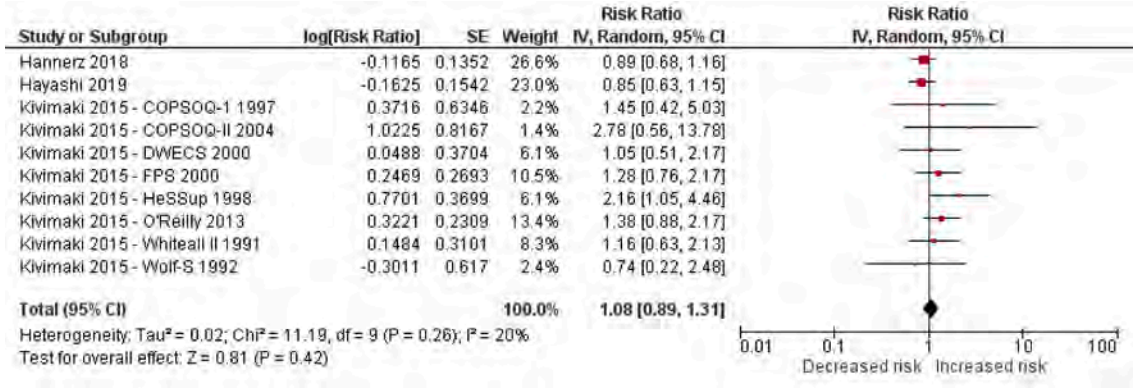

Fig. 13. Main meta-analysis (cohort studies). Outcome: Died from stroke (including mixed studies). Comparison; Worked ≥55 h/week compared with worked 35–40 h/week. Footnote: The approximate comparator for Hayashi 2019 was 35–45 h/day (vs. 45–55 h/week).

Table 6
Summary of results from subgroup analyses on long working hours and stroke, cohort studies (and case-control for type of stoke).

Acquired stroke (stroke incidence)		Died from stroke (stroke mortality)	
WHO region (7 studies)	p = 0.86	**WHO region** (10 studies)	p = 0.11
Americas	1.31 (0.94–1.83)	Europe	1.15 (0.93–1.41)
Europe	1.36 (1.11–1.67)	Western Pacific	0.85 (0.63–1.15)
Sex (2 studies)	p = 0.23		
Men	1.13 (0.79–1.62)		
Women	1.91 (1.04–1.80)		
Age (2 studies)	p = 0.57		
40–45 years	1.02 (0.59–1.76)		
45–50 years	1.51 (0.64–3.58)		
50–55 years	1.60 (1.10–2.33)		
55–60 years	1.68 (1.22–2.31)		
60–65 years	2.57 (0.57–11.56)		
SES (1 study)	p = 0.22	**SES**	p = 0.28
High SES	1.14 (0.18–1.58)	High SES	1.24 (0.67–2.30)
Intermediate SES	1.71 (1.19–2.49)	Intermediate SES	2.13 (1.31–3.48)
Low SES	1.68 (0.95–2.97)	Low SES	1.27 (0.70–2.30)
Industry (1 study)	p = 0.28		
Private sector	1.47 (1.10–1.96)		
Public sector	1.15 (0.82–1.62)		
Occupation (1 study)	p = 0.11		
Managers	1.00 (0.57–1.76)		
Professionals	1.29 (0.86–1.94)		
Technicians /Associate Professionals	1.27 (0.74–2.20)		
Clerical Support Workers	1.63 (0.66–4.02)		
Services and Sales Workers	1.11 (0.54–2.26)		
Craft and Related Trades Workers	3.07 (1.36–6.94)		
Plant and Machine Operators, and Assemblers	2.95 (1.26–6.87)		
Elementary Occupations	0.39 (0.09–1.64)		
Type of stroke (2 case-control studies)	**p = 0.65**		
Haemorrhagic stroke (odds ratio)	1.88 (1.42–2.49)		
Ischaemic stroke (odds ratio)	2.21 (1.16–4.21)		

Table 7
Summary of results from sensitivity analyses on long working hours and stroke, cohort studies.

Acquired stroke (incidence)		Died from stroke (mortality)	
Risk of bias (7studies)	p = 0.40	**Risk of bias** (10 studies)	p = 0.30
Any “high”/”probably high”	1.35 (0.60–1.97)	Any “high”/”probably high”	1.08 (0.89–1.31)
Only “low”/”probably low”	1.40 (1.17–1.67)	Only “low”/”probably low”	0.99 (0.80–1.70)
Effect estimate measurement (1 study)	p = 0.87		
Relative Risk	1.36 (1.10–1.67)		
Hazard Rate Ratio	1.27 (1.11–1.47)		
Odds Ratio	1.30 (1.13–1.48)		
Comparator (7 studies)	p = 0.86	**Comparator** (10 studies)	p = 0.11
All homogeneous comparator	1.35 (1.13–1.61)	All homogeneous comparator	1.08 (0.89–1.31)
Strict comparator	1.31 (0.94–1.83)	Strict comparator	1.15 (0.93–1.41)

We had serious concerns for imprecision, given large CI in the pooled effect estimate, and we therefore downgraded by one level (−1). We did not upgrade for a large effect estimate, but did for evidence of dose–response. All positive studies except one case-control study (Jeong et al., 2013) found a dose–response relationship; this caused us to upgrade the evidence by one level (+1). The Kivimaki 2015 systematic review (Kivimaki et al., 2015a) also showed that increasing the number of hours worked per week increased the risk. Fadel 2019 (Fadel et al., 2019) studied years of exposure and showed a significant gradient with a threshold at 10 years.

In summary, we neither downgraded nor upgraded our original classification and therefore concluded with the final rating of “moderate”.

4.6.1.3. Comparison: worked ≥55 h/week, compared with worked 35–40 h/week. As above, we had no serious concerns regarding risk of bias, inconsistency, indirectness, or publication bias (Fig. 14). We downgraded by one level for imprecision and upgraded for evidence of dose–response relationship and therefore concluded with a final rating of “moderate”.

4.6.2. Outcome: died from stroke

We had the same ratings of quality of evidence for all comparisons of fatal stroke: worked 41–48 h/week, 49–54 h/week, or ≥55 h/week compared with worked 35–40 h/week

For all comparisons, we had no serious concerns regarding risk of bias (only the first domain “selection bias” might raise a possible risk of bias). We had no serious concerns regarding inconsistency, indirectness, or publication bias (Fig. 15). We had serious concerns for imprecision, given the large CI in the pooled effect estimate, and we therefore downgraded by one level (−1). We did not upgrade for effect size or dose–response effect. In summary, for all these comparisons we started at “moderate” for observational studies and downgraded by one level (−1) for imprecision to a final rating of “low”.

