BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

The impact of fatigue and sleep deprivation on physician and patient outcomes: A systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021967
Article Type:	Research
Date Submitted by the Author:	26-Jan-2018
Complete List of Authors:	Gates, Michelle; University of Alberta, Pediatrics Wingert, Aireen; University of Alberta, Pediatrics Featherstone, Robin; University of Alberta, Pediatrics Samuels, Charles; Centre for Sleep and Human Performance Simon, Christopher; Canadian Medical Association Dyson, Michele; University of Alberta, Pediatrics
Keywords:	Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SLEEP MEDICINE

The impact of fatigue and sleep deprivation on physician and patient outcomes: A systematic review

Michelle Gates, Aireen Wingert, Robin Featherstone, Charles Samuels, Christopher Simon, Michele P Dyson

Michelle Gates, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486C, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Aireen Wingert, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-490B, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Robin Featherstone, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486D, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Research librarian

Charles Samuels, Centre for Sleep and Human Performance, Suite 106, 51 Sunpark Drive SE, Calgary, Alberta, Canada, T2X 3V4
Physician and medical director

Christopher Simon, Canadian Medical Association, 1867 Alta Vista Drive, Ottawa, Ontario, Canada, K1G 5W8
Advisor, physician health

Michele P Dyson, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-474, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Assistant professor

Correspondence to: MP Dyson mdyson@ualberta.ca

Word count (main text): 4,853

ABSTRACT

Objectives: For physicians in independent practice, we aimed to synthesize evidence relevant to the Canadian context on two questions: (1) what are the impacts of sleep loss and fatigue on physician health and performance, and patient safety; (2) what is the effectiveness of interventions that target sleep loss and fatigue, in terms of physician and patient outcomes?

Design: We conducted a systematic review of online literature. Following a pilot phase, one reviewer independently selected studies by title and abstract; full texts were then reviewed in duplicate. One reviewer extracted data; another independently assessed a random 10% sample. Two reviewers assessed risk of bias. We synthesized results narratively.

Data sources: We searched Medline, Embase, PsycINFO, CINAHL and PubMed for published studies in April 2016; we updated the Medline search in November 2017. To locate unpublished studies, we searched Embase for conference proceedings since 2000, and hand-searched relevant meeting abstracts and association and foundation websites.

Eligibility criteria for selecting studies: English or French language primary research studies of any design published from 2000 to 2017 that examined the effect of fatigue or sleep-related exposures or interventions on any outcome among physicians in independent practice and their patients were eligible.

Results: We included 47 quantitative studies of variable quality. 28 studies showed associations between fatigue or sleep deprivation and physician health and well-being, specifically burnout, stress, adverse mental health outcomes, and reduced life satisfaction. 21 studies showed no impact on surgical performance, and mixed findings for psychomotor performance, work performance, and medical errors. Six cohort studies showed little to no effect on patient outcomes related to surgical or obstetric procedures.

Conclusions: Fatigue and sleep deprivation have detrimental effects on the health of physicians in independent practice. Due to numerous methodological shortfalls, the current body of evidence is inadequate to inform strong practice recommendations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- This rigorously conducted and transparently reported systematic review is the first to synthesize evidence on the effects of fatigue and sleep loss on physicians in independent practice.
- The review is timely, given recent calls for research into individual and organisational solutions for burnout, and an increased focus on physician health.
- While we have identified a diverse body of evidence, we could not draw definitive conclusions due to methodological weaknesses and heterogeneous outcome measures in the included studies.
- We have focused on evidence from high income countries; our findings may not be generalizable to other settings.

BACKGROUND

The working hours of physicians, which have been historically long and unpredictable, have been a topic of debate for many years.[1] Beginning in the late 1980s, evidence indicating that medical resident fatigue could negatively impact their cognitive functioning and performance, resulting in an increased risk of medical error, began to accumulate.[2] In response, by the early 2000s physicians' regulatory bodies worldwide began to take action toward restricting the work hours of medical residents and ensuring adequate time for recovery between shifts.[3-5] Since their implementation in the United States by the Accreditation Council for Graduate Medical Education (ACGME), the impact of work hour regulations has been widely researched. Still, evidence for impacts on patient care, resident training and wellbeing remains equivocal.[6-9] This is likely because work hours are only one of many contributors to fatigue and physician wellbeing. In fact, the ACGME has recently reversed the 2011 changes that limited resident work hours to 16 hours per shift and the requirement for 8 hours of time off between shifts. This decision was made in favour of promoting "flexibility" for residency training program work hours and scheduling.

The focus on medical trainees has left physicians in independent practice as a relatively neglected group in research and policy. In Canada, there is no concrete regulation on the hours or patterns in which physicians choose to work.[10] In the absence of clear policies, physicians trained under traditional systems (i.e., prior to resident work hour regulations) may find it difficult to work shorter hours or take more frequent breaks.[1] Indeed, more than 40% of practicing physicians in the United States work in excess of 80 hours per week.[11] While long work hours remain a cultural norm in medicine, in comparable high-risk industries (e.g., aviation), work patterns and work hours are tightly regulated.[12] The need for similar evidence-based policies in medicine has become a topic of increased interest. Exemplar of this, an evidence-based guideline for fatigue risk management in emergency medical services,[13] informed by a comprehensive set of systematic reviews, has recently been published. For physicians, it has been argued that there is a need to adapt healthcare systems and provide support in identifying the signs of fatigue and mitigating its risks.[1]

Besides potentially affecting patient outcomes, fatigue can impact the health and wellbeing of physicians themselves. Burnout, just one outcome related to fatigue, has been described as epidemic among physicians[14-16] and ultimately affects recruitment and retention of physicians both in community and acute care settings. While the effect of physician wellbeing on the sustainability of healthcare systems has recently received increased attention,[17] evidence-based solutions to burnout remain relatively elusive.[18] What is clear, is that comprehensive organisational-level efforts are necessary to fully address the issue.[19] Research addressing the factors that influence burnout and overall physician wellness is needed to inform system- and individual-level strategies.[20, 21] To date, evidence of the effects of fatigue and the role of chronic sleep restriction on physicians in independent practice has not been synthesized, making it unclear what gaps in knowledge remain unaddressed.

Given this void, we undertook a systematic review of primary research relevant to the Canadian context, to examine the effects of fatigue and chronic sleep restriction on physicians in independent practice, and on interventions to combat these effects. Our review was guided by the following research questions: Among physicians in independent practice, (1) what are the impacts of fatigue and chronic sleep restriction on physician health, physician performance, and patient safety; and (2) what is the effectiveness of interventions that target fatigue and chronic sleep restriction loss, in terms of improving physician and patient outcomes?

METHODS

Review conduct

The conduct of this systematic review was guided by Cochrane standards.[22] The research team convened to plan the key research questions and methodology but did not register a formal protocol. The findings are reported in adherence with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement.[23] Ethical approval was not required for this study.

Patient involvement

Patients were not involved.

Literature search

An information specialist developed a search strategy that included concepts related to physicians, fatigue and sleep. On 13 April 2016 we searched the following online databases with coverage in the biomedical sciences and psychology: Medline, Embase, PsycINFO, CINAHL and PubMed, limited to English and French language articles published from 2000 to 2016 (Medline search updated in November 2017). Though fatigue among physicians is not a new phenomenon,[2] we limited our search to articles published post-2000 to include studies relevant to current physician practice. Work hour limitations have existed in European countries since 1993, but implementation in the United States (2003)[5] and Canada (2013) for residents is more recent.[24] We aimed to include studies published in this era of increased awareness about the potential impacts of long work hours. To locate unpublished studies, we searched Embase for conference proceedings since 2000 and hand-searched meeting abstracts of the Canadian Conference on Physician Health and the International Conference on Physician Health (2012 to 2016). We also searched the following association and foundation websites: American Medical Association, Australian Medical Association, British Medical Association, Canadian Medical Association, European Medical Association, National Sleep Foundation, Ontario Medical Association and the World Medical Association. The complete search strategy undertaken is reported in Supplementary file 1.

Inclusion criteria

Primary studies (quantitative or qualitative) of fatigue- or sleep-related exposures or interventions among physicians in independent practice were eligible for inclusion. We included physicians practicing in any medical specialty and in any healthcare setting within a high income country,[25] to identify practices comparable to the Canadian setting. Studies including physicians-in-training were included only if data for physicians in independent practice could be isolated. Exposures of interest included fatigue, sleep restriction, or sleepiness. We also included studies of any intervention that aimed to reduce fatigue or sleep restriction with any comparator (or no comparator). All reported outcomes, measured at any time, were eligible for inclusion.

We excluded commentaries, letters, editorials and dissertations. Systematic reviews, health technology assessments, economic evaluations and practice guidelines were excluded, although the reference lists were scanned for potential primary studies for inclusion. Studies that focused solely on physicians-in-training (e.g., trainees, residents, fellows, interns, medical students, junior doctors, registrars) were ineligible. To maintain the focused scope of the review, we excluded work hours, work load, and any other exposure or intervention that was indirectly related to fatigue or sleep restriction.

Study selection

Two reviewers piloted the selection criteria for title and abstract screening in duplicate on 300 records. Following the pilot phase, the reviewers applied the criteria independently to the remaining records. Then, we retrieved all records classified as “include” or “unsure” and the two reviewers assessed their full text for eligibility, in duplicate. Disagreements during the full-text screening phase were resolved by discussion or the involvement of a third reviewer, when needed.

Data extraction

Reviewers used a standardized form to extract data in Microsoft Office Excel (*v.* 2016, Microsoft Corporation, Redmond, WA). One reviewer independently extracted data from each included study and a second reviewer verified a random 10% sample. Since no major errors or omissions were noted, we did not undertake further verification.

We extracted the following data: country of publication; funding source; study design; inclusion and exclusion criteria; population characteristics (i.e., sample size, age and gender distribution, physician specialty); setting (i.e., physician workplace, urban or rural); exposure or intervention; definition of fatigue or sleep loss; sleep and fatigue scales used and timing of measurement; comparators (if applicable); and outcomes.

Risk of bias appraisal

Two reviewers independently assessed the risk of bias in each included study using standard tools. Disagreements were resolved via discussion or by consulting a third reviewer. We used the Cochrane Risk of Bias tool[22] to assess randomised controlled trials across seven domains: sequence generation; allocation concealment; blinding of participants and personnel; blinding of outcome assessors; incomplete outcome reporting; selective outcome reporting; and other sources of bias. Adapted versions of the tool developed by the Effective Practice and Organization of Care group[26] were used to assess before-after and time series studies. We used the Newcastle-Ottawa Quality Assessment Scale[27] to appraise cohort studies across three domains: sample selection; comparability; and outcome assessment. We adapted the scale to assess cross-sectional studies and the one non-comparative study.

Evidence synthesis

We found insufficient homogeneity in populations, exposures or interventions, and outcomes to pool the data via meta-analysis. We have presented the findings narratively and in summary tables.

RESULTS

We identified 16,083 unique records via the database searches, 56 grey literature sources, and 15 additional records in reference lists of systematic reviews. We excluded 15,016 citations by title and abstract, and another 1,091 by full text. Forty-seven studies[28-74] were eligible for inclusion. Figure 1 shows the flow of studies through the selection process.

Included study characteristics

A summary of the study characteristics is provided in Table 1. Supplementary file 2 presents descriptive information for each included study. There were 45 observational studies[28-34, 36-69, 71-74] and two intervention studies.[35, 70] All studies were quantitative. Nearly half (n = 20/47, 43%) of the studies took place in North America,[29, 31-33, 36, 40, 43, 52-55, 57, 58, 60-63, 67, 72, 74] and slightly more than one-third (n = 16/47, 34%) in Europe.[28, 30, 34, 35, 37, 41, 42, 45-48, 56, 59, 68, 70, 71]

Table 1. Summary characteristics of the included studies

Study characteristics	n	%	Physician characteristics	n	%	Exposures, interventions and outcomes	n	%
Study design			**Gender**			**Exposures (observational)**[a]	**45**	**96**
Cross-sectional	34	72	Reported[b]	38	81	Fatigue	15	32
Cohort	6	13	>50% male	30	79	Sleep deprivation	37	79
Before-after	3	6	**Age**			Overnight or extended shifts	18	38
RCT	2	4	Reported[b]	38	81	**Interventions (experimental)**	**2**	**4**
Time series	1	2	Range (years)	20 to >70		**Outcomes**		
Non-comparative	1	2	**Specialty area**[c]			Physician health and wellbeing	28	60
Region and country			Surgeons	13	28	Work and life satisfaction	9	19
North America	20	43	Anesthesiologists	10	21	Burnout	7	15
US	15	32	Generalists	7	15	Stress	8	17
Canada	4	9	ED or ICU physicians	3	6	Mental health and wellbeing	7	15
Canada, US & Mexico	1	2	Oncologists	2	4	Other health-related outcomes	5	11
Europe	16	34	Obstetrician-gynecologists	1	2	Physician performance, risk of error	21	45
France	4	9	Mixed groups	14	30	Psychomotor performance	7	15
Finland	3	6	**Work setting**[d]			Work ability and quality of care	5	11
Spain	2	4	Hospitals	37	78	Incidence of medical errors	5	11
Austria	2	4	Private practice	13	28	Surgical efficiency, effectiveness	5	11
Norway	2	4	Primary care centres, outpatient clinics	7	15	Patient outcomes	6	13
Denmark	1	2	Academic practice, training programs	5	11			
Germany	1	2	Other (e.g., industry, military)	11	23			
Malta	1	2	Not reported	3	6			
Japan	4	9	**Urban or rural**					
Australia	2	4	Reported[b]	16	34			
Israel	2	4	Urban	12	75			
New Zealand	2	4	Rural	2	13			
United Kingdom	1	2	Mixed	2	13			

ED: emergency department; ICU: intensive care unit; RCT: randomised controlled trial; US: United States of America

[a]Exposures that have been directly related to an outcome. Some studies included multiple exposures.

[b]Percentages presented using the total number of studies where the outcome was reported as the denominator.

[c]Anesthesiologists include physician anesthetists; generalists include primary care physicians, internists, and general practitioners; mixed groups refers to studies including more than one physician group or specialty (usually large-scale surveys). In some studies, multiple distinct groups were represented.

[d]As defined by the authors. Values for the settings will exceed 100% because studies may occur in more than one setting.

The 47 studies reported outcomes for 36,190 (range = 6 to 7,905) physicians and 69,809 (range = 270 to 38,978) adult patients. About half reported on surgeons (n = 13/47, 28%),[28, 33, 36, 40, 43, 49, 53, 57, 58, 61, 70, 72, 74] or anesthesiologists/physician anesthetists (n = 10/47, 21%).[30, 31, 34, 38, 39, 45-47, 54, 55] Where it was reported, the samples tended to be predominantly male (n = 30/38, 79%) and physician age varied widely. Hospitals were the most common setting (n = 37/47, 79%).[28-31, 33-38, 40-42, 44-47, 49, 52-56, 58-65, 68-70, 72-74] In the studies where it was reported (n = 16/47, 34%),[28, 32, 33, 35, 36, 38, 40, 45, 46, 50, 51, 53, 60, 71, 72, 74] all but four studies[50, 51, 72, 74] took place in solely an urban setting.

Fifteen (32%) studies reported on fatigue exposure,[29, 35, 40, 43, 52, 58-66, 68] while others (n = 37/47, 79%) reported on sleep deprivation or reduced sleep quality.[28, 30-42, 44-51, 53-57, 59, 62, 66, 67, 69-74] A few (n = 5/47, 11%) reported on both.[35, 40, 59, 62, 66] In some cases (n = 18/47, 38%), fatigue or sleep loss were related to overnight work or long on-call shifts.[28, 31, 33, 35, 36, 38, 40, 41, 45, 46, 48, 53-55, 58, 70, 72, 74] Measured outcomes varied widely and were ultimately organised into physician physical and mental health, physician performance and risk of error, and patient outcomes.

Risk of bias appraisal

The overall quality of the body of research was poor; 72% (n = 34/47) of studies were rated at unclear or high risk of bias. Of the two randomised controlled trials, one was rated as unclear overall risk of bias[70] and one as high risk.[35] All cohort studies were at low risk of bias (mean score: 8.4/9, range: 8-9).[33, 36, 53, 58, 72, 74] All of the before-after studies were rated as high risk of bias.[28, 40, 45] The single time series study was assessed at high risk of bias.[46] The cross-sectional studies varied in performance (mean score: 3.0/5, range: 1-4); only one-third (n = 12/34, 35%) were at low risk of bias.[34, 37, 39, 42, 47, 54, 55, 64, 65, 67, 69, 71] The one non-comparative study was at unclear risk of bias.[38] Detailed assessments for each study are shown in Supplementary file 3.

Physician health and wellbeing outcomes

Twenty-eight studies reported on physician health and wellbeing-related outcomes,[29, 30, 32, 34, 35, 37, 41-43, 45-52, 55, 57, 59, 62, 63, 65-67, 69, 71, 73] including burnout (n = 7), stress (n = 8), mental health and wellbeing (n = 7), life and job satisfaction (n = 9) and other markers of health (n = 5) (Supplementary file 4).

Seven cross-sectional studies demonstrated links between sleep deprivation and burnout among surgeons,[49, 57] anesthesiologists,[34] generalists[71] and other mixed groups.[65, 67, 69] Two studies reported on surgeons; the larger (n = 2,564, low risk of bias) study of neurosurgeons showed increased odds of burnout with sleep deprivation (OR 0.84, 95% CI 0.75-0.94, P = 0.002).[49] Among anesthesiologists one study (n = 565, low risk of bias) indicated that burnout was more prevalent among the sleep-deprived (47.6% vs. 16.3%, P < 0.001).[34] In one small (n = 11) study of generalists, those with burnout had poorer Pittsburgh Sleep Quality Index scores (7.24±4.17 vs. 2.72±2.22, P < 0.001).[71] In the two larger studies of mixed physician groups (low risk of bias), burnout was more prevalent among those who were sleep deprived (39.6% vs. 26.4%, P < 0.05),[67] and physical fatigue was correlated with burnout (r = 0.88, P < 0.05).[65]

Seven observational studies of varying methodological quality[29, 41, 42, 45, 47, 57, 59] and one intervention study at high risk of bias[35] reported on stress outcomes among surgeons,[57] anesthesiologists,[45, 47] emergency physicians,[35, 59] internal medicine physicians,[41] and mixed groups.[29, 42] In a small sample (n=20) of internal medicine physicians, a 24-hour call shift had no effect on biochemical or physiological stress parameters, except levels of thyroid stimulating hormone, which was higher post-shift (P = 0.049, data not reported).[41] The remaining observational studies suggested that there was a link between sleep deprivation or fatigue and stress. The one study of orthopedic surgeons (n = 264, high risk of bias) showed that sleep deprivation and psychological distress were correlated (data not reported, P < 0.001).[57] The two reports on anesthesiologists were of varied quality; the larger (n = 328) study that was at low risk of bias showed that stress symptoms were predicted by sleep deprivation (β = -0.269, P < 0.001).[47] Among the two studies reporting on mixed groups of physicians, the larger (n = 1,541, low risk of bias) study showed an association between sleep problems and psychological distress (β = 0.18, P < 0.001).[42] One RCT assessed the impact of sleep deprivation from shift work, showing that stress among emergency physicians (n = 17) was higher following the shift as compared to a control day (data not reported, P < 0.05).[35]

Seven cross-sectional studies of varying methodological quality reported on aspects of mental health including addiction or substance misuse,[30, 48, 66] depression,[73] thoughts of suicide,[47] mood disturbance[55, 66] and overall wellbeing.[62] One study,[48] which was at high risk of bias, showed no association between hours of sleep when on call and hazardous drinking behaviours. Meanwhile, the six other studies all showed deleterious effects of sleep deprivation and fatigue on mental health. Three studies reported on anesthetists,[30, 47, 55] with two large surveys showing increased odds of tobacco (OR 1.42, 95% CI 1.04-1.94) and tranquilizer/hypnotics (OR 3.26, 95% CI 2.12-5.02) dependency being predicted by sleep deprivation,[30] and sleep disturbance being associated with thoughts of suicide (P = 0.009).[47] A small study (n = 21) showed greater mood disturbance following a 17-hour night shift than a usual day (Profile of Mood States score 42.57±15.26 vs. 70.90±6.91, P < 0.001).[55] Among oncologists (n = 241), overall wellbeing was predicted by lower levels of fatigue after controlling for personal and professional characteristics (P = 0.002).[62] A large (n = 3,862, unclear risk of bias) study of physicians showed that sleep deprivation was associated with increased odds of depression (OR 2.70, 95% CI 1.82-4.03 for men; OR 2.38, 95% CI 1.11-5.10 for women).[73] In open-ended questions, senior physicians in one study (unclear risk of bias) attributed the development of mental illness to tiredness and stress at work.[66]

Nine cross-sectional studies of varying methodological quality reported on outcomes related to job satisfaction,[37, 42, 43, 50, 67, 69] life satisfaction[32, 57, 67] or work-life balance.[63, 67] All but two[37, 67] of these studies showed that sleep deprivation and fatigue were associated with reductions in satisfaction. The six studies that investigated job satisfaction were all at low risk of bias and generally included mixed groups of physicians[42, 67, 69]; one study reported on general practitioners,[50] and another on surgeons.[43] Three studies showed that reductions in sleep duration and/or quality[42, 43, 69] were associated with reduced job satisfaction. Meanwhile one showed no association between sleep deprivation and career satisfaction,[67] and another showed no relationship between earlier sleep disturbance and later job demands or job control.[37] A single study (n = 92) reporting on rural general practitioners indicated that frequent sleep disturbance predicted the intention to retire early (OR 2.91, 95% CI 1.11-7.6, P < 0.05).[50]

The three studies that reported on life satisfaction were of variable quality, but all demonstrated links between sleep deprivation or fatigue and reductions in life satisfaction.[32, 57, 67] Of two studies among mixed physician groups,[32, 67] the one larger (n = 840) study showed that sleep deprivation (less than 7 hours per day) was a predictor of reduced life satisfaction (OR 0.44, 95% CI 0.29-0.67, P < 0.05).[67] One study at high risk of bias reported on orthopedic surgeons (n = 264), showing that sleep deprivation was correlated with lower marital satisfaction (data not reported, P < 0.001).[57] Two large studies at low or unclear risk of bias reported on work-life balance.[63, 67] Among oncologists (n = 1,117), reduced satisfaction with work-life balance was predicted by high levels of fatigue, even when controlling for personal and work-related factors and burnout (OR 0.489, 95% CI 0.337-0.710, $P < 0.001$).[63] Among a mixed group of physicians (n = 840, low risk of bias), sleep deprivation predicted a reduced perception of having balanced personal and professional commitments (OR 0.46, 95% CI 0.31-0.71, $P \leq 0.05$).[67]

Five cross sectional studies at high or unclear risk of bias[32, 51, 52, 66] and one time series study at high risk of bias[46] reported on other health-related outcomes. Among a mixed group of physicians (n = 180), one study at high risk of bias showed that Epworth Sleepiness Scale scores were higher among physicians who worried about having a car accident while driving home (7.0 vs. 5.4, $P < 0.001$).[32] Among generalists (n = 578), almost 1 in 10 (8.7%) admitted to falling asleep while driving due to fatigue.[52] Also among generalists (n = 92), those with frequent work-related sleeping problems were at increased odds of sickness presenteeism (OR 2.92, 95% CI 1.19-7.16, $P = 0.02$).[51] The one time series study concluded that a single 24-h shift did not cause major chronodisruption among anesthetists (n = 10).[46] Meanwhile, open-ended comments from a large sample (n = 3,550) of senior physicians suggests that they attributed the development of physical health problems to a lifestyle of sleep deprivation, poor eating habits and lack of exercise imposed by their jobs.[66]

Physician performance and risk of errors

Twenty-one studies reported on physician performance and safety-related outcomes,[28, 31-33, 36, 38-42, 44, 45, 54, 56, 60, 61, 64, 66, 68, 70, 74] including surgical efficiency and effectiveness (n = 5), psychomotor performance (n = 7), work ability and quality of care (n = 5) and incidence of medical errors (n = 5) (Supplementary file 5).

Three cohort studies at low risk of bias,[33, 36, 74] one before-after study at high risk of bias[28] and one randomized controlled trial at high risk of bias[70] examined the effects of sleep deprivation from overnight work or extended shifts, during surgeries[33, 36, 74] or laparoscopic simulations.[28, 70] The cohort studies, which reported on 49,776 surgical procedures, found no adverse effects on any measure of surgical efficiency or effectiveness.[33, 36, 74] The small (n = 29) before-after study showed no impact of sleep deprivation from shift-work nor of sleep hours on performance on a laparoscopic simulation.[28] One small (n = 64) intervention study compared a 24-hour shift to a usual work day, also finding no detriment to performance on a laparoscopic simulation despite diminished sleep hours while working on-call.[70]

Two before-after studies at high risk of bias[40, 45] and five cross-sectional studies of variable methodological quality[31, 38, 41, 54, 56] reported on psychomotor performance outcomes among surgeons,[40] anesthesiologists,[31, 38, 45, 54] emergency physicians,[56] and internal medicine physicians.[41] Four studies[38, 40, 45, 54] showed an overall reduction in psychomotor performance in the fatigued state while the others had mixed results.[31, 56] Among a small group of surgeons (n = 9), performance on a virtual ring transfer task deteriorated after an on-call shift (data not reported, P < 0.05).[40] The four studies among anesthetists reported mixed findings. One small (n = 11) before-after study showed longer reaction times (690.8±73.4 vs. 746.5±113.7 milliseconds) and reduced concentration ability (26.4±23.5 vs. 56.3±23.0 on a 100-point scale, P = 0.007) following a 24-hour shift with sleep deprivation[45]; Two others found that sleep loss was associated with slower reaction times.[38, 54] Conversely, a small study (n = 11) found no effect of overnight shiftwork with sleep deprivation on any measure of psychomotor performance except Hopkin's Verbal Learning Test (t-score of 48.6±7.6 vs. 41.5±9.9, P = 0.04).[31] Among emergency physicians (n = 18), one study (unclear risk of bias) those who were sleep deprived had a reduced performance on most but not all psychomotor tests,[56] while among internal medicine physicians (n = 20, low risk of bias), neurocognitive parameters did not seem to worsen post-call.[41]

Five cross-sectional studies of variable methodological quality reported on associations between sleep deprivation or fatigue and work ability or perceived performance, all among mixed groups of physicians.[32, 42, 60, 64, 66] The two large studies at low risk of bias showed that sleep problems and fatigue had a negative impact on physicians' work.[42, 64] Among 1,541 physicians in Finland, sleeping problems were inversely associated with scores on the Work Ability Index (β = -0.29, P < 0.001),[42] while a study of 890 physicians from Israel demonstrated that perceived quality of care was predicted by fatigue even after controlling for components of burnout (β = 0.17, P < 0.05).[64] Similarly, in one study, comments from senior physicians suggested that continual tiredness and exhaustion negatively affected their competence.[66] The two studies[32, 60] that were at high risk of bias had conflicting findings.

Five cross-sectional studies of variable methodological quality reported on associations between sleep deprivation, fatigue and self-reported medical errors among surgeons,[61] anesthesiologists[39] and mixed groups of physicians.[32, 44, 68] Two studies showed that sleep disturbance was associated with an increased risk of errors,[39, 44] while the findings of the other studies were mixed.[32, 61, 68] A large (n = 7,905) study at unclear risk of bias showed that only 6.9% of surgeons reported fatigue as the most important contributor to medical errors.[61] Among anesthesiologists, a smaller study (n = 183) at low risk of bias showed that the risk of fatigue-related errors increased with more nights of work-related sleep disturbance (RR 1.25, 95% CI 1.06-1.49).[39] Two of the studies reporting on mixed groups of physicians had conflicting results,[32, 44] while another reported that physicians' opinions on the association between fatigue and prescribing errors differed by work setting.[68] One-third (34%) of community-based, 96% of hospital-based, and 8% of office-based physicians believed that there was a high or very high association between fatigue and prescribing errors ($P < 0.05$).[68]

Patient Outcomes

Six large (n = 270 to 38,978) cohort studies at low risk of bias reported on patient outcomes, all related to surgical[33, 36, 53, 58, 72, 74] or obstetric[53] procedures (Supplementary file 6). In these studies, sleep deprivation or fatigue were typically defined as overnight work prior to a daytime procedure[36, 53, 58, 72, 74]; though two studies measured sleep hours[33] or 'sleep opportunity'.[53] Overall there appeared to be little[33] to no[36, 58, 72, 74] effect of sleep deprivation from overnight work on adverse patient outcomes such as operative complications, length of stay, and mortality. One study showed that nighttime work prior to a daytime procedure did not affect complication rates, but that shorter sleep opportunity increased the odds of operative (OR 2.70, 95% CI 1.13-6.48, $P = 0.03$) but not obstetric complications.[53]

DISCUSSION

Fatigue and chronic sleep restriction are two potential drivers of reduced physician wellbeing[17, 19] that have thus far been understudied in physicians in independent practice. Burnout is becoming increasingly prevalent among physicians,[14-16] and recent research indicates that comprehensive individual- and system-level strategies are needed to address the problem.[6-9, 19, 21] We have systematically reviewed evidence from a heterogeneous array of available studies reporting on diverse outcomes related to physicians in independent practice and their patients. The included studies were often at high or unclear risk of bias, included small samples of physicians, and inconsistently measured and reported exposures and outcomes. As a result of these methodological shortcomings, the currently available evidence is inadequate to inform practice or policy recommendations.

Traditionally, much of the fatigue-related research has focused on hazards to patients. The current review included five cohort studies showing that sleep loss and/or fatigue did not seem to jeopardize patient safety. Despite these findings, evidence for psychomotor performance, surgical skills and errors suggest that there is indeed a potential for negative outcomes. The studies, like many of the others in this and other systematic reviews,[75] employed indirect definitions that make it difficult to classify sleep deprived physicians with certainty. In recent years there has been a shift away from the singular focus on patient safety toward a more comprehensive view that also considers the detrimental effects of fatigue, sleep loss and other occupational hazards on physician wellness.[76] Evidence from this review supports the negative effects that fatigue and sleep loss may have on physician health and wellbeing. It is now recognized that health systems cannot be sustained by a workforce that is facing an epidemic of burnout.[19, 77, 78]

In light of high rates of burnout, the ongoing dialogue about the need for a cultural shift in the practice of medicine[79, 80] is now more important than ever. Recognition of the potential effects of physician fatigue on patients, physicians, and healthcare systems as a whole must be emphasized at a systemic level, encouraging a shift in which the risks are viewed as unacceptable.[1, 20, 76] Likewise, although research to date has focused largely on individual-level approaches to address burnout, it is now clear that placing the burden of a system-level problem solely on the individual is unlikely to bring about significant and lasting change.[81] Recent research has highlighted physician burnout as a system-driven issue that will require corresponding national-scale multicomponent solutions.[1, 19, 77, 78] As such, in the past several years both the American and Canadian Medical Associations have developed policies and programs that address physician health.[77, 82] The Canadian Medical Association's new policy on physician health calls on broad stakeholder groups (e.g., policymakers, regional health authorities, governments) to take shared responsibility for the health of physicians and to make meaningful and concerted efforts towards promoting a healthy and sustainable workforce.[77]

Our systematic review indicates that the current evidence base is inadequate to inform decision-making. Correspondingly, a 2016 research summit on physician wellness and burnout outlined the need for timely, relevant and methodologically robust research to inform practice and policy.[21] We identified only two intervention studies, which supports the assertion that novel interventions with realistic budgets and timelines at both individual and organisation levels need to be tested.[21] The vast array of tools used by current studies to measure sleep, fatigue and various outcomes impedes evidence synthesis. It will be important to make use of exiting validated measures[83-85] consistently in future research. Identifying outcomes of importance to physicians and their patients should be prioritized, such that these may be collected within intervention studies. Reporting these consistently will allow for effective synthesis of findings and reduce research waste.[86] Integrated knowledge translation strategies involving multiple stakeholder groups (e.g., physicians, patients, medical schools, physicians' associations and governing bodies, policymakers) may help to ensure that the research is relevant and facilitates decision-making.[87]

Strengths and Limitations

Our systematic review is the first to synthesize evidence on the effects of fatigue and sleep loss on physicians in independent practice. The review is timely, given recent calls for research into individual and organisational solutions for burnout,[20, 21] and an increased focus on physician health.[76, 77] While we have identified a diverse body of evidence, we could not draw definitive conclusions due to methodological weaknesses and heterogeneous outcome measures in the included studies. The findings may have been influenced by publication bias, and may not be generalized to all settings, given our restriction to high income countries. Rigorously conducted and reported studies will be required to determine with confidence the potential impacts of fatigue on physicians and their patients, and to inform reasonable and sustainable solutions to the problem.

CONCLUSION

The evidence synthesized in this review suggests some detrimental impacts of fatigue and sleep deprivation on physician health and wellbeing, and mixed evidence for potential impacts on performance and safety outcomes. The evidence overall did not indicate any impact on patient outcomes. Our overall confidence in the findings is low, owing to a body of research that is hindered by methodological weaknesses, including small sample sizes and inconsistent measurement of fatigue exposure and outcomes. Further methodologically robust research that includes consistent outcomes that are of interest to physicians and their patients is needed to inform strong practice recommendations and policy decisions.

ACKNOWLEDGMENTS

We would like to thank Dr. Meghan Sebastianski and Andrea Milne for assistance with screening, and Dr. Allison Gates for screening, data verification, and quality assessment of the French studies.

COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Dr. Christopher Simon is employed by the Canadian Medical Association, who provided financial support for the research; there are no other relationships or activities that could appear to have influenced the submitted work.

CONTRIBUTOR STATEMENT

All authors contributed to the conception and design of the project. MG and AW contributed to the acquisition, analysis and interpretation of the data, and drafted the manuscript. RF contributed to acquisition of data. CSa, CSi and MPD contributed to interpretation of data and revised the manuscript for important intellectual content. All authors approved the final version of the manuscript as submitted.

FUNDING STATEMENT

This research was supported by the Canadian Medical Association. The funder had no role in the study design; collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

ROLE OF FUNDERS

Dr. Christopher Simon is employed by the Canadian Medical Association. The remaining authors are independent from the funders. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

TRANSPARENCY DECLARATION

The lead author (MG) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted; and all discrepancies from the study as planned have been explained.

DATA SHARING STATEMENT

All authors, external and internal, had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the interpretation.

DATA SHARING STATEMENT

The data pertaining to this systematic review are available from the corresponding author upon reasonable request.

FIGURE CAPTIONS

Figure 1. Flow of records through the selection process

REFERENCES

1. Greig P, Snow R. Fatigue and risk: are train drivers safer than doctors? *BMJ* 2017;359. doi: https://doi.org/10.1136/bmj.j5107
2. Parker JBR. The effects of fatigue on physician performance - an underestimated cause of physician impairment and increased patient risk. *Can J Anaesth* 1987;34(5):489-95.
3. Temple J. Resident duty hours around the globe: where are we now? BMC Med Educ 2014;14(Suppl 1):S8. https://doi.org/10.1186/1472-6920-14-S1-S8
4. Imrie K, Frank J, The National Steering Committee on Resient Duty Hours. Fatigue, risk, & excellence: towards a pan-Canadian consensus on resident duty hours. Ottawa, Ontario: The Royal College of Physicians and Surgeons of Canada, 2013. http://www.residentdutyhours.ca/final_report.php (accessed 12 Jan 2018).

5. Accreditation Council for Graduate Medical Education. History of duty hours. 2017. http://www.acgme.org/What-We-Do/Accreditation/Clinical-Experience-and-Education-formerly-Duty-Hours/History-of-Duty-Hours (accessed 12 Jan 2018).
6. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ* 2015;7(3):349-63.
7. Harris JD, Staheli G, LeClere L, Andersone D, McCormick F. What effects have resident work-hour changes had on education, quality of life, and safety? A systematic review. *Clin Orthop* 2015;473(5):1600-8.
8. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg* 2014;259(6):1041-53.
9. Peets A, Ayas NT. Restricting resident work hours: the good, the bad, and the ugly. *Crit Care Med* 2012;40(3):960-6.
10. Canadian Medical Association. CMA Policy: Management of physician fatigue. Ottawa, Canada: Canadian Medical Association; 2014.
11. Anim M, Markert RJ, Wood VC, Schuster BL. Physician practice patterns resemble ACGME duty hours. *Am J Med* 2009;122:587-93.
12. Transport Canada. Fatigue risk management system for the canadian aviation industry: policies and procedures development guidelines. Ottawa, Canada: Her Majesty the Queen in Right of Canada, as represented by the Minister of Transport, 2008. https://www.tc.gc.ca/eng/civilaviation/publications/TP14576-6042.htm (accessed 12 Jan 2018).
13. Patterson PD, Higgins JS, Lang ES, et al. Evidence-based guidelines for fatigue risk management in ems: formulating research questions and selecting outcomes. *Prehosp Emerg Care* 2017;21(2):149-56.
14. Boudreau RA, Grieco RL, Cahoon SL, Robertson RC, Wedel RJ. The pandemic from within: two surveys of physician burnout in Canada. *Can J Commun Ment Health* 2006;25(2):71-88.
15. Kumar S. Burnout and doctors: prevalence, prevention and intervention. *Healthcare* 2016;4(3):37. doi: 10.3390/healthcare4030037
16. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among us physicians relative to the general us population. *Arch Intern Med* 2012;172(18):1377-85.
17. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. *Acad Psychiatry* 2017. doi: 10.1007/s40596-017-0781-6

18. Fralick M, Flegel K. Physician burnout: who will protect us from ourselves? *CMAJ* 2014;186(10):731. doi: 10.1503/cmaj.140588
19. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. *Mayo Clin Proc* 2017;92(1):129-46.
20. Lemaire JB, Wallace JE. Burnout among doctors. BMJ 2017;358:j3360. doi: 10.1136/bmj.j3360
21. Dyrbye LN, Trockel M, Frank E, et al. Development of a research agenda to identify evidence-based strategies to improve physician wellness and reduce burnout. *Ann Intern Med* 2017;166(10):743-4.
22. Higgins JPT, Green S (editors). The Cochrane handbook for systematic reviews of interventions, version 5.1.0. London, UK: The Cochrane Collaboration, 2011. http://ww.handbook.cochrane.org (accessed 12 Jan 2018).
23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6(7):e1000100. https://doi.org/10.1371/journal.pmed.1000100
24. Pattani R, Wu PE, Dhalla IA. Resident duty hours in Canada: past, present and future. *CMAJ* 2014;186(10):761-5.
25. International Monetary Fund. World economic outlook: too slow for too long. Washington, DC: International Monetary Fund, 2016. http://www.imf.org/external/pubs/ft/weo/2016/01/ (accessed 12 Jan 2018).
26. Cochrane Effective Practice and Organisation of Care. Suggested risk of bias criteria for EPOC reviews. Oslo, Norway: Norwegian Knowledge Centre for the Health Services, 2016. http://epoc.cochrane.org/epoc-specific-resources-review-authors (accessed 12 Jan 2018).
27. Wells GA SB, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa, Canada: The Ottawa Hospital Research Institute, 2014. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 12 Jan 2018).
28. Amirian I, Andersen LT, Rosenberg J, Gogenur I. Laparoscopic skills and cognitive function are not affected in surgeons during a night shift. *J Surg Educ* 2014;71(4):543-50.
29. Aziz A. Sources of perceived stress among american medical doctors: a cross-cultural perspective. Cross Cultural Management 2004;11(4):28-39.
30. Beaujouan L, Czernichow S, Pourriat JL, Bonnet F. Prevalence and risk factors for substance abuse and dependence among anaesthetists: a national survey. *Ann Fr Anesth Réanim* 2005;24(5):471-9.

31. Chang LC, Mahoney JJ, 3rd, Raty SR, Ortiz J, Apodaca S, De La Garza R 2nd. Neurocognitive effects following an overnight call shift on faculty anesthesiologists. *Acta Anaesthesiol Scand* 2013;57(8):1051-7.

32. Chen I, Vorona R, Chiu R, Ware JC. A survey of subjective sleepiness and consequences in attending physicians. *Behav Sleep Med* 2008;6(1):1-15.

33. Chu MW, Stitt LW, Fox SA, Kiaii B, Quantz M, Guo L, et al. Prospective evaluation of consultant surgeon sleep deprivation and outcomes in more than 4000 consecutive cardiac surgical procedures. *Arch Surg* 2011;146(9):1080-5.

34. Doppia MA, Estryn-Béhar M, Fry C, Guetarni K, Lieutaud T, comité de pilotage de l'enquête SESMAT. Burnout in French doctors: a comparative study among anaesthesiologists and other specialists in French hospitals (SESMAT study). *Ann Fr Anesth Réanim* 2011;30(11):782-94.

35. Dutheil F, Trousselard M, Perrier C, et al. Urinary interleukin-8 is a biomarker of stress in emergency physicians, especially with advancing age--the JOBSTRESS randomized trial. *PLoS One* 2013;8(8):e71658. https://doi.org/10.1371/journal.pone.0071658

36. Ellman PI, Law MG, Tache-Leon C, et al. Sleep deprivation does not affect operative results in cardiac surgery. *Ann Thorac Surg* 2004;78(3):906-11.

37. Elovainio M, Heponiemi T, Jokela M, et al. Stressful work environment and wellbeing: what comes first? *J Occup Health Psychol* 2015;20(3):289-300.

38. Gander P, Millar M, Webster C, Merry A. Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiol Int* 2008;25(6):1077-91.

39. Gander PH, Merry A, Millar MM, Weller J. Hours of work and fatigue-related error: a survey of New Zealand anaesthetists. *Anaesth Intensive Care* 2000;28(2):178-83.

40. Gerdes J, Kahol K, Smith M, Leyba MJ, Ferrara JJ. Jack Barney award: the effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. *Am J Surg* 2008;196(6):813-9.

41. Harbeck B, Suefke S, Haas CS, Lehnert H, Kropp P, Moenig H. No stress after 24-hour on-call shifts? *J Occup Health* 2015;57(5):438-47.

42. Heponiemi T, Puttonen S, Elovainio M. On-call work and physicians' well-being: testing the potential mediators. *Occup Med* 2014;64(5):352-7.

43. Jackson TN, Pearcy CP, Khorgami Z, Agrawal V, Taubman KE, Truitt MS. The physician attrition crisis: a cross-sectional survey of the risk factors for reduced job satisfaction among US surgeons. *World J Surg* 2017;24. doi: 10.1007/s00268-017-4286-y

44. Kaneita Y, Ohida T. Association of current work and sleep situations with excessive daytime sleepiness and medical incidents among Japanese physicians. *J Clin Sleep Med* 2011;7(5):512-22.