4.7. Assessment of strength of evidence

According to our protocol we rated the strength of evidence based

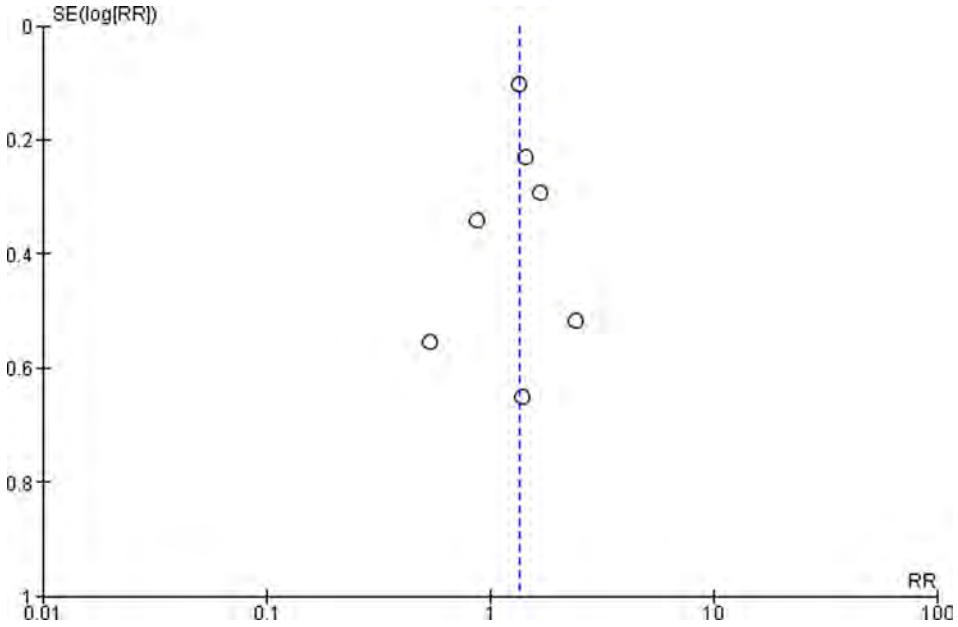

Fig. 14. Funnel plot of prioritized evidence (cohort studies). Outcome: Acquired stroke, Comparison: Worked ≥55 h/week or more, compared with worked 35–40 h/ week.

on a combination of four criteria outlined in the Navigation guide: (i) Quality of the entire body of evidence; (ii) Direction of the effect estimate; (iii) Confidence in the effect estimate; (iv) Other compelling attributes.

4.7.1. Quality of the entire body of evidence

Our meta-analyses are based on 20 cohort studies in total, include a very large number of participants in different WHO regions, take into account relevant confounders, and provide a body of evidence sufficient to assess the harmfulness of the exposure. The analyses document a moderately increased risk of incident stroke when working ≥55 h/ week compared with working 35–40 h/week, with the lower CI beyond 1.0 and a rather narrow overall CI, with dose–response relationship. The quality of studies is adequate, given similar study protocols, consistent measurement of exposure and outcome, and clear temporal distinction between exposure and outcome, including control of reverse causation by excluding studies with outcomes proximal to exposure assessment. Overall, risk of bias of these cohort studies is probably low. We did not consider the evidence of case-control studies in our assessment of quality and strength of evidence, given the lower confidence we have in this study design (though case-control studies give similar results supporting a threshold effect of working ≥55 h/week).

4.7.2. Direction of the effect estimate

The study results are sufficient to assess the direction of the effect estimate. For all three exposure categories (41–48 h/week; 49–54 h/ week; ≥55 h/week) no single study documented a negative effect estimate (with the higher CI below 1). In the first exposure category, all studies except one displayed effect estimates around 1.0, the result is significant at 49–54 h/ week (limit) and in the third exposure category five studies demonstrated positive effect estimates, with lower CIs beyond, or close to 1. Overall, heterogeneity was low.

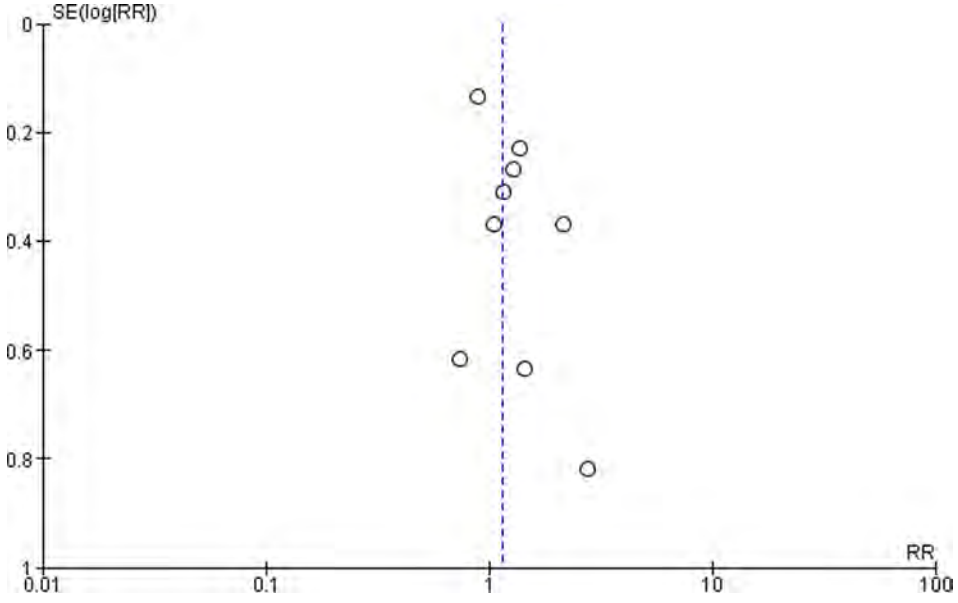

Fig. 15. Funnel plot of (cohort studies), Outcome: Died from stroke, Comparison: Worked ≥55 h/week, compared with worked 35–40 h/week.

Table 8
Summary of findings.