45. Lederer W, Kopp M, Hahn O, et al. Post-duty psychomotor performance in young and senior anaesthetists. *Eur J Anaesthesiol* 2006;23(3):251-6.

46. Leichtfried V, Putzer G, Perkhofer D, Schobersberger W, Benzer A. Circadian melatonin profiles during single 24-h shifts in anesthetists. *Sleep Breath* 2011;15(3):503-12.

47. Lindfors PM, Nurmi KE, Meretoja OA, et al. On-call stress among Finnish anaesthetists. *Anaesthesia* 2006;61(9):856-66.

48. Mahmood JI, Stoen Grotmol K, Tesli M, Vaglum P, Tyssen R. Contextual factors and mental distress as possible predictors of hazardous drinking in Norwegian medical doctors: a 15-year longitudinal, nationwide study. *Eur Addict Res* 2017;23(1):19-27.

49. Nishimura K, Nakamura F, Takegami M, et al. Cross-sectional survey of workload and burnout among Japanese physicians working in stroke care: the nationwide survey of acute stroke care capacity for proper designation of comprehensive stroke center in Japan (J-ASPECT) study. *Circ Cardiovasc Qual Outcomes* 2014;7(3):414-22.

50. Pit SW, Hansen V. Factors influencing early retirement intentions in Australian rural general practitioners. *Occup Med* 2014;64(4):297-304.

51. Pit SW, Hansen V. The relationship between lifestyle, occupational health, and work-related factors with presenteeism amongst general practitioners. *Arch Environ Occup Health* 2016;71(1):49-56.

52. Roberts DL, Shanafelt TD, Dyrbye LN, West CP. A national comparison of burnout and work-life balance among internal medicine hospitalists and outpatient general internists. *J Hosp Med* 2014;9(3):176-81.

53. Rothschild JM, Keohane CA, Rogers S, et al. Risks of complications by attending physicians after performing nighttime procedures. *JAMA* 2009;302(14):1565-72.

54. Saadat H, Bissonnette B, Tumin D, et al. Effects of partial sleep deprivation on reaction time in anesthesiologists. *Paediatr Anaesth* 2017;27(4):358-62.

55. Saadat H, Bissonnette B, Tumin D, et al. Time to talk about work-hour impact on anesthesiologists: The effects of sleep deprivation on Profile of Mood States and cognitive tasks. *Paediatr Anaesth* 2016;26(1):66-71.

56. Sanches I, Teixeira F, dos Santos JM, Ferreira AJ. Effects of acute sleep deprivation resulting from night shift work on young doctors. *Acta Med Port* 2015;28(4):457-62.

57. Sargent MC, Sotile W, Sotile MO, Rubash H, Barrack RL. Quality of life during orthopaedic training and academic practice. Part 1: orthopaedic surgery residents and faculty. *J Bone Joint Surg Am* 2009;91(10):2395-405.

58. Schieman C, MacLean AR, Buie WD, Rudmik LR, Ghali WA, Dixon E. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? *Am J Surg* 2008;195(5):684-7.

59. Sende J, Jbeili C, Schvahn S, Khalid M, Asaph J, Romano H, et al. Stress factors and stress consequences among emergency physicians: vational survey. *Annales Francaises de Médecine d'Urgence* 2012;2(4):224-31.

60. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. *Hum Perform Extrem Environ* 2001;6(1):6-11.

61. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg* 2010;251(6):995-1000.

62. Shanafelt TD, Novotny P, Johnson ME, et al. The well-being and personal wellness promotion strategies of medical oncologists in the North Central Cancer Treatment Group. *Oncology* 2005;68(1):23-32.

63. Shanafelt TD, Raymond M, Kosty M, et al. Satisfaction with work-life balance and the career and retirement plans of US oncologists. *J Clin Oncol* 2014;32(11):1127-35.

64. Shirom A, Nirel N, Vinokur AD. Overload, autonomy, and burnout as predictors of physicians' quality of care. *J Occup Health Psychol* 2006;11(4):328-42.

65. Shirom A, Nirel N, Vinokur AD. Work hours and caseload as predictors of physician burnout: the mediating effects by perceived workload and by autonomy. *J Appl Psychol* 2010;59(4):539-65.

66. Smith F, Goldacre MJ, Lambert TW. Adverse effects on health and wellbeing of working as a doctor: views of the UK medical graduates of 1974 and 1977 surveyed in 2014. *J R Soc Med* 2017;110(5):198-207.

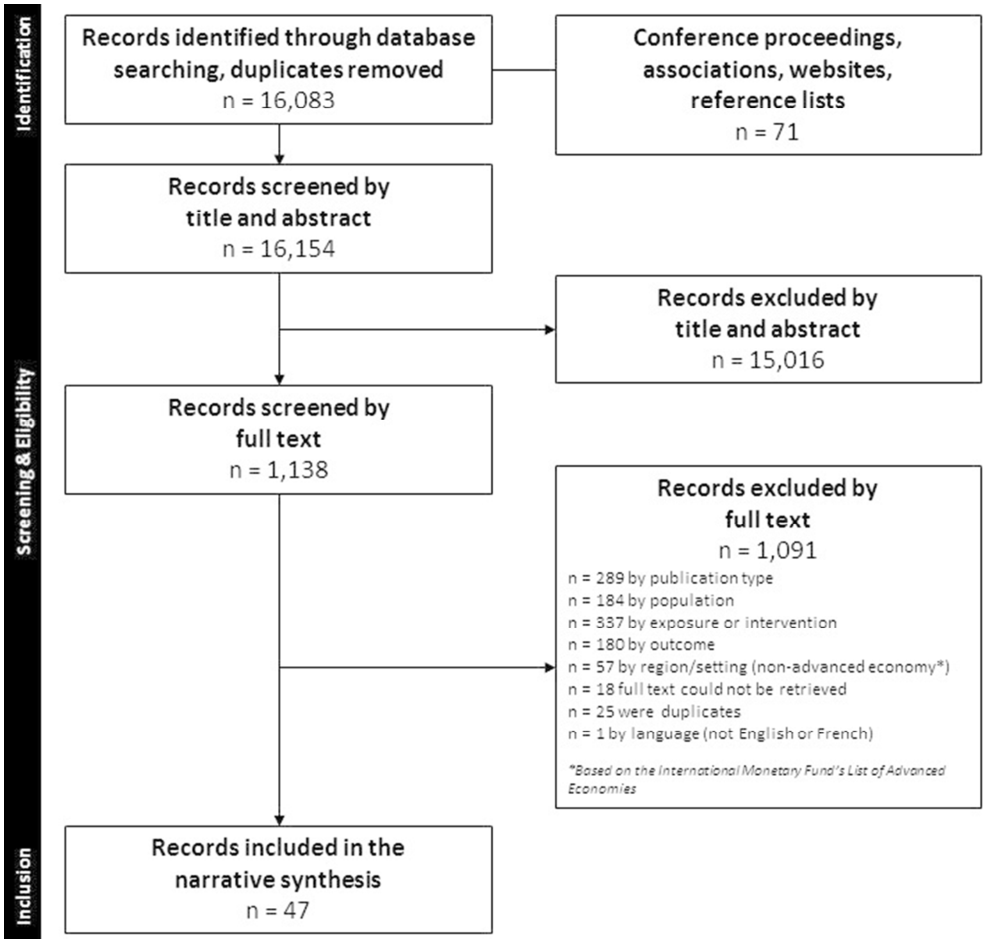
67. Starmer AJ, Frintner MP, Freed GL. Work-life balance, burnout, and satisfaction of early career pediatricians. *Pediatrics* 2016;e20153183. doi: 10.1542/peds.2015-3183

68. Tanti A, Camilleri M, Borg AA, et al. Opinions of Maltese doctors and pharmacists on medication errors. *Int J Saf Med* 2017;29(1-2):81-99.

69. Tokuda Y, Hayano K, Ozaki M, Bito S, Yanai H, Koizumi S. The interrelationships between working conditions, job satisfaction, burnout and mental health among hospital physicians in Japan: a path analysis. *Ind Health* 2009;47(2):166-72.

70. Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. *Am J Surg* 2005;189(6):753-7.
71. Vela-Bueno A, Moreno-Jimenez B, Rodriguez-Munoz A, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. *J Psychosom Res* 2008;64(4):435-42.
72. Vinden C, Nash DM, Rangrej J, et al. Complications of daytime elective laparoscopic cholecystectomies performed by surgeons who operated the night before. *Obstet Gynecol Surv* 2014;69(2):71-3.
73. Wada K, Yoshikawa T, Goto T, et al. National survey of the association of depressive symptoms with the number of off duty and on-call, and sleep hours among physicians working in Japanese hospitals: a cross sectional study. *BMC Public Health* 2010;10:127. doi: 10.1186/1471-2458-10-127
74. Govindarajan A, Urbach DR, Kumar M, et al. Outcomes of daytime procedures performed by attending surgeons after night work. *N Engl J Med* 2015;373(9):845-53.
75. Sturm L, Dawson D, Vaughan R, et al. Effects of fatigue on surgeon performance and surgical outcomes: a systematic review. *ANZ J Surg* 2011;81(7-8):502-9.
76. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet* 2009;374(9702):1714-21.
77. Canadian Medical Association. CMA policy: physician health. Ottawa, Canada: Canadian Medical Assocation, 2017. http://policybase.cma.ca/dbtw-wpd/Policypdf/PD18-01.pdf (accessed 12 Jan 2018).
78. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Intern Med* 2017;177(2):195-205.
79. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. *N Engl J Med* 2002;347(16):1249-55.
80. Ward S, Outram S. Medicine: in need of culture change. *Itern Med J* 2016;46(1):112-6.
81. Panagioti M, Panagopoulou E, Bower P, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med* 2017;177(2):195-205.
82. American Medical Association. STEPSforward™. Chicago, IL: American Medical Assocation, 2017. https://www.stepsforward.org/modules/joy-in-medicine (accessed 12 Jan 2018).
83. Jackson C. The Chalder Fatigue scale (CFQ 11). Occup Med 2015;65(1):86. doi: 10.1093/occmed/kqu168

84. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *J Psychosom Res* 1993;37(2):147-53.
85. Smets E, Garssen B, Bonke Bd, De Haes J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* 1995;39(3):315-25.
86. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383(9912):166-75.
87. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. *Implement Sci* 2016;11:38. https://doi.org/10.1186/s13012-016-0399-1



60x56mm (300 x 300 DPI)

Search Strategy

Database: In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

Date searched: 13 April 2016, updated 7 November 2017

Records retrieved: 5068 and 1442 in the update (removed duplicates retrieved in previous search)

1. Medical Staff, Hospital/
2. Physician Impairment/
3. exp Physicians/
4. allergist*.ti.
5. (an?esthetist* or an?esthesiologist*).ti.
6. cardiologist*.ti.
7. clinician*.ti.
8. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
9. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
10. dermatologist*.ti.
11. endocrinologist*.ti.
12. doctor*.ti.
13. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
14. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
15. family practitioner*.ti.
16. gastroenterologist*.ti.
17. (general practitioner* or GP*).ti.
18. (general adj2 physician*).ti.
19. geriatrician*.ti.
20. gyn?ecologist*.ti.

21. h?ematologist*.ti.

22. (health* adj2 (professional* or provider*)).ti.

23. hospitalist*.ti.

24. (house staff* or housestaff*).ti.

25. intensivist*.ti.

26. internist*.ti.

27. medical professional*.ti.

28. obstetrician*.ti.

29. oncologist*.ti.

30. ophthalmologist*.ti.

31. orthop?edist*.ti.

32. (otolaryngologist* or otorhinolaryngologist*).ti.

33. neonatologist*.ti.

34. nephrologist*.ti.

35. neurologist*.ti.

36. neuropsychiatrist*.ti.

37. neurosurgeon*.ti.

38. p?ediatrician*.ti.

39. perinatologist*.ti.

40. physician*.ti.

41. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

42. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.

43. primary care practitioner*.ti.

44. psychiatrist*.ti.

45. pulmonologist*.ti.

46. rheumatologist*.ti.

47. surgeon*.ti.

48. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

49. traumatologist*.ti.

50. urologist*.ti.

51. or/1-50 [Combined MeSH, title, and text word searches for physicians]

52. Burnout, Professional/

53. exp Circadian Rhythm/

54. exp Fatigue/

55. Occupational Health/

56. Rest/ph, px [Physiology, Psychology]

57. Sleep Deprivation/

58. Sleep Disorders, Circadian Rhythm/

59. Sleep Wake Disorders/

60. exp Stress, Psychological/

61. Workload/px [Psychology]

62. Work Schedule Tolerance/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw,kf.

64. biological rhythm*.tw,kf.

65. (burn out* or burned out* or burnt out* or burnout*).tw,kf.

66. circadian misalignment.tw,kf.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw,kf.

68. exhaust*.tw,kf.

69. fatigu*.tw,kf.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw,kf.

71. tired*.tw,kf.

72. weariness.tw,kf.

73. or/52-72 [Combined MeSH and text words for fatigue]

74. and/51,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (comment or editorial or letter).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. remove duplicates from 80

Database: Ovid Embase 1996 to 2016 Week 15

Date searched: 13 April 2016

Records retrieved: 8859

1. medical staff/
2. exp physician/
3. allergist*.ti.
4. (an?esthetist* or an?esthesiologist*).ti.
5. cardiologist*.ti.
6. clinician*.ti.
7. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
8. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
9. dermatologist*.ti.
10. endocrinologist*.ti.
11. doctor*.ti.
12. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
13. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
14. family practitioner*.ti.
15. gastroenterologist*.ti.
16. (general practitioner* or GP*).ti.
17. (general adj2 physician*).ti.

18. geriatrician*.ti.

19. gyn?ecologist*.ti.

20. h?ematologist*.ti.

21. (health* adj2 (professional* or provider*)).ti.

22. hospitalist*.ti.

23. (house staff* or housestaff*).ti.

24. intensivist*.ti.

25. internist*.ti.

26. medical professional*.ti.

27. obstetrician*.ti.

28. oncologist*.ti.

29. ophthalmologist*.ti.

30. orthop?edist*.ti.

31. (otolaryngologist* or otorhinolaryngologist*).ti.

32. neonatologist*.ti.

33. nephrologist*.ti.

34. neurologist*.ti.

35. neuropsychiatrist*.ti.

36. neurosurgeon*.ti.

37. p?ediatrician*.ti.

38. perinatologist*.ti.

39. physician*.ti.

40. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

41. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

42. primary care practitioner*.ti.

43. psychiatrist*.ti.

44. pulmonologist*.ti.

45. rheumatologist*.ti.

46. surgeon*.ti.

47. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

48. traumatologist*.ti.

49. urologist*.ti.

50. or/1-49 [Combined Emtree, title, and text word searches for physicians]

51. burnout/

52. circadian rhythm/

53. circadian rhythm sleep disorder/

54. fatigue/

55. mental stress/

56. occupational health/

57. sleep deprivation/

58. sleep waking cycle/

59. work capacity/

60. work schedule/

61. working time/

62. workload/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

64. biological rhythm*.tw.

65. (burn out* or burned out* or burnt out* or burnout*).tw.

66. circadian misalignment.tw.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw.

68. exhaust*.tw.

69. fatigu*.tw.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

71. tired*.tw.

72. weariness.tw.

73. or/51-72 [Combined Emtree and text words for fatigue]

74. and/50,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (conference* or editorial or letter or proceeding).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. limit 80 to embase

Database: Ovid PsycINFO 1987 to April Week 1 2016

Date searched: 13 April 2016

Records retrieved: 2094

1. exp Physicians/
2. allergist*.ti.
3. (an?esthetist* or an?esthesiologist*).ti.
4. cardiologist*.ti.
5. clinician*.ti.
6. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
7. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
8. dermatologist*.ti.
9. endocrinologist*.ti.
10. doctor*.ti.
11. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
12. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
13. family practitioner*.ti.
14. gastroenterologist*.ti.
15. (general practitioner* or GP*).ti.
16. (general adj2 physician*).ti.

17. geriatrician*.ti.

18. gyn?ecologist*.ti.

19. h?ematologist*.ti.

20. (health* adj2 (professional* or provider*)).ti.

21. hospitalist*.ti.

22. intensivist*.ti.

23. internist*.ti.

24. medical professional*.ti.

25. obstetrician*.ti.

26. oncologist*.ti.

27. ophthalmologist*.ti.

28. orthop?edist*.ti.

29. (otolaryngologist* or otorhinolaryngologist*).ti.

30. neonatologist*.ti.

31. nephrologist*.ti.

32. neurologist*.ti.

33. neuropsychiatrist*.ti.

34. neurosurgeon*.ti.

35. p?ediatrician*.ti.

36. perinatologist*.ti.

37. physician*.ti.

38. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

39. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

40. primary care practitioner*.ti.

41. psychiatrist*.ti.

42. pulmonologist*.ti.

43. rheumatologist*.ti.

44. surgeon*.ti.

45. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

46. traumatologist*.ti.

47. urologist*.ti.

48. or/1-47 [Combined thesaurus, title, and text word searches for physicians]

49. Compassion Fatigue/

50. Fatigue/

51. Human Biological Rhythms/

52. Occupational Health/

53. Occupational Stress/

54. Sleep/

55. Sleepiness/

56. Working Conditions/

57. Work Rest Cycles/

58. Work Week Length/

59. Work Scheduling/

60. Workday Shifts/

61. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

62. biological rhythm*.tw.

63. (burn out* or burned out* or burnt out* or burnout*).tw.

64. circadian misalignment.tw.

65. ((circadian or diurnam or ultradian) adj rhythm*).tw.

66. exhaust*.tw.

67. fatigu*.tw.

68. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

69. tired*.tw.

70. weariness.tw.

71. or/49-70 [Combined thesaurus and text words for fatigue]

72. and/48,71 [Combined concepts for physicians and fatigue]

73. limit 72 to yr="2000-Current"

74. limit 73 to (english or french)

Database: CINAHL Plus with Full Text (1937 to the present) via EBSCOhost

Date searched: 14 April 2016

Records retrieved: 3378

S1. (MH "Medical Staff, Hospital+")

S2. (MH "Physicians+")

S3. TI allertist*

S4. TI (anesthetist* or anaesthetist* or anesthesiologist* or anaesthesiologist*)

S5. TI cardiologist*

S6. TI clinician*

S7. clinician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S8. clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S9. TI dermatologist*

S10. TI endocrinologist*

S11. TI doctor*

S12. doctor* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S13. doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S14. TI "family practitioner*"

S15. TI gastroenterologist*

S16. TI ("general practitioner*" or GP*)

S17. TI (general N2 physician*)

S18. TI geriatrician*

S19. TI (gynaecologist* or gynecologist*)

S20. TI (haematologist* or hematologist*)

S21. TI hospitalist*

S22. TI ("house staff*" or housestaff*)

S23. TI intensivist*

S24. TI internist*

S25. TI obstetrician*

S26. TI oncologist*

S27. TI ophthalmologist*

S28. TI (orthopaedist* or orthopedist*)

S29. TI (otolaryngologist* or otorhinolaryngologist*)

S30. TI neonatologist*

S31. TI nephrologist*

S32. TI neurologist*

S33. TI neuropsychiatrist*

S34. TI neurosurgeon*

S35. TI (paediatrician* OR pediatrician*)

S36. TI perinatologist*

S37. TI physician*

S38. physician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S39. physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S40. TI "primary care practitioner*"

S41. TI psychiatrist*

S42. TI pulmonologist*

S43. TI rheumatologist*

S44. TI surgeon*

S45. surgeon* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S46. TI traumatologist*

S47. TI urologist*

S48. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28

OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47

S49. (MH "Circadian Rhythm")

S50. (MH "Fatigue")

S51. (MH "Impairment, Health Professional")

S52. (MH "Mental Fatigue")

S53. (MH "Occupational Health")

S54. (MH "Shiftwork")

S55. (MH "Sleep Deprivation")

S56. (MH "Sleep Disorders, Circadian Rhythm")

S57. (MH "Sleep-Wake Transition Disorders")

S58. (MH "Stress, Occupational+")

S59. (MH "Stress, Psychological")

S60. ("24 hour*" or "24 hr*" or "twenty four hour*" or "twentyfour hour*") N1 rhythm*

S61. "biological rhythm*"

S62. "burn out*" or "burned out*" or "burnt out*" or burnout*

S63. "circadian misalignment"

S64. (circadian or diurnam or ultradian) N1 rhythm*

S65. exhaust*

S66. fatigu*

S67. sleep* N3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)

S68. tired*

S69. weariness

S70. S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69

S71. S48 AND S70

S72. S48 AND S70 Limiters - Published Date: 20000101-20161231; Publication Type: Clinical Trial, Journal Article, Meta Analysis, Meta Synthesis, Practice Guidelines, Randomized Controlled Trial, Research, Review, Systematic Review; Language: English, French

Database: PubMed via NCBI Entrez

Date searched: 14 April 2016

Records retrieved: 92

(((("Medical Staff, Hospital"[mh:noexp] OR "Physician Impairment"[mh:noexp] OR "Physicians"[mh] OR allergist[ti] OR allergists[ti] OR anaesthetist[ti] OR anaesthetists[ti] OR anaesthesiologist[ti] OR anaesthesiologists[ti] OR anesthetist[ti] OR anesthetists[ti] OR anesthesiologist[ti] OR anesthesiologists[ti] OR cardiologist[ti] OR cardiologists[ti] OR clinician[ti] OR clinicians[ti] OR ((clinician[tiab] OR clinicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((clinician[tiab] OR clinicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR dermatologist[ti] OR dermatologists[ti] OR endocrinologist[ti] OR endocrinologists[ti] OR doctor[ti] OR doctors[ti] OR ((doctor[tiab] OR doctors[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((doctor[tiab] OR doctors[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "family practitioner"[ti] OR

"family practitioners"[ti] OR gastroenterologist[ti] OR gastroenterologists[ti] OR "general practice physician"[ti] OR "general practice physicians"[ti] OR "general practitioner"[ti] OR "general practitioners"[ti] OR geriatrician[ti] OR geriatricians[ti] OR gynaecologist[ti] OR gynaecologists[ti] OR gynecologist[ti] OR gynecologists[ti] OR haematologist[ti] OR haematologists[ti] OR hematologist[ti] OR hematologists[ti] OR "health care professional"[ti] OR "health care professionals"[ti] AND "health care provider"[ti] OR "health care providers" OR "health professional"[ti] OR "health professionals"[ti] OR "health provider"[ti] OR "health providers"[ti] OR "healthcare professional"[ti] OR "healthcare professionals"[ti] OR "healthcare provider"[ti] OR "healthcare providers"[ti] OR hospitalist[ti] OR hospitalists[ti] OR "house staff"[ti] OR "house staffs"[ti] OR housestaff[ti] OR housestaffs[ti] OR intensivist[ti] OR intensivists[ti] OR internist[ti] OR internists[ti] OR "medical professional"[ti] OR "medical professionals"[ti] OR obstetrician[ti] OR obstetricians[ti] OR oncologist[ti] OR oncologists[ti] OR ophthalmologist[ti] OR ophthalmologists[ti] OR orthopaedist[ti] OR orthopaedists[ti] OR orthopedist[ti] OR orthopedists[ti] OR otolaryngologist[ti] OR otolaryngologists[ti] OR otorhinolaryngologist[ti] OR otorhinolaryngologists[ti] OR neonatologist[ti] OR neonatologists[ti] OR nephrologist[ti] OR nephrologists[ti] OR neurologist[ti] OR neurologists[ti] OR neuropsychiatrist[ti] OR neuropsychiatrists[ti] OR neurosurgeon[ti] OR neurosurgeons[ti] OR paediatrician[ti] OR paediatricians[ti] OR pediatrician[ti] OR pediatricians[ti] OR perinatologist[ti] OR perinatologists[ti] OR physician[ti] OR physicians[ti] OR ((physician[tiab] OR physicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((physician[tiab] OR physicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "primary care practitioner"[ti] OR "primary care practitioners"[ti] OR psychiatrist[ti] OR psychiatrists[ti] OR pulmonologist[ti] OR pulmonologists[ti] OR rheumatologist[ti] OR rheumatologists[ti] OR surgeon[ti] OR surgeons[ti] OR ((surgeon[tiab] OR surgeons[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR

"burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR traumatologist[ti] OR traumatologists[ti] OR urologist[ti] OR urologists[ti]) AND ("Burnout, Professional"[mh:noexp] OR "Circadian Rhythm"[mh] OR "Fatigue"[mh] OR "Occupational Health"[mh:noexp] OR "Rest/physiology"[mh:noexp] OR "Rest/psychology"[mh:noexp] OR "Sleep Deprivation"[mh:noexp] OR "Sleep Disorders, Circadian Rhythm"[mh:noexp] OR "Stress, Psychological"[mh] OR "Workload/psychology"[mh] OR "Work Schedule Tolerance"[mh:noexp] OR "24 hour rhythm"[tiab] OR "24 hour rhythms"[tiab] OR "24 hr rhythm"[tiab] OR "24 hr rhythms"[tiab] OR alertness[tiab] OR "biological rhythm"[tiab] OR "biological rhythms"[tiab] OR "burn out"[tiab] OR "burned out"[tiab] OR "burnt out"[tiab] OR burnout[tiab] OR "circadian misalignment"[tiab] OR "circadian rhythm"[tiab] OR "circadian rhythms"[tiab] OR "diurnal rhythm"[tiab] OR "diurnal rhythms"[tiab] OR exhausted[tiab] OR exhaustion[tiab] OR exhausting[tiab] OR exhausts[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR (("Sleep"[mh:noexp] OR sleep[tiab] OR sleeping[tiab]) AND (deprivation[tiab] OR deprive[tiab] OR deprived[tiab] OR deprives[tiab] OR depriving[tiab] OR disorder[tiab] OR disorders[tiab] OR lack[tiab] OR lacked[tiab] OR lacking[tiab] OR lacks[tiab] OR loss[tiab] AND insufficient[tiab] OR problem[tiab] OR problems[tiab])) OR tired[tiab] OR tiredness[tiab] OR "twenty four hour rhythm"[tiab] OR "twenty four hour rhythms"[tiab] OR weariness[tiab] OR "ultradian rhythm"[tiab] OR "ultradian rhythms"[tiab])) NOT (((Animals[MESH] OR Animal Experimentation[MESH] OR "Models, Animal"[MESH] OR Vertebrates[MESH]) NOT (Humans[MESH] OR Human experimentation[MESH])) OR (((animals[tiab] OR animal model[tiab] OR rat[tiab] OR rats[tiab] OR mouse[tiab] OR mice[tiab] OR rabbit[tiab] OR rabbits[tiab] OR pig[tiab] OR pigs[tiab] OR porcine[tiab] OR swine[tiab] OR dog[tiab] OR dogs[tiab] OR hamster[tiab] OR hamsters[tiab] OR chicken[tiab] OR chickens[tiab] OR sheep[tiab]) AND (publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb])) NOT (human[ti] OR humans[ti] OR people[ti] OR children[ti] OR adults[ti] OR seniors[ti] OR patient[ti] OR patients[ti])))) NOT (editorial[pt] OR comment[pt] OR letter[pt] OR newspaper article[pt])) AND ((publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR (pubstatUSheadofprint))

Filters activated: Publication date from 2000/01/01 to 2016/12/31, English, French.

Supplementary table 1. Descriptive characteristics of the included studies

Study	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
Country	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Observational (exposure) studies (n=45)								
Cohort design								
Chu, 2011 [33] Canada	Surgeons	6	NR	Range: 32-55y	Tertiary care academic hospital	Urban	Sleep deprivation due to work on the night preceding surgery	Surgeon operative efficiency; patient postoperative mortality, adverse outcomes, length of stay
	Patients: cardiac surgery cases	4,047	NR	NR				
Ellman, 2004 [36] US	Surgeons	NR	NR	NR	University hospitals	Urban	Sleep deprivation due to work on the night preceding surgery	Surgeon operative efficiency; patient complications, in-hospital mortality, length of stay, need for blood products
	Patients: adult cardiac surgery cases	6,751	70%	S: 63.4±0.7y C: 63.5±0.1y				
Govindarajan, 2015 [74] Canada	Surgeons	1,448	NR	46.3±8.7	Academic and non-academic hospitals	Mixed	Sleep deprivation due to work on the night preceding a daytime surgery	Duration of surgery; Patient complications, mortality, readmissions, length of stay
	Patients: surgical cases	38,978	NR	56.4±16.6y				
Rothschild, 2009 [53] US	Surgeons Obstetrician/gynecologists	220	Surgeons: 84% OB/GYNs: 28%	Surgeons: 42.0±7.6y OB/GYNs: 42.0±9.0y	Tertiary care academic trauma centre/referral centre for high-risk obstetrics	Urban	Sleep deprivation due to work on the night preceding a daytime procedure	Patient complications, preventable complications
	Patients: surgical and obstetrics cases	Surg.: 4,471 Obst.: 4,902	Surg: S: 25% C: 28% Obst.: S: 0% C: 0%	Surg: S: 49.1±16.3y C: 50.0±16.3y Obst.: S: 32.9±5.2y C: 33.5±5.0y				
Schieman, 2007 [58] Canada	Colorectal surgeons	NR	NR	NR	University teaching hospitals	NR	Fatigue due to work on the night preceding surgery	Patient operative complications, length of stay, mortality, cancer recurrence
	Patients: undergoing anterior resection for rectal cancer	270	NR	S: 64.5y C: 64.4y				

Study	Physician and patient characteristics				Setting		Interventions or	Outcomes
Country	Type	n=	Sex (% male)	Age	Location	Urban or rural	exposures	
Vinden, 2014 [72] Canada	General surgeons	331	83%	48±10y	Community hospitals	Mixed	Sleep deprivation due to overnight work preceding daytime surgery	Patient mortality, operative complications
	Patients: Elective cholecystectomies	10,390	S: 27% C: 26%	S: 49±16y C: 49±16y				
Before-after design								
Amirian, 2014 [28] Denmark	Surgeons	29	55%	Median: 35y Range: 27-49y	Academic hospital	Urban	17-h night shift with sleep deprivation	Cognitive and psychomotor abilities on a laparoscopic simulation
Gerdes, 2008 [40] US	Surgeons	9	NR	NR	University Hospital	Urban	Fatigue; sleep deprivation from overnight call shift	Cognitive and psychomotor abilities
Lederer, 2006 [45] Austria	Senior anesthetists	11	82%	49.0±2.0y	Hospital	Urban	Sleep deprivation from 24-h call shift	Concentration ability; reaction time; performance on psychometric tasks
Time series design								
Leichtfried, 2011 [46] Austria	Anesthetists	10	100%	Mean: 32y Range: 29-35y	University Hospital	Urban	Sleep deprivation from 24-h shift; sleepiness, sleep hours	Melatonin metabolite profile
Cross-sectional design								
Aziz, 2004 [29] US	Family medicine physicians Various specialties	153	NR	NR	Hospitals	NR	Fatigue	Stress
Beaujouan, 2005 [30] France	Anesthesiologists	3,476	64%	≤35y: 9% 36-45y: 28% 46-55y: 49% 56-65y: 13%	Public sector General hospitals University hospitals Private hospitals	NR	Sleep deprivation	Substance abuse
Chang, 2013 [31] US	Anesthesiologists	11	64%	Mean: 38y IQR: 34-48y	Level 1 trauma centre	NR	Sleep deprivation due to 15-h overnight call shift; sleepiness	Cognitive performance; reaction time
Chen, 2008 [32] US	Psychiatrists Internists	180	77%	Academic: 79% 36-55y	Medical school Private practices	Urban	Sleep deprivation; sleepiness	Impact on personal and professional life;

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
	General practitioners Surgeons Obstetrician-gynecologists Radiologists Pediatricians Other			Private practice: 73% 36-65y				perceived risk of errors
Doppia, 2011 [34] France	Anesthesiologists	565	64%	<35y: 11% 35-54y: 63% >55y: 25%	Public hospitals Private hospitals Work-health environments Public health units	NR	Sleep deprivation	Burnout
Elovaino, 2015 [37] Finland	Physicians in various specialties	1,524	40%	Median: 49.7y Range: 24-69y	Hospitals Primary care Private practice Other unspecified	NR	Sleep difficulties	Job demands and control
Gander, 2000 [39] New Zealand	Anesthetists	183	NR	Mean: 46y	Combined public/private practice Other unspecified	NR	Work-related sleep disturbance	Risk of fatigue-related errors
Harbeck, 2015 [41] Germany	Internists	20	45%	Median: 32y Range: 26-42y	Hospital	NR	Sleep disturbance due to a 24-call shift	Biochemical and physiological parameters; neurocognitive function
Heponiemi, 2014 [42] Finland	Physicians in various specialties Non-specialized physicians	1,541	40%	49.80±9.49y, Range: 24-67y	Hospitals Primary care clinic Private practice Other unspecified	NR	Sleep difficulties	Job satisfaction; work ability; psychological distress
Jackson, 2017 [43] US	Surgeons in various subspecialties	993	61%	More; less satisfied: 30-39y: 23%;24% 40-49y: 32%;36% 50-59y:	Academic practice Non-academic practice	NR	Not feeling well rested	Job satisfaction

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
				23%;27% ≥60y: 23%;14%				
Kanieta, 2011 [44] Japan	Internists Surgeons Orthopedics Pediatricians Obstetrician-gynecologists Psychiatrists Dermatologists Urologists Opthalmologists Otorhinolaryngologists Other	3,486	66%	20-39y: 11% 40-49y: 25% 50-59y: 28% 60-69y: 16% ≥70y: 21%	Hospitals Clinics Other unspecified	NR	Sleep deprivation and difficulties; insomnia	Medical incidents
Lindfors, 2006 [47] Finland	Anesthetists	328	53%	47±7.8y Range: 32-69y	University hospitals Central and district hospitals Private sector	NR	Sleep disturbances; sleepiness	Stress; suicidal tendencies
Mahmood, 2016 [48] Norway	Generalists Internists Pediatricians Surgical specialties Anesthesiologists	450 (all time points)	41%	43y±2.8y	Public health system Private practice	NR	Sleep deprivation due to on-call shifts	Alcohol misuse
Nishimura, 2014 [49] Japan	Neurosurgeons and neurologists	2,564	NR	NR	Stroke care centres Teaching hospitals	NR	Sleep deprivation	Burnout
Pit, 2014 [50] Australia	General practitioners	92	60%	50±10.7y	NR	Rural	Work-related sleep disturbance	Early retirement intentions
Pit, 2016 [51] Australia	General practitioners	92	60%	50±10.7y	Private (solo) practice Group practice	Rural	Work-related sleep disturbance	Sickness presenteeism
Roberts, 2014 [52] US	General internists Internal medicine	578	58%	Hospitalists: 46.9±12.4y	Private practice Academic medical	NR	Fatigue	Falling asleep while driving

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
	hospitalists			Generalists: 53.6±10.2y	centre Veterans hospital Military practice Other			
Saadat, 2016 [55] US	Anesthesiologists	21	71%	30-40y: 57% 41-50y: 19% 51-55y: 24% Range: 32-56y	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Mood disturbances
Saadat, 2017 [54] US	Anesthesiologists	21	65%	Range: 32-56 years	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Reaction time
Sanches, 2015 [56] Spain	Emergency medicine physicians	18	28%	29.2±2.6y	Central hospital	NR	Sleep deprivation	Cognitive and psychomotor abilities
Sargent, 2009 [57] US	Orthopedic surgeons	264	92%	NR	Orthopedic surgery training programs	NR	Sleep deprivation	Burnout; psychological distress; marital satisfaction
Sende, 2012 [59] France	Emergency physicians	318	62%	39±8y	Hospitals Mobile emergency services Other unspecified	NR	Fatigue; sleep deprivation	Stress
Sexton, 2001 [60] US	Consulting physicians: Surgeons Anesthesiologists Pulmonary physicians Cardiologists Pediatricians	271	NR	NR	Teaching and non-teaching hospitals	Urban	Fatigue	Perceived performance effectiveness
Shanafelt, 2005 [62] US, Canada, Mexico	Oncologists	241	85%	>50y: 51%	Community clinics Hospitals Private practice Academic medical centres	NR	Fatigue; sleep deprivation	Quality of life/well-being
Shanafelt, 2010	Surgeons	7,905	87%	Median: 51y	Private practice	NR	Fatigue	Perceived major medical

Study Country	Physician and patient characteristics Type	 n=	 Sex (% male)	 Age	Setting Location	 Urban or rural	Interventions or exposures	Outcomes
[61] US				Q1: 43y Q2: 59y	Academic medical centres Veterans hospital Active military practice Retired or not in practice Other			errors
Shanafelt, 2014 [63] US	Oncologists	1,117	52%	Median: 52y	Private practice Academic practice Veteran's hospital Industry, other	NR	Fatigue	Satisfaction with work-life balance
Shirom, 2006 [64] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Perception of quality of patient care
Shirom, 2010 [65] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Burnout
Smith, 2017 [66] UK	General practitioners Surgeons Other unspecified specialties	3,550	63%	NR	NR (varied)	NR	Perceived fatigue, sleep deprivation	Physical and mental health; competence
Starmer, 2016 [67] US	General pediatricians Pediatric surgeons Pediatric hospitalists Pediatric specialists (unspecified)	840	40%	NR	NR (some in private practice)	NR	Sleep deprivation	Burnout; balanced personal and professional commitments; life and career satisfaction
Tanti, 2017 [68]	Physicians (unspecified)	204	62%	Median: 41y	Hospitals	NR	Fatigue	Prescribing errors

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Malta					Community Office-based			
Tokuda, 2009 [69] Japan	Hospital physicians: Generalists Other unspecified specialties	236	75%	40.9±7.8y Range: 26-76y	Hospitals with ≥20 inpatient beds	NR	Sleep deprivation	Burnout; job satisfaction
Vela-Bueno, 2008 [71] Spain	Primary care physicians	113	27%	41.4±8.0y	Primary care centres	Urban	Sleep problems, insomnia	Burnout
Wada, 2010 [73] Japan	Physicians (unspecified)	3,862	78%	M: 75% 30-59y F: 85% 30-59y	Hospitals	NR	Sleep deprivation	Depressive symptoms
Non-comparative design								
Gander, 2008 [38] New Zealand	Anesthetists	20	85%	Median: 44y	Hospitals	Urban	Sleep disturbance from consecutive working days or on-call work	Psychomotor performance
Intervention studies (n=2)								
Randomized controlled trials								
Dutheil, 2013 [35] France	Emergency physicians	17	35%	39.1y±6.9y	University hospital	Urban	Fatigue related to 14-h and 24-h shifts; sleep deprivation; low sleep quality;	Perceived stress; urine interleukine-8
Uchal, 2005 [70] Norway	Surgeons Gynecologists Orthopedic surgeons Urologists Vascular surgeons	64	67%	Median: Post-call: 33.0y Post-work: 38.0y	Government hospitals	NR	Sleep deprivation due to 24-h call shift	Product quality, procedure effectiveness of a surgical simulation

C: control group; F: female; h: hour(s); IQR: interquartile range; M: male; NR: not reported; S: study group; SD: standard deviation; Surg: surgical; Obst: obstetric; Q: quartile; UK: United Kingdom; US: United States of America; y: year(s)

Supplementary table 2. Summary of risk of bias assessments for randomized controlled trials (n=2)[a]

First Author, Year	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias	Overall risk of bias[b]
Dutheil, 2013	Low	Unclear	High	High	Low	Low	High	High
Uchal, 2005	Low	Low	Unclear	Low	Low	Low	Low	Unclear

[a]Assessed using the Cochrane Collaboration's Risk of Bias Tool

[b]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Supplementary table 3. Summary of quality assessments for cohort studies (n=6)[a]

First Author, Year	Selection					Comparability		Outcome				Total Score[b] /9
	Representa-tiveness of exposed cohort /1	Selection of non-exposed cohort /1	Ascertain-ment of exposure /1	Outcome not present at start /1	Total /4	Compara-bility of cohorts /2	Total /2	Assess-ment of outcome /1	Adequate length of follow-up /1	Adequate follow-up of cohorts /1	Total /1	
Chu, 2011	1	1	0	1	3	2	2	1	1	1	3	8
Ellman, 2004	1	1	1	1	4	1	1	1	1	1	3	8
Govindarajan, 2015	1	1	1	1	4	2	2	1	1	1	3	9
Rothschild, 2009	1	1	1	1	4	2	2	1	1	1	3	9
Schieman, 2008	1	1	1	1	4	1	1	1	1	1	3	8
Vinden, 2014	1	1	1	1	4	1	1	1	1	1	3	8

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale

[b]An overall score of 7 to 9 stars is considered as low risk of bias, 4 to 6 as unclear risk of bias, and 3 or less as high risk of bias

Supplementary table 4. Summary of risk of bias assessments for before-after studies (n=3)[a]

First Author, Year	Random sequence generation[b]	Allocation concealment[b]	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Amirian, 2014	NA	NA	High	High	Low	Low	High	High
Gerdes, 2008	NA	NA	High	High	Low	Low	High	High
Lederer, 2006	NA	NA	High	High	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for before-after studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as 'not applicable' (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Supplementary table 5. Summary of risk of bias assessments for time series studies (n=1)[a]

First Author, Year	Intervention independent of other changes	Intervention effect pre-specified	Intervention unlikely to affect data collection	Allocation concealment[a]	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Leitchfried, 2011	Low	High	Low	NA	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for interrupted time series studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as not applicable (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Supplementary table 6. Summary of quality assessments for cross-sectional studies (n=34)[a]

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Aziz, 2004	0	0	0	0	1	0	1	1
Beaujouan, 2005	1	0	1	0	1	0	1	2
Chang, 2013	1	0	1	0	1	1	2	3
Chen, 2008	1	0	1	0	1	0	1	2
Doppia, 2011	1	1	2	0	1	1	2	4
Elovaino, 2015	1	1	2	0	1	1	2	4
Gander, 2000	1	1	2	0	1	1	2	4
Harbeck, 2015	1	0	1	0	1	1	2	3
Heponiemi, 2014	1	1	2	0	1	1	2	4
Jackson, 2017	0	0	0	0	1	0	1	1
Kanieta, 2011	1	0	1	0	1	1	2	3
Lindfors, 2006	1	1	2	0	1	1	2	4
Mahmood, 2017	1	0	1	0	1	0	1	2
Nishimura, 2014	1	1	2	0	1	0	1	3
Pit, 2014	1	0	1	0	1	1	2	3
Pit, 2016	1	0	1	0	1	1	2	3
Roberts, 2014	1	1	2	0	1	0	1	3
Saadat, 2016	1	1	2	0	1	1	2	4
Saadat, 2017	1	1	2	0	1	1	2	4
Sanches, 2015	1	0	1	0	1	0	2	3
Sargent, 2009	1	0	1	0	1	0	1	2

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Sende, 2010	1	0	1	0	1	0	1	2
Sexton, 2001	1	0	1	0	1	0	1	2
Shanafelt, 2005	1	0	1	0	1	1	2	3
Shanafelt, 2010	1	0	1	0	1	0	1	2
Shanafelt, 2014	1	0	1	0	1	1	2	3
Shirom, 2006	1	1	2	0	1	1	2	4
Shirom, 2010	1	1	2	0	1	1	2	4
Smith, 2016	1	0	1	0	1	1	2	3
Starmer, 2016	1	1	2	0	1	1	2	4
Tanti, 2017	1	0	1	0	1	0	1	2
Tokuda, 2009	1	1	2	0	1	1	2	4
Vela-Bueno, 2008	1	1	2	0	1	1	2	4
Wada, 2010	1	1	2	0	1	0	1	3

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted for cross-sectional studies

[b]An overall score of 4 to 5 stars is considered as low risk of bias, 3 as unclear risk of bias, and 2 or less as high risk of bias. For response rate, ≥50% was used as the criterion to be awarded a star

Supplementary table 7. Summary of quality assessments for non-comparative studies (n=1)[a]

First Author, Year	Selection			Exposure		Outcome				Total Score[b] /6
	Adequacy of case definition /1	Representativeness of the sample /1	Total /2	Ascertainment of exposure	Total /1	Assessment of outcome /1	Same method of assessment for entire sample /1	Loss to follow-up /1	Total /3	
Gander, 2008	1	1	2	0	0	0	1	1	2	4

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted by the authors to be suitable to the non-comparative design

[b]An overall score of 5 to 6 stars is considered as low risk of bias, 3 to 4 as unclear risk of bias, and 2 or less as high risk of bias

Supplementary table 8. Physician health and wellness outcomes and associations with fatigue

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Exposures or interventions: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Jackson, 2017 RoB: high	CS	Not feeling well rested: self-reported as 'unhealthy' Time points NR	71% healthy, 28% unhealthy in terms of being well rested	Job satisfaction: Abridged Job in General Scale; grouped into more or less satisfied using the median Time points NR	Job satisfaction in those more vs. less satisfied: Healthy (well rested): 85% vs. 58%, p<0001; Unhealthy (not well rested): 15% vs. 42%, p<0.001.
Nishimura, 2014 RoB: unclear	CS	Sleep hours/night: self-reported (continuous) Time points NR	Mean±SD sleep: 5.94±1.08h	Burnout: Japanese MBI (severe: EE >4.0 and either DP >2.6 or PE <4.17) Time points NR	1) Mean±SD sleep for not burned out vs. mild to moderate vs. severe: 6.07±1.15 vs. 5.88±0.94 vs. 5.63±0.94, p<0.05; 2) Association between sleep and burnout (OR (95% CI)): bivariate 0.67 (0.61-0.73), p<0.001; multivariate including work characteristics and mental health: 0.84 (0.75-0.94), p=0.002.
Sargent, 2009 RoB: high	CS	Sleep deprivation: self-reported on a 4-point scale (none, a little, quite a bit, a lot) Time points NR	21% none, 48% a little, 23% quite a bit, 8% a lot	Burnout: MBI (norms NR); Marital satisfaction: RDAS; Psychological morbidity: GHQ-12 score ≥4 Time points NR	1) Positive correlation between sleep deprivation and EE, DP, psychological distress, lower marital satisfaction, all p<0.001. No relationship with PA.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift with on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Stress during duty: 4-point scale from 'calm' to 'very demanding' Assessed post-duty	1) Mean stress score during duty: 2.1.
Leitchtfried, 2011 RoB: high	TS	24-h shift; Sleepiness: ESS (range: 0-24); Sleep hours: self-reported	ESS (mean (range)): 7.4 (4-12); Mean±SD sleep hours: 1) pre-study: 7.74±1.35h; 2) Pre-24-h shift (11h00 on day	aMT6-s: urinalysis Assessed at 4-h intervals from 07:00 to 11:00	1) aMT6-s over shift, mean (95% CI): higher at 11:00AM pre- (12.2 (6.3-8.1)) and post-shift (9.3 (3.7-14.9)) vs. during, p=0.016; 2) Correlations between sleep and aMT6-s (data

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		(continuous) Sleepiness assessed pre-shit, sleep hours pre, during and post-shift	1: 0.13±0.35h, 19:00 on day 1: 6.99±0.68h); 3) During the 24-h shift (07h00 on day 2: 0.0±0.0h, 19h00 on day 2, 5.49±1.95h); 4) Post-24-h shift (11h00 on day 3: 0.5±0.71h, 19h00 on day 3: 7.06±1.18h).		NR): mild for sleep duration the night prior with aMT6-s at 3PM the following day; sleep on night 2 with aMT6-s at 3PM the next day; total sleep with aMT6-s at 11AM on third day; moderate for sleep on first night with aMT6-s at 7AM and 11AM pre-shift, 11PM during 24-h shift and 11AM post-shift; total sleep pre-shift and nocturnal sleep during 24-h shift with aMT6-s at 11PM during shift; total sleep with aMT6-s at 3PM on first and second day, 11PM on second day; 3) Correlations between ESS and aMT6-s: moderate for aMT6-s at 7AM during shift, 11AM on day off.
Beaujouan, 2005 RoB: high	CS	Sleep deprivation: 4-point scale (always, frequently, rarely, never) Time points NR	48.8% always or frequently feel sleep deprived	Substance abuse: 93-item addiction and substance abuse questionnaire Time points NR	1) 60.6% with drug dependence vs. 46.0% of those without reported sleep difficulties, p<0.001. 2) OR (95% CI) of addiction for frequently/always vs. rarely/never sleep deprived: tobacco 1.42 (1.04-1.94); tranquilizer/hypnotics 3.26 (2.12-5.02).
Doppia, 2011 RoB: low	CS	Insufficient sleep: 4-point scale (no, not really, sort of, yes) Time points NR	28.9% reported insufficient sleep during work time	Burnout: CBI (mild: 1-2.4, moderate: 2.5-3.5, severe: 3.6-5) Time points NR	1) Frequency of burnout by response for sleep sufficiency: 47.6% for no/not really, 16.3% for sort of/yes, p<0.001.
Lindfors, 2006 RoB: low	CS	Sleep hours/day: self-reported to the nearest 0.5h; Adequacy of sleep and rest: self-reported (yes/no) Time points NR	Sleep hours (mean (range)): 7 (5-9)	Stress: MOSQ on a 3-point scale (no, to some extent, clearly); Thoughts of suicide: 4-point scale ('never' to 'have tried') Time points NR	1) Sleep sufficiency predicted stress symptoms: bivariate β=-0.362, p<0.001; multivariate including gender, sick leave, suicide β=-0.269, p<0.00; 2) Sleep disturbance associated with thoughts of suicide, p=0.009.
Saadat, 2015	CS	Sleep deprivation (<7h/24-h) due to 17-h overnight shift;	Mean±SD sleepiness on a regular day vs. post-call day:	Simple cognitive tests: VAS from 0 (not at all) to 100	Regular day v. post-call day, mean±SD scores: 1) Simple cognitive tests: energetic 6.04±2.27 vs.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
RoB: low		Sleepiness and alertness: VAS from 0 (not at all) to 100 (extremely) All assessed on a regular day and a post-call day	2.99±2.18 vs. 6.79±2.30, p<0.001	(extremely); Mood disturbance: PMS (scoring NR) All assessed on a regular day and a post-call day	2.53±1.87, confident 7.03±1.83 vs. 4.98±2.29, irritable 2.03±1.94 vs. 4.86±2.16, sleepy 2.99±2.18 vs. 6.79±2.30, talkative 4.46±1.74 vs. 2.41±1.97, all p<0.001; jittery 1.44±1.74 vs. 3.12±2.34, p=0.003; anxiousness ns; 2) PMS: tension 13.48±2.71 vs. 15.43±4.46, p=0.049; anger 15.24±4.41 vs. 18.14±5.92, p=0.005; fatigue 10.14±2.63 vs. 20.05±6.87, p<0.001; confusion 10.57±1.69 vs. 12.57±4.24, p=0.025; vigor 24.05±6.75 vs.16.67±5.70, p<0.001; depression: ns; total mood disturbance: 42.57±15.26 vs. 70.90±6.91, p<0.001.
ER or ICU physicians					
Dutheil, 2013 RoB: high	RCT	14-h or 24-h shift; Sleep hours: self-reported sleep and wake time; Sleep quality: VAS from 1 (low) to 100 (high); Mental and physical fatigue: VAS from 1 (low) to 100 (high) Assessed on day prior to shift; during shift; each day of protocol (work, off, clerical, control)	1) Sleep duration and quality lower during shifts (14h and 24h) than any other day, and lower during the 24-h vs. 14-h shift (p<0.05); 2) Mental and physical fatigue higher after 14-h and 24-h shift vs. control day (data NR).	Stress: VAS from 0 (low) to 100 (high); IL-8: urinalysis Assessed at 08:30 and 18:30 on each day of protocol	1) Stress: higher following 14-h and 24-h shifts vs. the control day, p<0.05 (data NR); 2) IL-8: higher following 24-h shift vs. control (p=0.007) and 14-h shift (p=0.015); ns difference between 14-h shift and control day; 3) Correlations with IL-8: sleep hours pre-24-h shift, r=-0.627, p=0.007; poor sleep quality during 14-h and 24-h shifts, r=0.452, p=0.031; 4) Multivariable regression: 24-h shift increased IL-8 by 1.9ng vs. control day, p=0.007; ns association with 14-h shift, mental or physical fatigue, sleep deprivation, 14-h shift.
Sende, 2012 RoB: high	CS	Fatigue and sleep deprivation as sources of stress Time points NR	NR	Most important sources of stress among 4 categories (work-related, patient-related, organizational, individual) Time points NR	1) 78% indicated that sleep loss and fatigue were sources of stress.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Generalists[b]					
Harbeck, 2015 RoB: unclear	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Biochemical (laboratory values) and physiological (heart rate variability, skin resistance, blood pressure) stress parameters Assessed before a normal day shift, and after a 24-h on call shift	Before a normal shift vs. after overnight call shift: 1) Biochemical parameters: no changes in any parameter except for thyroid stimulating hormone which was higher after the on-call shift (p = 0.049, data NR); 2) Physiological parameters: no significant changes in any parameter
Pit, 2014 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day' Time points NR	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Early retirement (<65 years) intentions (yes/no) Time points NR	For sleep disturbance a few times a year to every day vs. never: 1) Intention to retire early: 74% vs. 26%, p<0.01; 2) Association with intention to retire early (OR (95% CI)): univariate 3.6 (1.47-8.80), p<0.01; multivariate including work, occupational, individual factors 2.91 (1.11-7.6), p<0.05; 4) RR (95% CI) for intention to retire early: 2.0 (1.18-3.49); attributable fraction: 50.0%; population attributable fraction: 37.1%.
Pit, 2016 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day' Time points NR	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Sickness presenteeism: 'yes' response indicated 1 or more days Assessed for the past 12 months	For sleep disturbance a few times a year to every day vs. never: 1) Sickness presenteeism: 32% vs. 68%, p=0.018; 2) Association with sickness presenteeism (OR (95% CI)): 2.92 (1.19-7.16), p=0.02.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Roberts, 2014 RoB: unclear	CS	Fatigue: LAS from 0 (low) to 10 (high) Assessed for the past week	Mean (SD) score: 5.8 (2.4) for hospitalists; 5.9 (2.4) for general internists	Impact of fatigue on daily activities (falling asleep while driving) (yes/no) Time points NR	1) 8.7% of hospitalists and 4.3% of outpatient general internists had fallen asleep while driving due to fatigue.
Vela-Bueno, 2008 RoB: low	CS	Sleep Quality: PSQI (Spanish): score ≥5 indicates low quality (range; 0 to 21); Insomnia: DSM-IV criteria Time points NR; insomnia symptoms in past month	Prevalence (% (95% CI)): 1) Sleep-onset latency >30 minutes: 8.4 (4.8-11.9); 2) Wake time after sleep onset >30 minutes: 15.4 (10.8-19.9); 3) Early morning awakening: 22.5 (19.5-30.4); 4) Nonrestorative sleep: 22.5 (17.2-27.7); 5) Daytime impairment for ≥5 days in past month: 14.2 (9.7-18.6); 6) Insomnia: 18.8 (13.8-23.7).	Burnout: PBM with a 7-point scale from 1 (never) to 7 (always) Time points NR	Low vs. high burnout, mean±SD: 1) Global PSQI: 2.72±2.22 vs. 7.24±4.17, p<0.001; 2) PSQI subscores: sleep quality: 0.54±0.57 vs. 1.40±0.83, p<0.001; sleep latency: 0.51±0.80 vs. 1.38±1.03, p=0.002; sleep duration: 0.45±0.64 vs. 1.16±0.92, p=0.003; sleep efficiency: 0.21±0.57 vs. 0.77±0.98, p=0.018; sleep disturbance: ns; use of medication: 0.14±0.49 vs. 0.57±0.83, p=0.032; daytime dysfunction: 0.52±0.73 vs. 1.57±0.88, p=0.002. 3) Prevalence (95% CI) of insomnia symptoms: sleep latency: 5.5% (2.5-11.5%) vs. 21.1% (10.5-31.6%), p=0.015; wake time >30 min after sleep onset: 9.4% (1.6-17.1%) vs. 25.5% (14.2-37.7%), p=0.029; early awakening: 14.5% (5.1-23.8%) vs. 45.6 (32.7-58.4%), p<0.001; somewhat/very dissatisfied with sleep: 5.5% (2.5-11.5%) vs. 50% (37.1-62.8%), p<0.001; day impairment: 5.5% (2.5-11.5%) vs. 38.2% (25.6-50.7%), p<0.001; insomnia: 7.3% (0.4-14%) vs. 39.7% (27.1-52.2%), p<0.001.
Oncologists					
Shanafelt, 2005 RoB: unclear	CS	Fatigue: LASA QOL ≤7; Sleep deprivation: 10-point Likert scale from 0 (not at all) to 10 (stressful as can be) Time points NR	75% had a high level of fatigue; Mean±SD sleep score: 4.5±2.65.	Wellbeing: 10-item LASA QOL, high ≥8 vs. low ≤7 Time points NR	1) Sleep deprivation for high vs. low overall well-being (mean±SD): 3.9±2.57 vs. 5.1±2.60, p=0.0004; 2) Lower fatigue predicted overall wellbeing in a multivariate model including personal and professional characteristics, p=0.002.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Shanafelt, 2014 RoB: unclear	CS	Fatigue: 10-point LAS (lower scores indicate greater fatigue) Time points NR	Mean±SD fatigue score: 5.7±2.4	Satisfaction with WLB: 5-point Likert scale from 'strongly agree' to 'strongly disagree' Time points NR	1) OR (95%CI) of lower satisfaction predicted by high fatigue (vs. not) in multivariate model including personal and work-related factors, and burnout: 0.489 (0.337-0.710), p<0.001.
Mixed groups of physicians					
Aziz, 2004 RoB: high	CS	Working while fatigued: 5-point scale from 'extreme' to 'a little' Time points NR	NR	Stress: 47-item questionnaire with a 5-point scale from 'extreme' to 'a little' Time points NR	1) Sources of stress: working while fatigued had a mean±SD score of 2.44±1.20, factor loading: 0.653, in factor analysis; 2) Inverse correlation between stress and working while fatigued: r=-0.270 (significance level NR).
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: worried about having a car accident while driving home post-call: 5.4 vs. 7.0, p<0.001; sleep loss has a major impact on personal life: 8.4 vs. 7.0, p=0.01; 3) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Elovaino, 2015 RoB: low	CS	Sleeping problems: Jenkins Scale with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006 and 2010	Mean±SD score: 2006: 2.30 (1.00); 2010: 2.35 (1.05).	Jobs demands: 5 items scored on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree); Job control: 3 items derived from the Karasek Job Questionnaire	There was no association between sleeping problems in 2006 and job demands or control in 2010.
Heponiemi, 2014 RoB: low	CS	Sleeping problems: Jenkins Scale[81] with a 6-point scale from 1 (never) to 6 (every	Mean±SD (range) score: 2.30±1.00 (1-6)	Psychological distress: GHQ-12 with a 4-point scale (low to high);	1) Sleeping problems associated with job satisfaction, β=-0.12, p<0.001, psychological distress, β=0.18, p<0.001;

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		night) Assessed in 2006		Job satisfaction: JDS with a Likert scale from 1 (strongly disagree) to 5 (strongly agree) Assessed in 2010	2) Total indirect effect of on-call duty through two mediators (sleeping problems, work interference with family) (R^2 (95% CI)): job satisfaction 0.06 (-0.059, -0.016), p<0.001; psychological distress 0.16 (0.023, 0.081), p<0.001.
Mahmood, 2016 RoB: high	CS	Sleep deprivation: self-reported mean hours of sleep when on call Assessed at 4 years, 10 years, and 15 years post-graduation	Mean±SD hours: 4 years: 4.52 (2.79); 10 years: 5.38 (6.36); 15 years: 6.41 (7.14).	Alcohol use disorders: Modified 9-item version of the Alcohol Use Disorder Identification Test (AUDIT) ≥6 for men and ≥5 for women. Assessed at 4 years, 10 years, and 15 years post-graduation	There was no association between hours of sleep when on call and hazardous drinking behaviours (p=0.732)
Shirom, 2010 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Burnout: SMBM on a 7-point scale from 1 (almost never) to 7 (always)	1) Correlation between physical fatigue subscale and overall burnout: 0.88, p<0.05; 2) In a predictive structural model for burnout, physical fatigue accounted for unique variance in the burnout items, not accounted for by total burnout (R^2=0.24).
Smith, 2017 RoB: unclear	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Mental and physical illness: self-reported via open-ended comments Time points NR	Some physicians reported developing mental illness (e.g., bipolar disorder, alcohol misuse) due to tiredness and stress at work; others developed physical health problems due to sleep deprivation, poor eating habits and lack of exercise.
Starmer, 2016 RoB: low	CS	Sleep deprivation: <7 hours sleep in a typical 24-h period (self-reported) Time points NR	27.7% sleep deprived	Burnout, satisfaction with career and life, balanced personal and professional commitments: Each on a 5-point Likert scale (strongly	≥7-h vs. <7-h sleep: 1) Burnout (% strongly agree/agree): 26.4% vs. 39.6%, p<0.05; career satisfaction (% strongly agree/agree): ns; life satisfaction (% completely/very satisfied): 76.4% vs. 55.9%,

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
				agree to strongly disagree) Time points NR	p<0.05; balanced personal and professional commitments (% completely/very satisfied): 49.7% vs. 26.1%. 2) <7-h sleep (vs. ≥7-h) (OR, 95% CI) associated with life satisfaction 0.44 (0.29-0.67), p<0.05; balanced personal/professional commitments 0.46 (0.31-0.71), p≤0.05, in a model including work and personal factors.
Tokuda, 2009 RoB: low	CS	Sleep hours/day: self-reported (continuous) Time points NR (included weekday and weekends)	Mean±SD (range) sleep hours/day: 6±0.9 (3-8)	Burnout: MBI (Japanese) with a 7-point Likert scale: 0 (none) to 6 (every day); Job satisfaction: JHPSS[86] with a 5-point Likert scale: 1 (strongly disagree) to 5 (strongly agree) Time points NR	Maximum likelihood estimates±SE: 1) Sleeping time to job satisfaction: group 0.990±0.458, p=0.031; ns for men; women 1.711±0.805, p=0.034; 2) Sleeping time to EE: group -0.219 ±0.070, p=0.002; men -0.215±0.082, p=0.009; ns for women.
Wada, 2010 RoB: unclear	CS	Sleep hours/day: Self-reported (continuous) Assessed for past month when not completing overnight work	<5 hours: 8.7% men, 9.9% women; 5 to <6 hours: 32.3% men, 34.6% women; 6 to <7 hours: 46.0% men, 43.7% women; ≥7 hours: 13.0% men, 11.8% women.	Depression: QIDS-SR; Japanese score <5 (no symptoms) to >20 (very severe symptoms) Assessed for past 7 days	1) Sleep hours for those with vs. without depressive symptoms: <5: 18.7% vs. 7.7% men, 20.5% vs. 8.7% women; 5 to <6: 33.7% vs. 32.2% men, 38.6% vs. 34.2% women; 6 to <7: 35.1% vs. 46.9% men; 31.8% vs. 45.1% women; 2) Association between <5h sleep (vs. 6-7h) and depressive symptoms (OR (95% CI)): univariate 2.79 (1.96-3.95) for men, 2.65 (1.47-4.78) for women; multivariate (including age and workload factors) 2.70 (1.82-4.03) for men, 2.38 (1.11-5.10) for women.

[a]Includes studies of anesthetists, where these were physicians.
[b]Includes primary care physicians, internal medicine physicians, and general practitioners.

AM: morning; aMT6-s: melatonin metabolite; BA: before-after; CI: confidence interval; CBI: Copenhagen Burnout Inventory; CS: cross-sectional; DP: depersonalization; DSM: Diagnostic and Statistical Manual of Mental Disorders; EE: emotional exhaustion; ER: emergency; ESS: Epworth Sleepiness Scale; GHQ: General Health Questionnaire; h: hour(s); ICU: intensive care unit; IL-8: interleukin-8; JDS: Job Diagnostic Survey; JHPSS: Japanese Hospital Physicians Satisfaction Scale; LAS: linear analog scale; LASA: linear analog assessment scales; MBI: Maslach Burnout Inventory; MOSQ: Modified Occupational Stress Questionnaire; min: minute(s); NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PA: personal achievement; PBM: Pines Burnout Measure; PE: professional efficacy; PM: afternoon; PMS: Profile of Mood States; PSQI: Pittsburgh Sleep Quality Index; QIDS-SR: Quick Inventory Depressive Scale – Self-Reported; QOL: Quality of Life; RCT: randomized controlled trial; RDAS: Revised Dyadic Adjustment Scale; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SMBM: Shirom-Melamed Burnout Measure; TS: time series; US: United States of America; VAS: visual analog scale; vs.: versus; WLB: work-life balance

Supplementary table 9. Performance and safety outcomes related to fatigue or sleep loss among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	 Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Uchal, 2005 RoB: unclear	RCT	Sleep deprivation from a 24-h call shift vs. 8-h work; Sleep hours: self-reported (continuous); Sleepiness: ESS (moderate: 10-15, severe: ≥16) Assessed post-call and post-work	Median (range) sleep hours: 1.5 (0-3) post-call vs. 6.5 (5-9) post-work, p<0.05; Median ESS score: 7.0 post-call vs. 5.5 post-work, ns.	Surgical performance: laparoscopic surgical simulator for product quality, procedure effectiveness Assessed post-call and post-work	Post call vs. post-work: 1) Product quality: no difference in accuracy error, tissue damage, leak rate; 2) Procedure effectiveness: no difference in goal-directed actions, non-goal directed actions, operating time.
Chu, 2011 RoB: low	CO	Sleep deprivation: self-reported hours, moderate (3-6h) or severe (<3h) Assessed the night before surgery	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived and 1,595 (39.4%) moderately sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	For 0-3 vs. 3-6 vs. >6 hours of sleep: no difference in CABG or ACC.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and another case in the next 24-h	Of 6,751 procedures, 339 (5%) performed by sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	Sleep deprived vs. non-sleep deprived: no difference in CABG or ACC.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Surgical performance: duration of surgery	Sleep deprived vs. non-sleep deprived: no difference in duration of surgery, even after stratification by type of procedure.
Amirian, 2014 RoB: high	BA	17-h night call shift; Sleep hours during the shift: Wrist-mounted Micro-Mini-Motionlogger; Sleepiness: KSS	Naps pre-call: 11 (37%) napped for median (IQR) 90 (58-128) min; Median (IQR) sleep: 91 (62-123) min on the pre-call night vs. 430 (329-449) on	Surgical performance: LapSimGyn laparoscopic simulation for time, blood loss, instrument path; D2 test of attention and concentration	Pre- vs. post-call: 1) LapSimGyn: no difference in total time, blood loss, instrument path length, instrument angular path; napping did not affect performance; 2) D2 test: improvement in concentration,

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Assessed on pre-call and on-call day; sleepiness assessed during shift	the on-call night, p<0.001; Sleep on-call: 12 (40%) slept for median (IQR) 98 (39-135) min; Significant development of sleepiness during shift (p<0.001), plateau score of 7 at 04:00 to 08:00.	Assessed on pre-call and on-call day	p<0.05. No changes in any other parameters; 3) ns difference in laparoscopic simulation time in those who slept during the shift vs. not.
Gerdes, 2008 RoB: high	BA	On-call shift; Fatigue: questionnaire designed by Behrenz & Monga, 1999; Sleep hours: self-reported (continuous) Assessed in 3 sessions pre- and post-call	Fatigue differential from pre- to post-call (range): 1-7 (units unclear); Sleep during call (range): 1-5h	Psychomotor performance: virtual ring transfer task for gesture-level proficiency, hand movement smoothness, tool movement smoothness, elapsed time Assessed in 3 sessions pre- and post-call	1) Pre- to post-call: decrease in all measures of psychomotor proficiency (p<0.05, data NR) except elapsed time; no change in number of psychomotor errors; increase cognitive errors (p<0.05, data NR); 2) Cognitive errors increased exponentially as fatigue ratings increased (R^2=0.9219) and as hours of sleep declined (R^2=0.933).
Shanafelt, 2010 RoB: high	CS	Degree of fatigue as a contributor to errors (self-reported) Assessed for the past 3 months	NR	Perceived recent major medical errors (self-reported) Assessed for the past 3 months	1) Prevalence of perceived recent major medical error: 8.9%; 2) Of those reporting an error, 6.9% listed degree of fatigue as the greatest contributing factor.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift, on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Psychomotor performance: reaction time, critical flicker fusion, response measure, peripheral awareness; Concentration ability: scale of 0 (low tiredness) to 100 (maximum tiredness) Assessed pre- and post-duty	Pre- vs. post-duty, mean±SD: 1) Psychometric testing: recognition reaction time (ms): 439.6±50.8 vs. 480.3±58.9; motor reaction time (ms): 252.8±39.3 vs. 465.4±65.0; total reaction time (ms): 690.8±73.4 vs. 746.5±113.7; critical flicker fusion (Hz): 29.0±2.3 vs. 28.7±3.7; response measure (pixels): 647.8±126.7 vs. 598.3±138.1, peripheral awareness task recognition time: 58.9±59.2 vs.

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
					51.6±47.5; 2) Concentration ability: 26.4±23.5 vs. 56.3±23.0, p=0.007.
Chang, 2013 RoB: unclear	CS	15-h in-house overnight call; Sleepiness pre-call: ESS ≥9; Sleep hours: self-reported (continuous) Sleepiness assessed pre-call, sleep hours during call	Median (IQR) ESS: 9 (9), 64% scored ≥9; Median (IQR) hours slept during shift: 1 (0-3).	Psychomotor performance: reaction time; CCPT II; N-back; HVLT (3 trials of 12 words) Assessed at baseline and pre- and post-call	1) Afternoon baseline vs. pre-call: no difference in reaction time, CCPT, N-back, of HVLT; Morning baseline vs. post-call: 1) No change in auditory or visual reaction time; 2) CCPT (t-scores): No change in detectability, response style, hit reaction time, omissions/commissions; 3) N-back % accuracy: no change for auditory, visual, or mean N-value; 4) HVLT (t-score): mean for trials 1-3: 48.6±7.6 vs. 41.5±9.9 (p=0.04); delayed recall: ns; 5) No correlation between ESS scores pre-call or sleep during shift and any measure of psychomotor performance.
Gander, 2000 RoB: low	CS	Nights of work-related sleep disturbance: self-reported (continuous) Assessed for the past 6 months	NR	Risk of fatigue-related errors: questionnaire modelled after Gravenstein et al., 1990 Assessed for the past 6 months	1) Risk of fatigue-related errors increased with increasing nights of work-related sleep disturbance: RR: 1.25, 95% CI: 1.06-1.49.
Saadat, 2017 RoB: low	CS	Sleep deprivation due to an overnight call shift	NR	Reaction time: PVT Assessed after an overnight call shift and the morning of a regular (non-call) day	Mean (SD) reaction time was slower post-call (297.76 (83.75)) vs. on a regular day (266.58 (38.35)), p=0.047.
Gander, 2008 RoB: unclear	NC	Sleep loss across consecutive working days or on-call work: Wrist-mounted Actiwatch (Mini Mitter, Bend, Oregon, US), sleep	≥2 hours sleep <baseline: 8% of 24-h periods that included day work vs. 14% that included day + call; Sleep hours: mean 0.6h less	Psychomotor performance: PVT Assessed within 2 hours pre- and post-call	1) In fixed model analysis for reaction time including sleep, time since waking, work hours: acute sleep loss associated with slower median reaction time, $F_{(1,184)}$=5.70, p<0.05; longer time since waking associated with poorer

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		and duty diary Assessed over a 2-week period including a weekend of rostered shifts or on-call	sleep when working day shifts (p=0.014) and 0.8h less sleep when working day shifts + call (p=0.013) vs. off.		performance on the slowest 10%, $F_{(1,185)}$=5.13, p<0.05; 2) Reaction time across 12 consecutive work days: no change in pre-duty reaction times but post-duty reaction times slowed linearly, median -3.38, p<0.001; decline in performance across 10 minutes became progressively steeper both pre- and post-duty, p=0.020.
ER or ICU physicians					
Sanches, 2015 RoB: unclear	CS	Acute sleep deprivation (<5h of night sleep after a night shift of 12h) Sleep hours: 7-day Actigraphy via SenseWear® Pro2 Armband; Sleepiness: ESS; Sleep quality: PSQI Assessed the week and night before the psychomotor tests	Non-sleep deprived vs. sleep deprived: PSQI >5: 0% vs. 33%, ns; ESS≥10: 11% vs. 67% Sleep time (mean±SD) in week before tests: duration and number of naps higher in sleep deprived group, but diurnal sleep hours lower, 428.6±30.1 vs. 375.8±55.9, p=0.038; Sleep quality (mean±SD): week before tests: 3.3±0.7 vs. 2.6±0.3, p=0.013; night before tests: 3.1±0.8 vs. 1.9±1.0, p=0.020.	Psychomotor performance via Battery Test Reaction 5 (v1): StimulTest, InstrucTest, MovemTest; TP test of visual attention Assessed on morning after night shift 8	Sleep deprived group vs. non-sleep deprived, mean±SD: 1) InstrucTest: correct answers: 169.4 (16.0) vs. 148.3 (28.3), p=0.070; wrong answers: ns; perfection index (%): 99.6 (0.3) vs. 98.9 (1.3), p=0.021; response latency (sec/click): ns; 2) StimulTest: correct answers: 170.7 (21.9) vs. 145.1 (17.9), p=0.022; wrong answers: ns; perfection index (%): ns; response latency (sec/click): 1.06 (0.1) vs. 1.24 (0.1), p=0.022; 3) MovemTest: ns for any parameter; 4) TP: omitted symbols: 34.2±18.4 vs. 62.7±44.0, p=0.034; concentration index (%): 14.1±8.9 vs. 30.0±25.9, p=0.019; quality index (%): 13.8±8.6 vs. 29.2±26.4, p=0.031; correct/wrong symbols: ns; Correlations between sleep and tests: 1) TP for sleep hours nights 1-6: omitted symbols: r=-0.686, p=0.011 for non-sleep-deprived, ns for sleep-deprived; concentration index (%): r=-0.359, p=0.037 for sleep-deprived, ns for non-sleep deprived; r=-0.359, p=0.037 for the group; no other significant correlations; 2) No correlation between PSQI, ESS and any of

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					the psychomotor tests.
Generalists[b]					
Harbeck, 2015	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Neurocognitive parameters: computerized attentional test (vigilance, alertness); D2 letter cancellation test (divided attention); Trail Making Test (visual attention, task switching); Digit Span, Digit Symbol Substitution Test, Weschler Memory Scale (memory functions) Assessed before a normal day shift, and after a 24-h on call shift	Intrinsic alertness, focused attention and vigilance were similar on both occasions; Phasic alertness improved following the on-call shift: mean (SD) 24.8 (15.6) vs. 38.3 (21.5), p = 0.022.
Mixed specialties or undefined populations					
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: written an incorrect order: 8.8 vs. 7.3, p=0.02; might fall asleep while examining a patient: 13.2 vs. 7.7, p=0.001; look forward to sleeping at grand rounds: 10.4 vs. 7.4, p=0.002; 3) No difference in ESS score for those who agree/strongly agree vs. other response: work is unaffected by sleep loss and fatigue, thinking is unaffected by sleep loss, sleep loss and fatigue affect my medical decisions, have heard of others making medical errors due to sleep loss and fatigue, never make errors in prescriptions

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					on post-call days, have made medical errors because of sleep loss and fatigue; 4) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Heponiemi, 2014 RoB: low	CS	Sleeping problems: 4-item Jenkins Scale on 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Work ability: Work Ability Index on scale from 1 (could not work at all) to 10 (best work ability) Assessed in 2010	1) On-call duty had an indirect effect on work ability (R^2=0.11, 95% CI: -0.122, -0.031, p<0.001) through two mediators (work interference with family, sleeping problems); 2) Sleeping problems inversely associated with work ability, β=-0.29, p<0.001.
Kanieta, 2011 RoB: unclear	CS	Sleep hours: self-reported (continuous) Sleepiness and sleep difficulties: 5-point scale from 1 (never) to 5 (always); Insomnia: ≥3 sleep difficulties Assessed for the past month	Insufficient rest: 32.5%; Daytime sleepiness: 3.5%; Insomnia: 20.0%; Sleep time (mean±SD min): 279.8±60.9	Self-reported medical incidents: 4-point scale from 1 (never) to 4 (often) Assessed for the past month	1) Prevalence of medical incidents (% (95% CI)): sleep deprived (26.8% (24.2, 29.4)) vs. not (15.2% (13.7, 16.7)), p<0.01; insomnia (24.8% (21.6, 28.0)) insomnia vs. not (17.6% (16.2, 19.0)), p<0.01; ≥6h sleep (18.3% (16.8, 19.8)) vs. <6h (21.7% (18.8, 24.6)), p=0.03; 2) Predictors of medical incidents in multivariate model including personal and work-related factors (OR (95% CI)): lacking rest due to sleep deprivation vs. not (1.65 (1.33-2.04)), p<0.01); insomnia vs. not (1.45 (1.16-1.82), p<0.01); ns for sleep hours.
Sexton, 2001 RoB: high	CS	Fatigue as a factor impacting performance Time points NR	NR	Performance effectiveness measured by 1 question: agree, neutral, disagree Time points NR	1) "When fatigued, I perform effectively during critical phases of operations/patient care": Anesthetic: 47% agree; 15% neutral; 38% disagree; Surgical: 70% agree; 12% neutral; 18% disagree.
Shirom, 2006 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always)	NR	Quality of care: Adapted 15-item SERVQUAL with a 5-point Likert scale from 1 (very small extent) to 5 (very large extent)	1) Quality of care positively predicted by fatigue in a model incorporating several other components of burnout, β=0.17, p<0.05.

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Time points NR	
Smith, 2017 RoB: moderate	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Perceived competence: self-reported via open-ended comments Time points NR	Some physicians indicated that continual tiredness and exhaustion led to concerns that it would affect their competence; some felt that professional performance was compromised at times of physical and mental fatigue.
Tanti, 2017 RoB: high	CS	Fatique: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	NR	Prescribing errors: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	Perception of the contribution of fatigue to prescribing errors differed by physician type ($p<0.05$): 34% of community doctors, 96% hospital doctors, 8% of office-working doctors perceived a very high or high association between fatigue and prescribing errors.

[a]Includes studies of anesthetists, where these were physicians.

[b]Includes primary care physicians, internal medicine physicians, and general practitioners.

ACC: aortic cross-clamp time; BA: before-after; CABG: cardiopulmonary bypass time; CCPT II: Connor's Continuous Performance Test II; CI: confidence interval; CO: cohort; CS: cross-sectional; ER: emergency; ESS: Epworth Sleepiness Scale; h: hour(s); HVLT: Hopkin's Verbal Learning Task; Hz: Hertz; ICU: intensive care unit; IQR: interquartile range; KSS: Karolinska Sleep Scale; min: minutes; ms: millisecond(s); N-back: Dual N-back test; NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; PVT: Psychomotor vigilance Performance Task; RR: risk ratio; RCT: randomized controlled trial; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SERVQUAL: Service Quality Measure; SMBM: Shirom-Melamed Burnout Measure; TP: Toulouse-Piéron test; TS: time series; US: United States of America; vs.: versus

Supplementary table 10. Patient outcomes related to fatigue or sleep restriction among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Surgeons					
Chu, 20 RoB: low	CO	Sleep deprivation: moderate (3-6 h) or severe (<3-h) sleep deprivation the night before surgery (self-reported hours)	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived, 1,595 (39.4%) by moderately sleep-deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) 0-3 vs. 3-6 vs. >6 hours of sleep: No difference in incidence of mortality, incidence of 10 major complications (except septicemia, 3.6% vs. 0.9% vs. 0.8%, p=0.03), ICU length of stay; in-hospital length of stay (days): 7.0 vs. 6.0 vs. 7.0, p<0.001.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and performed a subsequent case in the next 24-h	Of 6,751 procedures, 339 (5%) were performed by sleep deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, need for blood products, complications (operative, neurologic, renal, infectious, pulmonary), in-hospital length of stay.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Chart review: mortality, surgical complications, readmission, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, surgical complications, readmissions within 30 days, or length of stay.
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse surgical complications Assessed during and post-surgery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total number of complications, preventable complications, type of complications; 2) Operating room procedures with complications, OR (95% CI): 8.5% for 0-6h sleep vs. 3.1% for >6h sleep, 2.70 (1.13-6.48), p=0.03; 3) All procedures with complications, OR (95% CI): 6.2% for 0-6h sleep vs. 3.4% for >6h sleep, 1.72 (1.02-2.89), p=0.04.
Schieman, 2007 RoB: low	CO	Fatigue: surgeon billed for clinical work after 22:00 the night before surgery	Of 270 procedures, 22 (8%) were performed by fatigued surgeons	Chart review: surgical complications, length of stay, mortality, cancer recurrence	1) Fatigued vs. non-fatigued surgeons: no difference in intra- or post-operative complication rate, length of stay, in-hospital

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
				Assessed during and post-surgery	length of stay, cancer recurrence.
Vinden, 2014 RoB: low	CO	Sleep deprivation (at risk): surgeon worked 00:00 to 07:00 and performed surgery 07:00 to 18:00	Of 94,183 surgeries, 2,078 (2.2%) were performed by surgeons who were 'at risk'	Chart review: conversion to open procedure (from laparoscopic), iatrogenic injuries, mortality Assessed during and post-surgery	1) At risk vs. not at risk surgeon: no difference in incidence of conversion to open procedure, iatrogenic injuries, mortality, in either univariate or multivariate analyses.
Obstetricians					
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse obstetric complications Assessed during and post-delivery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total complications, preventable complications, type of complications; 2) No association between sleep deprivation and proportion of procedures with complications, nor difference for 0-6h vs. >6h of sleep opportunity.

CI: confidence interval; CO: cohort; h: hours; NR: not reported; OR: odds ratio; RoB: Risk of Bias; SD: standard deviation; US: United States of America; vs.: versus

Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	7
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	7, no meta-analysis

Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	Not applicable
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7-10, Table 1, Supplementary file 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	10, Supplementary file 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Supplementary files 4-6
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	10-15 (no meta-analysis)
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Not applicable
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	15-17
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	17
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	17
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	18

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org. Page 2 of 2

BMJ Open

The impact of fatigue and sleep restriction on physician and patient outcomes: A systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021967.R1
Article Type:	Research
Date Submitted by the Author:	27-Apr-2018
Complete List of Authors:	Gates, Michelle; University of Alberta, Pediatrics Wingert, Aireen; University of Alberta, Pediatrics Featherstone, Robin; University of Alberta, Pediatrics Samuels, Charles; Centre for Sleep and Human Performance Simon, Christopher; Canadian Medical Association Dyson, Michele; University of Alberta, Pediatrics
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Evidence based practice, Health services research, Medical education and training, Occupational and environmental medicine
Keywords:	Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SLEEP MEDICINE

The impact of fatigue and sleep restriction on physician and patient outcomes: A systematic review

Michelle Gates, Aireen Wingert, Robin Featherstone, Charles Samuels, Christopher Simon, Michele P Dyson

Michelle Gates, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486C, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Aireen Wingert, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-490B, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Robin Featherstone, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486D, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Research librarian

Charles Samuels, Centre for Sleep and Human Performance, Suite 106, 51 Sunpark Drive SE, Calgary, Alberta, Canada, T2X 3V4
Physician and medical director

Christopher Simon, Canadian Medical Association, 1867 Alta Vista Drive, Ottawa, Ontario, Canada, K1G 5W8
Advisor, physician health

Michele P Dyson, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-474, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Assistant professor

Correspondence to: MP Dyson mdyson@ualberta.ca

Word count (main text): 5,929

ABSTRACT

Objectives: For physicians in independent practice, we synthesized evidence on the (a) impacts of sleep restriction and fatigue on health and performance, and patient safety; (b) effectiveness of interventions targeting sleep restriction and fatigue.

Design: We systematically reviewed online literature. After piloting, one reviewer selected studies by title and abstract; full texts were then reviewed in duplicate. One reviewer extracted data; another verified a random 10% sample. Two reviewers assessed risk of bias. We pooled findings via meta-analysis when appropriate, or narratively.

Data sources: We searched Medline, Embase, PsycINFO, CINAHL and PubMed for published studies in April 2016; Medline was updated in November 2017. We searched Embase for conference proceedings, and hand-searched meeting abstracts, association and foundation websites.

Eligibility criteria for selecting studies: English or French language primary research studies published from 2000-2017 examining the effect of fatigue or sleep-related exposures or interventions on any outcome among physicians in independent practice and their patients.

Results: Of 16,154 records identified, we included 47 quantitative studies of variable quality. 28 studies showed associations between fatigue or sleep restriction and physician health and well-being outcomes. 21 studies showed no association with surgical performance, and mixed findings for psychomotor performance, work performance, and medical errors. We pooled data from six cohort studies for patient outcomes. For sleep deprived versus non-sleep deprived surgeons, we found no difference in patient mortality (n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, p = 0.82, $I^2 = 0\%$), intraoperative complications (n = 19,798, RR 1.35, 95% CI 0.82 to 2.21), postoperative complications (n = 60,201, RR 0.99, 95% CI 0.95 to 1.03) or length of stay (n = 50,046, MD -0.33, 95% CI -1.03 to 0.36).