Effect of exposure to long working hours on stroke among workers
Population: all ≥15 years workers
Settings: all countries and work settings
Exposure: worked 41–48, 49–54 or ≥55 h/week (or equivalent)
Comparator: worked 35–40 h/ week

Outcomes	Exposure category	Illustrative comparative risks (95% CI)		Relative effect (95% CI)	No. of participants (studies)	Quality of the evidence	Strength of Evidence for Human Evidence	Comments
		Assumed risk Unexposed workers (worked 35–40 h/ week)	Corresponding risk Workers in the exposure category					
Has stroke	**–**	**–**	**–**	**–**	–	–		No evidence was found on this outcome.
Acquired Stroke (measured with administrative record or self-report) Follow-up: 8–20 years	**Worked 41–48 h/week**	**150 cases per 100,000 person years** [a]	**165 per 100,000 person years** (141–192)	**RR 1.04** (0.94–1.14)	277,202 (18 studies)	⊖ Low [b]	Inadequate evidence of harmfulness	Better indicated by lower values Additional evidence from one additional cohort study and a case-control study also provided no evidence for an effect for this comparison on this outcome. We are very uncertain about the effect of this exposure category on this outcome.
	Worked 49–54 h/week		**191 per 100,000 person years** (155–235)	**RR 1.13** (1.00–1.28), P = 0.04	275,139 (17 studies)	⊕⊖ Moderate [c]	Limited evidence of harmfulness	Better indicated by lower values Additional evidence from one additional cohort study and a case-control study also provided a small but possible evidence for an effect for this comparison on this outcome. We are very uncertain about the effect of this exposure category on this outcome.
	Worked ≥ 55 h/week		**203 per 100,000 person years** (179 to 242)	**RR 1.35** (1.13–1.61)	162,644 (7 studies)	⊕⊖ Moderate [c]	Sufficient evidence of toxicity/ harmfulness	Better indicated by lower values Additional evidence from two case-control studies also suggests a small increase in the risk for the outcome for this comparison. Compared with working 35–40 h/ week, working ≥ 55 h/week may have led to an increase in having acquired stroke.
Died from stroke (mortality) (measured with administrative record) Follow-up: 8–20 years	**41–48 h/w**	**150 cases per 100,000 person years** [a]	**152 per 100,000 person years** (137–168)	**RR 1.01** (0.91–1.12)	265,937 (12 studies)	⊖ Low [b]	Inadequate evidence of toxicity/ harmfulness	Better indicated by lower values We are very uncertain about the effect of this exposure category on this outcome
	49–54 h/w		**170 per 100,000 person years** (149–194)	**RR 1.13** (0.99–1.29)	256,129 (11 studies)	⊖ Low [b]	Inadequate evidence of toxicity/ harmfulness	Better indicated by lower values. We are very uncertain about the effect of this exposure category on this outcome
	≥55 h/w		**167 per 10,000 person years** (157–205)	**RR 1.08** (0.89–1.31)	726,803 (10 studies)	⊖ Low [b]	Inadequate evidence of toxicity/ harmfulness	Better indicated by lower values Compared with working 35–40 h/week, working ≥ 55 h/week may have led to an increase in dying due to stroke but

CI: confidence interval; **RR**: relative risk.
Navigation Guide quality of evidence ratings
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.

[a] We extracted the risk of any stroke events among workers working 35–40 h/week from Hannerz 2018 as the assumed risk. (Note that this study provided one baseline risk for both non-fatal and/or fatal stroke, so that it was not possible to differentiate assumed risk for exclusively non-fatal events and fatal events separately.)
[b] Downgraded by one grade, because of serious imprecision (i.e., large CIs in the pooled effect estimate).
[c] Downgraded by one grade, because of serious imprecision (i.e., large CIs in the pooled effect estimate), and upgraded for a dose-response relationship.

4.7.3. Confidence in the effect estimate

There is limited evidence to determine the level of confidence in the effect estimate for several reasons. First, while studies included analysis of several relevant confounders that in part can also act as mediators, no additional data are available on causal pathways linking exposure to the health outcome under study. Indirect supportive evidence comes from studies dealing with health-adverse working conditions other than long working hours, with conditions that implicate identical pathways from exposure to outcome, such as adverse health behaviours or chronic psychosocial stress with pathophysiological effects on stroke. Second, the assumption of a dose–response relationship between the three exposure categories and the outcome was slightly supported by our findings. There was no indication of an effect estimate at the lowest exposure category and perhaps a slightly larger effect at the next lowest exposure category. An effect estimate with the lower CI above 1 was found at the third exposure category, ≥55 h/week, indicating a possible dose response relationship. Third, the magnitude of the effect estimate was modest, given an overall pooled RR lower than two. Although even a modest increase in risk in a serious health event can be relevant for policy when the exposure is highly prevalent (which is certainly the case with long working hours), this low magnitude of the effect estimate does not increase our confidence in a causal association. Fourth, no intervention studies are available that demonstrate a reduction of the effect estimate following exposure reduction.

4.7.4. Other compelling attributes

We were not able to access data that could offer evidence for a discussion of other compelling attributes in assessing the strength of evidence. In summary, we conclude that there is limited evidence of the toxicity of long working hours, defined as ≥55 h/week, for elevated risk of fatal or non-fatal stroke.

Additional assessment of strength of evidence based on the Bradford Hill criteria is included in Appendix 9 in the Supplementary data.

4.7.5. Rating by outcome and comparison

Based on the considerations presented above, we judged the existing bodies of evidence as:

- *Inadequate* evidence for harmfulness for all exposure categories for stroke prevalence and mortality, and for the exposure category 40–48 h/week for stroke incidence.
- *Limited* evidence for harmfulness for the exposure category 49–54 h/week for stroke incidence.
- *Sufficient* evidence for harmfulness for the exposure category ≥55 h/week for stroke incidence.

5. Discussion

5.1. Summary of evidence

As shown in the table of summary of findings (Table 8), our systematic review found no eligible study on the outcome of stroke prevalence. It found low quality of evidence of weak or no associations between the exposure categories of working 41–48 h/week and working 49–54 h/week and the outcomes of stroke incidence and mortality, when compared to 35–40 h/week. Based on the other considerations for evaluating the strength of evidence, we concluded that there was inadequate evidence of harmfulness based on human evidence, including the effects of working ≥55 h/week on risk of fatal stroke. We found moderate quality evidence of clinically meaningful associations of working ≥55 h/week with elevated risk of incident stroke and concluded there is sufficient evidence of toxicity from the human evidence. Particularly, findings based on cohort studies documented modest, but relatively robust effects of working ≥55 h/week on risk of stroke, given the large sample size, the standardized adjustment for confounding, and the probably low risk of bias. A risk elevation by 35 percent is considered modest, but in view of the high prevalence of long working hours and considerable incidence/mortality rates of stroke in working populations, this risk deserves attention in terms of preventive occupational health measures. Overall, the heterogeneity of findings is low, and sensitivity analyses confirm the robustness of results.

5.2. Comparison to previous systematic review evidence

The results of our systematic review are similar to the results of the only comprehensive prior systematic review and its update (Kivimaki et al., 2015a; Virtanen and Kivimaki 2018). Both these reviews reported a potential increased risk of incident fatal and/or non-fatal stroke from exposure to working long hours, with exposure to ≥55 h/week associated an increased RR of 1.33 (95% CI 1.11–1.61) in the 2015 review of 15 studies, and an RR of 1.21 (95% CI 1.01–1.45) in the 2018 update of 16 studies (Virtanen and Kivimaki 2018). Our systematic review and meta-analysis added evidence from two primary studies published in 2019 (Fadel et al., 2019; Hayashi et al., 2019). Our systematic review found that working ≥55 h/week may have led to a moderately, clinically meaningful increase in the risk of incident stroke (defined as either "pure" non-fatal or mixed incident stroke events, but excluding "pure" fatal events), when followed up between one year and 20 years (RR 1.35, 95% CI 1.13–1.61, 7 studies, 162,644 participants, I^2 3%, moderate quality of evidence). Applying the Navigation Guide (Woodruff and Sutton 2014) systematic review methodological framework, and taking all systematic review steps from 1 (protocol development and publication) through 7 (strength of evidence assessment), our systematic review concluded that the reviewed body of research provides sufficient evidence for harmfulness of the highest exposure category.