Conclusions: Fatigue and sleep deprivation may be associated with negative physician health outcomes. Current evidence is inadequate to inform practice recommendations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The review was informed by the methods outlined by Cochrane and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.
- The review was limited by the quality of the included studies, which was often poor. We could not draw definitive conclusions due to methodological weaknesses and heterogeneous outcome measures in the included studies.
- We have focused on evidence from high income countries; our findings may not be generalizable to other settings.

BACKGROUND

The working hours of physicians have been a topic of debate for many years.[1] Beginning in the late 1980s, evidence indicating that medical resident fatigue could negatively impact their cognitive functioning and performance, resulting in an increased risk of medical error, began to accumulate.[2] In response, by the early 2000s physicians' regulatory bodies worldwide began to take action toward restricting the work hours of medical residents and ensuring adequate time for recovery between shifts.[3-5] Since their implementation in the United States by the Accreditation Council for Graduate Medical Education (ACGME), the impact of work hour regulations has been widely researched. Still, evidence for impacts on patient care, resident training and wellbeing remains equivocal.[6-9] This is likely because work hours are only one of many contributors to fatigue and physician wellbeing. In fact, the ACGME has recently reversed the 2011 changes that limited resident work hours to 16 hours per shift and the requirement for 8 hours of time off between shifts. This decision was made in favour of promoting "flexibility" for residency training program work hours and scheduling.

The focus on medical trainees has left physicians in independent practice as a relatively neglected group in research and policy. In Canada, there is no concrete regulation on the hours or patterns in which physicians choose to work.[10] In the absence of clear policies, physicians trained under traditional systems may find it difficult to work shorter hours or take more frequent breaks.[1] Indeed, more than 40% of practicing physicians in the United States work in excess of 80 hours per week.[11] While long work hours remain a cultural norm in medicine, in comparable high-risk industries (e.g., aviation), work patterns and work hours are tightly regulated.[12] The need for similar evidence-based policies in medicine has become a topic of increased interest. Exemplar of this, an evidence-based guideline for fatigue risk management in emergency medical services,[13] informed by a comprehensive set of systematic reviews, has recently been published. For physicians, it has been argued that there is a need to adapt healthcare systems and provide support in identifying the signs of fatigue and mitigating its risks.[1]

Besides potentially affecting patient outcomes, fatigue can impact the health and wellbeing of physicians themselves. Burnout, just one outcome related to fatigue, has been described as epidemic among physicians[14-16] and ultimately affects recruitment and retention of physicians both in community and acute care settings. While the effect of physician wellbeing on the sustainability of healthcare systems has recently received increased attention,[17] evidence-based solutions to burnout

remain relatively elusive.[18] What is clear, is that comprehensive organisational-level efforts are necessary to fully address the issue.[19] Research addressing the factors that influence burnout and overall physician wellness is needed to inform system- and individual-level strategies.[20, 21] To date, evidence of the effects of fatigue and the role of chronic sleep restriction on physicians in independent practice has not been synthesized, making it unclear what gaps in knowledge remain unaddressed.

Given this void, we undertook a systematic review focusing broadly on primary research relevant to the Canadian context as a fundamental starting point to examine the effects of fatigue and chronic sleep restriction on physicians in independent practice, and on interventions to combat these effects. Our review was guided by the following research questions: Among physicians in independent practice, (1) what are the impacts of fatigue and chronic sleep restriction on physician health, physician performance, and patient safety; and (2) what is the effectiveness of interventions that target fatigue and chronic sleep restriction loss, in terms of improving physician and patient outcomes?

METHODS

Review conduct

The conduct of this systematic review was guided by Cochrane standards.[22] The research team convened to plan the key research questions and methodology but did not register a formal protocol. The findings are reported in adherence with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement.[23] Ethical approval was not required for this study.

Patient involvement

Patients were not involved.

Literature search

An information specialist developed a search strategy that included concepts related to physicians, fatigue and sleep. On 13 April 2016 we searched the following online databases with coverage in the biomedical sciences and psychology: Medline, Embase, PsycINFO, CINAHL and PubMed, limited to English and French language articles published from 2000 to 2016. We updated the Medline search in November 2017, as this database offered the highest precision. Though fatigue among physicians is not a new phenomenon,[2] we limited our search to articles published post-2000 to include studies relevant to current physician practice. Work hour limitations have existed in European countries since 1993, but

implementation in the United States (2003)[5] and Canada (2013) for residents is more recent.[24] We aimed to include studies published in this era of increased awareness about the potential impacts of long work hours. To locate unpublished studies, we searched Embase for conference proceedings since 2000 and hand-searched meeting abstracts of the Canadian Conference on Physician Health and the International Conference on Physician Health (2012 to 2016). We also searched the following association and foundation websites: American Medical Association, Australian Medical Association, British Medical Association, Canadian Medical Association, European Medical Association, National Sleep Foundation, Ontario Medical Association and the World Medical Association. The complete search strategy undertaken is reported in Supplementary file 1.

Inclusion criteria

Primary studies (quantitative or qualitative) of fatigue- or sleep-related exposures or interventions among physicians in independent practice were eligible for inclusion. We included physicians practicing in any medical specialty and in any healthcare setting within a high income country,[25] to identify practices comparable to the Canadian setting. Studies including physicians-in-training were included only if data for physicians in independent practice could be isolated. Exposures of interest included fatigue, sleep restriction, or sleepiness. We also included studies of any intervention that aimed to reduce fatigue or sleep restriction with any comparator (or no comparator). All reported outcomes, measured at any time, were eligible for inclusion.

We excluded commentaries, letters, editorials and dissertations. Systematic reviews, health technology assessments, economic evaluations and practice guidelines were excluded, although the reference lists of these as well as the included studies were scanned for potential primary studies. Studies that focused solely on physicians-in-training (e.g., trainees, residents, fellows, interns, medical students, junior doctors, registrars) were ineligible. To maintain the focused scope of the review, we excluded work hours, work load, and any other exposure or intervention that was indirectly related to fatigue or sleep restriction.

Study selection

The study team piloted the selection criteria, which were then applied by two independent reviewers following a two-phase process. We first screened titles and abstracts for potential relevance. Then, we retrieved all records classified as “include” or “unsure” and reviewed their full text for eligibility. Any

disagreements between reviewers were resolved by discussion or third-reviewer consultation when necessary.

Data extraction

Reviewers used a standardized form to extract data in Microsoft Office Excel (*v.* 2016, Microsoft Corporation, Redmond, WA). One reviewer independently extracted data from each included study and a second reviewer verified a random 10% sample. Since no major errors or omissions were noted, we did not undertake further verification.

We extracted the following data: country of publication; funding source; study design; inclusion and exclusion criteria; population characteristics (i.e., sample size, age and gender distribution, physician specialty); setting (i.e., physician workplace, urban or rural); exposure or intervention; definition of fatigue or sleep loss; sleep and fatigue scales used and timing of measurement; comparators (if applicable); and outcomes.

Risk of bias appraisal

Two reviewers independently assessed the risk of bias in each included study using standard tools. Disagreements were resolved via discussion or by consulting a third reviewer. We used the Cochrane Risk of Bias tool[22] to assess randomised controlled trials. Adapted versions of the tool developed by the Effective Practice and Organization of Care group[26] were used to assess before-after and time series studies. We used the Newcastle-Ottawa Quality Assessment Scale[27] to appraise cohort studies. We adapted the scale to assess cross-sectional studies and the one non-comparative study.

Evidence synthesis

We considered clinical and methodological heterogeneity in our decision on whether to proceed with meta-analysis for the outcomes identified. For most outcomes, we found insufficient homogeneity in study design, populations, exposures or interventions, and outcome measures to pool the data via meta-analysis. Thus, we have presented the findings for most outcomes narratively and in summary tables.

When statistical pooling was appropriate, this was undertaken using Review Manager (RevMan v.5.3, Copenhagen: The Nordic Cochrane Centre, the Cochrane Collaboration, 2014) via pairwise meta-analysis

using the DerSimonian and Laird random effects model (given expected heterogeneity).[28] We pooled dichotomous outcomes using the relative risk (95% confidence interval (CI)) and continuous outcomes using the mean difference (95% CI) since the units across studies were consistent (i.e., minutes). When meta-analysis was conducted, we assessed statistical heterogeneity using the chi-square test (using P = 0.05 as the threshold for significance), and quantified the extent of heterogeneity using the I^2 statistic.[29] Subgroup and sensitivity analyses were conducted when appropriate to explore heterogeneity. We intended to assess small study bias visually by inspecting funnel plots and statistically using Egger's regression test, but did not due to the small number (i.e., less than 8) of studies included in the meta-analyses.[30]

When data were not presented in the format required for meta-analysis, we estimated means or standard deviations (SDs) using standard equations. We used the median instead of the mean for one study[31] for the outcomes of length of stay and operating time. Additionally, for one study[32] in the length of stay analysis where the SD could not be estimated, we substituted the mean variance of other studies within the meta-analysis.[33]

RESULTS

We identified 16,083 unique records via the database searches, 56 grey literature sources, and 14 additional records in reference lists of systematic reviews. We excluded 15,016 citations by title and abstract, and another 1,090 by full text. Forty-seven studies[31, 32, 34-78] were eligible for inclusion, and 6[31, 32, 41, 58, 63, 77] were included in meta-analysis for the outcomes of operating time, intra- and post-operative complications, patient mortality and length of hospital stay. Figure 1 shows the flow of studies through the selection process.

Included study characteristics

A summary of the study characteristics is provided in Table 1. Supplementary file 2 presents descriptive information for each included study. There were 45 observational studies[31, 32, 34-39, 41-74, 76-78] and two intervention studies.[40, 75] All studies were quantitative. Nearly half (n = 20/47, 43%) of the studies took place in North America,[31, 32, 35, 37, 38, 41, 45, 48, 57-60, 62, 63, 65-68, 72, 77] and slightly more than one-third (n = 16/47, 34%) in Europe.[34, 36, 39, 40, 42, 46, 47, 50-53, 61, 64, 73, 75, 76]

Table 1. Summary characteristics of the included studies

Study characteristics	n	%	Physician characteristics	n	%	Exposures, interventions and outcomes	n	%
Study design			**Gender**			**Exposures (observational)**[a]	**45**	**96**
Cross-sectional	34	72	Reported[b]	38	81	Fatigue-related	15	32
Cohort	6	13	>50% male	30	79	Sleep-related	37	79
Before-after	3	6	**Age**			Overnight or extended shifts	18	38
RCT	2	4	Reported[b]	38	81	**Interventions (experimental)**	**2**	**4**
Time series	1	2	Range (years)	20 to >70		**Outcomes**		
Non-comparative	1	2	**Specialty area**[c]			Physician health and wellbeing	28	60
Region and country			Surgeons	13	28	Work and life satisfaction	9	19
North America	20	43	Anesthesiologists	10	21	Burnout	7	15
US	15	32	Generalists	7	15	Stress	8	17
Canada	4	9	ED or ICU physicians	3	6	Mental health and wellbeing	7	15
Canada, US & Mexico	1	2	Oncologists	2	4	Other health-related outcomes	5	11
Europe	16	34	Obstetrician-gynecologists	1	2	Physician performance, risk of error	21	45
France	4	9	Mixed groups	14	30	Psychomotor performance	7	15
Finland	3	6	**Work setting**[d]			Work ability and quality of care	5	11
Spain	2	4	Hospitals	37	79	Incidence of medical errors	5	11
Austria	2	4	Private practice	13	28	Surgical efficiency, effectiveness	6	13
Norway	2	4	Primary care centres, outpatient clinics	7	15	Patient outcomes	6	13
Denmark	1	2	Academic practice, training programs	5	11			
Germany	1	2	Other (e.g., industry, military)	11	23			
Malta	1	2	Not reported	3	6			
Japan	4	9	**Urban or rural**					
Australia	2	4	Reported[b]	16	34			
Israel	2	4	Urban	12	75			
New Zealand	2	4	Rural	2	13			
United Kingdom	1	2	Mixed	2	13			

ED: emergency department; ICU: intensive care unit; RCT: randomised controlled trial; US: United States of America

[a]Exposures that have been directly related to an outcome. Some studies included multiple exposures.

[b]Percentages presented using the total number of studies where the outcome was reported as the denominator.
[c]Anesthesiologists include physician anesthetists; generalists include primary care physicians, internists, and general practitioners; mixed groups refers to studies including more than one physician group or specialty (usually large-scale surveys). In some studies, multiple distinct groups were represented.
[d]As defined by the authors. Values for the settings will exceed 100% because studies may occur in more than one setting.

The 47 studies reported outcomes for 36,190 (range = 6 to 7,905) physicians and 69,809 (range = 270 to 38,978) adult patients. About half reported on surgeons (n = 13/47, 28%),[31, 32, 34, 41, 45, 48, 54, 58, 62, 63, 66, 75, 77] or anesthesiologists/physician anesthetists (n = 10/47, 21%).[36, 37, 39, 43, 44, 50-52, 59, 60] Where it was reported, the samples tended to be predominantly male (n = 30/38, 79%) and physician age varied widely. Hospitals were the most common setting (n = 37/47, 79%).[31, 32, 34-37, 39-43, 45-47, 49-52, 54, 57-61, 63-70, 73-75, 77, 78] In the studies where it was reported (n = 16/47, 34%),[31, 32, 34, 38, 40, 41, 43, 45, 50, 51, 55, 56, 58, 65, 76, 77] all but four studies[31, 55, 56, 77] took place in solely an urban setting.

Fifteen (32%) studies reported on fatigue-related exposures (e.g., as a source of stress, exhaustion, physical fatigue; hereafter referred to as 'fatigue'),[35, 40, 45, 48, 57, 63-71, 73] while others (n = 37/47, 79%) reported on sleep-related exposures (e.g., sleep hours, sleep restriction, sleep deprivation, sleep disruption, sleepiness; hereafter referred to as 'sleep restriction').[31, 32, 34, 36-47, 49-56, 58-62, 64, 67, 71, 72, 74-78] A few (n = 5/47, 11%) reported on both.[40, 45, 64, 67, 71] In some cases (n = 18/47, 38%), fatigue or sleep restriction were related to overnight work or long on-call shifts.[31, 32, 34, 37, 40, 41, 43, 45, 46, 50, 51, 53, 58-60, 63, 75, 77] Measured outcomes varied widely and were ultimately organised into physician physical and mental health, physician performance and risk of error, and patient outcomes.

Risk of bias appraisal

The overall quality of the body of research was poor; 62% (n = 29/47) of studies were rated at unclear or high risk of bias. Of the two randomised controlled trials, one was rated as unclear overall risk of bias[75] and one as high risk.[40] All cohort studies were at low risk of bias (mean score: 8.4/9, range: 8-9).[31, 32, 41, 58, 63, 77] All of the before-after studies were rated as high risk of bias.[34, 45, 50] The single time series study was assessed at high risk of bias.[51] The cross-sectional studies varied in performance (mean score: 3.0/5, range: 1-4); only one-third (n = 12/34, 35%) were at low risk of bias.[39, 42, 44, 47, 52, 59, 60, 69, 70, 72, 74, 76] The one non-comparative study was at unclear risk of bias.[43] Detailed assessments of the sources of bias per study are shown in Supplementary file 3.

Physician health and wellbeing outcomes

Twenty-eight studies reported on physician health and wellbeing-related outcomes,[35, 36, 38-40, 42, 46-48, 50-57, 60, 62, 64, 67, 68, 70-72, 74, 76, 78] including burnout (n = 7), stress (n = 8), mental health and wellbeing (n = 7), life and job satisfaction (n = 9) and other markers of health (n = 5) (Supplementary file 4).

Seven cross-sectional studies reported on burnout (5 low[39, 70, 72, 74, 76], 1 unclear[54], 1 high risk of bias[62]) among surgeons,[54, 62] anesthesiologists,[39] generalists,[76] and other mixed groups.[70, 72, 74] Two studies reported on surgeons; the larger (n = 2,564, low risk of bias) study of neurosurgeons showed increased odds of burnout with sleep deprivation (hours of sleep per night; OR 0.84, 95% CI 0.75 to 0.94, P = 0.002).[54] Among anesthesiologists one study (n = 565, low risk of bias) indicated that burnout (measured via Maslach Burnout Inventory) was more prevalent among the sleep-deprived ('lack of sleep' on one question; 47.6% vs. 16.3%, P < 0.001).[39] In one small (n = 11) study of generalists, those with burnout (measured via Pines Burnout Measure) had poorer Pittsburgh Sleep Quality Index scores (7.24±4.17 vs. 2.72±2.22, P < 0.001).[76] In the two larger studies of mixed physician groups (low risk of bias), burnout (measured via 5-point scale) was more prevalent among those who were sleep deprived (<7 hours of sleep per 24 hours; 39.6% vs. 26.4%, P < 0.05),[72] and physical fatigue ('feeling tired' on a 7-point scale) was correlated with burnout (Shirom-Melamed Burnout Measure; r = 0.88, P < 0.05).[70] In summary, evidence from 7 cross-sectional studies (71% at low risk of bias), showed associations between sleep restriction and burnout.

Six cross-sectional studies (2 low[47, 52], 1 unclear[46], 3 high risk of bias[35, 62, 64]), one uncontrolled before-after study (high risk of bias[50]), and one intervention study (high risk of bias[40]) reported on stress outcomes among surgeons,[62] anesthesiologists,[50, 52] emergency physicians,[40, 64] internal medicine physicians,[46] and mixed groups.[35, 47] In a small sample (n = 20) of internal medicine physicians, sleep restriction related to a 24-hour call shift showed no association with biochemical or physiological stress parameters, except levels of thyroid stimulating hormone, which was higher post-shift (P = 0.049, data not reported).[46] The remaining observational studies suggested that there was an association between sleep restriction or fatigue and stress. The one study of orthopedic surgeons (n = 264, high risk of bias) showed that sleep restriction (measured on a 3-point scale) and psychological distress (measured via General Health Questionnaire-12) were correlated (data not reported, P < 0.001).[62] The two reports on anesthesiologists were of varied quality; the larger (n = 328, low risk of bias) study showed that stress symptoms (measured via Modified Occupational Stress Questionnaire) were predicted by sleep sufficiency (self-reported on one question, β = -0.269, P < 0.001).[52] Among the two studies reporting on mixed groups of physicians, the larger (n = 1,541, low risk of bias) study showed an association between sleep problems (4 questions derived from Jenkins scale) and psychological distress (General Health Questionnaire-12; β = 0.18, P < 0.001).[47] One RCT assessed the impact of sleep restriction from shift work (14-hour or 24-hour shifts), showing that stress (on a visual analog scale) among emergency physicians (n = 17) was higher following the shift as compared to a control day (data not reported, P < 0.05).[40] In summary, evidence from one intervention study at high risk of bias and all but one of the 7 observational studies (29% at low risk of bias) identified supported an inverse association between fatigue or sleep deprivation and stress.

Seven cross-sectional studies (2 low,[52, 60] 3 unclear,[67, 71, 78] 2 high risk of bias[36, 53]) reported on aspects of mental health including addiction or substance misuse,[36, 53, 71] depression,[78] thoughts of suicide,[52] mood disturbance,[60, 71] and overall wellbeing.[67] One study,[53] which was at high risk of bias, showed no association between hours of sleep when on call and hazardous drinking behaviours (via Alcohol Use Disorder Identification Test). Meanwhile, the six other studies all showed associations between sleep restriction and fatigue and reduced mental health. Three studies reported on anesthetists,[36, 52, 60] with two large surveys showing increased odds of tobacco (OR 1.42, 95% CI 1.04 to 1.94) and tranquilizer/hypnotics (OR 3.26, 95% CI 2.12 to 5.02) dependency being predicted by sleep deprivation (measured by one question),[36] and sleep disturbance being associated with thoughts of suicide (using a 4-point scale; P = 0.009).[52] A small study (n = 21) showed greater mood disturbance following a 17-hour night shift than a usual day (Profile of Mood States score 42.57±15.26 vs. 70.90±6.91, P < 0.001).[60] Among oncologists (n = 241), overall wellbeing was predicted by lower levels of fatigue after controlling for personal and professional characteristics (assessed via linear analog scale quality of life survey, P = 0.002).[67] A large (n = 3,862, unclear risk of bias) study of physicians showed that sleep restriction (lower sleep hours when not at work in the past month) was associated with increased odds of depression (Quick Inventory Depressive Scale; OR 2.70, 95% CI 1.82 to 4.03 for men; OR 2.38, 95% CI 1.11 to 5.10 for women).[78] In open-ended questions, senior physicians in one study (unclear risk of bias) attributed the development of mental illness to tiredness and stress at work.[71] In summary, 7 cross-sectional studies (29% at low risk of bias) were identified, and of these 6 supported an association between sleep restriction or fatigue and negative mental health outcomes.

Nine cross-sectional studies (4 low,[42, 47, 72, 74] 2 unclear,[55, 68] 3 high risk of bias[38, 48, 62]) reported on outcomes related to job satisfaction,[42, 47, 48, 55, 72, 74] life satisfaction,[38, 62, 72] or work-life balance.[68, 72] The six studies that investigated job satisfaction were all at low risk of bias and generally included mixed groups of physicians;[47, 72, 74] one study reported on general practitioners,[55], another on surgeons,[48] and one on mixed specialties.[42] Three studies showed that reductions in sleep duration and/or quality[47, 48, 74] were associated with reduced job satisfaction. Meanwhile one showed no association between sleep restriction (<7 hours per 24-hour period) and career satisfaction (measured on a 5-point Likert scale),[72] and another showed no relationship between earlier sleep disturbance (Jenkins Scale) and later job demands or job control (measured via 5-point scale).[42] A single study (n = 92) reporting on rural general practitioners indicated that frequent sleep disturbance (measured on a 7-point scale) predicted the intention to retire early (OR 2.91, 95% CI 1.11 to 7.6, P < 0.05).[55] In summary, 6 cross-sectional studies (all at low risk of bias) were identified, and all but two[42, 72] of these studies showed that sleep restriction and fatigue were associated with reductions in satisfaction.

The three studies reported on life satisfaction.[38, 62, 72] Of two studies among mixed physician groups,[38, 72] the one larger (n = 840) study showed that sleep restriction (< 7 hours per day) was a predictor of reduced life satisfaction (measured on a 5-point Likert scale; OR 0.44, 95% CI 0.29 to 0.67, P ≤ 0.05).[72] One study at high risk of bias reported on orthopedic surgeons (n = 264), showing that sleep deprivation (measured via 3-point scale) was correlated with lower marital satisfaction (Revised Dyadic Adjustment Scale; data not reported, P < 0.001).[62] Two large studies at low or unclear risk of bias reported on work-life balance.[68, 72] Among oncologists (n = 1,117), reduced satisfaction with work-life balance (measured on a 5-point Likert scale) was predicted by high levels of fatigue (measured via 10-point visual analog scale), even when controlling for personal and work-related factors and burnout (OR 0.489, 95% CI 0.337 to 0.710, P < 0.001).[68] Among a mixed group of physicians (n = 840, low risk of bias), sleep restriction (<7 hours in a typical 24-hour period) predicted a reduced perception of having balanced personal and professional commitments (5-point Likert scale; OR 0.46, 95% CI 0.31 to 0.71, P ≤ 0.05).[72] In summary, 3 cross-sectional studies (all unclear or high risk of bias) supported an association between sleep restriction or fatigue and reduced life satisfaction, and 2 cross-sectional studies (50% low risk of bias) supported an association with reduced work-life balance.

Four cross sectional studies (3 unclear,[56, 57, 71] 1 high risk of bias[38]) and one time series study (high risk of bias[51]) reported on other health-related outcomes. Among a mixed group of physicians (n = 180), one study at high risk of bias showed that Epworth Sleepiness Scale scores were higher among physicians who worried about having a car accident while driving home (7.0 vs. 5.4, P < 0.001).[38] Among generalists (n = 578), almost 1 in 10 (8.7%) admitted to falling asleep while driving due to fatigue.[57] Also among generalists (n = 92), those with frequent work-related sleep disturbance (measured on a 7-point scale) were at increased odds of sickness presenteeism (OR 2.92, 95% CI 1.19 to 7.16, P = 0.02).[56] The one time series study concluded that a single 24-h shift did not cause major chronodisruption (based on serum melatonin measurement) among anesthetists (n = 10).[51] Meanwhile, open-ended comments from a large sample (n = 3,550) of senior physicians suggests that they attributed the development of physical health problems to a lifestyle of sleep restriction, poor eating habits and lack of exercise imposed by their jobs.[71] In summary, 5 cross sectional studies (0% at low risk of bias) supported associations between sleep restriction and fatigue and varied deleterious health outcomes (i.e., car accidents, sickness presenteeism, physical health problems). One time series study at high risk of bias did not support such a relationship.

Physician performance and risk of errors

Twenty-one studies reported on physician performance and safety-related outcomes,[31, 32, 34, 37, 38, 41, 43-47, 49, 50, 59, 61, 65, 66, 69, 71, 73, 75] including surgical efficiency and effectiveness (n = 6), psychomotor performance (n = 7), work ability and quality of care (n = 5) and medical errors (n = 5) (Supplementary file 4).

Four cohort studies (all low risk of bias[31, 32, 41, 63]), one before-after study (high risk of bias[34]) and one randomized controlled trial (high risk of bias[75]) examined the effects of sleep restriction from overnight work or extended shifts, during surgeries[31, 32, 41] or laparoscopic simulations.[34, 75] We pooled the data from these studies[31, 32, 41, 63] via meta-analysis, which showed no difference in operating time (sometimes referred to as surgeon efficiency) between sleep deprived and non-sleep deprived surgeons (Figure 2; n = 50,046, MD -0.14, 95% CI -1.60 to 1.33, P = 0.86, I^2 = 0%). Of studies not meta-analysed, the small (n = 29) before-after study showed no impact of sleep deprivation from shift-work nor of sleep hours on performance on a laparoscopic simulation (LapSimGyn).[34] One small (n = 64) intervention study compared a 24-hour shift to a usual work day, also finding no detriment to performance on a laparoscopic simulation (Minimally Invasive Surgical Trainer-Virtual Reality) despite diminished sleep hours while working on-call.[75] In summary, pooled data from 4 cohort studies (100% low risk of bias) showed no effect of sleep restriction on surgical efficiency. Additional data from one RCT (high risk of bias) and one before-after study (high risk of bias) also showed no association between sleep restriction and performance on laparoscopic simulations.

Two before-after studies (high risk of bias[45, 50]) and five cross-sectional studies (2 low,[43, 59] 3 unclear,[37, 46] 1 high risk of bias [61]) reported on psychomotor performance outcomes among surgeons,[45] anesthesiologists,[37, 43, 50, 59] emergency physicians,[61] and internal medicine physicians.[46] Among a small group of surgeons (n = 9), performance on a virtual ring transfer task deteriorated after an on-call shift (data not reported, P < 0.05).[45] The four studies among anesthetists reported mixed findings. One small (n = 11) before-after study showed longer reaction times (690.8±73.4 vs. 746.5±113.7 milliseconds) and reduced concentration ability (26.4±23.5 vs. 56.3±23.0 on a 100-point scale, P = 0.007) following a 24-hour shift with sleep restriction;[50] Two others found that sleep restriction due to overnight shifts was associated with slower reaction times.[43, 59] Conversely, a small study (n = 11) found no effect of overnight shiftwork with sleep restriction on any measure of psychomotor performance except Hopkin's Verbal Learning Test (t-score of 48.6±7.6 vs. 41.5±9.9, P = 0.04).[37] Among emergency physicians (n = 18), one study (high risk of bias) showed that those who were sleep deprived (<5 hours sleep after a 24-hour shift) had a reduced performance on most but not all psychomotor tests (Battery Test Reaction 5),[61] while among internal medicine physicians (n = 20, low risk of bias), neurocognitive parameters did not seem to worsen post-call.[46] In summary, two before-after (0% low risk of bias) and 5 cross-sectional studies (40% low risk of bias) showed mixed results for the association between fatigue or sleep restriction and psychomotor performance.

Five cross-sectional studies (2 low,[47, 69] 1 unclear,[71] 2 high risk of bias[38, 65]) reported on associations between sleep deprivation or fatigue and work ability or perceived performance, all among mixed groups of physicians.[38, 47, 65, 69, 71] The two large studies at low risk of bias showed that sleep problems and fatigue were inversely associated with physicians' perceived quality of work.[47, 69] Among 1,541 physicians in Finland, sleeping problems (measured by 4 questions from the Jenkins Scale) were inversely associated with scores on the Work Ability Index (β = -0.29, P < 0.001),[47] while a study of 890 physicians from Israel demonstrated that perceived quality of care was predicted by fatigue (1 item on the Shirom-Melamed Burnout Measure) even after controlling for components of burnout (β = 0.17, P < 0.05).[69] Similarly, in one study, comments from senior physicians suggested that continual

tiredness and exhaustion negatively affected their perceived competence.[71] The two studies[38, 65] that were at high risk of bias had conflicting findings. In summary, 5 cross-sectional studies (40% at low risk of bias) reported on perceived work performance; those that were at low risk of bias supported an association between fatigue or sleep restriction and reduced performance.

Five cross-sectional studies (1 low,[44] 2 unclear,[49, 66] 2 high risk of bias[38, 73]) reported on associations between sleep restriction or fatigue and self-reported medical errors among surgeons,[66] anesthesiologists[44] and mixed groups of physicians.[38, 49, 73] A large (n = 7,905) study at unclear risk of bias showed that only 6.9% of surgeons reported fatigue as the most important contributor to medical errors.[66] Among anesthesiologists, a smaller study (n = 183) at low risk of bias showed that the risk of self-reported fatigue-related errors increased with more nights of work-related sleep disturbance (RR 1.25, 95% CI 1.06 to 1.49).[44] Two of the studies reporting on mixed groups of physicians had conflicting results,[38, 49] while another reported that physicians' opinions on the association between fatigue and prescribing errors differed by work setting.[73] One-third (34%) of community-based, 96% of hospital-based, and 8% of office-based physicians believed that there was a high or very high association between fatigue and prescribing errors ($P < 0.05$).[73] In summary, 5 cross-sectional studies (20% at low risk of bias) reported on self-reported errors, and these showed mixed findings for associations with fatigue or sleep restriction.

Patient Outcomes

Six large (n = 270 to 38,978) cohort studies at low risk of bias reported on patient outcomes, all related to surgical[31, 32, 41, 58, 63, 77] or obstetric[58] procedures (Supplementary file 4). In these studies, sleep restriction or fatigue were typically defined as overnight work prior to a daytime procedure[31, 41, 58, 63, 77]; though two studies measured sleep hours[32] or 'sleep opportunity'.[58] We pooled data for procedures performed by sleep deprived versus non-sleep deprived surgeons (or obstetrician-gynecologists in one case[58]). Analyses showed no difference in the rate of intra-operative complications (Figure 3, 3 studies,[58, 63, 77] n = 19,798, RR 1.35, 95% CI 0.82 to 2.21, p=0.24, $I^2 = 82\%$), post-operative complications (Figure 4; 5 studies,[31, 32, 41, 63, 77] n = 60,201, RR 0.99, 95% CI 0.95 to 1.03, p = 0.51, $I^2 = 0\%$), patient mortality (Figure 5; 5 studies,[31, 32, 41, 63, 77] n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, p = 0.82, $I^2 = 0\%$), or length of hospital stay in days (Figure 6; 4 studies,[31, 32, 41, 63] n = 50,046, MD -0.33, 95% CI -1.03 to 0.36, p = 0.35, $I^2 = 86\%$). One study[77] in the mortality analysis reported the number of deaths only as ≤5. We assumed 2 events for this study (midpoint between 0 and

5); sensitivity analysis using the lowest (i.e., 0) and highest (i.e., 5) possible number of events did not change the overall result (Supplementary file 5). We imputed the average variance for one study[32] in the length of stay analysis; sensitivity analysis using either the highest or lowest SD did not change the results (Supplementary file 5). Subgroup analysis by type of surgery did not explain the substantial between-study heterogeneity detected for length of stay, nor intraoperative complications, though it may be noted that the types of complications reported varied by study.

DISCUSSION

Fatigue and chronic sleep restriction are two potential drivers of reduced physician wellbeing[17, 19] that have thus far been understudied in physicians in independent practice. Burnout is becoming increasingly prevalent among physicians,[14-16] and recent research indicates that comprehensive individual- and system-level strategies are needed to address the problem.[6-9, 19, 21] We have systematically reviewed evidence from a heterogeneous array of available studies reporting on diverse outcomes related to physicians in independent practice and their patients. The included studies were often at high or unclear risk of bias, included small samples of physicians, and inconsistently measured and reported exposures and outcomes. The key message gleaned from this review is that despite growing interest in the topic of physician wellness, the robust evidence needed to inform individual and systems-level fatigue management strategies is lacking.

Traditionally, much of the fatigue-related research has focused on hazards to patients. The current review included six cohort studies showing that sleep restriction and/or fatigue did not seem to result in increased rates of patient morality, operative complications, or length of hospital stay. Despite these findings, evidence for psychomotor performance, surgical skills and errors suggest that there is indeed a potential for negative outcomes. The included studies, like many of the others in this and other systematic reviews,[79] employed indirect definitions that make it difficult to classify sleep deprived physicians with certainty. In recent years there has been a shift away from the singular focus on patient safety toward a more comprehensive view that also considers the detrimental effects of fatigue, sleep loss and other occupational hazards on physician wellness.[80] Evidence from this review supports that fatigue and sleep restriction may be negatively associated with physician health and wellbeing. It is now recognized that health systems cannot be sustained by a workforce that is facing an epidemic of burnout.[19, 81, 82]

In light of high rates of burnout, the ongoing dialogue about the need for a cultural shift in the practice of medicine[83, 84] is now more important than ever. Recognition of the potential effects of physician fatigue on patients, physicians, and healthcare systems as a whole must be emphasized at a systemic level, encouraging a shift in which the risks are viewed as unacceptable.[1, 20, 80] Likewise, although research to date has focused largely on individual-level approaches to address burnout, it is now clear that placing the burden of a system-level problem solely on the individual is unlikely to bring about significant and lasting change.[85] Recent research has highlighted physician burnout as a system-driven issue that will require corresponding national-scale multicomponent solutions.[1, 19, 81, 82] As such, in the past several years both the American and Canadian Medical Associations have developed policies and programs that address physician health.[81, 86] The Canadian Medical Association's new policy on physician health calls on broad stakeholder groups (e.g., policymakers, regional health authorities, governments) to take shared responsibility for the health of physicians and to make meaningful and concerted efforts towards promoting a healthy and sustainable workforce.[81]

The most salient finding of this review is that the current evidence is insufficient to inform policy and practice. Correspondingly, a 2016 research summit on physician wellness and burnout outlined the need for timely, relevant and methodologically robust research to inform practice and policy.[21] The findings herein may be used by researchers and practitioners to develop and design methodologically strong research programs related to physician fatigue, inform successful research grant proposals, and lobby healthcare organizations to increase the focus on physician fatigue management programs. It will be important to make use of existing validated measures[87-89] consistently in future research. Identifying outcomes of importance to physicians and their patients should be prioritized, such that these may be collected within intervention studies. Reporting these consistently will allow for the effective synthesis of findings and reduce research waste.[90] Integrated knowledge translation strategies involving multiple stakeholder groups (e.g., physicians, patients, medical schools, physicians' associations and governing bodies, policymakers) may help to ensure that the research is relevant and facilitates decision-making.[91]

Strengths and Limitations

Our systematic review is the first to synthesize evidence on the effects of fatigue and sleep restriction on physicians in independent practice. The review is timely, given recent calls for research into individual and organisational solutions for burnout,[20, 21] and an increased focus on physician health.[80, 81]

While we have identified a diverse body of evidence, we could not draw definitive conclusions due to methodological weaknesses (e.g., 62% at high risk of bias, reliance primarily on cross-sectional designs and uncontrolled studies, subjective measurement of exposures and outcomes, small sample sizes, inclusion of predominantly male physicians within urban settings) and heterogeneous outcome measures in the included studies. Given that the 2017 update search was limited to one database, it is possible that a small number of relevant studies could have been missed. We believe that the likelihood that these might alter the conclusions of the review is low. The findings may have been influenced by publication bias, and may not be generalized to all settings, given our restriction to high income countries.

CONCLUSION

The evidence synthesized in this review suggests that fatigue and sleep restriction are associated with some detrimental physician health and wellbeing outcomes; the evidence for potential associations with performance and safety outcomes was mixed. Meta-analyses for patient outcomes did not show any significant associations with physician sleep deprivation. Our overall confidence in the findings is low, owing to a body of research that is hindered by methodological weaknesses. Further methodologically robust research that includes consistent outcomes that are of interest to physicians and their patients is needed to inform strong practice recommendations and policy decisions.

ACKNOWLEDGMENTS

We would like to thank Dr. Meghan Sebastianski and Andrea Milne for assistance with screening, and Dr. Allison Gates for screening, data verification, and quality assessment of the French studies.

COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Dr. Christopher Simon is employed by the Canadian Medical Association, who provided financial support for the research; there are no other relationships or activities that could appear to have influenced the submitted work.

CONTRIBUTOR STATEMENT

All authors contributed to the conception and design of the project. MG and AW contributed to the acquisition, analysis and interpretation of the data, and drafted the manuscript. RF contributed to acquisition of data. CSa, CSi and MPD contributed to interpretation of data and revised the manuscript for important intellectual content. All authors approved the final version of the manuscript as submitted.

FUNDING STATEMENT

This research was supported by the Canadian Medical Association. The funder had no role in the study design; collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

ROLE OF FUNDERS

Dr. Christopher Simon is employed by the Canadian Medical Association. The remaining authors are independent from the funders. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

TRANSPARENCY DECLARATION

The lead author (MG) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted; and all discrepancies from the study as planned have been explained.

DATA ACCESS STATEMENT

All authors, external and internal, had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the interpretation.

DATA SHARING STATEMENT

The data pertaining to this systematic review are available from the corresponding author upon reasonable request.

FIGURE CAPTIONS

Figure 1. Flow of records through the selection process

Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

Figure 3. Forest plot for intra-operative complications among procedures performed by sleep deprived and non-sleep deprived surgeons or obstetrician-gynecologists

Legend: Rothschild 2009 reported the total number of procedures with complications; Schieman 2008 reported the intraoperative complication rate; Vinden 2013 reported conversion to open procedure

Figure 4. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

Legend: Vinden 2013 reported iatrogenic injuries; Schieman 2008, Govindarajan 2015, and Chu 2011 reported post-operative complication rate; Ellman 2004 reported post-operative complications (other types of complications reported not included in the analysis)

Figure 5. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

Figure 6. Forest plot for patient length of hospital stay (days) among surgeries performed by sleep deprived and non-sleep deprived surgeons

REFERENCES

1. Greig P, Snow R. Fatigue and risk: are train drivers safer than doctors? *BMJ* 2017;359. doi: https://doi.org/10.1136/bmj.j5107
2. Parker JBR. The effects of fatigue on physician performance - an underestimated cause of physician impairment and increased patient risk. *Can J Anaesth* 1987;34(5):489-95.
3. Temple J. Resident duty hours around the globe: where are we now? BMC Med Educ 2014;14(Suppl 1):S8. https://doi.org/10.1186/1472-6920-14-S1-S8
4. Imrie K, Frank J, The National Steering Committee on Resient Duty Hours. Fatigue, risk, & excellence: towards a pan-Canadian consensus on resident duty hours. Ottawa, Ontario: The Royal College of Physicians and Surgeons of Canada, 2013. http://www.residentdutyhours.ca/final_report.php (accessed 12 Jan 2018).
5. Accreditation Council for Graduate Medical Education. History of duty hours. 2017. http://www.acgme.org/What-We-Do/Accreditation/Clinical-Experience-and-Education-formerly-Duty-Hours/History-of-Duty-Hours (accessed 12 Jan 2018).
6. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ* 2015;7(3):349-63.
7. Harris JD, Staheli G, LeClere L, Andersone D, McCormick F. What effects have resident work-hour changes had on education, quality of life, and safety? A systematic review. *Clin Orthop* 2015;473(5):1600-8.
8. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg* 2014;259(6):1041-53.
9. Peets A, Ayas NT. Restricting resident work hours: the good, the bad, and the ugly. *Crit Care Med* 2012;40(3):960-6.
10. Canadian Medical Association. CMA Policy: Management of physician fatigue. Ottawa, Canada: Canadian Medical Association; 2014.
11. Anim M, Markert RJ, Wood VC, Schuster BL. Physician practice patterns resemble ACGME duty hours. *Am J Med* 2009;122:587-93.
12. Transport Canada. Fatigue risk management system for the canadian aviation industry: policies and procedures development guidelines. Ottawa, Canada: Her Majesty the Queen in Right of Canada, as represented by the Minister of Transport, 2008. https://www.tc.gc.ca/eng/civilaviation/publications/TP14576-6042.htm (accessed 12 Jan 2018).

13. Patterson PD, Higgins JS, Lang ES, et al. Evidence-based guidelines for fatigue risk management in ems: formulating research questions and selecting outcomes. *Prehosp Emerg Care* 2017;21(2):149-56.

14. Boudreau RA, Grieco RL, Cahoon SL, Robertson RC, Wedel RJ. The pandemic from within: two surveys of physician burnout in Canada. *Can J Commun Ment Health* 2006;25(2):71-88.

15. Kumar S. Burnout and doctors: prevalence, prevention and intervention. *Healthcare* 2016;4(3):37. doi: 10.3390/healthcare4030037

16. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among us physicians relative to the general us population. *Arch Intern Med* 2012;172(18):1377-85.

17. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. *Acad Psychiatry* 2017. doi: 10.1007/s40596-017-0781-6

18. Fralick M, Flegel K. Physician burnout: who will protect us from ourselves? *CMAJ* 2014;186(10):731. doi: 10.1503/cmaj.140588

19. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. *Mayo Clin Proc* 2017;92(1):129-46.

20. Lemaire JB, Wallace JE. Burnout among doctors. BMJ 2017;358:j3360. doi: 10.1136/bmj.j3360

21. Dyrbye LN, Trockel M, Frank E, et al. Development of a research agenda to identify evidence-based strategies to improve physician wellness and reduce burnout. *Ann Intern Med* 2017;166(10):743-4.

22. Higgins JPT, Green S (editors). The Cochrane handbook for systematic reviews of interventions, version 5.1.0. London, UK: The Cochrane Collaboration, 2011. http://ww.handbook.cochrane.org (accessed 12 Jan 2018).

23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6(7):e1000100. https://doi.org/10.1371/journal.pmed.1000100

24. Pattani R, Wu PE, Dhalla IA. Resident duty hours in Canada: past, present and future. *CMAJ* 2014;186(10):761-5.

25. International Monetary Fund. World economic outlook: too slow for too long. Washington, DC: International Monetary Fund, 2016. http://www.imf.org/external/pubs/ft/weo/2016/01/ (accessed 12 Jan 2018).

26. Cochrane Effective Practice and Organisation of Care. Suggested risk of bias criteria for EPOC reviews. Oslo, Norway: Norwegian Knowledge Centre for the Health Services, 2016. http://epoc.cochrane.org/epoc-specific-resources-review-authors (accessed 12 Jan 2018).

27. Wells GA SB, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa, Canada: The Ottawa Hospital Research Institute, 2014. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 12 Jan 2018).
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-60.
30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.
31. Govindarajan A, Urbach DR, Kumar M, et al. Outcomes of daytime procedures performed by attending surgeons after night work. *N Engl J Med* 2015;373(9):845-53.
32. Chu MW, Stitt LW, Fox SA, Kiaii B, Quantz M, Guo L, et al. Prospective evaluation of consultant surgeon sleep deprivation and outcomes in more than 4000 consecutive cardiac surgical procedures. *Arch Surg* 2011;146(9):1080-5.
33. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59(1):7-10.
34. Amirian I, Andersen LT, Rosenberg J, Gogenur I. Laparoscopic skills and cognitive function are not affected in surgeons during a night shift. *J Surg Educ* 2014;71(4):543-50.
35. Aziz A. Sources of perceived stress among american medical doctors: a cross-cultural perspective. Cross Cultural Management 2004;11(4):28-39.
36. Beaujouan L, Czernichow S, Pourriat JL, Bonnet F. Prevalence and risk factors for substance abuse and dependence among anaesthetists: a national survey. *Ann Fr Anesth Réanim* 2005;24(5):471-9.
37. Chang LC, Mahoney JJ, 3rd, Raty SR, Ortiz J, Apodaca S, De La Garza R 2nd. Neurocognitive effects following an overnight call shift on faculty anesthesiologists. *Acta Anaesthesiol Scand* 2013;57(8):1051-7.
38. Chen I, Vorona R, Chiu R, Ware JC. A survey of subjective sleepiness and consequences in attending physicians. *Behav Sleep Med* 2008;6(1):1-15.
39. Doppia MA, Estryn-Béhar M, Fry C, Guetarni K, Lieutaud T, comité de pilotage de l'enquête SESMAT. Burnout in French doctors: a comparative study among anaesthesiologists and other specialists in French hospitals (SESMAT study). *Ann Fr Anesth Réanim* 2011;30(11):782-94.

40. Dutheil F, Trousselard M, Perrier C, et al. Urinary interleukin-8 is a biomarker of stress in emergency physicians, especially with advancing age--the JOBSTRESS randomized trial. *PLoS One* 2013;8(8):e71658. https://doi.org/10.1371/journal.pone.0071658
41. Ellman PI, Law MG, Tache-Leon C, et al. Sleep deprivation does not affect operative results in cardiac surgery. *Ann Thorac Surg* 2004;78(3):906-11.
42. Elovainio M, Heponiemi T, Jokela M, et al. Stressful work environment and wellbeing: what comes first? *J Occup Health Psychol* 2015;20(3):289-300.
43. Gander P, Millar M, Webster C, Merry A. Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiol Int* 2008;25(6):1077-91.
44. Gander PH, Merry A, Millar MM, Weller J. Hours of work and fatigue-related error: a survey of New Zealand anaesthetists. *Anaesth Intensive Care* 2000;28(2):178-83.
45. Gerdes J, Kahol K, Smith M, Leyba MJ, Ferrara JJ. Jack Barney award: the effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. *Am J Surg* 2008;196(6):813-9.
46. Harbeck B, Suefke S, Haas CS, Lehnert H, Kropp P, Moenig H. No stress after 24-hour on-call shifts? *J Occup Health* 2015;57(5):438-47.
47. Heponiemi T, Puttonen S, Elovainio M. On-call work and physicians' well-being: testing the potential mediators. *Occup Med* 2014;64(5):352-7.
48. Jackson TN, Pearcy CP, Khorgami Z, Agrawal V, Taubman KE, Truitt MS. The physician attrition crisis: a cross-sectional survey of the risk factors for reduced job satisfaction among US surgeons. *World J Surg* 2017;24. doi: 10.1007/s00268-017-4286-y
49. Kaneita Y, Ohida T. Association of current work and sleep situations with excessive daytime sleepiness and medical incidents among Japanese physicians. *J Clin Sleep Med* 2011;7(5):512-22.
50. Lederer W, Kopp M, Hahn O, et al. Post-duty psychomotor performance in young and senior anaesthetists. *Eur J Anaesthesiol* 2006;23(3):251-6.
51. Leichtfried V, Putzer G, Perkhofer D, Schobersberger W, Benzer A. Circadian melatonin profiles during single 24-h shifts in anesthetists. *Sleep Breath* 2011;15(3):503-12.
52. Lindfors PM, Nurmi KE, Meretoja OA, et al. On-call stress among Finnish anaesthetists. *Anaesthesia* 2006;61(9):856-66.
53. Mahmood JI, Stoen Grotmol K, Tesli M, Vaglum P, Tyssen R. Contextual factors and mental distress as possible predictors of hazardous drinking in Norwegian medical doctors: a 15-year longitudinal, nationwide study. *Eur Addict Res* 2017;23(1):19-27.

54. Nishimura K, Nakamura F, Takegami M, et al. Cross-sectional survey of workload and burnout among Japanese physicians working in stroke care: the nationwide survey of acute stroke care capacity for proper designation of comprehensive stroke center in Japan (J-ASPECT) study. *Circ Cardiovasc Qual Outcomes* 2014;7(3):414-22.

55. Pit SW, Hansen V. Factors influencing early retirement intentions in Australian rural general practitioners. *Occup Med* 2014;64(4):297-304.

56. Pit SW, Hansen V. The relationship between lifestyle, occupational health, and work-related factors with presenteeism amongst general practitioners. *Arch Environ Occup Health* 2016;71(1):49-56.

57. Roberts DL, Shanafelt TD, Dyrbye LN, West CP. A national comparison of burnout and work-life balance among internal medicine hospitalists and outpatient general internists. *J Hosp Med* 2014;9(3):176-81.

58. Rothschild JM, Keohane CA, Rogers S, et al. Risks of complications by attending physicians after performing nighttime procedures. *JAMA* 2009;302(14):1565-72.

59. Saadat H, Bissonnette B, Tumin D, et al. Effects of partial sleep deprivation on reaction time in anesthesiologists. *Paediatr Anaesth* 2017;27(4):358-62.

60. Saadat H, Bissonnette B, Tumin D, et al. Time to talk about work-hour impact on anesthesiologists: The effects of sleep deprivation on Profile of Mood States and cognitive tasks. *Paediatr Anaesth* 2016;26(1):66-71.

61. Sanches I, Teixeira F, dos Santos JM, Ferreira AJ. Effects of acute sleep deprivation resulting from night shift work on young doctors. *Acta Med Port* 2015;28(4):457-62.

62. Sargent MC, Sotile W, Sotile MO, Rubash H, Barrack RL. Quality of life during orthopaedic training and academic practice. Part 1: orthopaedic surgery residents and faculty. *J Bone Joint Surg Am* 2009;91(10):2395-405.

63. Schieman C, MacLean AR, Buie WD, Rudmik LR, Ghali WA, Dixon E. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? *Am J Surg* 2008;195(5):684-7.

64. Sende J, Jbeili C, Schvahn S, Khalid M, Asaph J, Romano H, et al. Stress factors and stress consequences among emergency physicians: vational survey. *Annales Francaises de Médecine d'Urgence* 2012;2(4):224-31.

65. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. *Hum Perform Extrem Environ* 2001;6(1):6-11.

66. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg* 2010;251(6):995-1000.

67. Shanafelt TD, Novotny P, Johnson ME, et al. The well-being and personal wellness promotion strategies of medical oncologists in the North Central Cancer Treatment Group. *Oncology* 2005;68(1):23-32.
68. Shanafelt TD, Raymond M, Kosty M, et al. Satisfaction with work-life balance and the career and retirement plans of US oncologists. *J Clin Oncol* 2014;32(11):1127-35.
69. Shirom A, Nirel N, Vinokur AD. Overload, autonomy, and burnout as predictors of physicians' quality of care. *J Occup Health Psychol* 2006;11(4):328-42.
70. Shirom A, Nirel N, Vinokur AD. Work hours and caseload as predictors of physician burnout: the mediating effects by perceived workload and by autonomy. *J Appl Psychol* 2010;59(4):539-65.
71. Smith F, Goldacre MJ, Lambert TW. Adverse effects on health and wellbeing of working as a doctor: views of the UK medical graduates of 1974 and 1977 surveyed in 2014. *J R Soc Med* 2017;110(5):198-207.
72. Starmer AJ, Frintner MP, Freed GL. Work-life balance, burnout, and satisfaction of early career pediatricians. *Pediatrics* 2016;e20153183. doi: 10.1542/peds.2015-3183
73. Tanti A, Camilleri M, Borg AA, et al. Opinions of Maltese doctors and pharmacists on medication errors. *Int J Saf Med* 2017;29(1-2):81-99.
74. Tokuda Y, Hayano K, Ozaki M, Bito S, Yanai H, Koizumi S. The interrelationships between working conditions, job satisfaction, burnout and mental health among hospital physicians in Japan: a path analysis. *Ind Health* 2009;47(2):166-72.
75. Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. *Am J Surg* 2005;189(6):753-7.
76. Vela-Bueno A, Moreno-Jimenez B, Rodriguez-Munoz A, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. *J Psychosom Res* 2008;64(4):435-42.
77. Vinden C, Nash DM, Rangrej J, et al. Complications of daytime elective laparoscopic cholecystectomies performed by surgeons who operated the night before. *Obstet Gynecol Surv* 2014;69(2):71-3.
78. Wada K, Yoshikawa T, Goto T, et al. National survey of the association of depressive symptoms with the number of off duty and on-call, and sleep hours among physicians working in Japanese hospitals: a cross sectional study. *BMC Public Health* 2010;10:127. doi: 10.1186/1471-2458-10-127
79. Sturm L, Dawson D, Vaughan R, et al. Effects of fatigue on surgeon performance and surgical outcomes: a systematic review. *ANZ J Surg* 2011;81(7-8):502-9.

80. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet* 2009;374(9702):1714-21.

81. Canadian Medical Association. CMA policy: physician health. Ottawa, Canada: Canadian Medical Assocation, 2017. http://policybase.cma.ca/dbtw-wpd/Policypdf/PD18-01.pdf (accessed 12 Jan 2018).

82. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Intern Med* 2017;177(2):195-205.

83. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. *N Engl J Med* 2002;347(16):1249-55.

84. Ward S, Outram S. Medicine: in need of culture change. *Itern Med J* 2016;46(1):112-6.

85. Panagioti M, Panagopoulou E, Bower P, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med* 2017;177(2):195-205.

86. American Medical Association. STEPSforward™. Chicago, IL: American Medical Assocation, 2017. https://www.stepsforward.org/modules/joy-in-medicine (accessed 12 Jan 2018).

87. Jackson C. The Chalder Fatigue scale (CFQ 11). Occup Med 2015;65(1):86. doi: 10.1093/occmed/kqu168

88. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *J Psychosom Res* 1993;37(2):147-53.

89. Smets E, Garssen B, Bonke Bd, De Haes J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* 1995;39(3):315-25.

90. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383(9912):166-75.

91. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. *Implement Sci* 2016;11:38. https://doi.org/10.1186/s13012-016-0399-1

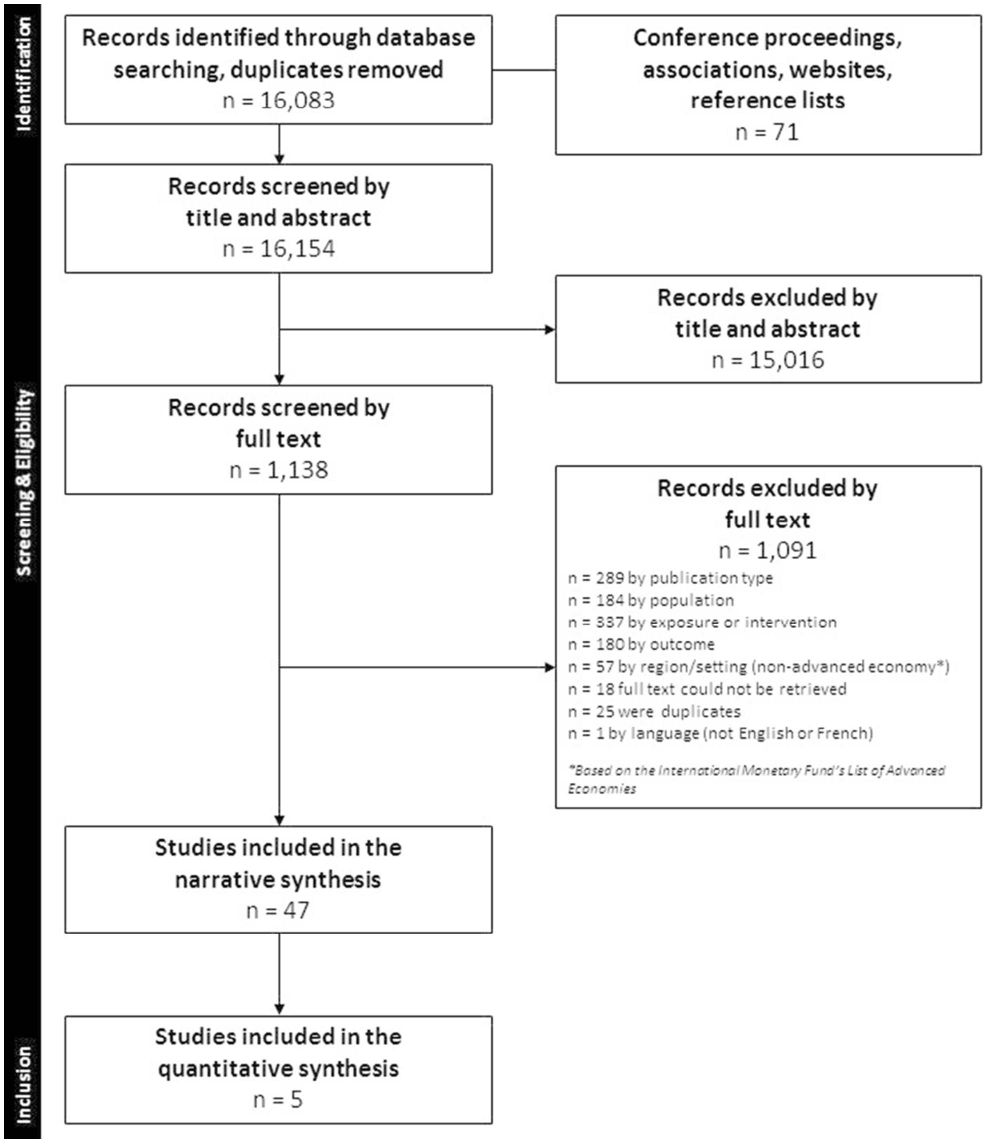


Figure 1. Flow of records through the selection process

60x69mm (300 x 300 DPI)

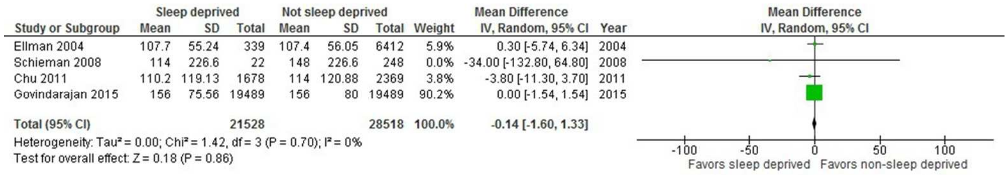


Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

262x46mm (96 x 96 DPI)

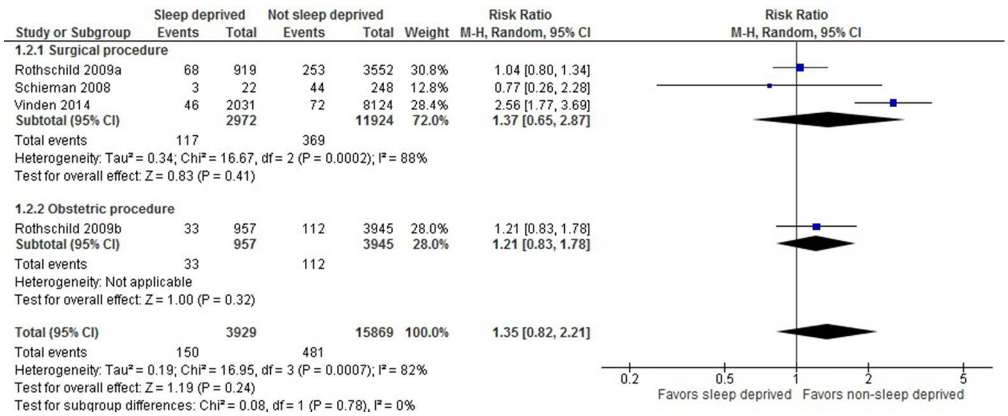
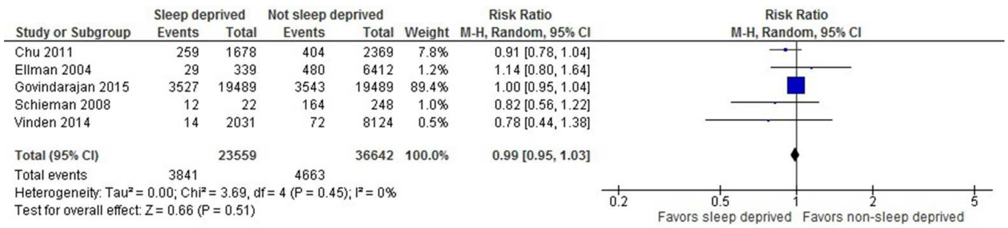


Figure 3. Forest plot for intra-operative complications among procedures performed by sleep deprived and non-sleep deprived surgeons or obstetrician-gynecologists

239x97mm (96 x 96 DPI)



Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI
Chu 2011	259	1678	404	2369	7.8%	0.91 [0.78, 1.04]
Ellman 2004	29	339	480	6412	1.2%	1.14 [0.80, 1.64]
Govindarajan 2015	3527	19489	3543	19489	89.4%	1.00 [0.95, 1.04]
Schieman 2008	12	22	164	248	1.0%	0.82 [0.56, 1.22]
Vinden 2014	14	2031	72	8124	0.5%	0.78 [0.44, 1.38]
Total (95% CI)		**23559**		**36642**	**100.0%**	**0.99 [0.95, 1.03]**
Total events	3841		4663			

Heterogeneity: Tau² = 0.00; Chi² = 3.69, df = 4 (P = 0.45); I² = 0%
Test for overall effect: Z = 0.66 (P = 0.51)

Figure 4. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

239x55mm (96 x 96 DPI)

Figure 5. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

249x55mm (96 x 96 DPI)

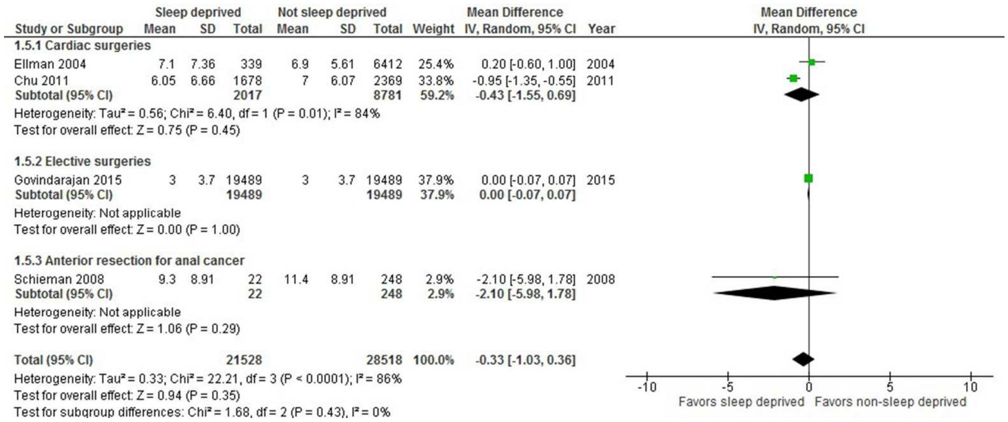


Figure 6. Forest plot for patient length of hospital stay (days) among surgeries performed by sleep deprived and non-sleep deprived surgeons

254x105mm (96 x 96 DPI)

Supplementary file 1. Search Strategy

Database: In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

Date searched: 13 April 2016, updated 7 November 2017

Records retrieved: 5068 and 1442 in the update (removed duplicates retrieved in previous search)

1. Medical Staff, Hospital/
2. Physician Impairment/
3. exp Physicians/
4. allergist*.ti.
5. (an?esthetist* or an?esthesiologist*).ti.
6. cardiologist*.ti.
7. clinician*.ti.
8. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
9. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
10. dermatologist*.ti.
11. endocrinologist*.ti.
12. doctor*.ti.
13. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
14. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
15. family practitioner*.ti.
16. gastroenterologist*.ti.
17. (general practitioner* or GP*).ti.
18. (general adj2 physician*).ti.
19. geriatrician*.ti.
20. gyn?ecologist*.ti.

21. h?ematologist*.ti.

22. (health* adj2 (professional* or provider*)).ti.

23. hospitalist*.ti.

24. (house staff* or housestaff*).ti.

25. intensivist*.ti.

26. internist*.ti.

27. medical professional*.ti.

28. obstetrician*.ti.

29. oncologist*.ti.

30. ophthalmologist*.ti.

31. orthop?edist*.ti.

32. (otolaryngologist* or otorhinolaryngologist*).ti.

33. neonatologist*.ti.

34. nephrologist*.ti.

35. neurologist*.ti.

36. neuropsychiatrist*.ti.

37. neurosurgeon*.ti.

38. p?ediatrician*.ti.

39. perinatologist*.ti.

40. physician*.ti.

41. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

42. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.

43. primary care practitioner*.ti.

44. psychiatrist*.ti.

45. pulmonologist*.ti.

46. rheumatologist*.ti.

47. surgeon*.ti.

48. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

49. traumatologist*.ti.

50. urologist*.ti.

51. or/1-50 [Combined MeSH, title, and text word searches for physicians]

52. Burnout, Professional/

53. exp Circadian Rhythm/

54. exp Fatigue/

55. Occupational Health/

56. Rest/ph, px [Physiology, Psychology]

57. Sleep Deprivation/

58. Sleep Disorders, Circadian Rhythm/

59. Sleep Wake Disorders/

60. exp Stress, Psychological/

61. Workload/px [Psychology]

62. Work Schedule Tolerance/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw,kf.

64. biological rhythm*.tw,kf.

65. (burn out* or burned out* or burnt out* or burnout*).tw,kf.

66. circadian misalignment.tw,kf.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw,kf.

68. exhaust*.tw,kf.

69. fatigu*.tw,kf.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw,kf.

71. tired*.tw,kf.

72. weariness.tw,kf.

73. or/52-72 [Combined MeSH and text words for fatigue]

74. and/51,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (comment or editorial or letter).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. remove duplicates from 80

Database: Ovid Embase 1996 to 2016 Week 15

Date searched: 13 April 2016

Records retrieved: 8859

1. medical staff/
2. exp physician/
3. allergist*.ti.
4. (an?esthetist* or an?esthesiologist*).ti.
5. cardiologist*.ti.
6. clinician*.ti.
7. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
8. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
9. dermatologist*.ti.
10. endocrinologist*.ti.
11. doctor*.ti.
12. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
13. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
14. family practitioner*.ti.
15. gastroenterologist*.ti.
16. (general practitioner* or GP*).ti.
17. (general adj2 physician*).ti.

18. geriatrician*.ti.

19. gyn?ecologist*.ti.

20. h?ematologist*.ti.

21. (health* adj2 (professional* or provider*)).ti.

22. hospitalist*.ti.

23. (house staff* or housestaff*).ti.

24. intensivist*.ti.

25. internist*.ti.

26. medical professional*.ti.

27. obstetrician*.ti.

28. oncologist*.ti.

29. ophthalmologist*.ti.

30. orthop?edist*.ti.

31. (otolaryngologist* or otorhinolaryngologist*).ti.

32. neonatologist*.ti.

33. nephrologist*.ti.

34. neurologist*.ti.

35. neuropsychiatrist*.ti.

36. neurosurgeon*.ti.

37. p?ediatrician*.ti.

38. perinatologist*.ti.

39. physician*.ti.

40. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

41. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

42. primary care practitioner*.ti.

43. psychiatrist*.ti.

44. pulmonologist*.ti.

45. rheumatologist*.ti.

46. surgeon*.ti.

47. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

48. traumatologist*.ti.

49. urologist*.ti.

50. or/1-49 [Combined Emtree, title, and text word searches for physicians]

51. burnout/

52. circadian rhythm/

53. circadian rhythm sleep disorder/

54. fatigue/

55. mental stress/

56. occupational health/

57. sleep deprivation/

58. sleep waking cycle/

59. work capacity/

60. work schedule/

61. working time/

62. workload/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

64. biological rhythm*.tw.

65. (burn out* or burned out* or burnt out* or burnout*).tw.

66. circadian misalignment.tw.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw.

68. exhaust*.tw.

69. fatigu*.tw.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

71. tired*.tw.

72. weariness.tw.

73. or/51-72 [Combined Emtree and text words for fatigue]

74. and/50,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (conference* or editorial or letter or proceeding).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. limit 80 to embase

Database: Ovid PsycINFO 1987 to April Week 1 2016

Date searched: 13 April 2016

Records retrieved: 2094

1. exp Physicians/
2. allergist*.ti.
3. (an?esthetist* or an?esthesiologist*).ti.
4. cardiologist*.ti.
5. clinician*.ti.
6. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
7. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
8. dermatologist*.ti.
9. endocrinologist*.ti.
10. doctor*.ti.
11. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
12. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
13. family practitioner*.ti.
14. gastroenterologist*.ti.
15. (general practitioner* or GP*).ti.
16. (general adj2 physician*).ti.

17. geriatrician*.ti.

18. gyn?ecologist*.ti.

19. h?ematologist*.ti.

20. (health* adj2 (professional* or provider*)).ti.

21. hospitalist*.ti.

22. intensivist*.ti.

23. internist*.ti.

24. medical professional*.ti.

25. obstetrician*.ti.

26. oncologist*.ti.

27. ophthalmologist*.ti.

28. orthop?edist*.ti.

29. (otolaryngologist* or otorhinolaryngologist*).ti.

30. neonatologist*.ti.

31. nephrologist*.ti.

32. neurologist*.ti.

33. neuropsychiatrist*.ti.

34. neurosurgeon*.ti.

35. p?ediatrician*.ti.

36. perinatologist*.ti.

37. physician*.ti.

38. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

39. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

40. primary care practitioner*.ti.

41. psychiatrist*.ti.

42. pulmonologist*.ti.

43. rheumatologist*.ti.

44. surgeon*.ti.

45. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

46. traumatologist*.ti.

47. urologist*.ti.

48. or/1-47 [Combined thesaurus, title, and text word searches for physicians]

49. Compassion Fatigue/

50. Fatigue/

51. Human Biological Rhythms/

52. Occupational Health/

53. Occupational Stress/

54. Sleep/

55. Sleepiness/

56. Working Conditions/

57. Work Rest Cycles/

58. Work Week Length/

59. Work Scheduling/

60. Workday Shifts/

61. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

62. biological rhythm*.tw.

63. (burn out* or burned out* or burnt out* or burnout*).tw.

64. circadian misalignment.tw.

65. ((circadian or diurnam or ultradian) adj rhythm*).tw.

66. exhaust*.tw.

67. fatigu*.tw.

68. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

69. tired*.tw.

70. weariness.tw.

71. or/49-70 [Combined thesaurus and text words for fatigue]

72. and/48,71 [Combined concepts for physicians and fatigue]

73. limit 72 to yr="2000-Current"

74. limit 73 to (english or french)

Database: CINAHL Plus with Full Text (1937 to the present) via EBSCOhost

Date searched: 14 April 2016

Records retrieved: 3378

S1. (MH "Medical Staff, Hospital+")

S2. (MH "Physicians+")

S3. TI allertist*

S4. TI (anesthetist* or anaesthetist* or anesthesiologist* or anaesthesiologist*)

S5. TI cardiologist*

S6. TI clinician*

S7. clinician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S8. clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S9. TI dermatologist*

S10. TI endocrinologist*

S11. TI doctor*

S12. doctor* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S13. doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S14. TI "family practitioner*"

S15. TI gastroenterologist*

S16. TI ("general practitioner*" or GP*)

S17. TI (general N2 physician*)

S18. TI geriatrician*

S19. TI (gynaecologist* or gynecologist*)

S20. TI (haematologist* or hematologist*)

S21. TI hospitalist*

S22. TI ("house staff*" or housestaff*)

S23. TI intensivist*

S24. TI internist*

S25. TI obstetrician*

S26. TI oncologist*

S27. TI ophthalmologist*

S28. TI (orthopaedist* or orthopedist*)

S29. TI (otolaryngologist* or otorhinolaryngologist*)

S30. TI neonatologist*

S31. TI nephrologist*

S32. TI neurologist*

S33. TI neuropsychiatrist*

S34. TI neurosurgeon*

S35. TI (paediatrician* OR pediatrician*)

S36. TI perinatologist*

S37. TI physician*

S38. physician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S39. physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S40. TI "primary care practitioner*"

S41. TI psychiatrist*

S42. TI pulmonologist*

S43. TI rheumatologist*

S44. TI surgeon*

S45. surgeon* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S46. TI traumatologist*

S47. TI urologist*

S48. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28

OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47

S49. (MH "Circadian Rhythm")

S50. (MH "Fatigue")

S51. (MH "Impairment, Health Professional")

S52. (MH "Mental Fatigue")

S53. (MH "Occupational Health")

S54. (MH "Shiftwork")

S55. (MH "Sleep Deprivation")

S56. (MH "Sleep Disorders, Circadian Rhythm")

S57. (MH "Sleep-Wake Transition Disorders")

S58. (MH "Stress, Occupational+")

S59. (MH "Stress, Psychological")

S60. ("24 hour*" or "24 hr*" or "twenty four hour*" or "twentyfour hour*") N1 rhythm*

S61. "biological rhythm*"

S62. "burn out*" or "burned out*" or "burnt out*" or burnout*

S63. "circadian misalignment"

S64. (circadian or diurnam or ultradian) N1 rhythm*

S65. exhaust*

S66. fatigu*

S67. sleep* N3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)

S68. tired*

S69. weariness

S70. S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69

S71. S48 AND S70

S72. S48 AND S70 Limiters - Published Date: 20000101-20161231; Publication Type: Clinical Trial, Journal Article, Meta Analysis, Meta Synthesis, Practice Guidelines, Randomized Controlled Trial, Research, Review, Systematic Review; Language: English, French

Database: PubMed via NCBI Entrez

Date searched: 14 April 2016

Records retrieved: 92

(((("Medical Staff, Hospital"[mh:noexp] OR "Physician Impairment"[mh:noexp] OR "Physicians"[mh] OR allergist[ti] OR allergists[ti] OR anaesthetist[ti] OR anaesthetists[ti] OR anaesthesiologist[ti] OR anaesthesiologists[ti] OR anesthetist[ti] OR anesthetists[ti] OR anesthesiologist[ti] OR anesthesiologists[ti] OR cardiologist[ti] OR cardiologists[ti] OR clinician[ti] OR clinicians[ti] OR ((clinician[tiab] OR clinicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((clinician[tiab] OR clinicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR dermatologist[ti] OR dermatologists[ti] OR endocrinologist[ti] OR endocrinologists[ti] OR doctor[ti] OR doctors[ti] OR ((doctor[tiab] OR doctors[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((doctor[tiab] OR doctors[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "family practitioner"[ti] OR

"family practitioners"[ti] OR gastroenterologist[ti] OR gastroenterologists[ti] OR "general practice physician"[ti] OR "general practice physicians"[ti] OR "general practitioner"[ti] OR "general practitioners"[ti] OR geriatrician[ti] OR geriatricians[ti] OR gynaecologist[ti] OR gynaecologists[ti] OR gynecologist[ti] OR gynecologists[ti] OR haematologist[ti] OR haematologists[ti] OR hematologist[ti] OR hematologists[ti] OR "health care professional"[ti] OR "health care professionals"[ti] AND "health care provider"[ti] OR "health care providers" OR "health professional"[ti] OR "health professionals"[ti] OR "health provider"[ti] OR "health providers"[ti] OR "healthcare professional"[ti] OR "healthcare professionals"[ti] OR "healthcare provider"[ti] OR "healthcare providers"[ti] OR hospitalist[ti] OR hospitalists[ti] OR "house staff"[ti] OR "house staffs"[ti] OR housestaff[ti] OR housestaffs[ti] OR intensivist[ti] OR intensivists[ti] OR internist[ti] OR internists[ti] OR "medical professional"[ti] OR "medical professionals"[ti] OR obstetrician[ti] OR obstetricians[ti] OR oncologist[ti] OR oncologists[ti] OR ophthalmologist[ti] OR ophthalmologists[ti] OR orthopaedist[ti] OR orthopaedists[ti] OR orthopedist[ti] OR orthopedists[ti] OR otolaryngologist[ti] OR otolaryngologists[ti] OR otorhinolaryngologist[ti] OR otorhinolaryngologists[ti] OR neonatologist[ti] OR neonatologists[ti] OR nephrologist[ti] OR nephrologists[ti] OR neurologist[ti] OR neurologists[ti] OR neuropsychiatrist[ti] OR neuropsychiatrists[ti] OR neurosurgeon[ti] OR neurosurgeons[ti] OR paediatrician[ti] OR paediatricians[ti] OR pediatrician[ti] OR pediatricians[ti] OR perinatologist[ti] OR perinatologists[ti] OR physician[ti] OR physicians[ti] OR ((physician[tiab] OR physicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((physician[tiab] OR physicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "primary care practitioner"[ti] OR "primary care practitioners"[ti] OR psychiatrist[ti] OR psychiatrists[ti] OR pulmonologist[ti] OR pulmonologists[ti] OR rheumatologist[ti] OR rheumatologists[ti] OR surgeon[ti] OR surgeons[ti] OR ((surgeon[tiab] OR surgeons[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR

"burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR traumatologist[ti] OR traumatologists[ti] OR urologist[ti] OR urologists[ti]) AND ("Burnout, Professional"[mh:noexp] OR "Circadian Rhythm"[mh] OR "Fatigue"[mh] OR "Occupational Health"[mh:noexp] OR "Rest/physiology"[mh:noexp] OR "Rest/psychology"[mh:noexp] OR "Sleep Deprivation"[mh:noexp] OR "Sleep Disorders, Circadian Rhythm"[mh:noexp] OR "Stress, Psychological"[mh] OR "Workload/psychology"[mh] OR "Work Schedule Tolerance"[mh:noexp] OR "24 hour rhythm"[tiab] OR "24 hour rhythms"[tiab] OR "24 hr rhythm"[tiab] OR "24 hr rhythms"[tiab] OR alertness[tiab] OR "biological rhythm"[tiab] OR "biological rhythms"[tiab] OR "burn out"[tiab] OR "burned out"[tiab] OR "burnt out"[tiab] OR burnout[tiab] OR "circadian misalignment"[tiab] OR "circadian rhythm"[tiab] OR "circadian rhythms"[tiab] OR "diurnal rhythm"[tiab] OR "diurnal rhythms"[tiab] OR exhausted[tiab] OR exhaustion[tiab] OR exhausting[tiab] OR exhausts[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR (("Sleep"[mh:noexp] OR sleep[tiab] OR sleeping[tiab]) AND (deprivation[tiab] OR deprive[tiab] OR deprived[tiab] OR deprives[tiab] OR depriving[tiab] OR disorder[tiab] OR disorders[tiab] OR lack[tiab] OR lacked[tiab] OR lacking[tiab] OR lacks[tiab] OR loss[tiab] AND insufficient[tiab] OR problem[tiab] OR problems[tiab])) OR tired[tiab] OR tiredness[tiab] OR "twenty four hour rhythm"[tiab] OR "twenty four hour rhythms"[tiab] OR weariness[tiab] OR "ultradian rhythm"[tiab] OR "ultradian rhythms"[tiab])) NOT (((Animals[MESH] OR Animal Experimentation[MESH] OR "Models, Animal"[MESH] OR Vertebrates[MESH]) NOT (Humans[MESH] OR Human experimentation[MESH])) OR (((animals[tiab] OR animal model[tiab] OR rat[tiab] OR rats[tiab] OR mouse[tiab] OR mice[tiab] OR rabbit[tiab] OR rabbits[tiab] OR pig[tiab] OR pigs[tiab] OR porcine[tiab] OR swine[tiab] OR dog[tiab] OR dogs[tiab] OR hamster[tiab] OR hamsters[tiab] OR chicken[tiab] OR chickens[tiab] OR sheep[tiab]) AND (publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb])) NOT (human[ti] OR humans[ti] OR people[ti] OR children[ti] OR adults[ti] OR seniors[ti] OR patient[ti] OR patients[ti])))) NOT (editorial[pt] OR comment[pt] OR letter[pt] OR newspaper article[pt])) AND ((publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR (pubstatUSheadofprint))

Filters activated: Publication date from 2000/01/01 to 2016/12/31, English, French.