5.3. Limitations and strengths of this systematic review

Our systematic review has several limitations. Although we conducted a broad and sensitive search, that included many academic and grey literature databases and consultation with additional experts, we may have missed eligible studies. Considering the large number of included studies, the number of participants, and the number of disease events, it seems unlikely that the overall results would have been affected.

Second, we did not receive a substantial amount of the missing data we requested from the authors of the studies included in this systematic review. (Appendix 1) Some uncertainties in the body of evidence could have been resolved by these missing data. Future systematic reviews and meta-analyses might benefit from being granted access to more missing data from included studies.

There were some minor differences in risk of bias assessments in our study compared to the previous Kivimaki et al., 2015a review. In our review, we also considered proportion of population inclusion and sample selection. Their study rated self-reports less favourably, while we assigned a rating of "probably low" risk of bias. Supporting our assessment of low risk, Kivimaki et al. (2015a) found no difference in their sensitivity analyses between self-reported and administrative records, which was similar to Wong et al. (2019), and the fair accuracy reported by other studies (Woodfield et al., 2015a, 2015b).

Reverse causality (or health selection) may perhaps be an alternative explanation for the elevated risk to stroke incidence that we found in our systematic review and meta-analysis. However, the Kivimaki et al. (Kivimaki et al., 2015a) systematic review and meta-analysis performed a sensitivity analysis that compared studies that excluded participants with the outcome within three years of the beginning of the study (lower risk of reverse causation) with studies that included these participants (higher risk of reverse causation). This sensitivity analysis found no evidence of any meaningful differences. Most studies included in the systematic review excluded participants

with the outcome at or just after baseline. The Fadel study (Fadel et al., 2019), for example, excluded participants with the outcome in the first 5 years. It also seems unlikely that workers who suffer a stroke would work longer hours after the event than before. The probability that reverse causation could explain the elevated risks found in this systematic review is low.

Adjustments for various confounders, mediators and/or effect modifiers may also have limited this systematic review. The question of variable adjustment was considered, especially the indirect pathway described in Fig. 1. In the four studies that allowed adjustments on other cardiovascular risk factors, unadjusted risk estimates were similar (though lower), without over-adjustment effects and reassured us on the lack of residual confounding. Other confounders, mediators, and potential pathways through personality, working condition (night/shift, psychosocial factors), age difference, and income are possible, as are different effects of these conditions on different type of stroke (ischemic versus haemorrhagic). Much more work is needed to determine potential mechanisms of the observed associations.

5.3.1. Strengths

Our systematic review and meta-analysis have a number of strengths, including:

- Our systematic review and meta-analysis has followed the recommended steps (Woodruff and Sutton, 2014, Fig. 1), of having a pre-published protocol and assessments for strength of evidence, and is a model representation of systematic review methods on the topic.
- Previous systematic reviews have not sought to differentiate stroke prevalence from stroke incidence (non-fatal events) and stroke mortality (fatal events), but our systematic review improves accuracy by differentiating these three different outcomes.
- Previous systematic reviews have not comprehensively provided detailed analyses across all analytic steps of the systematic review and meta-analysis for comparisons of standard categories of exposure to long working hours compared with standard working hours, but we have provided such analyses for three such comparisons commonly used in the epidemiological literature across all steps of the systematic review, improving systematic review evidence on this topic.
- Whereas previous systematic reviews have not comprehensively assessed risk of bias and quality of evidence using established systematic review frameworks with dedicated tools and approaches, we have rigorously applied the Navigation Guide framework in this systematic review, ensuring transparency.
- In previous systematic reviews, strength of the evidence was not commonly assessed, but in our systematic review we have applied pre-specified criteria to rate the strength of evidence for each include comparison for each included outcome, and again this is a novel contribution to the systematic review and meta-analytic body of evidence on the topic.
- Finally, to our knowledge, this is the first systematic review and meta-analysis conducted specifically for a global occupational burden of disease study, and as such it provides a model for future systematic reviews that will help ensure that these global health estimates adhere fully with GATHER (Stevenset al., 2016).

6. Use of evidence for burden of disease estimation

This systematic review and meta-analysis was conducted by WHO and ILO, supported by a large network of individual experts for the development of the WHO/ILO Joint Estimates (Ryder, 2017). More specifically, it provides the crucial evidence base for these organizations to produce estimates of the burden of deaths and DALYs from stroke attributable to exposure to long working hours. The systematic review found a large body of evidence for the comparison of 49–54 working hours/week compared with 35–40 working hours/week for the outcomes of stroke incidence; this body of evidence was judged to be of low quality and to provide limited evidence for toxicity/harmfulness. The systematic review also found a large body of evidence from several prospective cohort studies for the comparison of ≥55 working hours/week compared with 35–40 working hours/week for the outcome of stroke incidence; this body of evidence were judged to be of moderate quality and to provide sufficient evidence for toxicity/harmfulness. Producing estimates of the burden for stroke attributable to exposure to the categories of working 49–54 and ≥55 working hours/week appears evidence-based and warranted, and the parameters reviewed (including the pooled RRs from the meta-analyses for these comparisons) appear suitable as input data for WHO/ILO modelling of work-related burden of disease and injury.

7. Conclusions

We judged the existing bodies of evidence as inadequate evidence for harmfulness for the exposure categories 41–48, 48–54 and ≥55 h/week for stroke prevalence and mortality, as well as for the exposure category 41–48 h/week for stroke incidence. Evidence on the exposure category 49–54 h/week for stroke incidence was judged to provide limited evidence of toxicity/harmfulness. Evidence on exposure to working ≥55 h/week was judged as sufficient evidence of harmfulness for stroke incidence. The RRs for the comparisons 49–54 h/week and ≥55 h/week, compared with 35–40 h/week, are suitable as input data for WHO/ILO modelling of work-related burden of disease and injury.

8. Differences between protocol and systematic review

- Our protocol did not specify how to deal with studies with outcomes definitions being “mixed” in terms of including both fatal and non-fatal events. We added such criteria for dealing with these studies with the outcome definition being “mixed”. Our approach was the same as that used in another systematic review (Li et al., 2020).
- The protocol did not include stroke prevalence as an eligible outcome, but we included this outcome as it may be relevant for estimating the burden of disease; we found that there was insufficient evidence to evaluate this outcome.
- We had planned to use Ryyan software for selecting studies, but used Covidence software instead, due to some specific preferred features of the software and consistent author access across reviews.
- In the protocol, we defined the eligible comparator in the unit of hours per week only. In the systemic review we included studies in which the comparator was defined using a different unit (e.g. hours per day), as long the units could be converted to hours worked per week in comparable categories.

9. Financial support

All authors are salaried staff members of their respective institutions. AD is also paid as the Editor-in-Chief of *Les Archives de Maladies Professionnelles et de l'Environnement*. The publication was prepared with financial support from the World Health Organization cooperative agreement with the Centres for Disease Control and Prevention National Institute for Occupational Safety and Health of the United States of America (Grant 1 E11 OH0010676-02; Grant 6 NE11 OH010461-02-01; and Grant 5 NE11 OH010461-03-00).

Sponsors

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

Author contributions

Had the idea for this systematic review: FPe, Ivan Ivanov (WHO), Nancy Leppink (ILO).

Coordinated the entire series of systematic reviews: FPe, Ivan Ivanov, Nancy Leppink.

Gathered the reviewers: FPe, YU.

Was the lead reviewer: AD, SI, GS.

Led the design of the systematic review including the standard methods: FPe.

Contributed substantially to the design of the systematic review: AD, SI, GS, YU, LG, DP, RR, JL, JS, FPi, MB, BE, YR, LLMH.

Conducted the search: DG, JP, GS.

Selected studies: FB, CDT, DG, MR, AM, LLMH, AT, AD, AO.

Extracted data: FB, DG, MR, AT, AD.

Requested missing data: AD, JL, JS, RR, LG, MS, SKK, BJK.

Reanalysed dataset: SKK, BJK, LG, MS, AD.

Assessed risk of bias: CD, AO, MB, JS, RR, LG, DP, GS.

Conducted the meta-analyses: AD, FPe, GS.

Assessed quality of evidence: CD, MB, AO, YR, FPe, GS, DG.

Assessed strength of evidence: AD, SI, GS, DG.

Developed the standards and wrote the template for all systematic reviews in the series: FPe.

Wrote the first draft of the manuscript using the template: AD, GS.

Revising the manuscript critically for important intellectual content: FPe, AD, GS, SI, DG, BE.

Ensured tailoring of the systematic review for WHO/ILO estimation purposes: FPe.

Ensured harmonization across systematic reviews in the series: FPe.

Technical editing: GS, BE, FPe.

Approved the final version of the systematic review to be published: All authors.

Agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: All authors.

Acknowledgments

We thank Paul Whaley (Associate Editor – Systematic Review, *Environment International*, and Lancaster University) and Professor Tim Driscoll (University of Sydney) for the editorial guidance and support.

We thank Lisa Bero (University of Sydney), Rebecca Morgan (McMaster University) and Tracey Woodruff (University of California, San Francisco) for their constructive critical comments on the methods of this systematic review and an earlier version of the manuscript.

The following individuals or organizations shared data from primary included studies with us: CONSTANCES Cohort Study team, Marcel Goldberg, Harald Hannerz, Mika Kivimäki, and Marie Zins.

The authors alone are responsible for the views expressed in this article, and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2020.105746.

References

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

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

Arditi, C., Burnand, B., Peytremann-Bridevaux, I., 2016. Adding non-randomised studies to a Cochrane review brings complementary information for healthcare stakeholders: an augmented systematic review and meta-analysis. BMC Health Serv.. Res 16, 598.

Aziz, F., Talhelm, L., Keefer, J., Krawiec, C., 2019. Vascular surgery residents spend one fifth of their time on electronic health records after duty hours. J. Vasc. Surg. 69, 1574–1579.

Balshem, H., Helfand, M., Schunemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E., Falck-Ytter, Y., Meerpohl, J., Norris, S., Guyatt, G.H., 2011. GRADE guidelines: 3. Rating the quality of evidence. J. Clin. Epidemiol. 64, 401–406.

Barroga, E.F., Kojima, T., 2013. Research study designs: an appraisal for peer reviewers and science editors. Eur. Sci. Ed. 44–45.

Bejot, Y., Daubail, B., Giroud, M., 2016. Epidemiology of stroke and transient ischemic attacks: Current knowledge and perspectives. Rev. Neurol. (Paris) 172, 59–68.

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

Chandola, T., Heraclides, A., Kumari, M., 2010. Psychophysiological biomarkers of workplace stressors. Neurosci. Biobehav. Rev. 35, 51–57.

Covidence systematic review software, V.H.I., Melbourne, Australia. Available at www.covidence.org.

Deeks, J., Higgins, J., Altman, D., 2011. Chapter 9: Analysing data and undertaking meta-analyses in: Higgins J., Green S., (Eds.). Cochrane Handbook for Systematic Reviews of Interventions Version 510 [updated March 2011] The Cochrane Collaboration, 2011 Available from wwwhandbookcochraneorg.

Descatha, A., Sembajwe, G., Baer, M., Boccuni, F., Di Tecco, C., Duret, C., Evanoff, B.A., Gagliardi, D., Ivanov, I.D., Leppink, N., Magnusson Hanson, L.L., Marinaccio, A., Ozguler, A., Pega, F., Pico, F., Prüss-Üstün, A.M., Ronchetti, M., Roquelaure, Y., Sabbath, E., Stevens, G.A., Tsutsumi, A., Ujita, Y., Iavicoli, S., 2018. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on stroke. Environ. Int. 19, 366–378.

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

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

Ezzati, M., Lopez, A.D., Rodgers, A., Murray, C.J.L., 2004. Comparative Quantification of Health Risks: Global and Regional Burdedn of Disease Attributable to Selected Major Risk Factors. World Health Organization, Geneva, Switzerland.

Fadel, M., Sembajwe, G., Gagliardi, D., Pico, F., Li, J., Ozguler, A., Siegrist, J., Evanoff, B.A., Baer, M., Tsutsumi, A., Iavicoli, S., Leclerc, A., Roquelaure, Y., Descatha, A., 2019. Association Between reported long working hours and history of stroke in the CONSTANCES cohort. Stroke 50, 1879–1882.

Figueroa, J.L., 2014. Distributional effects of Oportunidades on early child development. Soc. Sci. Med. 113, 42–49.

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

Godderis, L., Bakusic, J., Boonen, E., Delvaux, E., Ivanov, I.D., Lambrechts, M.-C., Latorraca, C.O., Leppink, N., Martimbianco, A.L., Pega, F., Prüss-Üstün, A.M., Riera, R., Ujita, Y., Pachito, D.V., 2018. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on alcohol use and alcohol use disorder. Environ. Int. 120, 22–33.