Supplementary table 1. Descriptive characteristics of the included studies

Study	Physician and patient characteristics				Setting		Interventions or	Outcomes
Country	Type	n=	Sex (% male)	Age	Location	Urban or rural	exposures	
Observational (exposure) studies (n=45)								
Cohort design								
Chu, 2011 [32] Canada	Surgeons	6	NR	Range: 32-55y	Tertiary care academic hospital	Urban	Sleep restriction due to work on the night preceding surgery	Length of surgery; patient postoperative mortality, complications, length of stay
	Patients: cardiac surgery cases	4,047	NR	NR				
Ellman, 2004 [41] US	Surgeons	NR	NR	NR	University hospitals	Urban	Sleep restriction due to work on the night preceding surgery	Length of surgery; patient complications, in-hospital mortality, length of stay, need for blood products
	Patients: adult cardiac surgery cases	6,751	70%	S: 63.4±0.7y C: 63.5±0.1y				
Govindarajan, 2015 [31] Canada	Surgeons	1,448	NR	46.3±8.7	Academic and non-academic hospitals	Mixed	Sleep deprivation due to work on the night preceding a daytime surgery	Length of surgery; Patient complications, mortality, readmissions, length of stay
	Patients: surgical cases	38,978	NR	56.4±16.6y				
Rothschild, 2009 [58] US	Surgeons Obstetrician/gynecologists	220	Surgeons: 84% OB/GYNs: 28%	Surgeons: 42.0±7.6y OB/GYNs: 42.0±9.0y	Tertiary care academic trauma centre/referral centre for high-risk obstetrics	Urban	Sleep deprivation due to work on the night preceding a daytime procedure	Patient complications, preventable complications
	Patients: surgical and obstetrics cases	Surg.: 4,471 Obst.: 4,902	Surg: S: 25% C: 28% Obst.: S: 0% C: 0%	Surg: S: 49.1±16.3y C: 50.0±16.3y Obst.: S: 32.9±5.2y C: 33.5±5.0y				
Schieman, 2007 [63] Canada	Colorectal surgeons	NR	NR	NR	University teaching hospitals	NR	Fatigue due to work on the night preceding surgery	Length of surgery; patient operative complications, length of stay, mortality, cancer recurrence
	Patients: undergoing anterior resection for rectal cancer	270	NR	S: 64.5y C: 64.4y				

Study Country	Physician and patient characteristics Type	 n=	 Sex (% male)	 Age	Setting Location	 Urban or rural	Interventions or exposures	Outcomes
Vinden, 2014 [77] Canada	General surgeons	331	83%	48±10y	Community hospitals	Mixed	Sleep deprivation due to overnight work preceding daytime surgery	Patient mortality, operative complications
	Patients: Elective cholecystectomies	10,390	S: 27% C: 26%	S: 49±16y C: 49±16y				
Before-after design								
Amirian, 2014 [34] Denmark	Surgeons	29	55%	Median: 35y Range: 27-49y	Academic hospital	Urban	17-h night shift with sleep deprivation	Cognitive and psychomotor abilities on a laparoscopic simulation
Gerdes, 2008 [45] US	Surgeons	9	NR	NR	University Hospital	Urban	Fatigue; sleep deprivation from overnight call shift	Cognitive and psychomotor abilities
Lederer, 2006 [50] Austria	Senior anesthetists	11	82%	49.0±2.0y	Hospital	Urban	Sleep deprivation from 24-h call shift	Concentration ability; reaction time; performance on psychometric tasks
Time series design								
Leichtfried, 2011 [51] Austria	Anesthetists	10	100%	Mean: 32y Range: 29-35y	University Hospital	Urban	Sleep deprivation from 24-h shift; sleepiness, sleep hours	Melatonin metabolite profile
Cross-sectional design								
Aziz, 2004 [35] US	Family medicine physicians Various specialties	153	NR	NR	Hospitals	NR	Fatigue	Stress
Beaujouan, 2005 [36] France	Anesthesiologists	3,476	64%	≤35y: 9% 36-45y: 28% 46-55y: 49% 56-65y: 13%	Public sector General hospitals University hospitals Private hospitals	NR	Sleep deprivation	Substance abuse
Chang, 2013 [37] US	Anesthesiologists	11	64%	Mean: 38y IQR: 34-48y	Level 1 trauma centre	NR	Sleep deprivation due to 15-h overnight call shift; sleepiness	Cognitive performance; reaction time

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Chen, 2008 [38] US	Psychiatrists Internists General practitioners Surgeons Obstetrician-gynecologists Radiologists Pediatricians Other	180	77%	Academic: 79% 36-55y Private practice: 73% 36-65y	Medical school Private practices	Urban	Sleep deprivation; sleepiness	Impact on personal and professional life; perceived risk of errors
Doppia, 2011 [39] France	Anesthesiologists	565	64%	<35y: 11% 35-54y: 63% >55y: 25%	Public hospitals Private hospitals Work-health environments Public health units	NR	Sleep deprivation	Burnout
Elovaino, 2015 [42] Finland	Physicians in various specialties	1,524	40%	Median: 49.7y Range: 24-69y	Hospitals Primary care Private practice Other unspecified	NR	Sleep difficulties	Job demands and control
Gander, 2000 [43] New Zealand	Anesthetists	183	NR	Mean: 46y	Combined public/private practice Other unspecified	NR	Work-related sleep disturbance	Risk of fatigue-related errors
Harbeck, 2015 [46] Germany	Internists	20	45%	Median: 32y Range: 26-42y	Hospital	NR	Sleep disturbance due to a 24-call shift	Biochemical and physiological parameters; neurocognitive function
Heponiemi, 2014 [47] Finland	Physicians in various specialties Non-specialized physicians	1,541	40%	49.80±9.49y, Range: 24-67y	Hospitals Primary care clinic Private practice Other unspecified	NR	Sleep difficulties	Job satisfaction; work ability; psychological distress

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Jackson, 2017 [48] US	Surgeons in various subspecialties	993	61%	More; less satisfied: 30-39y: 23%;24% 40-49y: 32%;36% 50-59y: 23%;27% ≥60y: 23%;14%	Academic practice Non-academic practice	NR	Not feeling well rested	Job satisfaction
Kanieta, 2011 [49] Japan	Internists Surgeons Orthopedics Pediatricians Obstetrician-gynecologists Psychiatrists Dermatologists Urologists Opthalmologists Otorhinolaryngologists Other	3,486	66%	20-39y: 11% 40-49y: 25% 50-59y: 28% 60-69y: 16% ≥70y: 21%	Hospitals Clinics Other unspecified	NR	Sleep deprivation and difficulties; insomnia	Medical incidents
Lindfors, 2006 [52] Finland	Anesthetists	328	53%	47±7.8y Range: 32-69y	University hospitals Central and district hospitals Private sector	NR	Sleep disturbances; sleepiness	Stress; suicidal tendencies
Mahmood, 2016 [53] Norway	Generalists Internists Pediatricians Surgical specialties Anesthesiologists	450 (all time points)	41%	43y±2.8y	Public health system Private practice	NR	Sleep deprivation due to on-call shifts	Alcohol misuse

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Nishimura, 2014 [54] Japan	Neurosurgeons and neurologists	2,564	NR	NR	Stroke care centres Teaching hospitals	NR	Sleep deprivation	Burnout
Pit, 2014 [55] Australia	General practitioners	92	60%	50±10.7y	NR	Rural	Work-related sleep disturbance	Early retirement intentions
Pit, 2016 [56] Australia	General practitioners	92	60%	50±10.7y	Private (solo) practice Group practice	Rural	Work-related sleep disturbance	Sickness presenteeism
Roberts, 2014 [57] US	General internists Internal medicine hospitalists	578	58%	Hospitalists: 46.9±12.4y Generalists: 53.6±10.2y	Private practice Academic medical centre Veterans hospital Military practice Other	NR	Fatigue	Falling asleep while driving
Saadat, 2016 [60] US	Anesthesiologists	21	71%	30-40y: 57% 41-50y: 19% 51-55y: 24% Range: 32-56y	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Mood disturbances
Saadat, 2017 [59] US	Anesthesiologists	21	65%	Range: 32-56 years	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Reaction time
Sanches, 2015 [61] Spain	Emergency medicine physicians	18	28%	29.2±2.6y	Central hospital	NR	Sleep deprivation	Cognitive and psychomotor abilities
Sargent, 2009 [62] US	Orthopedic surgeons	264	92%	NR	Orthopedic surgery training programs	NR	Sleep deprivation	Burnout; psychological distress; marital satisfaction
Sende, 2012 [64] France	Emergency physicians	318	62%	39±8y	Hospitals Mobile emergency services Other unspecified	NR	Fatigue; sleep deprivation	Stress

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Sexton, 2001 [65] US	Consulting physicians: Surgeons Anesthesiologists Pulmonary physicians Cardiologists Pediatricians	271	NR	NR	Teaching and non-teaching hospitals	Urban	Fatigue	Perceived performance effectiveness
Shanafelt, 2005 [67] US, Canada, Mexico	Oncologists	241	85%	>50y: 51%	Community clinics Hospitals Private practice Academic medical centres	NR	Fatigue; sleep deprivation	Quality of life/well-being
Shanafelt, 2010 [66] US	Surgeons	7,905	87%	Median: 51y Q1: 43y Q2: 59y	Private practice Academic medical centres Veterans hospital Active military practice Retired or not in practice Other	NR	Fatigue	Perceived major medical errors
Shanafelt, 2014 [68] US	Oncologists	1,117	52%	Median: 52y	Private practice Academic practice Veteran's hospital Industry, other	NR	Fatigue	Satisfaction with work-life balance
Shirom, 2006 [69] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Perception of quality of patient care

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Shirom, 2010 [70] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Burnout
Smith, 2017 [71] UK	General practitioners Surgeons Other unspecified specialties	3,550	63%	NR	NR (varied)	NR	Perceived fatigue; sleep deprivation	Physical and mental health; competence
Starmer, 2016 [72] US	General pediatricians Pediatric surgeons Pediatric hospitalists Pediatric specialists (unspecified)	840	40%	NR	NR (some in private practice)	NR	Sleep deprivation	Burnout; balanced personal and professional commitments; life and career satisfaction
Tanti, 2017 [73] Malta	Physicians (unspecified)	204	62%	Median: 41y	Hospitals Community Office-based	NR	Fatigue	Prescribing errors
Tokuda, 2009 [74] Japan	Hospital physicians: Generalists Other unspecified specialties	236	75%	40.9±7.8y Range: 26-76y	Hospitals with ≥20 inpatient beds	NR	Sleep deprivation	Burnout; job satisfaction
Vela-Bueno, 2008 [76] Spain	Primary care physicians	113	27%	41.4±8.0y	Primary care centres	Urban	Sleep problems, insomnia	Burnout
Wada, 2010 [78] Japan	Physicians (unspecified)	3,862	78%	M: 75% 30-59y F: 85% 30-59y	Hospitals	NR	Sleep deprivation	Depressive symptoms

Study Country	Physician and patient characteristics Type	 n=	 Sex (% male)	 Age	Setting Location	 Urban or rural	Interventions or exposures	Outcomes
Non-comparative design								
Gander, 2008 [43] New Zealand	Anesthetists	20	85%	Median: 44y	Hospitals	Urban	Sleep disturbance from consecutive working days on-call work	Psychomotor performance
Intervention studies (n=2)								
Randomized controlled trials								
Dutheil, 2013 [40] France	Emergency physicians	17	35%	39.1y±6.9y	University hospital	Urban	Fatigue related to 14-h and 24-h shifts; sleep deprivation; low sleep quality;	Perceived stress; urine interleukine-8
Uchal, 2005 [75] Norway	Surgeons Gynecologists Orthopedic surgeons Urologists Vascular surgeons	64	67%	Median: Post-call: 33.0y Post-work: 38.0y	Government hospitals	NR	Sleep deprivation due to 24-h call shift	Product quality, procedure effectiveness of a surgical simulation

C: control group; F: female; h: hour(s); IQR: interquartile range; M: male; NR: not reported; S: study group; SD: standard deviation; Surg: surgical; Obst: obstetric; Q: quartile; UK: United Kingdom; US: United States of America; y: year(s)

Supplementary file 3. Risk of bias assessments

Summary of risk of bias assessments for randomized controlled trials (n=2)[a]

First Author, Year	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias	Overall risk of bias[b]
Dutheil, 2013	Low	Unclear	High	High	Low	Low	High	High
Uchal, 2005	Low	Low	Unclear	Low	Low	Low	Low	Unclear

[a]Assessed using the Cochrane Collaboration's Risk of Bias Tool

[b]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cohort studies (n=6)[a]

First Author, Year	Selection					Comparability		Outcome				Total
	Representa-tiveness of exposed cohort /1	Selection of non-exposed cohort /1	Ascertain-ment of exposure /1	Outcome not present at start /1	Total /4	Compara-bility of cohorts /2	Total /2	Assess-ment of outcome /1	Adequate length of follow-up /1	Adequate follow-up of cohorts /1	Total /1	Score[b] /9
Chu, 2011	1	1	0	1	3	2	2	1	1	1	3	8
Ellman, 2004	1	1	1	1	4	1	1	1	1	1	3	8
Govindarajan, 2015	1	1	1	1	4	2	2	1	1	1	3	9
Rothschild, 2009	1	1	1	1	4	2	2	1	1	1	3	9
Schieman, 2008	1	1	1	1	4	1	1	1	1	1	3	8
Vinden, 2014	1	1	1	1	4	1	1	1	1	1	3	8

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale

[b]An overall score of 7 to 9 stars is considered as low risk of bias, 4 to 6 as unclear risk of bias, and 3 or less as high risk of bias

Summary of risk of bias assessments for before-after studies (n=3)[a]

First Author, Year	Random sequence generation[b]	Allocation concealment[b]	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Amirian, 2014	NA	NA	High	High	Low	Low	High	High
Gerdes, 2008	NA	NA	High	High	Low	Low	High	High
Lederer, 2006	NA	NA	High	High	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for before-after studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as 'not applicable' (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of risk of bias assessments for time series studies (n=1)[a]

First Author, Year	Intervention independent of other changes	Intervention effect pre-specified	Intervention unlikely to affect data collection	Allocation concealment[a]	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Leitchfried, 2011	Low	High	Low	NA	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for interrupted time series studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as not applicable (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cross-sectional studies (n=34)[a]

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Aziz, 2004	0	0	0	0	1	0	1	1
Beaujouan, 2005	1	0	1	0	1	0	1	2
Chang, 2013	1	0	1	0	1	1	2	3
Chen, 2008	1	0	1	0	1	0	1	2
Doppia, 2011	1	1	2	0	1	1	2	4
Elovaino, 2015	1	1	2	0	1	1	2	4
Gander, 2000	1	1	2	0	1	1	2	4
Harbeck, 2015	1	0	1	0	1	1	2	3
Heponiemi, 2014	1	1	2	0	1	1	2	4
Jackson, 2017	0	0	0	0	1	0	1	1
Kanieta, 2011	1	0	1	0	1	1	2	3
Lindfors, 2006	1	1	2	0	1	1	2	4
Mahmood, 2017	1	0	1	0	1	0	1	2
Nishimura, 2014	1	1	2	0	1	0	1	3
Pit, 2014	1	0	1	0	1	1	2	3
Pit, 2016	1	0	1	0	1	1	2	3
Roberts, 2014	1	1	2	0	1	0	1	3
Saadat, 2016	1	1	2	0	1	1	2	4
Saadat, 2017	1	1	2	0	1	1	2	4
Sanches, 2015	1	0	1	0	1	0	1	2
Sargent, 2009	1	0	1	0	1	0	1	2

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Sende, 2010	1	0	1	0	1	0	1	2
Sexton, 2001	1	0	1	0	1	0	1	2
Shanafelt, 2005	1	0	1	0	1	1	2	3
Shanafelt, 2010	1	1	2	0	1	0	1	3
Shanafelt, 2014	1	0	1	0	1	1	2	3
Shirom, 2006	1	1	2	0	1	1	2	4
Shirom, 2010	1	1	2	0	1	1	2	4
Smith, 2016	1	0	1	0	1	1	2	3
Starmer, 2016	1	1	2	0	1	1	2	4
Tanti, 2017	1	0	1	0	1	0	1	2
Tokuda, 2009	1	1	2	0	1	1	2	4
Vela-Bueno, 2008	1	1	2	0	1	1	2	4
Wada, 2010	1	1	2	0	1	0	1	3

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted for cross-sectional studies

[b]An overall score of 4 to 5 stars is considered as low risk of bias, 3 as unclear risk of bias, and 2 or less as high risk of bias. For response rate, ≥50% was used as the criterion to be awarded a star

Summary of quality assessments for non-comparative studies (n=1)[a]

First Author, Year	Selection			Exposure		Outcome				Total
	Adequacy of case definition /1	Representat-iveness of the sample /1	Total /2	Ascertain-ment of exposure	Total /1	Assessment of outcome /1	Same method of assessment for entire sample /1	Loss to follow-up /1	Total /3	Score[b] /6
Gander, 2008	1	1	2	0	0	0	1	1	2	4

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted by the authors to be suitable to the non-comparative design

[b]An overall score of 5 to 6 stars is considered as low risk of bias, 3 to 4 as unclear risk of bias, and 2 or less as high risk of bias

Supplementary file 4. Detailed study outcomes

Physician health and wellness outcomes and associations with fatigue

Study	Study	Exposures or interventions		Outcomes	Associations between exposure and outcome
Risk of Bias (RoB)	design	Assessment measure and time points	Baseline	Assessment measure and time points	
Surgeons					
Jackson, 2017 RoB: high	CS	Not feeling well rested: self-reported as 'unhealthy' Time points NR	71% healthy, 28% unhealthy in terms of being well rested	Job satisfaction: Abridged Job in General Scale; grouped into more or less satisfied using the median Time points NR	Job satisfaction in those more vs. less satisfied: Healthy (well rested): 85% vs. 58%, p<0001; Unhealthy (not well rested): 15% vs. 42%, p<0.001.
Nishimura, 2014 RoB: unclear	CS	Sleep hours/night: self-reported (continuous) Time points NR	Mean±SD sleep: 5.94±1.08h	Burnout: Japanese MBI (severe: EE >4.0 and either DP >2.6 or PE <4.17) Time points NR	1) Mean±SD sleep for not burned out vs. mild to moderate vs. severe: 6.07±1.15 vs. 5.88±0.94 vs. 5.63±0.94, p<0.05; 2) Association between sleep and burnout (OR (95% CI)): bivariate 0.67 (0.61-0.73), p<0.001; multivariate including work characteristics and mental health: 0.84 (0.75-0.94), p=0.002.
Sargent, 2009 RoB: high	CS	Sleep deprivation: self-reported on a 4-point scale (none, a little, quite a bit, a lot) Time points NR	21% none, 48% a little, 23% quite a bit, 8% a lot	Burnout: MBI (norms NR); Marital satisfaction: RDAS; Psychological morbidity: GHQ-12 score ≥4 Time points NR	1) Positive correlation between sleep deprivation and EE, DP, psychological distress, lower marital satisfaction, all p<0.001. No relationship with PA.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift with on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Stress during duty: 4-point scale from 'calm' to 'very demanding' Assessed post-duty	1) Mean stress score during duty: 2.1.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Leitchtfried, 2011 RoB: high	TS	24-h shift; Sleepiness: ESS (range: 0-24); Sleep hours: self-reported (continuous) Sleepiness assessed pre-shit, sleep hours pre, during and post-shift	ESS (mean (range)): 7.4 (4-12); Mean±SD sleep hours: 1) pre-study: 7.74±1.35h; 2) Pre-24-h shift (11h00 on day 1: 0.13±0.35h, 19:00 on day 1: 6.99±0.68h); 3) During the 24-h shift (07h00 on day 2: 0.0±0.0h, 19h00 on day 2, 5.49±1.95h); 4) Post-24-h shift (11h00 on day 3: 0.5±0.71h, 19h00 on day 3: 7.06±1.18h).	aMT6-s: urinalysis Assessed at 4-h intervals from 07:00 to 11:00	1) aMT6-s over shift, mean (95% CI): higher at 11:00AM pre- (12.2 (6.3-8.1)) and post-shift (9.3 (3.7-14.9)) vs. during, p=0.016; 2) Correlations between sleep and aMT6-s (data NR): mild for sleep duration the night prior with aMT6-s at 3PM the following day; sleep on night 2 with aMT6-s at 3PM the next day; total sleep with aMT6-s at 11AM on third day; moderate for sleep on first night with aMT6-s at 7AM and 11AM pre-shift, 11PM during 24-h shift and 11AM post-shift; total sleep pre-shift and nocturnal sleep during 24-h shift with aMT6-s at 11PM during shift; total sleep with aMT6-s at 3PM on first and second day, 11PM on second day; 3) Correlations between ESS and aMT6-s: moderate for aMT6-s at 7AM during shift, 11AM on day off.
Beaujouan, 2005 RoB: high	CS	Sleep deprivation: 4-point scale (always, frequently, rarely, never) Time points NR	48.8% always or frequently feel sleep deprived	Substance abuse: 93-item addiction and substance abuse questionnaire Time points NR	1) 60.6% with drug dependence vs. 46.0% of those without reported sleep difficulties, p<0.001. 2) OR (95% CI) of addiction for frequently/always vs. rarely/never sleep deprived: tobacco 1.42 (1.04-1.94); tranquilizer/hypnotics 3.26 (2.12-5.02).
Doppia, 2011 RoB: low	CS	Insufficient sleep: 4-point scale (no, not really, sort of, yes) Time points NR	28.9% reported insufficient sleep during work time	Burnout: CBI (mild: 1-2.4, moderate: 2.5-3.5, severe: 3.6-5) Time points NR	1) Frequency of burnout by response for sleep sufficiency: 47.6% for no/not really, 16.3% for sort of/yes, p<0.001.
Lindfors, 2006 RoB: low	CS	Sleep hours/day: self-reported to the nearest 0.5h; Adequacy of sleep and rest: self-reported (yes/no)	Sleep hours (mean (range)): 7 (5-9)	Stress: MOSQ on a 3-point scale (no, to some extent, clearly); Thoughts of suicide: 4-point scale (‘never’ to ‘have tried’)	1) Sleep sufficiency predicted stress symptoms: bivariate β=-0.362, p<0.001; multivariate including gender, sick leave, suicide β=-0.269, p<0.001; 2) Sleep disturbance associated with thoughts of suicide, p=0.009.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Time points NR	
Saadat, 2015 RoB: low	CS	Sleep deprivation (<7h/24-h) due to 17-h overnight shift; Sleepiness and alertness: VAS from 0 (not at all) to 100 (extremely) All assessed on a regular day and a post-call day	Mean±SD sleepiness on a regular day vs. post-call day: 2.99±2.18 vs. 6.79±2.30, p<0.001	Simple cognitive tests: VAS from 0 (not at all) to 100 (extremely); Mood disturbance: PMS (scoring NR) All assessed on a regular day and a post-call day	Regular day v. post-call day, mean±SD scores: 1) Simple cognitive tests: energetic 6.04±2.27 vs. 2.53±1.87, confident 7.03±1.83 vs. 4.98±2.29, irritable 2.03±1.94 vs. 4.86±2.16, sleepy 2.99±2.18 vs. 6.79±2.30, talkative 4.46±1.74 vs. 2.41±1.97, all p<0.001; jittery 1.44±1.74 vs. 3.12±2.34, p=0.003; anxiousness ns; 2) PMS: tension 13.48±2.71 vs. 15.43±4.46, p=0.049; anger 15.24±4.41 vs. 18.14±5.92, p=0.005; fatigue 10.14±2.63 vs. 20.05±6.87, p<0.001; confusion 10.57±1.69 vs. 12.57±4.24, p=0.025; vigor 24.05±6.75 vs.16.67±5.70, p<0.001; depression: ns; total mood disturbance: 42.57±15.26 vs. 70.90±6.91, p<0.001.
ER or ICU physicians					
Dutheil, 2013 RoB: high	RCT	14-h or 24-h shift; Sleep hours: self-reported sleep and wake time; Sleep quality: VAS from 1 (low) to 100 (high); Mental and physical fatigue: VAS from 1 (low) to 100 (high) Assessed on day prior to shift; during shift; each day of protocol (work, off, clerical, control)	1) Sleep duration and quality lower during shifts (14h and 24h) than any other day, and lower during the 24-h vs. 14-h shift (p<0.05); 2) Mental and physical fatigue higher after 14-h and 24-h shift vs. control day (data NR).	Stress: VAS from 0 (low) to 100 (high); IL-8: urinalysis Assessed at 08:30 and 18:30 on each day of protocol	1) Stress: higher following 14-h and 24-h shifts vs. the control day, p<0.05 (data NR); 2) IL-8: higher following 24-h shift vs. control (p=0.007) and 14-h shift (p=0.015); ns difference between 14-h shift and control day; 3) Correlations with IL-8: sleep hours pre-24-h shift, r=-0.627, p=0.007; poor sleep quality during 14-h and 24-h shifts, r=0.452, p=0.031; 4) Multivariable regression: 24-h shift increased IL-8 by 1.9ng vs. control day, p=0.007; ns association with 14-h shift, mental or physical fatigue, sleep deprivation, 14-h shift.
Sende, 2012 RoB: high	CS	Fatigue and sleep deprivation as sources of stress	NR	Most important sources of stress among 4 categories (work-related, patient-	1) 78% indicated that sleep loss and fatigue were sources of stress.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	 Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		related, organizational, individual) Time points NR	
Generalists[b]					
Harbeck, 2015 RoB: unclear	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Biochemical (laboratory values) and physiological (heart rate variability, skin resistance, blood pressure) stress parameters Assessed before a normal day shift, and after a 24-h on call shift	Before a normal shift vs. after overnight call shift: 1) Biochemical parameters: no changes in any parameter except for thyroid stimulating hormone which was higher after the on-call shift (p = 0.049, data NR); 2) Physiological parameters: no significant changes in any parameter
Pit, 2014 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day' Time points NR	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Early retirement (<65 years) intentions (yes/no) Time points NR	For sleep disturbance a few times a year to every day vs. never: 1) Intention to retire early: 74% vs. 26%, p<0.01; 2) Association with intention to retire early (OR (95% CI)): univariate 3.6 (1.47-8.80), p<0.01; multivariate including work, occupational, individual factors 2.91 (1.11-7.6), p<0.05; 4) RR (95% CI) for intention to retire early: 2.0 (1.18-3.49); attributable fraction: 50.0%; population attributable fraction: 37.1%.
Pit, 2016 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day'	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Sickness presenteeism: 'yes' response indicated 1 or more days	For sleep disturbance a few times a year to every day vs. never: 1) Sickness presenteeism: 32% vs. 68%, p=0.018;

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Assessed for the past 12 months	2) Association with sickness presenteeism (OR (95% CI)): 2.92 (1.19-7.16), p=0.02.
Roberts, 2014 RoB: unclear	CS	Fatigue: LAS from 0 (low) to 10 (high) Assessed for the past week	Mean (SD) score: 5.8 (2.4) for hospitalists; 5.9 (2.4) for general internists	Impact of fatigue on daily activities (falling asleep while driving) (yes/no) Time points NR	1) 8.7% of hospitalists and 4.3% of outpatient general internists had fallen asleep while driving due to fatigue.
Vela-Bueno, 2008 RoB: low	CS	Sleep Quality: PSQI (Spanish): score ≥5 indicates low quality (range; 0 to 21); Insomnia: DSM-IV criteria Time points NR; insomnia symptoms in past month	Prevalence (% (95% CI)): 1) Sleep-onset latency >30 minutes: 8.4 (4.8-11.9); 2) Wake time after sleep onset >30 minutes: 15.4 (10.8-19.9); 3) Early morning awakening: 22.5 (19.5-30.4); 4) Nonrestorative sleep: 22.5 (17.2-27.7); 5) Daytime impairment for ≥5 days in past month: 14.2 (9.7-18.6); 6) Insomnia: 18.8 (13.8-23.7).	Burnout: PBM with a 7-point scale from 1 (never) to 7 (always) Time points NR	Low vs. high burnout, mean±SD: 1) Global PSQI: 2.72±2.22 vs. 7.24±4.17, p<0.001; 2) PSQI subscores: sleep quality: 0.54±0.57 vs. 1.40±0.83, p<0.001; sleep latency: 0.51±0.80 vs. 1.38±1.03, p=0.002; sleep duration: 0.45±0.64 vs. 1.16±0.92, p=0.003; sleep efficiency: 0.21±0.57 vs. 0.77±0.98, p=0.018; sleep disturbance: ns; use of medication: 0.14±0.49 vs. 0.57±0.83, p=0.032; daytime dysfunction: 0.52±0.73 vs. 1.57±0.88, p=0.002. 3) Prevalence (95% CI) of insomnia symptoms: sleep latency: 5.5% (2.5-11.5%) vs. 21.1% (10.5-31.6%), p=0.015; wake time >30 min after sleep onset: 9.4% (1.6-17.1%) vs. 25.5% (14.2-37.7%), p=0.029; early awakening: 14.5% (5.1-23.8%) vs. 45.6 (32.7-58.4%), p<0.001; somewhat/very dissatisfied with sleep: 5.5% (2.5-11.5%) vs. 50% (37.1-62.8%), p<0.001; day impairment: 5.5% (2.5-11.5%) vs. 38.2% (25.6-50.7%), p<0.001; insomnia: 7.3% (0.4-14%) vs. 39.7% (27.1-52.2%), p<0.001.
Oncologists					
Shanafelt, 2005 RoB: unclear	CS	Fatigue: LASA QOL ≤7; Sleep deprivation: 10-point Likert scale from 0 (not at all) to 10 (stressful as can be)	75% had a high level of fatigue; Mean±SD sleep score: 4.5±2.65.	Wellbeing: 10-item LASA QOL, high ≥8 vs. low ≤7 Time points NR	1) Sleep deprivation for high vs. low overall well-being (mean±SD): 3.9±2.57 vs. 5.1±2.60, p=0.0004; 2) Lower fatigue predicted overall wellbeing in a multivariate model including personal and professional characteristics, p=0.002.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR			
Shanafelt, 2014 RoB: unclear	CS	Fatigue: 10-point LAS (lower scores indicate greater fatigue) Time points NR	Mean±SD fatigue score: 5.7±2.4	Satisfaction with WLB: 5-point Likert scale from 'strongly agree' to 'strongly disagree' Time points NR	1) OR (95%CI) of lower satisfaction predicted by high fatigue (vs. not) in multivariate model including personal and work-related factors, and burnout: 0.489 (0.337-0.710), p<0.001.
Mixed groups of physicians					
Aziz, 2004 RoB: high	CS	Working while fatigued: 5-point scale from 'extreme' to 'a little' Time points NR	NR	Stress: 47-item questionnaire with a 5-point scale from 'extreme' to 'a little' Time points NR	1) Sources of stress: working while fatigued had a mean±SD score of 2.44±1.20, factor loading: 0.653, in factor analysis; 2) Inverse correlation between stress and working while fatigued: r=-0.270 (significance level NR).
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: worried about having a car accident while driving home post-call: 5.4 vs. 7.0, p<0.001; sleep loss has a major impact on personal life: 8.4 vs. 7.0, p=0.01; 3) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Elovaino, 2015 RoB: low	CS	Sleeping problems: Jenkins Scale with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006 and 2010	Mean±SD score: 2006: 2.30 (1.00); 2010: 2.35 (1.05).	Jobs demands: 5 items scored on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree); Job control: 3 items derived from the Karasek Job Questionnaire	There was no association between sleeping problems in 2006 and job demands or control in 2010.

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Heponiemi, 2014 RoB: low	CS	Sleeping problems: Jenkins Scale[81] with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Psychological distress: GHQ-12 with a 4-point scale (low to high); Job satisfaction: JDS with a Likert scale from 1 (strongly disagree) to 5 (strongly agree) Assessed in 2010	1) Sleeping problems associated with job satisfaction, β=-0.12, p<0.001, psychological distress, β=0.18, p<0.001; 2) Total indirect effect of on-call duty through two mediators (sleeping problems, work interference with family) (R^2 (95% CI)): job satisfaction 0.06 (-0.059, -0.016), p<0.001; psychological distress 0.16 (0.023, 0.081), p<0.001.
Mahmood, 2016 RoB: high	CS	Sleep deprivation: self-reported mean hours of sleep when on call Assessed at 4 years, 10 years, and 15 years post-graduation	Mean±SD hours: 4 years: 4.52 (2.79); 10 years: 5.38 (6.36); 15 years: 6.41 (7.14).	Alcohol use disorders: Modified 9-item version of the Alcohol Use Disorder Identification Test (AUDIT) ≥6 for men and ≥5 for women. Assessed at 4 years, 10 years, and 15 years post-graduation	There was no association between hours of sleep when on call and hazardous drinking behaviours (p=0.732)
Shirom, 2010 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Burnout: SMBM on a 7-point scale from 1 (almost never) to 7 (always)	1) Correlation between physical fatigue subscale and overall burnout: 0.88, p<0.05; 2) In a predictive structural model for burnout, physical fatigue accounted for unique variance in the burnout items, not accounted for by total burnout (R^2=0.24).
Smith, 2017 RoB: unclear	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Mental and physical illness: self-reported via open-ended comments Time points NR	Some physicians reported developing mental illness (e.g., bipolar disorder, alcohol misuse) due to tiredness and stress at work; others developed physical health problems due to sleep deprivation, poor eating habits and lack of exercise.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Starmer, 2016 RoB: low	CS	Sleep deprivation: <7 hours sleep in a typical 24-h period (self-reported) Time points NR	27.7% sleep deprived	Burnout, satisfaction with career and life, balanced personal and professional commitments: Each on a 5-point Likert scale (strongly agree to strongly disagree) Time points NR	≥7-h vs. <7-h sleep: 1) Burnout (% strongly agree/agree): 26.4% vs. 39.6%, p<0.05; career satisfaction (% strongly agree/agree): ns; life satisfaction (% completely/very satisfied): 76.4% vs. 55.9%, p<0.05; balanced personal and professional commitments (% completely/very satisfied): 49.7% vs. 26.1%. 2) <7-h sleep (vs. ≥7-h) (OR, 95% CI) associated with life satisfaction 0.44 (0.29-0.67), p<0.05; balanced personal/professional commitments 0.46 (0.31-0.71), p≤0.05, in a model including work and personal factors.
Tokuda, 2009 RoB: low	CS	Sleep hours/day: self-reported (continuous) Time points NR (included weekday and weekends)	Mean±SD (range) sleep hours/day: 6±0.9 (3-8)	Burnout: MBI (Japanese) with a 7-point Likert scale: 0 (none) to 6 (every day); Job satisfaction: JHPSS with a 5-point Likert scale: 1 (strongly disagree) to 5 (strongly agree) Time points NR	Maximum likelihood estimates±SE: 1) Sleeping time to job satisfaction: group 0.990±0.458, p=0.031; ns for men; women 1.711±0.805, p=0.034; 2) Sleeping time to EE: group -0.219 ±0.070, p=0.002; men -0.215±0.082, p=0.009; ns for women.
Wada, 2010 RoB: unclear	CS	Sleep hours/day: Self-reported (continuous) Assessed for past month when not completing overnight work	<5 hours: 8.7% men, 9.9% women; 5 to <6 hours: 32.3% men, 34.6% women; 6 to <7 hours: 46.0% men, 43.7% women; ≥7 hours: 13.0% men, 11.8% women.	Depression: QIDS-SR; Japanese score <5 (no symptoms) to >20 (very severe symptoms) Assessed for past 7 days	1) Sleep hours for those with vs. without depressive symptoms: <5: 18.7% vs. 7.7% men, 20.5% vs. 8.7% women; 5 to <6: 33.7% vs. 32.2% men, 38.6% vs. 34.2% women; 6 to <7: 35.1% vs. 46.9% men; 31.8% vs. 45.1% women; 2) Association between <5h sleep (vs. 6-7h) and depressive symptoms (OR (95% CI)): univariate 2.79 (1.96-3.95) for men, 2.65 (1.47-4.78) for women; multivariate (including age and workload

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					factors) 2.70 (1.82-4.03) for men, 2.38 (1.11-5.10) for women.

[a]Includes studies of anesthetists, where these were physicians.

[b]Includes primary care physicians, internal medicine physicians, and general practitioners.

AM: morning; aMT6-s: melatonin metabolite; BA: before-after; CI: confidence interval; CBI: Copenhagen Burnout Inventory; CS: cross-sectional; DP: depersonalization; DSM: Diagnostic and Statistical Manual of Mental Disorders; EE: emotional exhaustion; ER: emergency; ESS: Epworth Sleepiness Scale; GHQ: General Health Questionnaire; h: hour(s); ICU: intensive care unit; IL-8: interleukin-8; JDS: Job Diagnostic Survey; JHPSS: Japanese Hospital Physicians Satisfaction Scale; LAS: linear analog scale; LASA: linear analog assessment scales; MBI: Maslach Burnout Inventory; MOSQ: Modified Occupational Stress Questionnaire; min: minute(s); NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PA: personal achievement; PBM: Pines Burnout Measure; PE: professional efficacy; PM: afternoon; PMS: Profile of Mood States; PSQI: Pittsburgh Sleep Quality Index; QIDS-SR: Quick Inventory Depressive Scale – Self-Reported; QOL: Quality of Life; RCT: randomized controlled trial; RDAS: Revised Dyadic Adjustment Scale; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SMBM: Shirom-Melamed Burnout Measure; TS: time series; US: United States of America; VAS: visual analog scale; vs.: versus; WLB: work-life balance

Performance and safety outcomes related to fatigue or sleep loss among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Exposures or intervention: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Uchal, 2005 RoB: unclear	RCT	Sleep deprivation from a 24-h call shift vs. 8-h work; Sleep hours: self-reported (continuous); Sleepiness: ESS (moderate: 10-15, severe: ≥16) Assessed post-call and post-work	Median (range) sleep hours: 1.5 (0-3) post-call vs. 6.5 (5-9) post-work, p<0.05; Median ESS score: 7.0 post-call vs. 5.5 post-work, ns.	Surgical performance: laparoscopic surgical simulator(Minimally Invasivs Surgical Trainer-Virtual Reality) for product quality, procedure effectiveness Assessed post-call and post-work	Post call vs. post-work: 1) Product quality: no difference in accuracy error, tissue damage, leak rate; 2) Procedure effectiveness: no difference in goal-directed actions, non-goal directed actions, operating time.
Chu, 2011 RoB: low	CO	Sleep deprivation: self-reported hours, moderate (3-6h) or severe (<3h) Assessed the night before surgery	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived and 1,595 (39.4%) moderately sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	For 0-3 vs. 3-6 vs. >6 hours of sleep: no difference in CABG or ACC.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and another case in the next 24-h	Of 6,751 procedures, 339 (5%) performed by sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	Sleep deprived vs. non-sleep deprived: no difference in CABG or ACC.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Surgical performance: duration of surgery	Sleep deprived vs. non-sleep deprived: no difference in duration of surgery, even after stratification by type of procedure.
Amirian, 2014 RoB: high	BA	17-h night call shift; Sleep hours during the shift: Wrist-mounted Micro-Mini-Motionlogger; Sleepiness: KSS	Naps pre-call: 11 (37%) napped for median (IQR) 90 (58-128) min; Median (IQR) sleep: 91 (62-123) min on the pre-call night vs. 430 (329-449) on	Surgical performance: LapSimGyn laparoscopic simulation for time, blood loss, instrument path; D2 test of attention and concentration	Pre- vs. post-call: 1) LapSimGyn: no difference in total time, blood loss, instrument path length, instrument angular path; napping did not affect performance;

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Exposures or intervention: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
		Assessed on pre-call and on-call day; sleepiness assessed during shift	the on-call night, p<0.001; Sleep on-call: 12 (40%) slept for median (IQR) 98 (39-135) min; Significant development of sleepiness during shift (p<0.001), plateau score of 7 at 04:00 to 08:00.	Assessed on pre-call and on-call day	2) D2 test: improvement in concentration, p<0.05. No changes in any other parameters; 3) ns difference in laparoscopic simulation time in those who slept during the shift vs. not.
Gerdes, 2008 RoB: high	BA	On-call shift; Fatigue: questionnaire designed by Behrenz & Monga, 1999; Sleep hours: self-reported (continuous) Assessed in 3 sessions pre- and post-call	Fatigue differential from pre- to post-call (range): 1-7 (units unclear); Sleep during call (range): 1-5h	Psychomotor performance: virtual ring transfer task for gesture-level proficiency, hand movement smoothness, tool movement smoothness, elapsed time Assessed in 3 sessions pre- and post-call	1) Pre- to post-call: decrease in all measures of psychomotor proficiency (p<0.05, data NR) except elapsed time; no change in number of psychomotor errors; increase cognitive errors (p<0.05, data NR); 2) Cognitive errors increased exponentially as fatigue ratings increased (R^2=0.9219) and as hours of sleep declined (R^2=0.933).
Shanafelt, 2010 RoB: unclear	CS	Degree of fatigue as a contributor to errors (self-reported) Assessed for the past 3 months	NR	Perceived recent major medical errors (self-reported) Assessed for the past 3 months	1) Prevalence of perceived recent major medical error: 8.9%; 2) Of those reporting an error, 6.9% listed degree of fatigue as the greatest contributing factor.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift, on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Psychomotor performance: reaction time, critical flicker fusion, response measure, peripheral awareness; Concentration ability: scale of 0 (low tiredness) to 100 (maximum tiredness) Assessed pre- and post-duty	Pre- vs. post-duty, mean±SD: 1) Psychometric testing: recognition reaction time (ms): 439.6±50.8 vs. 480.3±58.9; motor reaction time (ms): 252.8±39.3 vs. 465.4±65.0; total reaction time (ms): 690.8±73.4 vs. 746.5±113.7; critical flicker fusion (Hz): 29.0±2.3 vs. 28.7±3.7; response measure (pixels): 647.8±126.7 vs. 598.3±138.1,

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					peripheral awareness task recognition time: 58.9±59.2 vs. 51.6±47.5; 2) Concentration ability: 26.4±23.5 vs. 56.3±23.0, p=0.007.
Chang, 2013 RoB: unclear	CS	15-h in-house overnight call; Sleepiness pre-call: ESS ≥9; Sleep hours: self-reported (continuous) Sleepiness assessed pre-call, sleep hours during call	Median (IQR) ESS: 9 (9), 64% scored ≥9; Median (IQR) hours slept during shift: 1 (0-3).	Psychomotor performance: reaction time; CCPT II; N-back; HVLT (3 trials of 12 words) Assessed at baseline and pre- and post-call	1) Afternoon baseline vs. pre-call: no difference in reaction time, CCPT, N-back, of HVLT; Morning baseline vs. post-call: 1) No change in auditory or visual reaction time; 2) CCPT (t-scores): No change in detectability, response style, hit reaction time, omissions/commissions; 3) N-back % accuracy: no change for auditory, visual, or mean N-value; 4) HVLT (t-score): mean for trials 1-3: 48.6±7.6 vs. 41.5±9.9 (p=0.04); delayed recall: ns; 5) No correlation between ESS scores pre-call or sleep during shift and any measure of psychomotor performance.
Gander, 2000 RoB: low	CS	Nights of work-related sleep disturbance: self-reported (continuous) Assessed for the past 6 months	NR	Risk of fatigue-related errors: questionnaire modelled after Gravenstein et al., 1990 Assessed for the past 6 months	1) Risk of fatigue-related errors increased with increasing nights of work-related sleep disturbance: RR: 1.25, 95% CI: 1.06-1.49.
Saadat, 2017 RoB: low	CS	Sleep deprivation due to an overnight call shift	NR	Reaction time: PVT Assessed after an overnight call shift and the morning of a regular (non-call) day	Mean (SD) reaction time was slower post-call (297.76 (83.75)) vs. on a regular day (266.58 (38.35)), p=0.047.

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Gander, 2008 RoB: unclear	NC	Sleep loss across consecutive working days or on-call work: Wrist-mounted Actiwatch (Mini Mitter, Bend, Oregon, US), sleep and duty diary Assessed over a 2-week period including a weekend of rostered shifts or on-call	≥2 hours sleep <baseline: 8% of 24-h periods that included day work vs. 14% that included day + call; Sleep hours: mean 0.6h less sleep when working day shifts (p=0.014) and 0.8h less sleep when working day shifts + call (p=0.013) vs. off.	Psychomotor performance: PVT Assessed within 2 hours pre- and post-call	1) In fixed model analysis for reaction time including sleep, time since waking, work hours: acute sleep loss associated with slower median reaction time, $F_{(1,184)}$=5.70, p<0.05; longer time since waking associated with poorer performance on the slowest 10%, $F_{(1,185)}$=5.13, p<0.05; 2) Reaction time across 12 consecutive work days: no change in pre-duty reaction times but post-duty reaction times slowed linearly, median -3.38, p<0.001; decline in performance across 10 minutes became progressively steeper both pre- and post-duty, p=0.020.
ER or ICU physicians					
Sanches, 2015 RoB: high	CS	Acute sleep deprivation (<5h of night sleep after a night shift of 12h) Sleep hours: 7-day Actigraphy via SenseWear® Pro2 Armband; Sleepiness: ESS; Sleep quality: PSQI Assessed the week and night before the psychomotor tests	Non-sleep deprived vs. sleep deprived: PSQI >5: 0% vs. 33%, ns; ESS≥10: 11% vs. 67% Sleep time (mean±SD) in week before tests: duration and number of naps higher in sleep deprived group, but diurnal sleep hours lower, 428.6±30.1 vs. 375.8±55.9, p=0.038; Sleep quality (mean±SD): week before tests: 3.3±0.7 vs. 2.6±0.3, p=0.013; night before tests: 3.1±0.8 vs. 1.9±1.0, p=0.020.	Psychomotor performance via Battery Test Reaction 5 (v1): StimulTest, InstrucTest, MovemTest; TP test of visual attention Assessed on morning after night shift 8	Sleep deprived group vs. non-sleep deprived, mean±SD: 1) InstrucTest: correct answers: 169.4 (16.0) vs. 148.3 (28.3), p=0.070; wrong answers: ns; perfection index (%): 99.6 (0.3) vs. 98.9 (1.3), p=0.021; response latency (sec/click): ns; 2) StimulTest: correct answers: 170.7 (21.9) vs. 145.1 (17.9), p=0.022; wrong answers: ns; perfection index (%): ns; response latency (sec/click): 1.06 (0.1) vs. 1.24 (0.1), p=0.022; 3) MovemTest: ns for any parameter; 4) TP: omitted symbols: 34.2±18.4 vs. 62.7±44.0, p=0.034; concentration index (%): 14.1±8.9 vs. 30.0±25.9, p=0.019; quality index (%): 13.8±8.6 vs. 29.2±26.4, p=0.031; correct/wrong symbols: ns; Correlations between sleep and tests: 1) TP for sleep hours nights 1-6: omitted symbols: r=-0.686, p=0.011 for non-sleep-

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					deprived, ns for sleep-deprived; concentration index (%): r=-0.359, p=0.037 for sleep-deprived, ns for non-sleep deprived; r=-0.359, p=0.037 for the group; no other significant correlations; 2) No correlation between PSQI, ESS and any of the psychomotor tests.
Generalists[b]					
Harbeck, 2015	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Neurocognitive parameters: computerized attentional test (vigilance, alertness); D2 letter cancellation test (divided attention); Trail Making Test (visual attention, task switching); Digit Span, Digit Symbol Substitution Test, Weschler Memory Scale (memory functions) Assessed before a normal day shift, and after a 24-h on call shift	Intrinsic alertness, focused attention and vigilance were similar on both occasions; Phasic alertness improved following the on-call shift: mean (SD) 24.8 (15.6) vs. 38.3 (21.5), p = 0.022.
Mixed specialties or undefined populations					
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: written an incorrect order: 8.8 vs. 7.3, p=0.02; might fall asleep while examining a patient: 13.2 vs. 7.7, p=0.001; look forward to sleeping at grand rounds: 10.4 vs. 7.4, p=0.002;

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					3) No difference in ESS score for those who agree/strongly agree vs. other response: work is unaffected by sleep loss and fatigue, thinking is unaffected by sleep loss, sleep loss and fatigue affect my medical decisions, have heard of others making medical errors due to sleep loss and fatigue, never make errors in prescriptions on post-call days, have made medical errors because of sleep loss and fatigue; 4) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Heponiemi, 2014 RoB: low	CS	Sleeping problems: 4-item Jenkins Scale on 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Work ability: Work Ability Index on scale from 1 (could not work at all) to 10 (best work ability) Assessed in 2010	1) On-call duty had an indirect effect on work ability (R^2=0.11, 95% CI: -0.122, -0.031, p<0.001) through two mediators (work interference with family, sleeping problems); 2) Sleeping problems inversely associated with work ability, β=-0.29, p<0.001.
Kanieta, 2011 RoB: unclear	CS	Sleep hours: self-reported (continuous) Sleepiness and sleep difficulties: 5-point scale from 1 (never) to 5 (always); Insomnia: ≥3 sleep difficulties Assessed for the past month	Insufficient rest: 32.5%; Daytime sleepiness: 3.5%; Insomnia: 20.0%; Sleep time (mean±SD min): 279.8±60.9	Self-reported medical incidents: 4-point scale from 1 (never) to 4 (often) Assessed for the past month	1) Prevalence of medical incidents (% (95% CI)): sleep deprived (26.8% (24.2, 29.4)) vs. not (15.2% (13.7, 16.7)), p<0.01; insomnia (24.8% (21.6, 28.0)) insomnia vs. not (17.6% (16.2, 19.0)), p<0.01; ≥6h sleep (18.3% (16.8, 19.8)) vs. <6h (21.7% (18.8, 24.6)), p=0.03; 2) Predictors of medical incidents in multivariate model including personal and work-related factors (OR (95% CI)): lacking rest due to sleep deprivation vs. not (1.65 (1.33-2.04)), p<0.01); insomnia vs. not (1.45 (1.16-1.82), p<0.01); ns for sleep hours.