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

Hannerz, H., Albertsen, K., Burr, H., Nielsen, M.L., Garde, A.H., Larsen, A.D., Pejtersen, J.H., 2018. Long working hours and stroke among employees in the general workforce of Denmark. Scand. J. Public Health 46, 368–374.

Hayashi, R., Iso, H., Yamagishi, K., Yatsuya, H., Saito, I., Kokubo, Y., Eshak, E.S., Sawada, N., Tsugane, S., 2019. Japan Public Health Center-Based Prospective Study, G. Working Hours and Risk of Acute Myocardial Infarction and Stroke Among Middle-Aged Japanese Men- The Japan Public Health Center-Based Prospective Study Cohort II. Circ. J. 83, 1072–1079.

Hayashi, T., Kobayashi, Y., Yamaoka, K., Yano, E., 1996. Effect of overtime work on 24-hour ambulatory blood pressure. J. Occup. Environ. Med. 38, 1007–1011.

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

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

Huang, S.S., Zhang, Z.J., 2006. Effects of psychosocial factors on work-related musculoskeletal disorders. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi 24, 248–250.

Huang, T.Y., Moser, D.K., Hsieh, Y.S., Gau, B.S., Chiang, F.T., Hwang, S.L., 2013. Moderating effect of psychosocial factors for dyspnea in Taiwanese and American heart failure patients. J. Nurs. Res. 21, 49–58.

Hulshof, C.T.J., Colosio, C., Daams, J.G., Ivanov, I.D., Prakash, K.C., Kuijer, P., Leppink, N., Mandic-Rajcevic, S., Masci, F., van der Molen, H.F., Neupane, S., Nygard, C.H., Oakman, J., Pega, F., Proper, K., Pruss-Ustun, A.M., Ujita, Y., Frings-Dresen, M.H.W., 2019. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of exposure to occupational ergonomic risk factors and of the effect of