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Sexton, 2001 RoB: high	CS	Fatigue as a factor impacting performance Time points NR	NR	Performance effectiveness measured by 1 question: agree, neutral, disagree Time points NR	1) "When fatigued, I perform effectively during critical phases of operations/patient care": Anesthetic: 47% agree; 15% neutral; 38% disagree; Surgical: 70% agree; 12% neutral; 18% disagree.
Shirom, 2006 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Quality of care: Adapted 15-item SERVQUAL with a 5-point Likert scale from 1 (very small extent) to 5 (very large extent) Time points NR	1) Quality of care positively predicted by fatigue in a model incorporating several other components of burnout, β=0.17, $p<0.05$.
Smith, 2017 RoB: moderate	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Perceived competence: self-reported via open-ended comments Time points NR	Some physicians indicated that continual tiredness and exhaustion led to concerns that it would affect their competence; some felt that professional performance was compromised at times of physical and mental fatigue.
Tanti, 2017 RoB: high	CS	Fatique: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	NR	Prescribing errors: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	Perception of the contribution of fatigue to prescribing errors differed by physician type ($p<0.05$): 34% of community doctors, 96% hospital doctors, 8% of office-working doctors perceived a very high or high association between fatigue and prescribing errors.

[a]Includes studies of anesthetists, where these were physicians.
[b]Includes primary care physicians, internal medicine physicians, and general practitioners.
ACC: aortic cross-clamp time; BA: before-after; CABG: cardiopulmonary bypass time; CCPT II: Connor's Continuous Performance Test II; CI: confidence interval; CO: cohort; CS: cross-sectional; ER: emergency; ESS: Epworth Sleepiness Scale; h: hour(s); HVLT: Hopkin's Verbal Learning Task; Hz: Hertz; ICU: intensive care unit; IQR: interquartile range; KSS: Karolinska Sleep Scale; min: minutes; ms: millisecond(s); N-back: Dual N-back test; NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; PVT: Psychomotor vigilance Performance Task; RR: risk ratio; RCT: randomized controlled trial; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SERVQUAL: Service Quality Measure; SMBM: Shirom-Melamed Burnout Measure; TP: Toulouse-Piéron test; TS: time series; US: United States of America; vs.: versus

Patient outcomes related to fatigue or sleep restriction among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Surgeons					
Chu, 2011 RoB: low	CO	Sleep deprivation: moderate (3-6 h) or severe (<3-h) sleep deprivation the night before surgery (self-reported hours)	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived, 1,595 (39.4%) by moderately sleep-deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) 0-3 vs. 3-6 vs. >6 hours of sleep: No difference in incidence of mortality, incidence of 10 major complications (except septicemia, 3.6% vs. 0.9% vs. 0.8%, p=0.03), ICU length of stay; in-hospital length of stay (days): 7.0 vs. 6.0 vs. 7.0, p<0.001.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and performed a subsequent case in the next 24-h	Of 6,751 procedures, 339 (5%) were performed by sleep deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, need for blood products, complications (operative, neurologic, renal, infectious, pulmonary), in-hospital length of stay.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Chart review: mortality, surgical complications, readmission, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, surgical complications, readmissions within 30 days, or length of stay.
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse surgical complications Assessed during and post-surgery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total number of complications, preventable complications, type of complications; 2) Operating room procedures with complications, OR (95% CI): 8.5% for 0-6h sleep vs. 3.1% for >6h sleep, 2.70 (1.13-6.48), p=0.03; 3) All procedures with complications, OR (95% CI): 6.2% for 0-6h sleep vs. 3.4% for >6h sleep, 1.72 (1.02-2.89), p=0.04.

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Schieman, 2007 RoB: low	CO	Fatigue: surgeon billed for clinical work after 22:00 the night before surgery	Of 270 procedures, 22 (8%) were performed by fatigued surgeons	Chart review: surgical complications, length of stay, mortality, cancer recurrence Assessed during and post-surgery	1) Fatigued vs. non-fatigued surgeons: no difference in intra- or post-operative complication rate, length of stay, in-hospital length of stay, cancer recurrence.
Vinden, 2014 RoB: low	CO	Sleep deprivation (at risk): surgeon worked 00:00 to 07:00 and performed surgery 07:00 to 18:00	Of 94,183 surgeries, 2,078 (2.2%) were performed by surgeons who were 'at risk'	Chart review: conversion to open procedure (from laparoscopic), iatrogenic injuries, mortality Assessed during and post-surgery	1) At risk vs. not at risk surgeon: no difference in incidence of conversion to open procedure, iatrogenic injuries, mortality, in either univariate or multivariate analyses.
Obstetricians					
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse obstetric complications Assessed during and post-delivery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total complications, preventable complications, type of complications; 2) No association between sleep deprivation and proportion of procedures with complications, nor difference for 0-6h vs. >6h of sleep opportunity.

CI: confidence interval; CO: cohort; h: hours; NR: not reported; OR: odds ratio; RoB: Risk of Bias; SD: standard deviation; US: United States of America; vs.: versus

Supplementary file 5. Statistical analyses

Dichotomous outcomes

Outcome or subgroup	Number of studies	Number of participants	Pooled risk ratio (95% CI)	I^2
1.1 Patient mortality	5	60,436	0.98 (0.84, 1.15)	0%
1.2 Intra-operative complications	3	19,798	1.35 (0.82, 2.21)	82%
1.2.1 Surgical procedure	3[a]	14,896	1.37 (0.65, 2.87)	88%
1.2.2 Obstetric procedure	1[a]	4,902	1.21 (0.83, 1.78)	NA
1.3 Post-operative complications	5	60,201	0.99 (0.95, 1.03)	0%

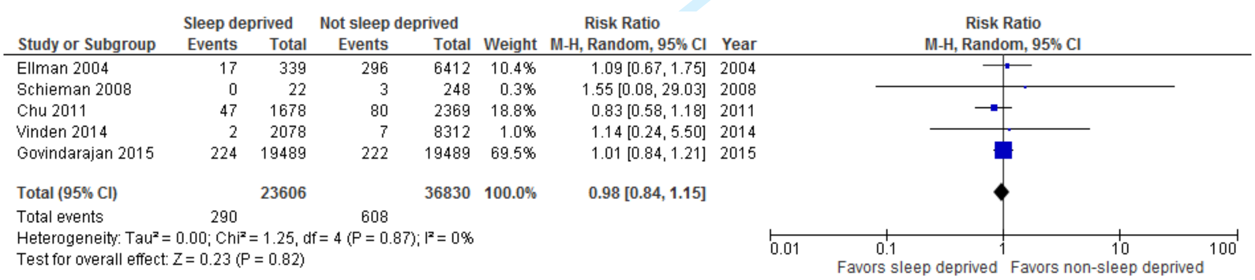
[a] Rothschild, 2009 is represented in both analyses

Continuous outcomes

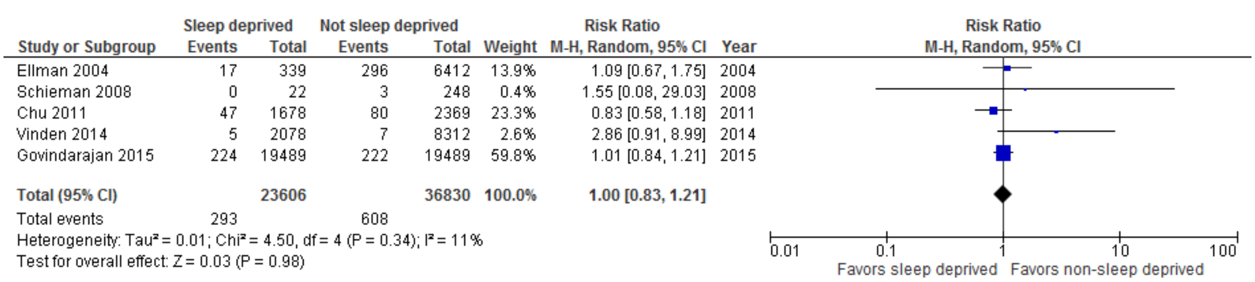
Outcome or subgroup	Number of studies	Number of participants	Pooled mean difference (95% CI)	I^2
1.4 Operating time (minutes)	4	50,046	-0.14 (-1.60, 1.33)	0%
1.5 Length of hospital stay (days)	4	50,046	-0.33 (-1.03, 0.36)	86%
1.5.1 Cardiac surgeries	2	10,798	-0.43 (-1.55, 0.69)	84%
1.5.2 Elective surgeries	1	38,978	0.00 (-0.07, 0.07)	NA
1.5.3 Anterior resection for anal cancer	1	270	-2.10 (-5.98, 1.78)	NA

CPBT: cardiopulmonary bypass time; NA: not applicable

1.1 Patient mortality

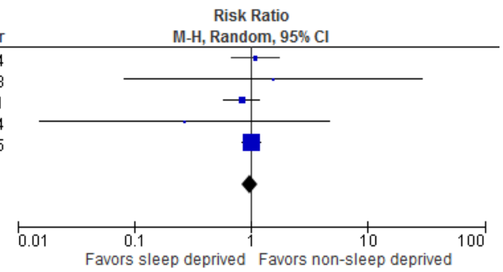


Sensitivity analysis using highest possible number of events for Vinden 2014



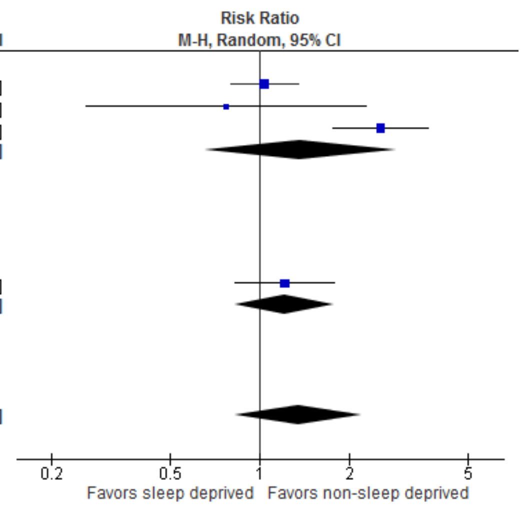
Sensitivity analysis using lowest possible number of events for Vinden 2014

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Ellman 2004	17	339	296	6412	10.5%	1.09 [0.67, 1.75]	2004
Schieman 2008	0	22	3	248	0.3%	1.55 [0.08, 29.03]	2008
Chu 2011	47	1678	80	2369	18.9%	0.83 [0.58, 1.18]	2011
Vinden 2014	0	2078	7	8312	0.3%	0.27 [0.02, 4.67]	2014
Govindarajan 2015	224	19489	222	19489	70.0%	1.01 [0.84, 1.21]	2015
Total (95% CI)		**23606**		**36830**	**100.0%**	**0.98 [0.84, 1.14]**	
Total events	288		608				
Heterogeneity: Tau² = 0.00; Chi² = 2.01, df = 4 (P = 0.73); I² = 0%							
Test for overall effect: Z = 0.29 (P = 0.77)							



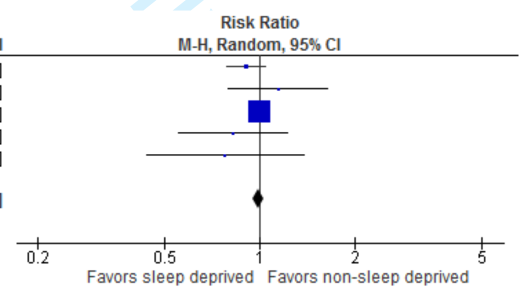
1.2 Intra-operative complications

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI
1.2.1 Surgical procedure						
Rothschild 2009a	68	919	253	3552	30.8%	1.04 [0.80, 1.34]
Schieman 2008	3	22	44	248	12.8%	0.77 [0.26, 2.28]
Vinden 2014	46	2031	72	8124	28.4%	2.56 [1.77, 3.69]
Subtotal (95% CI)		**2972**		**11924**	**72.0%**	**1.37 [0.65, 2.87]**
Total events	117		369			
Heterogeneity: Tau² = 0.34; Chi² = 16.67, df = 2 (P = 0.0002); I² = 88%						
Test for overall effect: Z = 0.83 (P = 0.41)						
1.2.2 Obstetric procedure						
Rothschild 2009b	33	957	112	3945	28.0%	1.21 [0.83, 1.78]
Subtotal (95% CI)		**957**		**3945**	**28.0%**	**1.21 [0.83, 1.78]**
Total events	33		112			
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.00 (P = 0.32)						
Total (95% CI)		**3929**		**15869**	**100.0%**	**1.35 [0.82, 2.21]**
Total events	150		481			
Heterogeneity: Tau² = 0.19; Chi² = 16.95, df = 3 (P = 0.0007); I² = 82%						
Test for overall effect: Z = 1.19 (P = 0.24)						
Test for subgroup differences: Chi² = 0.08, df = 1 (P = 0.78), I² = 0%						



1.3 Post-operative complications

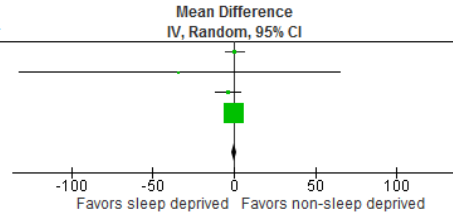
Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI
Chu 2011	259	1678	404	2369	7.8%	0.91 [0.78, 1.04]
Ellman 2004	29	339	480	6412	1.2%	1.14 [0.80, 1.64]
Govindarajan 2015	3527	19489	3543	19489	89.4%	1.00 [0.95, 1.04]
Schieman 2008	12	22	164	248	1.0%	0.82 [0.56, 1.22]
Vinden 2014	14	2031	72	8124	0.5%	0.78 [0.44, 1.38]
Total (95% CI)		**23559**		**36642**	**100.0%**	**0.99 [0.95, 1.03]**
Total events	3841		4663			
Heterogeneity: Tau² = 0.00; Chi² = 3.69, df = 4 (P = 0.45); I² = 0%						
Test for overall effect: Z = 0.66 (P = 0.51)						



1.4 Operating time (minutes)

	Sleep deprived			Not sleep deprived				Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year
Ellman 2004	107.7	55.24	339	107.4	56.05	6412	5.9%	0.30 [-5.74, 6.34]	2004
Schieman 2008	114	226.6	22	148	226.6	248	0.0%	-34.00 [-132.80, 64.80]	2008
Chu 2011	110.2	119.13	1678	114	120.88	2369	3.8%	-3.80 [-11.30, 3.70]	2011
Govindarajan 2015	156	75.56	19489	156	80	19489	90.2%	0.00 [-1.54, 1.54]	2015
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.14 [-1.60, 1.33]**	

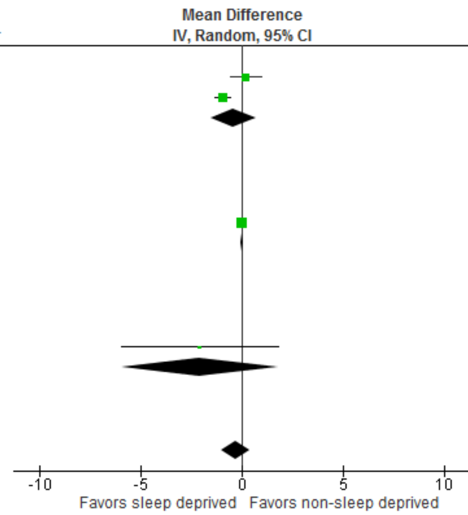
Heterogeneity: Tau² = 0.00; Chi² = 1.42, df = 3 (P = 0.70); I² = 0%
Test for overall effect: Z = 0.18 (P = 0.86)



1.5 Length of hospital stay (days)

	Sleep deprived			Not sleep deprived				Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year
1.5.1 Cardiac surgeries									
Ellman 2004	7.1	7.36	339	6.9	5.61	6412	25.4%	0.20 [-0.60, 1.00]	2004
Chu 2011	6.05	6.66	1678	7	6.07	2369	33.8%	-0.95 [-1.35, -0.55]	2011
Subtotal (95% CI)			**2017**			**8781**	**59.2%**	**-0.43 [-1.55, 0.69]**	
Heterogeneity: Tau² = 0.56; Chi² = 6.40, df = 1 (P = 0.01); I² = 84%									
Test for overall effect: Z = 0.75 (P = 0.45)									
1.5.2 Elective surgeries									
Govindarajan 2015	3	3.7	19489	3	3.7	19489	37.9%	0.00 [-0.07, 0.07]	2015
Subtotal (95% CI)			**19489**			**19489**	**37.9%**	**0.00 [-0.07, 0.07]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 0.00 (P = 1.00)									
1.5.3 Anterior resection for anal cancer									
Schieman 2008	9.3	8.91	22	11.4	8.91	248	2.9%	-2.10 [-5.98, 1.78]	2008
Subtotal (95% CI)			**22**			**248**	**2.9%**	**-2.10 [-5.98, 1.78]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 1.06 (P = 0.29)									
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.33 [-1.03, 0.36]**	

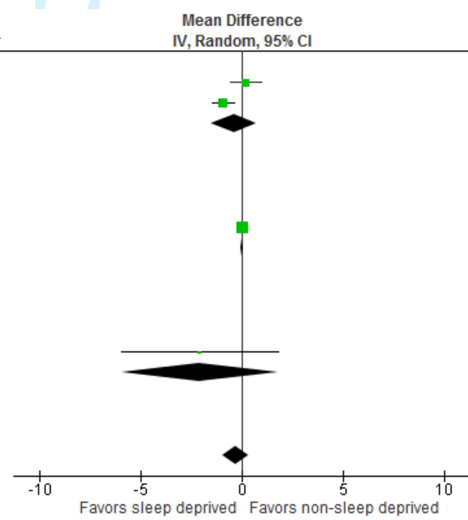
Heterogeneity: Tau² = 0.33; Chi² = 22.21, df = 3 (P < 0.0001); I² = 86%
Test for overall effect: Z = 0.94 (P = 0.35)
Test for subgroup differences: Chi² = 1.68, df = 2 (P = 0.43), I² = 0%



Sensitivity analysis imputing the highest standard deviation

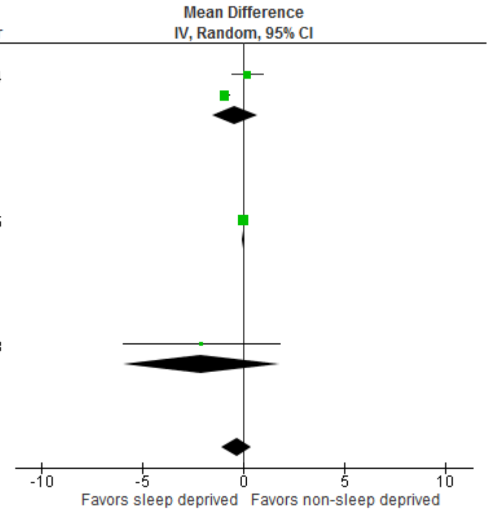
	Sleep deprived			Not sleep deprived				Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year
1.5.1 Cardiac surgeries									
Ellman 2004	7.1	7.36	339	6.9	5.61	6412	25.1%	0.20 [-0.60, 1.00]	2004
Chu 2011	6.05	8.91	1678	7	8.91	2369	31.3%	-0.95 [-1.51, -0.39]	2011
Subtotal (95% CI)			**2017**			**8781**	**56.4%**	**-0.41 [-1.54, 0.71]**	
Heterogeneity: Tau² = 0.54; Chi² = 5.39, df = 1 (P = 0.02); I² = 81%									
Test for overall effect: Z = 0.72 (P = 0.47)									
1.5.2 Elective surgeries									
Govindarajan 2015	3	3.7	19489	3	3.7	19489	41.0%	0.00 [-0.07, 0.07]	2015
Subtotal (95% CI)			**19489**			**19489**	**41.0%**	**0.00 [-0.07, 0.07]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 0.00 (P = 1.00)									
1.5.3 Anterior resection for anal cancer									
Schieman 2008	9.3	8.91	22	11.4	8.91	248	2.5%	-2.10 [-5.98, 1.78]	2008
Subtotal (95% CI)			**22**			**248**	**2.5%**	**-2.10 [-5.98, 1.78]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 1.06 (P = 0.29)									
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.30 [-0.94, 0.34]**	

Heterogeneity: Tau² = 0.26; Chi² = 12.36, df = 3 (P = 0.006); I² = 76%
Test for overall effect: Z = 0.93 (P = 0.35)
Test for subgroup differences: Chi² = 1.63, df = 2 (P = 0.44), I² = 0%



Sensitivity analysis imputing the lowest standard deviation

Study or Subgroup	Sleep deprived Mean	SD	Total	Not sleep deprived Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
1.5.1 Cardiac surgeries									
Ellman 2004	7.1	7.36	339	6.9	5.61	6412	25.7%	0.20 [-0.60, 1.00]	2004
Chu 2011	6.05	3.7	1678	7	3.7	2369	34.9%	-0.95 [-1.18, -0.72]	2011
Subtotal (95% CI)			**2017**			**8781**	**60.6%**	**-0.44 [-1.56, 0.68]**	
Heterogeneity: Tau² = 0.57; Chi² = 7.40, df = 1 (P = 0.007); I² = 86%									
Test for overall effect: Z = 0.77 (P = 0.44)									
1.5.2 Elective surgeries									
Govindarajan 2015	3	3.7	19489	3	3.7	19489	36.0%	0.00 [-0.07, 0.07]	2015
Subtotal (95% CI)			**19489**			**19489**	**36.0%**	**0.00 [-0.07, 0.07]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 0.00 (P = 1.00)									
1.5.3 Anterior resection for anal cancer									
Schieman 2008	9.3	8.91	22	11.4	8.91	248	3.4%	-2.10 [-5.98, 1.78]	2008
Subtotal (95% CI)			**22**			**248**	**3.4%**	**-2.10 [-5.98, 1.78]**	
Heterogeneity: Not applicable									
Test for overall effect: Z = 1.06 (P = 0.29)									
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.35 [-1.10, 0.40]**	
Heterogeneity: Tau² = 0.40; Chi² = 60.35, df = 3 (P < 0.00001); I² = 95%									
Test for overall effect: Z = 0.92 (P = 0.36)									
Test for subgroup differences: Chi² = 1.71, df = 2 (P = 0.42), I² = 0%									



Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8

Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-11, Table 1, Supplementary file 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	11, Supplementary file 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	p. 12-18; Supplementary file 4; figures 2-6
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	p. 12-18, figures 2-6
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Supplementary file 5
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-19
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	19-20
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	20
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	21

Appendix 1. PRISMA checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**. Page 2 of 2

BMJ Open

The impact of fatigue and insufficient sleep on physician and patient outcomes: A systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021967.R2
Article Type:	Research
Date Submitted by the Author:	28-Jun-2018
Complete List of Authors:	Gates, Michelle; University of Alberta, Pediatrics Wingert, Aireen; University of Alberta, Pediatrics Featherstone, Robin; University of Alberta, Pediatrics Samuels, Charles; Centre for Sleep and Human Performance Simon, Christopher; Canadian Medical Association Dyson, Michele; University of Alberta, Pediatrics
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Evidence based practice, Health services research, Medical education and training, Occupational and environmental medicine
Keywords:	Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SLEEP MEDICINE

The impact of fatigue and insufficient sleep on physician and patient outcomes: A systematic review

Michelle Gates, Aireen Wingert, Robin Featherstone, Charles Samuels, Christopher Simon, Michele P Dyson

Michelle Gates, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486C, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Aireen Wingert, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-490B, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Robin Featherstone, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486D, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Research librarian

Charles Samuels, Centre for Sleep and Human Performance, Suite 106, 51 Sunpark Drive SE, Calgary, Alberta, Canada, T2X 3V4
Physician and medical director

Christopher Simon, Canadian Medical Association, 1867 Alta Vista Drive, Ottawa, Ontario, Canada, K1G 5W8
Advisor, physician health

Michele P Dyson, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-474, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Assistant professor

Correspondence to: MP Dyson mdyson@ualberta.ca

Word count (main text): 5,929

ABSTRACT

Objectives: For physicians in independent practice, we synthesized evidence on the (a) impacts of insufficient sleep and fatigue on health and performance, and patient safety; (b) effectiveness of interventions targeting insufficient sleep and fatigue.

Design: We systematically reviewed online literature. After piloting, one reviewer selected studies by title and abstract; full texts were then reviewed in duplicate. One reviewer extracted data; another verified a random 10% sample. Two reviewers assessed risk of bias. We pooled findings via meta-analysis when appropriate, or narratively.

Data sources: We searched Medline, Embase, PsycINFO, CINAHL and PubMed for published studies in April 2016; Medline was updated in November 2017. We searched Embase for conference proceedings, and hand-searched meeting abstracts, association and foundation websites.

Eligibility criteria for selecting studies: English or French language primary research studies published from 2000-2017 examining the effect of fatigue or sleep-related exposures or interventions on any outcome among physicians in independent practice and their patients.

Results: Of 16,154 records identified, we included 47 quantitative studies of variable quality. 28 studies showed associations between fatigue or insufficient sleep and physician health and well-being outcomes. 21 studies showed no association with surgical performance, and mixed findings for psychomotor performance, work performance, and medical errors. We pooled data from six cohort studies for patient outcomes. For sleep deprived versus non-sleep deprived surgeons, we found no difference in patient mortality (n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, p = 0.82, I^2 = 0%) nor postoperative complications (n = 60,201, RR 0.99, 95% CI 0.95 to 1.03, I^2 = 0%). The findings for intraoperative complications and length of stay were considerably heterogeneous.

Conclusions: Fatigue and insufficient sleep may be associated with negative physician health outcomes. Current evidence is inadequate to inform practice recommendations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The review was informed by the methods outlined by Cochrane and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.
- The review was limited by the quality of the included studies, which was often poor. We could not draw definitive conclusions due to methodological weaknesses and heterogeneous outcome measures in the included studies.
- We have focused on evidence from high income countries; our findings may not be generalizable to other settings.

BACKGROUND

The working hours of physicians have been a topic of debate for many years.[1] Beginning in the late 1980s, evidence indicating that medical resident fatigue could negatively impact their cognitive functioning and performance, resulting in an increased risk of medical error, began to accumulate.[2] In response, by the early 2000s physicians' regulatory bodies worldwide began to take action toward restricting the work hours of medical residents and ensuring adequate time for recovery between shifts.[3-5] Since their implementation in the United States by the Accreditation Council for Graduate Medical Education (ACGME), the impact of work hour regulations has been widely researched. Still, evidence for impacts on patient care, resident training and wellbeing remains equivocal.[6-9] This is likely because work hours are only one of many contributors to fatigue and physician wellbeing. In fact, the ACGME has recently reversed the 2011 changes that limited resident work hours to 16 hours per shift and the requirement for 8 hours of time off between shifts. This decision was made in favour of promoting "flexibility" for residency training program work hours and scheduling.

The focus on medical trainees has left physicians in independent practice as a relatively neglected group in research and policy. In Canada, there is no concrete regulation on the hours or patterns in which physicians choose to work.[10] In the absence of clear policies, physicians trained under traditional systems may find it difficult to work shorter hours or take more frequent breaks.[1] Indeed, more than 40% of practicing physicians in the United States work in excess of 80 hours per week.[11] While long work hours remain a cultural norm in medicine, in comparable high-risk industries (e.g., aviation), work patterns and work hours are tightly regulated.[12] The need for similar evidence-based policies in medicine has become a topic of increased interest. Exemplar of this, an evidence-based guideline for fatigue risk management in emergency medical services,[13] informed by a comprehensive set of systematic reviews, has recently been published. For physicians, it has been argued that there is a need to adapt healthcare systems and provide support in identifying the signs of fatigue and mitigating its risks.[1]

Besides potentially affecting patient outcomes, fatigue can impact the health and wellbeing of physicians themselves. Burnout, just one outcome related to fatigue, has been described as epidemic among physicians[14-16] and ultimately affects recruitment and retention of physicians both in community and acute care settings. While the effect of physician wellbeing on the sustainability of healthcare systems has recently received increased attention,[17] evidence-based solutions to burnout

remain relatively elusive.[18] What is clear, is that comprehensive organisational-level efforts are necessary to fully address the issue.[19] Research addressing the factors that influence burnout and overall physician wellness is needed to inform system- and individual-level strategies.[20, 21] To date, evidence of the effects of fatigue and the role of chronic insufficient sleep on physicians in independent practice has not been synthesized, making it unclear what gaps in knowledge remain unaddressed.

Given this void, we undertook a systematic review focusing broadly on primary research relevant to the Canadian context as a fundamental starting point to examine the effects of fatigue and chronic insufficient sleep on physicians in independent practice, and on interventions to combat these effects. Our review was guided by the following research questions: Among physicians in independent practice, (1) what are the impacts of fatigue and chronic insufficient sleep on physician health, physician performance, and patient safety; and (2) what is the effectiveness of interventions that target fatigue and chronic insufficient sleep, in terms of improving physician and patient outcomes?

METHODS

Review conduct

The conduct of this systematic review was guided by Cochrane standards.[22] The research team convened to plan the key research questions and methodology but did not register a formal protocol. The findings are reported in adherence with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement.[23] Ethical approval was not required for this study.

Patient involvement

Patients were not involved.

Literature search

An information specialist developed a search strategy that included concepts related to physicians, fatigue and sleep. On 13 April 2016 we searched the following online databases with coverage in the biomedical sciences and psychology: Medline, Embase, PsycINFO, CINAHL and PubMed, limited to English and French language articles published from 2000 to 2016. We updated the Medline search in November 2017, as this database offered the highest precision. Though fatigue among physicians is not a new phenomenon,[2] we limited our search to articles published post-2000 to include studies relevant to current physician practice. Work hour limitations have existed in European countries since 1993, but

implementation in the United States (2003)[5] and Canada (2013) for residents is more recent.[24] We aimed to include studies published in this era of increased awareness about the potential impacts of long work hours. To locate unpublished studies, we searched Embase for conference proceedings since 2000 and hand-searched meeting abstracts of the Canadian Conference on Physician Health and the International Conference on Physician Health (2012 to 2016). We also searched the following association and foundation websites: American Medical Association, Australian Medical Association, British Medical Association, Canadian Medical Association, European Medical Association, National Sleep Foundation, Ontario Medical Association and the World Medical Association. The complete search strategy undertaken is reported in Supplementary file 1.

Inclusion criteria

Primary studies (quantitative or qualitative) of fatigue- or sleep-related exposures or interventions among physicians in independent practice were eligible for inclusion. We included physicians practicing in any medical specialty and in any healthcare setting within a high income country,[25] to identify practices comparable to the Canadian setting. Studies including physicians-in-training were included only if data for physicians in independent practice could be isolated. Exposures of interest included fatigue, insufficient sleep, or sleepiness. We also included studies of any intervention that aimed to reduce fatigue or sleep loss with any comparator (or no comparator). All reported outcomes, measured at any time, were eligible for inclusion.

We excluded commentaries, letters, editorials and dissertations. Systematic reviews, health technology assessments, economic evaluations and practice guidelines were excluded, although the reference lists of these as well as the included studies were scanned for potential primary studies. Studies that focused solely on physicians-in-training (e.g., trainees, residents, fellows, interns, medical students, junior doctors, registrars) were ineligible. To maintain the focused scope of the review, we excluded work hours, work load, and any other exposure or intervention that was indirectly related to fatigue or sleep.

Study selection

The study team piloted the selection criteria, which were then applied by two independent reviewers following a two-phase process. We first screened titles and abstracts for potential relevance. Then, we retrieved all records classified as “include” or “unsure” and reviewed their full text for eligibility. Any

disagreements between reviewers were resolved by discussion or third-reviewer consultation when necessary.

Data extraction

Reviewers used a standardized form to extract data in Microsoft Office Excel (*v.* 2016, Microsoft Corporation, Redmond, WA). One reviewer independently extracted data from each included study and a second reviewer verified a random 10% sample. Since no major errors or omissions were noted, we did not undertake further verification.

We extracted the following data: country of publication; funding source; study design; inclusion and exclusion criteria; population characteristics (i.e., sample size, age and gender distribution, physician specialty); setting (i.e., physician workplace, urban or rural); exposure or intervention; definition of fatigue or insufficient sleep; sleep and fatigue scales used and timing of measurement; comparators (if applicable); and outcomes.

Risk of bias appraisal

Two reviewers independently assessed the risk of bias in each included study using standard tools. Disagreements were resolved via discussion or by consulting a third reviewer. We used the Cochrane Risk of Bias tool[22] to assess randomised controlled trials. Adapted versions of the tool developed by the Effective Practice and Organization of Care group[26] were used to assess before-after and time series studies. We used the Newcastle-Ottawa Quality Assessment Scale[27] to appraise cohort studies. We adapted the scale to assess cross-sectional studies and the one non-comparative study.

Evidence synthesis

We considered clinical and methodological heterogeneity in our decision on whether to proceed with meta-analysis for the outcomes identified. For most outcomes, we found high levels of heterogeneity in study design, populations, exposures or interventions, and outcome measures and chose not pool the data via meta-analysis. Thus, we have presented the findings for most outcomes narratively and in summary tables.

When statistical pooling was appropriate, this was undertaken using Review Manager (RevMan v.5.3, Copenhagen: The Nordic Cochrane Centre, the Cochrane Collaboration, 2014) via pairwise meta-analysis

using the DerSimonian and Laird random effects model (given expected heterogeneity).[28] We pooled dichotomous outcomes using the relative risk (95% confidence interval (CI)) and continuous outcomes using the mean difference (95% CI) since the units across studies were consistent (i.e., minutes). When meta-analysis was conducted, we assessed statistical heterogeneity using the chi-square test (using P = 0.10 as the threshold for significance), and quantified the extent of heterogeneity using the I^2 statistic.[29] We considered an I^2 value of 0% to 40% to be low (potentially unimportant), 30% to 60% to be moderate, 50% to 90% to be substantial, and 75% to 100% to be considerable heterogeneity.[22] Subgroup and sensitivity analyses were conducted when appropriate to explore heterogeneity. We intended to assess small study bias visually by inspecting funnel plots and statistically using Egger's regression test, but did not due to the small number (i.e., less than 8) of studies included in the meta-analyses.[30]

When data were not presented in the format required for meta-analysis, we estimated means or standard deviations (SDs) using standard equations. We used the median instead of the mean for one study[31] for the outcomes of length of stay and operating time. Additionally, for one study[32] in the length of stay analysis where the SD could not be estimated, we substituted the mean variance of other studies within the meta-analysis.[33]

RESULTS

We identified 16,083 unique records via the database searches, 56 grey literature sources, and 14 additional records in reference lists of systematic reviews. We excluded 15,016 citations by title and abstract, and another 1,090 by full text. Forty-seven studies[31, 32, 34-78] were eligible for inclusion, and 6[31, 32, 41, 58, 63, 77] were included in meta-analysis for the outcomes of operating time, intra- and post-operative complications, patient mortality and length of hospital stay. Figure 1 shows the flow of studies through the selection process.

Included study characteristics

A summary of the study characteristics is provided in Table 1. Supplementary file 2 presents descriptive information for each included study. There were 45 observational studies[31, 32, 34-39, 41-74, 76-78] and two intervention studies.[40, 75] All studies were quantitative. Nearly half (n = 20/47, 43%) of the studies took place in North America,[31, 32, 35, 37, 38, 41, 45, 48, 57-60, 62, 63, 65-68, 72, 77] and slightly more than one-third (n = 16/47, 34%) in Europe.[34, 36, 39, 40, 42, 46, 47, 50-53, 61, 64, 73, 75, 76]

Table 1. Summary characteristics of the included studies

Study characteristics	n	%	Physician characteristics	n	%	Exposures, interventions and outcomes	n	%
Study design			**Gender**			**Exposures (observational)**[a]	**45**	**96**
Cross-sectional	34	72	Reported[b]	38	81	Fatigue-related	15	32
Cohort	6	13	>50% male	30	79	Sleep-related	37	79
Before-after	3	6	**Age**			Overnight or extended shifts	18	38
RCT	2	4	Reported[b]	38	81	**Interventions (experimental)**	**2**	**4**
Time series	1	2	Range (years)	20 to >70		**Outcomes**		
Non-comparative	1	2	**Specialty area**[c]			Physician health and wellbeing	28	60
Region and country			Surgeons	13	28	Work and life satisfaction	9	19
North America	20	43	Anesthesiologists	10	21	Burnout	7	15
US	15	32	Generalists	7	15	Stress	8	17
Canada	4	9	ED or ICU physicians	3	6	Mental health and wellbeing	7	15
Canada, US & Mexico	1	2	Oncologists	2	4	Other health-related outcomes	5	11
Europe	16	34	Obstetrician-gynecologists	1	2	Physician performance, risk of error	21	45
France	4	9	Mixed groups	14	30	Psychomotor performance	7	15
Finland	3	6	**Work setting**[d]			Work ability and quality of care	5	11
Spain	2	4	Hospitals	37	79	Incidence of medical errors	5	11
Austria	2	4	Private practice	13	28	Surgical efficiency, effectiveness	6	13
Norway	2	4	Primary care centres, outpatient clinics	7	15	Patient outcomes	6	13
Denmark	1	2	Academic practice, training programs	5	11			
Germany	1	2	Other (e.g., industry, military)	11	23			
Malta	1	2	Not reported	3	6			
Japan	4	9	**Urban or rural**					
Australia	2	4	Reported[b]	16	34			
Israel	2	4	Urban	12	75			
New Zealand	2	4	Rural	2	13			
United Kingdom	1	2	Mixed	2	13			

ED: emergency department; ICU: intensive care unit; RCT: randomised controlled trial; US: United States of America

[a]Exposures that have been directly related to an outcome. Some studies included multiple exposures.

[b]Percentages presented using the total number of studies where the outcome was reported as the denominator.

[c]Anesthesiologists include physician anesthetists; generalists include primary care physicians, internists, and general practitioners; mixed groups refers to studies including more than one physician group or specialty (usually large-scale surveys). In some studies, multiple distinct groups were represented.

[d]As defined by the authors. Values for the settings will exceed 100% because studies may occur in more than one setting.

The 47 studies reported outcomes for 36,190 (range = 6 to 7,905) physicians and 69,809 (range = 270 to 38,978) adult patients. About half reported on surgeons (n = 13/47, 28%),[31, 32, 34, 41, 45, 48, 54, 58, 62, 63, 66, 75, 77] or anesthesiologists/physician anesthetists (n = 10/47, 21%).[36, 37, 39, 43, 44, 50-52, 59, 60] Where it was reported, the samples tended to be predominantly male (n = 30/38, 79%) and physician age varied widely. Hospitals were the most common setting (n = 37/47, 79%).[31, 32, 34-37, 39-43, 45-47, 49-52, 54, 57-61, 63-70, 73-75, 77, 78] In the studies where it was reported (n = 16/47, 34%),[31, 32, 34, 38, 40, 41, 43, 45, 50, 51, 55, 56, 58, 65, 76, 77] all but four studies[31, 55, 56, 77] took place in solely an urban setting.

Fifteen (32%) studies reported on fatigue-related exposures (e.g., as a source of stress, exhaustion, physical fatigue; hereafter referred to as 'fatigue'),[35, 40, 45, 48, 57, 63-71, 73] while others (n = 37/47, 79%) reported on sleep-related exposures (e.g., sleep hours, insufficient sleep, sleep deprivation, sleep disruption, sleepiness; hereafter referred to as 'insufficient sleep').[31, 32, 34, 36-47, 49-56, 58-62, 64, 67, 71, 72, 74-78] A few (n = 5/47, 11%) reported on both.[40, 45, 64, 67, 71] In some cases (n = 18/47, 38%), fatigue or insufficient sleep were related to overnight work or long on-call shifts.[31, 32, 34, 37, 40, 41, 43, 45, 46, 50, 51, 53, 58-60, 63, 75, 77] Measured outcomes varied widely and were ultimately organised into physician physical and mental health, physician performance and risk of error, and patient outcomes.

Risk of bias appraisal

The overall quality of the body of research was poor; 62% (n = 29/47) of studies were rated at unclear or high risk of bias. Of the two randomised controlled trials, one was rated as unclear overall risk of bias[75] and one as high risk.[40] All cohort studies were at low risk of bias (mean score: 8.4/9, range: 8-9).[31, 32, 41, 58, 63, 77] All of the before-after studies were rated as high risk of bias.[34, 45, 50] The single time series study was assessed at high risk of bias.[51] The cross-sectional studies varied in performance (mean score: 3.0/5, range: 1-4); only one-third (n = 12/34, 35%) were at low risk of bias.[39, 42, 44, 47, 52, 59, 60, 69, 70, 72, 74, 76] The one non-comparative study was at unclear risk of bias.[43] Detailed assessments of the sources of bias per study are shown in Supplementary file 3.

Physician health and wellbeing outcomes

Twenty-eight studies reported on physician health and wellbeing-related outcomes,[35, 36, 38-40, 42, 46-48, 50-57, 60, 62, 64, 67, 68, 70-72, 74, 76, 78] including burnout (n = 7), stress (n = 8), mental health and wellbeing (n = 7), life and job satisfaction (n = 9) and other markers of health (n = 5) (Supplementary file 4).