exposure to occupational ergonomic risk factors on osteoarthritis of hip or knee and selected other musculoskeletal diseases. Environ. Int. 125, 554–566.
Imai, T., Kuwahara, K., Miyamoto, T., Okazaki, H., Nishihara, A., Kabe, I., Mizoue, T., Dohi, S., Japan Epidemiology Collaboration on Occupational Health Study, G., 2016. Validity and reproducibility of self-reported working hours among Japanese male employees. J. Occup. Health 58, 340–346.
International Labour Organization, 1987. ISCO–88: International Standard Classification of Occupations. Geneva: International Labour Organization.
International Labour Organization, 1999. ILO estimates over 1 million work-related fatalities each year. Geneva: International Labour Organization.
International Labour Organization, 2014. Safety and health at work: a vision for sustainable prevention: XX World Congress on Safety and Health at Work 2014: Global Forum for Prevention, 24-27 August 2014, Frankfurt, Germany. Geneva: International Labour Organization.
Jarczok, M.N., Jarczok, M., Mauss, D., Koenig, J., Li, J., Herr, R.M., Thayer, J.F., 2013. Autonomic nervous system activity and workplace stressors–a systematic review. Neurosci. Biobehav. Rev. 37, 1810–1823.
Jeong, I., Rhie, J., Kim, I., Ryu, I., Jung, P.K., Park, Y.S., Lim, Y.S., Kim, H.R., Park, S.G., Im, H.J., Lee, M.Y., Won, J.U., 2013. Working hours and cardiovascular disease in Korean workers: a case-control study. J. Occup. Health 55, 385–391.
Johnson, P.I., Koustas, E., Vesterinen, H.M., Sutton, P., Atchley, D.S., Kim, A.N., Campbell, M., Donald, J.M., Sen, S., Bero, L., Zeise, L., Woodruff, T.J., 2016. Application of the Navigation Guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan. Environ. Int. 92–93, 716–728.
Johnson, P.I., Sutton, P., Atchley, D.S., Koustas, E., Lam, J., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1028–1039.
Kang, B.K., Park, T.Y., Lee, J.A., Moon, T.W., Ko, M.M., Choi, J., Lee, M.S., 2012a. Reliability and validity of the Korean Standard Pattern Identification for Stroke (K-SPI-Stroke) questionnaire. BMC Complement. Altern. Med. 12, 55.
Kang, M.Y., Park, H., Seo, J.C., Kim, D., Lim, Y.H., Lim, S., Cho, S.H., Hong, Y.C., 2012b. Long working hours and cardiovascular disease: a meta-analysis of epidemiologic studies. J. Occup. Environ. Med. 54, 532–537.
Kim, B.J., Lee, S.H., Ryu, W.S., Kim, C.K., Chung, J.W., Kim, D., Park, H.K., Bae, H.J., Park, B.J., Yoon, B.W., Investigators, A.S., 2013. Excessive work and risk of haemorrhagic stroke: a nationwide case-control study. Int. J. Stroke 8 Suppl A100, 56–61.
Kivimaki, M., Jokela, M., Nyberg, S.T., Singh-Manoux, A., Fransson, E.I., Alfredsson, L., Bjorner, J.B., Borritz, M., Burr, H., Casini, A., Clays, E., De Bacquer, D., Dragano, N., Erbel, R., Geuskens, G.A., Hamer, M., Hooftman, W.E., Houtman, I.L., Jockel, K.H., Kittel, F., Knutsson, A., Koskenvuo, M., Lunau, T., Madsen, I.E., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Rugulies, R., Salo, P., Shipley, M.J., Siegrist, J., Steptoe, A., Suominen, S.B., Theorell, T., Vahtera, J., Westerholm, P.J., Westerlund, H., O'Reilly, D., Kumari, M., Batty, G.D., Ferrie, J.E., Virtanen, M., Consortium, I.P.-W., 2015a. Long working hours and risk of coronary heart disease and stroke: a systematic review and meta-analysis of published and unpublished data for 603,838 individuals. Lancet 386, 1739–1746.
Kivimaki, M., Kawachi, I., 2015. Work stress as a risk factor for cardiovascular disease. Curr. Cardiol. Rep. 17, 630.
Kivimaki, M., Singh-Manoux, A., Virtanen, M., Ferrie, J.E., Batty, G.D., Rugulies, R., 2015b. IPD-Work consortium: pre-defined meta-analyses of individual-participant data strengthen evidence base for a link between psychosocial factors and health. Scand. J. Work Environ. Health 41, 312–321.
Kivimaki, M., Steptoe, A., 2018. Effects of stress on the development and progression of cardiovascular disease. Nat. Rev. Cardiol. 15, 215–229.
Kivimaki, M., Virtanen, M., Kawachi, I., Nyberg, S.T., Alfredsson, L., Batty, G.D., Bjorner, J.B., Borritz, M., Brunner, E.J., Burr, H., Dragano, N., Ferrie, J.E., Fransson, E.I., Hamer, M., Heikkila, K., Knutsson, A., Koskenvuo, M., Madsen, I.E.H., Nielsen, M.L., Nordin, M., Oksanen, T., Pejtersen, J.H., Pentti, J., Rugulies, R., Salo, P., Siegrist, J., Steptoe, A., Suominen, S., Theorell, T., Vahtera, J., Westerholm, P.J.M., Westerlund, H., Singh-Manoux, A., Jokela, M., 2015c. Long working hours, socioeconomic status, and the risk of incident type 2 diabetes: a meta-analysis of published and unpublished data from 222 120 individuals. Lancet Diabetes Endocrinol. 3, 27–34.
Koustas, E., Lam, J., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1015–1027.
Krause, N., Brand, R.J., Kauhanen, J., Kaplan, G.A., Syme, S.L., Wong, C.C., Salonen, J.T., 2009. Work time and 11-year progression of carotid atherosclerosis in middle-aged Finnish men. Prev. Chronic Dis. 6, A13.
Lam, J., Koustas, E., Sutton, P., Johnson, P.I., Atchley, D.S., Sen, S., Robinson, K.A., Axelrad, D.A., Woodruff, T.J., 2014. The Navigation Guide - evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ. Health Perspect. 122, 1040–1051.
Lam, J., Lanphear, B., Bellinger, D., Axelrad, D., McPartland, J., Sutton, P., Davidson, L.I., Daniels, N., Sen, S., Woodruff, T.J., 2017. Developmental PBDE exposure and IQ/ADHD in childhood: a systematic review and meta-analysis. Environ. Health Perspect. 125.
Lam, J., Sutton, P., Halladay, A., Davidson, L.I., Lawler, C., Newschaffer, C.J., Kalkbrenner, A., Joseph J. Zilber School of Public Health, Windham, G.C., Daniels, N., Sen, S., Woodruff, T.J., 2016a. Applying the Navigation Guide Systematic Review Methodology Case Study #4: Association between Developmental Exposures to Ambient Air Pollution and Autism. PLoS One 21.
Lam, J., Sutton, P., Padula, A.M., Cabana, M.D., Koustas, E., Vesterinen, H.M., Whitaker, E., Skalla, L., Daniels, N., Woodruff, T.J., 2016b. Applying the Navigation Guide Systematic Review Methodology Case Study #6: Association between Formaldehyde Exposure and Asthma: A Systematic Review of the Evidence: (Protocol registered in PROSPERO, CRD42016038766). University of California at San Francisco, San Francisco, CA.
Li, J., Brisson, C., Clays, E., Ferrario, M.M., Ivanov, I.D., Landsbergis, P., Leppink, N., Pega, F., Pikhart, H., Pruss-Ustun, A., Rugulies, R., Schnall, P.L., Stevens, G., Tsutsumi, A., Ujita, Y., Siegrist, J., 2018. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on ischaemic heart disease. Environ. Int. 119 (558–569), 9.
Li, Y., Yatsuya, H., Iso, H., Yamagishi, K., Saito, I., Kokubo, Y., Sawada, N., Tsugane, S., 2019. Body mass index and risks of incident ischemic stroke subtypes: the Japan public health center-based prospective (JPHC) study. J. Epidemiol. 5 (29), 325–333.
Li, J., Pega, F., Ujita, Y., Brisson, C., Clayse, E., Descatha, A., Ferrarioi, M., Godderis, L., Iavicoli, S., Landsbergis, P., Metzendorf, M.-I., Pachito, D., Pikhart, H., Richter, B., Roncaioli, M., Rugulies, R., Schnall, P., Sembajwe, G., Trudel, X., Tsutsumi, A., Siegrist, J., 2020. The effect of exposure to long working hours on ischaemic heart disease: A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. Environ. Int. https://doi.org/10.1016/j.envint.2020.105739.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. 6, e1000100.
Mandrioli, D., Schlünssen, V., Adam, B., Cohen, R.A., Chen, W., Colosio, C., Fischer, A., Godderis, L., Göen, T., Ivanov, I.D., Leppink, N., Mandic-Rajcevic, S., Masci, F., Nemery, B., Pega, F., Prüss-Üstün, A.M., Sgargi, D., Ujita, Y., Van der Mierden, S., Zungu, M., Scheepers, P., 2018. WHO/ILO work-related burden of disease and injury: protocols for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. Environ. Int. 119, 174–185.
McEwen, B.S., 1998a. Protective and damaging effects of stress mediators. N. Engl. J. Med. 338, 171–179.
McEwen, B.S., 1998b. Stress, adaptation, and disease. allostasis and allostatic load. Ann. N. Y. Acad. Sci. 840, 33–44.
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., Group, P.-P., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4, 1.
Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., Guyatt, G., Hooijmans, C., Langendam, M., Mandrioli, D., Mustafa, R.A., Rehfuess, E.A., Rooney, A.A., Shea, B., Silbergeld, E.K., Sutton, P., Wolfe, M.S., Woodruff, T.J., Verbeek, J.H., Holloway, A.C., Santesso, N., Schunemann, H.J., 2016. GRADE: Assessing the quality of evidence in environmental and occupational health. Environ. Int. 92–93, 611–616.
Morgan, R.L., Whaley, P., Thayer, K.A., Schunemann, H.J., 2018. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environ. Int. 121, 1027–1031.
Mukherjee, D., Patil, C.G., 2011. Epidemiology and the global burden of stroke. World Neurosurg. 76, S85–S90.
Murray, C.J.L., Ezzati, M., Lopez, A.D., Rodgers, A., Vander Hoorn, S., 2004. Comparative Quantification of Health Risks: Conceptual Framework and Methodological Issues. In: Ezzati M., Lopez A.D., Rodgers A., Murray C.J.L., (Eds.) Comparative Quantification of Health Risks: Global and Regional Burdedn of Disease Attributable to Selected Major Risk Factors. World Health Organization, Geneva.
Nakata, A., 2012. Psychosocial job stress and immunity: a systematic review. Methods Mol. Biol. 934, 39–75.
National Academies of Sciences, E., Medicine, 2017. Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. National Academies Press (US), Washington (DC).
O'Reilly, D., Rosato, M., 2013. Worked to death? A census-based longitudinal study of the relationship between the numbers of hours spent working and mortality risk. Int. J. Epidemiol. 42, 1820–1830.
Paulo, M.S., Adam, B., Akagwu, C., Akparibo, I., Al-Rifai, R.H., Bazrafshan, S., Gobba, F., Green, A.C., Ivanov, I., Kezic, S., Leppink, N., Loney, T., Modenese, A., Pega, F., Peters, C.E., Pruss-Ustun, A.M., Tenkate, T., Ujita, Y., Wittlich, M., John, S.M., 2019. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on melanoma and non-melanoma skin cancer. Environ. Int. 126, 804–815.
Pega, F., Blakely, T., Glymour, M.M., Carter, K.N., Kawachi, I., 2016. Using marginal structural modeling to estimate the cumulative impact of an unconditional tax credit on self-rated health. Am. J. Epidemiol. 183, 315–324.
Pega, F., Liu, S.Y., Walter, S., Lhachimi, S.K., 2015. Unconditional cash transfers for assistance in humanitarian disasters: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst. Rev. 9, CD011247.
Pega, F., Liu, S.Y., Walter, S., Pabayo, R., Saith, R., Lhachimi, S.K., 2017. Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst. Rev. 11, CD011135.
Pruss-Ustun, A., Wolf, J., Corvalan, C., Bos, R., Neira, M., 2017. Preventing disease through healthy environments: a global assessment of the burden of disease from enviornmental risks in: Department of Public Health E.a.S.D.o.H., ed. World Health Organization, Geneva.
Pega, F., Norris, S., Backes, C., Bero, L., Descatha, A., Gagliardi, D., Godderis, L., Loney, T., Modenese, A., Morgan, R., Pachito, D., Paulo, M., Scheepers, P., Schlünssen, V.,

Sgargi, D., Silbergeld, E., Sørensen, K., Sutton, P., Tenkate, T., Torreão, D., Ujita, Y., van Deventer, E., Woodruff, T., Mandrioli, D., 2020. RoB-SPEO: A tool for assessing risk of bias in studies estimating the prevalence of exposure to occupational risk factors from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. Environ. Int. 135, 105039. https://doi.org/10.1016/j.envint.2019.105039. Epub 2019 Dec 18. PubMed PMID: 31864023.

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

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

Rugulies, R., Ando, E., Ayuso-Mateos, J.L., Bonafede, M., Cabello, M., Di Tecco, C., Dragano, N., Durand-Moreau, Q., Eguchi, H., Gao, J., Garde, A.H., Iavicoli, S., Ivanov, I.D., Leppink, N., Madsen, I.E.H., Pega, F., Pruss-Ustun, A.M., Rondinone, B.M., Sorensen, K., Tsuno, K., Ujita, Y., Zadow, A., 2019. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on depression. Environ. Int. 125, 515–528.

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

Schunemann, H., Hill, S., Guyatt, G., Akl, E.A., Ahmed, F., 2011. The GRADE approach and Bradford Hill's criteria for causation. J. Epidemiol. Community Health 65, 392–395.

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

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

Sonnentag, S., Venz, L., Casper, A., 2017. Advances in recovery research: What have we learned? What should be done next? J. Occup. Health Psychol. 22, 365–380.

Sparks, K, C.C., Fried, Y., Shirom, A., 1997. The effects of hours of work on health: a meta-analytic review. J. Occup. Organ. Psychol. 70, 391–408.

Saunders, D.L., Kehoe, K.C., Rinehart, V.H., Berg, B.W., 2005. Self-reporting of internal medicine house staff work hours. Hawaii Med. J. 64, 14–16.

Steptoe, A., Kivimaki, M., 2012. Stress and cardiovascular disease. Nat. Rev. Cardiol. 9, 360–370.

Steptoe, A., Kivimaki, M., 2013. Stress and cardiovascular disease: an update on current knowledge. Annu. Rev. Public Health 34, 337–354.

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

Taris, T.W., Ybema, J.F., Beckers, D.G., Verheijden, M.W., Geurts, S.A., Kompier, M.A., 2011. Investigating the associations among overtime work, health behaviors, and health: a longitudinal study among full-time employees. Int J Behav Med 18, 352–360.

Teixeira, L.R., Azevedo, T.M., Bortkiewicz, A., Correa da Silva, D.T., de Abreu, W., de Almeida, M.S., de Araujo, M.A.N., Gadzicka, E., Ivanov, I.D., Leppink, N., Macedo, M. R.V., de, S.M.E.M.G., Pawlaczyk-Luszczynska, M., Pega, F., Pruss-Ustun, A.M., Siedlecka, J., Stevens, G.A., Ujita, Y., Braga, J.U., 2019. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to occupational noise and of the effect of exposure to occupational noise on cardiovascular disease. Environ. Int. 125, 567–578.

Tenkate, T., Adam, B., Al-Rifai, R.H., Chou, B.R., Gobba, F., Ivanov, I.D., Leppink, N., Loney, T., Pega, F., Peters, C.E., Pruss-Ustun, A.M., Silva Paulo, M., Ujita, Y., Wittlich, M., Modenese, A., 2019. WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on cataract. Environ. Int. 125, 542–553.

Todd, S.R., Fahy, B.N., Paukert, J.L., Mersinger, D., Johnson, M.L., Bass, B.L., 2010. How accurate are self-reported resident duty hours? J. Surg. Educ. 67, 103–107.

Tsugane, S., Sobue, T., 2001. Baseline survey of JPHC study–design and participation rate. Japan Public Health Center-based Prospective Study on Cancer and Cardiovascular Diseases. J Epidemiol. 11, S24–S29.

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

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

Virtanen, M., Heikkila, K., Jokela, M., Ferrie, J.E., Batty, G.D., Vahtera, J., Kivimaki, M., 2012. Long working hours and coronary heart disease: a systematic review and meta-analysis. Am. J. Epidemiol. 176, 586–596.

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

Virtanen, M., Kivimaki, M., 2018. Long working hours and risk of cardiovascular disease. Curr. Cardiol. Rep. 20, 123.

Viswanathan, M., Ansari, M.T., Berkman, N.D., Chang, S., Hartling, L., McPheeters, M., Santaguida, P.L., Shamliyan, T., Singh, K., Tsertsvadze, A., Treadwell, J.R., 2008. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD).

Woodfield, R., Grant, I., Group, U.K.B.S.O., Follow-Up, U.K.B., Outcomes Working, G., Sudlow, C.L., 2015a. Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from the UK Biobank Stroke Outcomes Group. PLoS One 10, e0140533.

Woodfield, R., Group, U.K.B.S.O., Follow-up, U.K.B., Outcomes Working, G., Sudlow, C. L., 2015b. Accuracy of patient self-report of stroke: a systematic review from the UK biobank stroke outcomes group. PLoS One 10, e0137538.

Wong, K., Chan, A.H.S., Ngan, S.C., 2019. The effect of long working hours and overtime on occupational health: a meta-analysis of evidence from 1998 to 2018. Int. J. Environ. Res. Public Health 16.

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

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

World Health Organization, 2018. Global Health Estimates 2016: Disease Burden by Cause, Age, Sex, by Country and by Region, 2000–2016. World Health Organization, Geneva, Switzerland.

World Health Organization, 2020. Disease burden and mortality estimates: disease burden, 2000–2016. World Health Organization, Geneva, Switzerland.

Ziganshina, L.E., Abakumova, T., Vernay, L., 2017. 9. Cochrane Database Syst. Rev. 4, CD007026.