Seven cross-sectional studies reported on burnout (5 low[39, 70, 72, 74, 76], 1 unclear[54], 1 high risk of bias[62]) among surgeons,[54, 62] anesthesiologists,[39] generalists,[76] and other mixed groups.[70, 72, 74] Two studies reported on surgeons; the larger (n = 2,564, low risk of bias) study of neurosurgeons showed increased odds of burnout with sleep deprivation (hours of sleep per night; OR 0.84, 95% CI 0.75 to 0.94, P = 0.002).[54] Among anesthesiologists one study (n = 565, low risk of bias) indicated that burnout (measured via Maslach Burnout Inventory) was more prevalent among the sleep-deprived ('lack of sleep' on one question; 47.6% vs. 16.3%, P < 0.001).[39] In one small (n = 11) study of generalists, those with burnout (measured via Pines Burnout Measure) had poorer Pittsburgh Sleep Quality Index scores (7.24±4.17 vs. 2.72±2.22, P < 0.001).[76] In the two larger studies of mixed physician groups (low risk of bias), burnout (measured via 5-point scale) was more prevalent among those who were sleep deprived (<7 hours of sleep per 24 hours; 39.6% vs. 26.4%, P < 0.05),[72] and physical fatigue ('feeling tired' on a 7-point scale) was correlated with burnout (Shirom-Melamed Burnout Measure; r = 0.88, P < 0.05).[70] In summary, evidence from 7 cross-sectional studies (71% at low risk of bias), showed associations between insufficient sleep and burnout.

Six cross-sectional studies (2 low[47, 52], 1 unclear[46], 3 high risk of bias[35, 62, 64]), one uncontrolled before-after study (high risk of bias[50]), and one intervention study (high risk of bias[40]) reported on stress outcomes among surgeons,[62] anesthesiologists,[50, 52] emergency physicians,[40, 64] internal medicine physicians,[46] and mixed groups.[35, 47] In a small sample (n = 20) of internal medicine physicians, insufficient sleep related to a 24-hour call shift showed no association with biochemical or physiological stress parameters, except levels of thyroid stimulating hormone, which was higher post-shift (P = 0.049, data not reported).[46] The remaining observational studies suggested that there was an association between insufficient sleep or fatigue and stress. The one study of orthopedic surgeons (n = 264, high risk of bias) showed that insufficient sleep (measured on a 3-point scale) and psychological distress (measured via General Health Questionnaire-12) were correlated (data not reported, P < 0.001).[62] The two reports on anesthesiologists were of varied quality; the larger (n = 328, low risk of bias) study showed that stress symptoms (measured via Modified Occupational Stress Questionnaire) were predicted by sleep sufficiency (self-reported on one question, β = -0.269, P < 0.001).[52] Among the two studies reporting on mixed groups of physicians, the larger (n = 1,541, low risk of bias) study showed an association between sleep problems (4 questions derived from Jenkins scale) and psychological distress (General Health Questionnaire-12; β = 0.18, P < 0.001).[47] One RCT assessed the impact of insufficient sleep from shift work (14-hour or 24-hour shifts), showing that stress (on a visual analog scale) among emergency physicians (n = 17) was higher following the shift as compared to a control day (data not reported, P < 0.05).[40] In summary, evidence from one intervention study at high risk of bias and all but one of the 7 observational studies (29% at low risk of bias) identified supported an inverse association between fatigue or sleep deprivation and stress.

Seven cross-sectional studies (2 low,[52, 60] 3 unclear,[67, 71, 78] 2 high risk of bias[36, 53]) reported on aspects of mental health including addiction or substance misuse,[36, 53, 71] depression,[78] thoughts of suicide,[52] mood disturbance,[60, 71] and overall wellbeing.[67] One study,[53] which was at high risk of bias, showed no association between hours of sleep when on call and hazardous drinking behaviours (via Alcohol Use Disorder Identification Test). Meanwhile, the six other studies all showed associations between insufficient sleep and fatigue and reduced mental health. Three studies reported on anesthetists,[36, 52, 60] with two large surveys showing increased odds of tobacco (OR 1.42, 95% CI 1.04 to 1.94) and tranquilizer/hypnotics (OR 3.26, 95% CI 2.12 to 5.02) dependency being predicted by sleep deprivation (measured by one question),[36] and sleep disturbance being associated with thoughts of suicide (using a 4-point scale; P = 0.009).[52] A small study (n = 21) showed greater mood disturbance following a 17-hour night shift than a usual day (Profile of Mood States score 42.57±15.26 vs. 70.90±6.91, P < 0.001).[60] Among oncologists (n = 241), overall wellbeing was predicted by lower levels of fatigue after controlling for personal and professional characteristics (assessed via linear analog scale quality of life survey, P = 0.002).[67] A large (n = 3,862, unclear risk of bias) study of physicians showed that insufficient sleep (lower sleep hours when not at work in the past month) was associated with increased odds of depression (Quick Inventory Depressive Scale; OR 2.70, 95% CI 1.82 to 4.03 for men; OR 2.38, 95% CI 1.11 to 5.10 for women).[78] In open-ended questions, senior physicians in one study (unclear risk of bias) attributed the development of mental illness to tiredness and stress at work.[71] In summary, 7 cross-sectional studies (29% at low risk of bias) were identified, and of these 6 supported an association between insufficient sleep or fatigue and negative mental health outcomes.

Nine cross-sectional studies (4 low,[42, 47, 72, 74] 2 unclear,[55, 68] 3 high risk of bias[38, 48, 62]) reported on outcomes related to job satisfaction,[42, 47, 48, 55, 72, 74] life satisfaction,[38, 62, 72] or work-life balance.[68, 72] The six studies that investigated job satisfaction were all at low risk of bias and generally included mixed groups of physicians;[47, 72, 74] one study reported on general practitioners,[55], another on surgeons,[48] and one on mixed specialties.[42] Three studies showed that reductions in sleep duration and/or quality[47, 48, 74] were associated with reduced job satisfaction. Meanwhile one showed no association between insufficient sleep (<7 hours per 24-hour period) and career satisfaction (measured on a 5-point Likert scale),[72] and another showed no relationship between earlier sleep disturbance (Jenkins Scale) and later job demands or job control (measured via 5-point scale).[42] A single study (n = 92) reporting on rural general practitioners indicated that frequent sleep disturbance (measured on a 7-point scale) predicted the intention to retire early (OR 2.91, 95% CI 1.11 to 7.6, $P < 0.05$).[55] In summary, 6 cross-sectional studies (all at low risk of bias) were identified, and all but two[42, 72] of these studies showed that insufficient sleep and fatigue were associated with reductions in satisfaction.

The three studies reported on life satisfaction.[38, 62, 72] Of two studies among mixed physician groups,[38, 72] the one larger (n = 840) study showed that insufficient sleep (< 7 hours per day) was a predictor of reduced life satisfaction (measured on a 5-point Likert scale; OR 0.44, 95% CI 0.29 to 0.67, $P \leq 0.05$).[72] One study at high risk of bias reported on orthopedic surgeons (n = 264), showing that sleep deprivation (measured via 3-point scale) was correlated with lower marital satisfaction (Revised Dyadic Adjustment Scale; data not reported, $P < 0.001$).[62] Two large studies at low or unclear risk of bias reported on work-life balance.[68, 72] Among oncologists (n = 1,117), reduced satisfaction with work-life balance (measured on a 5-point Likert scale) was predicted by high levels of fatigue (measured via 10-point visual analog scale), even when controlling for personal and work-related factors and burnout (OR 0.489, 95% CI 0.337 to 0.710, $P < 0.001$).[68] Among a mixed group of physicians (n = 840, low risk of bias), insufficient sleep (<7 hours in a typical 24-hour period) predicted a reduced perception of having balanced personal and professional commitments (5-point Likert scale; OR 0.46, 95% CI 0.31 to 0.71, $P \leq 0.05$).[72] In summary, 3 cross-sectional studies (all unclear or high risk of bias) supported an association between insufficient sleep or fatigue and reduced life satisfaction, and 2 cross-sectional studies (50% low risk of bias) supported an association with reduced work-life balance.

Four cross sectional studies (3 unclear,[56, 57, 71] 1 high risk of bias[38]) and one time series study (high risk of bias[51]) reported on other health-related outcomes. Among a mixed group of physicians (n = 180), one study at high risk of bias showed that Epworth Sleepiness Scale scores were higher among physicians who worried about having a car accident while driving home (7.0 vs. 5.4, $P < 0.001$).[38] Among generalists (n = 578), almost 1 in 10 (8.7%) admitted to falling asleep while driving due to fatigue.[57] Also among generalists (n = 92), those with frequent work-related sleep disturbance (measured on a 7-point scale) were at increased odds of sickness presenteeism (OR 2.92, 95% CI 1.19 to 7.16, P = 0.02).[56] The one time series study concluded that a single 24-h shift did not cause major chronodisruption (based on serum melatonin measurement) among anesthetists (n = 10).[51] Meanwhile, open-ended comments from a large sample (n = 3,550) of senior physicians suggests that they attributed the development of physical health problems to a lifestyle of insufficient sleep, poor eating habits and lack of exercise imposed by their jobs.[71] In summary, 5 cross sectional studies (0% at low risk of bias) supported associations between insufficient sleep and fatigue and varied deleterious health outcomes (i.e., car accidents, sickness presenteeism, physical health problems). One time series study at high risk of bias did not support such a relationship.

Physician performance and risk of errors

Twenty-one studies reported on physician performance and safety-related outcomes,[31, 32, 34, 37, 38, 41, 43-47, 49, 50, 59, 61, 65, 66, 69, 71, 73, 75] including surgical efficiency and effectiveness (n = 6), psychomotor performance (n = 7), work ability and quality of care (n = 5) and medical errors (n = 5) (Supplementary file 4).

Four cohort studies (all low risk of bias[31, 32, 41, 63]), one before-after study (high risk of bias[34]) and one randomized controlled trial (high risk of bias[75]) examined the effects of insufficient sleep from overnight work or extended shifts, during surgeries[31, 32, 41] or laparoscopic simulations.[34, 75] We pooled the data from these studies[31, 32, 41, 63] via meta-analysis, which showed no difference in operating time (sometimes referred to as surgeon efficiency) between sleep deprived and non-sleep deprived surgeons (Figure 2; n = 50,046, MD -0.14, 95% CI -1.60 to 1.33, P = 0.86, $I^2 = 0\%$). Of studies not meta-analysed, the small (n = 29) before-after study showed no impact of sleep deprivation from shift-work nor of sleep hours on performance on a laparoscopic simulation (LapSimGyn).[34] One small (n = 64) intervention study compared a 24-hour shift to a usual work day, also finding no detriment to performance on a laparoscopic simulation (Minimally Invasive Surgical Trainer-Virtual Reality) despite diminished sleep hours while working on-call.[75] In summary, pooled data from 4 cohort studies (100% low risk of bias) showed no effect of insufficient sleep on surgical efficiency. Additional data from one RCT (high risk of bias) and one before-after study (high risk of bias) also showed no association between insufficient sleep and performance on laparoscopic simulations.

Two before-after studies (high risk of bias[45, 50]) and five cross-sectional studies (2 low,[43, 59] 3 unclear,[37, 46] 1 high risk of bias [61]) reported on psychomotor performance outcomes among surgeons,[45] anesthesiologists,[37, 43, 50, 59] emergency physicians,[61] and internal medicine physicians.[46] Among a small group of surgeons (n = 9), performance on a virtual ring transfer task deteriorated after an on-call shift (data not reported, P < 0.05).[45] The four studies among anesthetists reported mixed findings. One small (n = 11) before-after study showed longer reaction times (690.8±73.4 vs. 746.5±113.7 milliseconds) and reduced concentration ability (26.4±23.5 vs. 56.3±23.0 on a 100-point scale, P = 0.007) following a 24-hour shift with insufficient sleep;[50] Two others found that insufficient sleep due to overnight shifts was associated with slower reaction times.[43, 59] Conversely, a small study (n = 11) found no effect of overnight shiftwork with insufficient sleep on any measure of psychomotor performance except Hopkin's Verbal Learning Test (t-score of 48.6±7.6 vs. 41.5±9.9, P = 0.04).[37] Among emergency physicians (n = 18), one study (high risk of bias) showed that those who were sleep deprived (<5 hours sleep after a 24-hour shift) had a reduced performance on most but not all psychomotor tests (Battery Test Reaction 5),[61] while among internal medicine physicians (n = 20, low risk of bias), neurocognitive parameters did not seem to worsen post-call.[46] In summary, two before-after (0% low risk of bias) and 5 cross-sectional studies (40% low risk of bias) showed mixed results for the association between fatigue or insufficient sleep and psychomotor performance.

Five cross-sectional studies (2 low,[47, 69] 1 unclear,[71] 2 high risk of bias[38, 65]) reported on associations between sleep deprivation or fatigue and work ability or perceived performance, all among mixed groups of physicians.[38, 47, 65, 69, 71] The two large studies at low risk of bias showed that sleep problems and fatigue were inversely associated with physicians' perceived quality of work.[47, 69] Among 1,541 physicians in Finland, sleeping problems (measured by 4 questions from the Jenkins Scale) were inversely associated with scores on the Work Ability Index (β = -0.29, P < 0.001),[47] while a study of 890 physicians from Israel demonstrated that perceived quality of care was predicted by fatigue (1 item on the Shirom-Melamed Burnout Measure) even after controlling for components of burnout (β = 0.17, P < 0.05).[69] Similarly, in one study, comments from senior physicians suggested that continual

tiredness and exhaustion negatively affected their perceived competence.[71] The two studies[38, 65] that were at high risk of bias had conflicting findings. In summary, 5 cross-sectional studies (40% at low risk of bias) reported on perceived work performance; those that were at low risk of bias supported an association between fatigue or insufficient sleep and reduced performance.

Five cross-sectional studies (1 low,[44] 2 unclear,[49, 66] 2 high risk of bias[38, 73]) reported on associations between insufficient sleep or fatigue and self-reported medical errors among surgeons,[66] anesthesiologists[44] and mixed groups of physicians.[38, 49, 73] A large (n = 7,905) study at unclear risk of bias showed that only 6.9% of surgeons reported fatigue as the most important contributor to medical errors.[66] Among anesthesiologists, a smaller study (n = 183) at low risk of bias showed that the risk of self-reported fatigue-related errors increased with more nights of work-related sleep disturbance (RR 1.25, 95% CI 1.06 to 1.49).[44] Two of the studies reporting on mixed groups of physicians had conflicting results,[38, 49] while another reported that physicians' opinions on the association between fatigue and prescribing errors differed by work setting.[73] One-third (34%) of community-based, 96% of hospital-based, and 8% of office-based physicians believed that there was a high or very high association between fatigue and prescribing errors ($P < 0.05$).[73] In summary, 5 cross-sectional studies (20% at low risk of bias) reported on self-reported errors, and these showed mixed findings for associations with fatigue or insufficient sleep.

Patient Outcomes

Six large (n = 270 to 38,978) cohort studies at low risk of bias reported on patient outcomes, all related to surgical[31, 32, 41, 58, 63, 77] or obstetric[58] procedures (Supplementary file 4). In these studies, insufficient sleep or fatigue were typically defined as overnight work prior to a daytime procedure[31, 41, 58, 63, 77]; though two studies measured sleep hours[32] or 'sleep opportunity'.[58] We pooled data for procedures performed by sleep deprived versus non-sleep deprived surgeons (or obstetrician-gynecologists in one case[58]). Analyses showed no difference in the rate of post-operative complications (Figure 3; 5 studies,[31, 32, 41, 63, 77] n = 60,201, RR 0.99, 95% CI 0.95 to 1.03, $p = 0.51$, $I^2 = 0\%$) nor patient mortality (Figure 4; 5 studies,[31, 32, 41, 63, 77] n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, $p = 0.82$, $I^2 = 0\%$). One study[77] in the mortality analysis reported the number of deaths only as ≤5. We assumed 2 events for this study (midpoint between 0 and 5); sensitivity analysis using the lowest (i.e., 0) and highest (i.e., 5) possible number of events did not change the overall result (Supplementary file 5). We found considerable between-study heterogeneity in the analyses for intraoperative complications ($I^2 = 82\%$) and length of stay ($I^2 = 86\%$), which could not be explained via subgroup analyses by procedure type, thus we have suppressed the average estimates of effect. For length of stay, the results of one study on cardiac surgeries favoured sleep deprived surgeons,[32] while the others[31,41,63] had null results. For intraoperative complications, the findings of one study[63] favoured non-sleep deprived surgeons, but the others[58,77] had null results.

DISCUSSION

Fatigue and chronic insufficient sleep are two potential drivers of reduced physician wellbeing[17, 19] that have thus far been understudied in physicians in independent practice. Burnout is becoming increasingly prevalent among physicians,[14-16] and recent research indicates that comprehensive individual- and system-level strategies are needed to address the problem.[6-9, 19, 21] We have systematically reviewed evidence from a heterogeneous array of available studies reporting on diverse outcomes related to physicians in independent practice and their patients. The included studies were often at high or unclear risk of bias, included small samples of physicians, and inconsistently measured and reported exposures and outcomes. The key message gleaned from this review is that despite growing interest in the topic of physician wellness, the robust evidence needed to inform individual and systems-level fatigue management strategies is lacking.

Traditionally, much of the fatigue-related research has focused on hazards to patients. The current review included six cohort studies showing that insufficient sleep and/or fatigue did not seem to result in increased rates of patient morality or post-operative complications; findings for length of stay and intra-operative complications were inconclusive. Evidence for psychomotor performance, surgical skills and errors suggest that there is indeed a potential for negative outcomes. The included studies, like many of the others in this and other systematic reviews,[79] employed indirect definitions that make it difficult to classify sleep deprived physicians with certainty. In recent years there has been a shift away from the singular focus on patient safety toward a more comprehensive view that also considers the detrimental effects of fatigue, sleep loss and other occupational hazards on physician wellness.[80] Evidence from this review supports that fatigue and insufficient sleep may be negatively associated with physician health and wellbeing. It is now recognized that health systems cannot be sustained by a workforce that is facing an epidemic of burnout.[19, 81, 82]

In light of high rates of burnout, the ongoing dialogue about the need for a cultural shift in the practice of medicine[83, 84] is now more important than ever. Recognition of the potential effects of physician fatigue on patients, physicians, and healthcare systems as a whole must be emphasized at a systemic level, encouraging a shift in which the risks are viewed as unacceptable.[1, 20, 80] Likewise, although research to date has focused largely on individual-level approaches to address burnout, it is now clear that placing the burden of a system-level problem solely on the individual is unlikely to bring about significant and lasting change.[85] Recent research has highlighted physician burnout as a system-driven issue that will require corresponding national-scale multicomponent solutions.[1, 19, 81, 82] As such, in the past several years both the American and Canadian Medical Associations have developed policies and programs that address physician health.[81, 86] The Canadian Medical Association's new policy on physician health calls on broad stakeholder groups (e.g., policymakers, regional health authorities, governments) to take shared responsibility for the health of physicians and to make meaningful and concerted efforts towards promoting a healthy and sustainable workforce.[81]

The most salient finding of this review is that the current evidence is insufficient to inform policy and practice. Correspondingly, a 2016 research summit on physician wellness and burnout outlined the need for timely, relevant and methodologically robust research to inform practice and policy.[21] The findings herein may be used as motivation for researchers and practitioners to develop and design methodologically strong research programs related to physician fatigue, inform successful research grant proposals, and lobby healthcare organizations to increase the focus on physician fatigue management programs. It will be important to make use of existing validated measures[87-89] consistently in future research. Identifying outcomes of importance to physicians and their patients should be prioritized, such that these may be collected within intervention studies. Reporting these consistently will allow for the effective synthesis of findings and reduce research waste.[90] Integrated knowledge translation strategies involving multiple stakeholder groups (e.g., physicians, patients, medical schools, physicians' associations and governing bodies, policymakers) may help to ensure that the research is relevant and facilitates decision-making.[91]

Strengths and Limitations

Our systematic review is the first to synthesize evidence on the effects of fatigue and insufficient sleep on physicians in independent practice. The review is timely, given recent calls for research into individual and organisational solutions for burnout,[20, 21] and an increased focus on physician health.[80, 81] While we have identified a diverse body of evidence, we could not draw definitive conclusions due to methodological weaknesses (e.g., 62% at high risk of bias, reliance primarily on cross-sectional designs and uncontrolled studies, subjective measurement of exposures and outcomes, small sample sizes, inclusion of predominantly male physicians within urban settings) and heterogeneous outcome measures in the included studies. Given that the 2017 update search was limited to one database, it is possible that a small number of relevant studies could have been missed. We believe that the likelihood that these might alter the conclusions of the review is low. The findings may have been influenced by publication bias, and may not be generalized to all settings, given our restriction to high income countries.

CONCLUSION

The evidence synthesized in this review suggests that fatigue and insufficient sleep are associated with some detrimental physician health and wellbeing outcomes; the evidence for potential associations with performance and safety outcomes was mixed. Meta-analyses for patient outcomes demonstrated that in many cases, potential relationships with physician sleep deprivation remain unclear. Our overall confidence in the findings is low, owing to a body of research that is hindered by methodological weaknesses. Further methodologically robust research that includes consistent outcomes that are of interest to physicians and their patients is needed to inform strong practice recommendations and policy decisions.

ACKNOWLEDGMENTS

We would like to thank Dr. Meghan Sebastianski and Andrea Milne for assistance with screening, and Dr. Allison Gates for screening, data verification, and quality assessment of the French studies.

COPYRIGHT/LICENSE FOR PUBLICATION

COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Dr. Christopher Simon is employed by the Canadian Medical Association, who provided financial support for the research; there are no other relationships or activities that could appear to have influenced the submitted work.

CONTRIBUTOR STATEMENT

All authors contributed to the conception and design of the project. MG and AW contributed to the acquisition, analysis and interpretation of the data, and drafted the manuscript. RF contributed to acquisition of data. CSa, CSi and MPD contributed to interpretation of data and revised the manuscript for important intellectual content. All authors approved the final version of the manuscript as submitted.

FUNDING STATEMENT

This research was supported by the Canadian Medical Association. The funder had no role in the study design; collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

ROLE OF FUNDERS

Dr. Christopher Simon is employed by the Canadian Medical Association. The remaining authors are independent from the funders. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

TRANSPARENCY DECLARATION

The lead author (MG) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted; and all discrepancies from the study as planned have been explained.

DATA ACCESS STATEMENT

All authors, external and internal, had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the interpretation.

DATA SHARING STATEMENT

The data pertaining to this systematic review are available from the corresponding author upon reasonable request.

FIGURE CAPTIONS

Figure 1. Flow of records through the selection process

Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

Figure 3. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

Legend: Vinden 2013 reported iatrogenic injuries; Schieman 2008, Govindarajan 2015, and Chu 2011 reported post-operative complication rate; Ellman 2004 reported post-operative complications (other types of complications reported not included in the analysis)

Figure 4. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

REFERENCES

1. Greig P, Snow R. Fatigue and risk: are train drivers safer than doctors? *BMJ* 2017;359. doi: https://doi.org/10.1136/bmj.j5107
2. Parker JBR. The effects of fatigue on physician performance - an underestimated cause of physician impairment and increased patient risk. *Can J Anaesth* 1987;34(5):489-95.
3. Temple J. Resident duty hours around the globe: where are we now? BMC Med Educ 2014;14(Suppl 1):S8. https://doi.org/10.1186/1472-6920-14-S1-S8
4. Imrie K, Frank J, The National Steering Committee on Resient Duty Hours. Fatigue, risk, & excellence: towards a pan-Canadian consensus on resident duty hours. Ottawa, Ontario: The Royal College of Physicians and Surgeons of Canada, 2013. http://www.residentdutyhours.ca/final_report.php (accessed 12 Jan 2018).
5. Accreditation Council for Graduate Medical Education. History of duty hours. 2017. http://www.acgme.org/What-We-Do/Accreditation/Clinical-Experience-and-Education-formerly-Duty-Hours/History-of-Duty-Hours (accessed 12 Jan 2018).
6. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ* 2015;7(3):349-63.
7. Harris JD, Staheli G, LeClere L, Andersone D, McCormick F. What effects have resident work-hour changes had on education, quality of life, and safety? A systematic review. *Clin Orthop* 2015;473(5):1600-8.
8. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg* 2014;259(6):1041-53.
9. Peets A, Ayas NT. Restricting resident work hours: the good, the bad, and the ugly. *Crit Care Med* 2012;40(3):960-6.
10. Canadian Medical Association. CMA Policy: Management of physician fatigue. Ottawa, Canada: Canadian Medical Association; 2014.
11. Anim M, Markert RJ, Wood VC, Schuster BL. Physician practice patterns resemble ACGME duty hours. *Am J Med* 2009;122:587-93.
12. Transport Canada. Fatigue risk management system for the canadian aviation industry: policies and procedures development guidelines. Ottawa, Canada: Her Majesty the Queen in Right of Canada, as

represented by the Minister of Transport, 2008. https://www.tc.gc.ca/eng/civilaviation/publications/TP14576-6042.htm (accessed 12 Jan 2018).

13. Patterson PD, Higgins JS, Lang ES, et al. Evidence-based guidelines for fatigue risk management in ems: formulating research questions and selecting outcomes. *Prehosp Emerg Care* 2017;21(2):149-56.
14. Boudreau RA, Grieco RL, Cahoon SL, Robertson RC, Wedel RJ. The pandemic from within: two surveys of physician burnout in Canada. *Can J Commun Ment Health* 2006;25(2):71-88.
15. Kumar S. Burnout and doctors: prevalence, prevention and intervention. *Healthcare* 2016;4(3):37. doi: 10.3390/healthcare4030037
16. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among us physicians relative to the general us population. *Arch Intern Med* 2012;172(18):1377-85.
17. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. *Acad Psychiatry* 2017. doi: 10.1007/s40596-017-0781-6
18. Fralick M, Flegel K. Physician burnout: who will protect us from ourselves? *CMAJ* 2014;186(10):731. doi: 10.1503/cmaj.140588
19. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. *Mayo Clin Proc* 2017;92(1):129-46.
20. Lemaire JB, Wallace JE. Burnout among doctors. BMJ 2017;358:j3360. doi: 10.1136/bmj.j3360
21. Dyrbye LN, Trockel M, Frank E, et al. Development of a research agenda to identify evidence-based strategies to improve physician wellness and reduce burnout. *Ann Intern Med* 2017;166(10):743-4.
22. Higgins JPT, Green S (editors). The Cochrane handbook for systematic reviews of interventions, version 5.1.0. London, UK: The Cochrane Collaboration, 2011. http://www.handbook.cochrane.org (accessed 12 Jan 2018).
23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6(7):e1000100. https://doi.org/10.1371/journal.pmed.1000100
24. Pattani R, Wu PE, Dhalla IA. Resident duty hours in Canada: past, present and future. *CMAJ* 2014;186(10):761-5.
25. International Monetary Fund. World economic outlook: too slow for too long. Washington, DC: International Monetary Fund, 2016. http://www.imf.org/external/pubs/ft/weo/2016/01/ (accessed 12 Jan 2018).

26. Cochrane Effective Practice and Organisation of Care. Suggested risk of bias criteria for EPOC reviews. Oslo, Norway: Norwegian Knowledge Centre for the Health Services, 2016. http://epoc.cochrane.org/epoc-specific-resources-review-authors (accessed 12 Jan 2018).
27. Wells GA SB, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa, Canada: The Ottawa Hospital Research Institute, 2014. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 12 Jan 2018).
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-60.
30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.
31. Govindarajan A, Urbach DR, Kumar M, et al. Outcomes of daytime procedures performed by attending surgeons after night work. *N Engl J Med* 2015;373(9):845-53.
32. Chu MW, Stitt LW, Fox SA, Kiaii B, Quantz M, Guo L, et al. Prospective evaluation of consultant surgeon sleep deprivation and outcomes in more than 4000 consecutive cardiac surgical procedures. *Arch Surg* 2011;146(9):1080-5.
33. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59(1):7-10.
34. Amirian I, Andersen LT, Rosenberg J, Gogenur I. Laparoscopic skills and cognitive function are not affected in surgeons during a night shift. *J Surg Educ* 2014;71(4):543-50.
35. Aziz A. Sources of perceived stress among american medical doctors: a cross-cultural perspective. Cross Cultural Management 2004;11(4):28-39.
36. Beaujouan L, Czernichow S, Pourriat JL, Bonnet F. Prevalence and risk factors for substance abuse and dependence among anaesthetists: a national survey. *Ann Fr Anesth Réanim* 2005;24(5):471-9.
37. Chang LC, Mahoney JJ, 3rd, Raty SR, Ortiz J, Apodaca S, De La Garza R 2nd. Neurocognitive effects following an overnight call shift on faculty anesthesiologists. *Acta Anaesthesiol Scand* 2013;57(8):1051-7.
38. Chen I, Vorona R, Chiu R, Ware JC. A survey of subjective sleepiness and consequences in attending physicians. *Behav Sleep Med* 2008;6(1):1-15.

39. Doppia MA, Estryn-Béhar M, Fry C, Guetarni K, Lieutaud T, comité de pilotage de l'enquête SESMAT. Burnout in French doctors: a comparative study among anaesthesiologists and other specialists in French hospitals (SESMAT study). *Ann Fr Anesth Réanim* 2011;30(11):782-94.
40. Dutheil F, Trousselard M, Perrier C, et al. Urinary interleukin-8 is a biomarker of stress in emergency physicians, especially with advancing age--the JOBSTRESS randomized trial. *PLoS One* 2013;8(8):e71658. https://doi.org/10.1371/journal.pone.0071658
41. Ellman PI, Law MG, Tache-Leon C, et al. Sleep deprivation does not affect operative results in cardiac surgery. *Ann Thorac Surg* 2004;78(3):906-11.
42. Elovainio M, Heponiemi T, Jokela M, et al. Stressful work environment and wellbeing: what comes first? *J Occup Health Psychol* 2015;20(3):289-300.
43. Gander P, Millar M, Webster C, Merry A. Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiol Int* 2008;25(6):1077-91.
44. Gander PH, Merry A, Millar MM, Weller J. Hours of work and fatigue-related error: a survey of New Zealand anaesthetists. *Anaesth Intensive Care* 2000;28(2):178-83.
45. Gerdes J, Kahol K, Smith M, Leyba MJ, Ferrara JJ. Jack Barney award: the effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. *Am J Surg* 2008;196(6):813-9.
46. Harbeck B, Suefke S, Haas CS, Lehnert H, Kropp P, Moenig H. No stress after 24-hour on-call shifts? *J Occup Health* 2015;57(5):438-47.
47. Heponiemi T, Puttonen S, Elovainio M. On-call work and physicians' well-being: testing the potential mediators. *Occup Med* 2014;64(5):352-7.
48. Jackson TN, Pearcy CP, Khorgami Z, Agrawal V, Taubman KE, Truitt MS. The physician attrition crisis: a cross-sectional survey of the risk factors for reduced job satisfaction among US surgeons. *World J Surg* 2017;24. doi: 10.1007/s00268-017-4286-y
49. Kaneita Y, Ohida T. Association of current work and sleep situations with excessive daytime sleepiness and medical incidents among Japanese physicians. *J Clin Sleep Med* 2011;7(5):512-22.
50. Lederer W, Kopp M, Hahn O, et al. Post-duty psychomotor performance in young and senior anaesthetists. *Eur J Anaesthesiol* 2006;23(3):251-6.
51. Leichtfried V, Putzer G, Perkhofer D, Schobersberger W, Benzer A. Circadian melatonin profiles during single 24-h shifts in anesthetists. *Sleep Breath* 2011;15(3):503-12.
52. Lindfors PM, Nurmi KE, Meretoja OA, et al. On-call stress among Finnish anaesthetists. *Anaesthesia* 2006;61(9):856-66.

53. Mahmood JI, Stoen Grotmol K, Tesli M, Vaglum P, Tyssen R. Contextual factors and mental distress as possible predictors of hazardous drinking in Norwegian medical doctors: a 15-year longitudinal, nationwide study. *Eur Addict Res* 2017;23(1):19-27.
54. Nishimura K, Nakamura F, Takegami M, et al. Cross-sectional survey of workload and burnout among Japanese physicians working in stroke care: the nationwide survey of acute stroke care capacity for proper designation of comprehensive stroke center in Japan (J-ASPECT) study. *Circ Cardiovasc Qual Outcomes* 2014;7(3):414-22.
55. Pit SW, Hansen V. Factors influencing early retirement intentions in Australian rural general practitioners. *Occup Med* 2014;64(4):297-304.
56. Pit SW, Hansen V. The relationship between lifestyle, occupational health, and work-related factors with presenteeism amongst general practitioners. *Arch Environ Occup Health* 2016;71(1):49-56.
57. Roberts DL, Shanafelt TD, Dyrbye LN, West CP. A national comparison of burnout and work-life balance among internal medicine hospitalists and outpatient general internists. *J Hosp Med* 2014;9(3):176-81.
58. Rothschild JM, Keohane CA, Rogers S, et al. Risks of complications by attending physicians after performing nighttime procedures. *JAMA* 2009;302(14):1565-72.
59. Saadat H, Bissonnette B, Tumin D, et al. Effects of partial sleep deprivation on reaction time in anesthesiologists. *Paediatr Anaesth* 2017;27(4):358-62.
60. Saadat H, Bissonnette B, Tumin D, et al. Time to talk about work-hour impact on anesthesiologists: The effects of sleep deprivation on Profile of Mood States and cognitive tasks. *Paediatr Anaesth* 2016;26(1):66-71.
61. Sanches I, Teixeira F, dos Santos JM, Ferreira AJ. Effects of acute sleep deprivation resulting from night shift work on young doctors. *Acta Med Port* 2015;28(4):457-62.
62. Sargent MC, Sotile W, Sotile MO, Rubash H, Barrack RL. Quality of life during orthopaedic training and academic practice. Part 1: orthopaedic surgery residents and faculty. *J Bone Joint Surg Am* 2009;91(10):2395-405.
63. Schieman C, MacLean AR, Buie WD, Rudmik LR, Ghali WA, Dixon E. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? *Am J Surg* 2008;195(5):684-7.
64. Sende J, Jbeili C, Schvahn S, Khalid M, Asaph J, Romano H, et al. Stress factors and stress consequences among emergency physicians: vational survey. *Annales Francaises de Médecine d'Urgence* 2012;2(4):224-31.

65. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. *Hum Perform Extrem Environ* 2001;6(1):6-11.

66. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg* 2010;251(6):995-1000.

67. Shanafelt TD, Novotny P, Johnson ME, et al. The well-being and personal wellness promotion strategies of medical oncologists in the North Central Cancer Treatment Group. *Oncology* 2005;68(1):23-32.

68. Shanafelt TD, Raymond M, Kosty M, et al. Satisfaction with work-life balance and the career and retirement plans of US oncologists. *J Clin Oncol* 2014;32(11):1127-35.

69. Shirom A, Nirel N, Vinokur AD. Overload, autonomy, and burnout as predictors of physicians' quality of care. *J Occup Health Psychol* 2006;11(4):328-42.

70. Shirom A, Nirel N, Vinokur AD. Work hours and caseload as predictors of physician burnout: the mediating effects by perceived workload and by autonomy. *J Appl Psychol* 2010;59(4):539-65.

71. Smith F, Goldacre MJ, Lambert TW. Adverse effects on health and wellbeing of working as a doctor: views of the UK medical graduates of 1974 and 1977 surveyed in 2014. *J R Soc Med* 2017;110(5):198-207.

72. Starmer AJ, Frintner MP, Freed GL. Work-life balance, burnout, and satisfaction of early career pediatricians. *Pediatrics* 2016;e20153183. doi: 10.1542/peds.2015-3183

73. Tanti A, Camilleri M, Borg AA, et al. Opinions of Maltese doctors and pharmacists on medication errors. *Int J Saf Med* 2017;29(1-2):81-99.

74. Tokuda Y, Hayano K, Ozaki M, Bito S, Yanai H, Koizumi S. The interrelationships between working conditions, job satisfaction, burnout and mental health among hospital physicians in Japan: a path analysis. *Ind Health* 2009;47(2):166-72.

75. Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. *Am J Surg* 2005;189(6):753-7.

76. Vela-Bueno A, Moreno-Jimenez B, Rodriguez-Munoz A, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. *J Psychosom Res* 2008;64(4):435-42.

77. Vinden C, Nash DM, Rangrej J, et al. Complications of daytime elective laparoscopic cholecystectomies performed by surgeons who operated the night before. *Obstet Gynecol Surv* 2014;69(2):71-3.

78. Wada K, Yoshikawa T, Goto T, et al. National survey of the association of depressive symptoms with the number of off duty and on-call, and sleep hours among physicians working in Japanese hospitals: a cross sectional study. *BMC Public Health* 2010;10:127. doi: 10.1186/1471-2458-10-127

79. Sturm L, Dawson D, Vaughan R, et al. Effects of fatigue on surgeon performance and surgical outcomes: a systematic review. *ANZ J Surg* 2011;81(7-8):502-9.

80. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet* 2009;374(9702):1714-21.

81. Canadian Medical Association. CMA policy: physician health. Ottawa, Canada: Canadian Medical Assocation, 2017. http://policybase.cma.ca/dbtw-wpd/Policypdf/PD18-01.pdf (accessed 12 Jan 2018).

82. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Intern Med* 2017;177(2):195-205.

83. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. *N Engl J Med* 2002;347(16):1249-55.

84. Ward S, Outram S. Medicine: in need of culture change. *Itern Med J* 2016;46(1):112-6.

85. Panagioti M, Panagopoulou E, Bower P, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med* 2017;177(2):195-205.

86. American Medical Association. STEPSforward™. Chicago, IL: American Medical Assocation, 2017. https://www.stepsforward.org/modules/joy-in-medicine (accessed 12 Jan 2018).

87. Jackson C. The Chalder Fatigue scale (CFQ 11). Occup Med 2015;65(1):86. doi: 10.1093/occmed/kqu168

88. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *J Psychosom Res* 1993;37(2):147-53.

89. Smets E, Garssen B, Bonke Bd, De Haes J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* 1995;39(3):315-25.

90. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383(9912):166-75.

91. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. *Implement Sci* 2016;11:38. https://doi.org/10.1186/s13012-016-0399-1

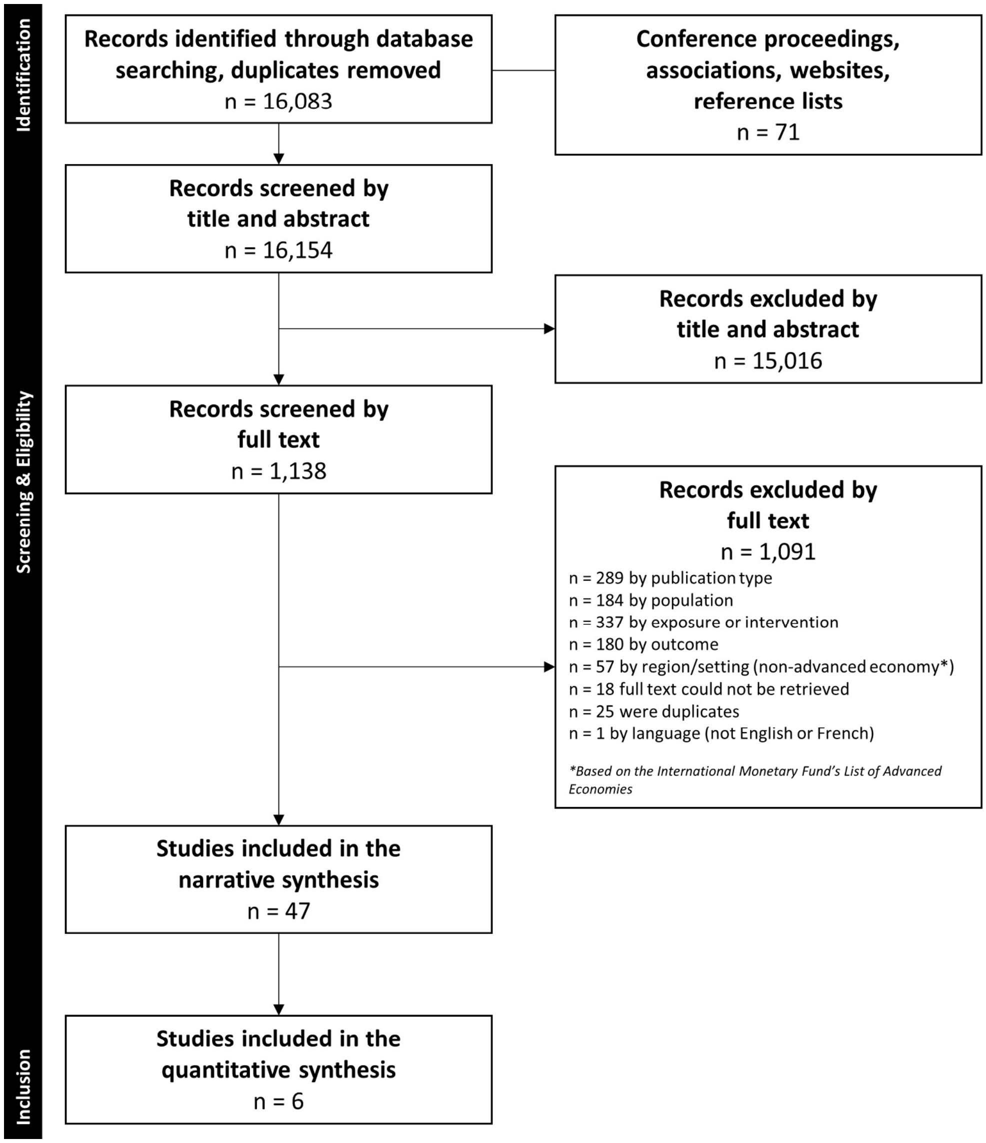


Figure 1. Flow of records through the selection process

190x215mm (300 x 300 DPI)

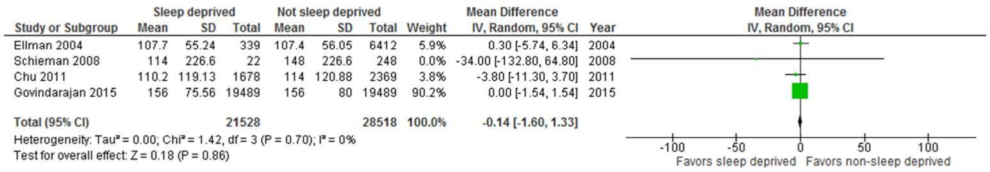


Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

381x101mm (300 x 300 DPI)

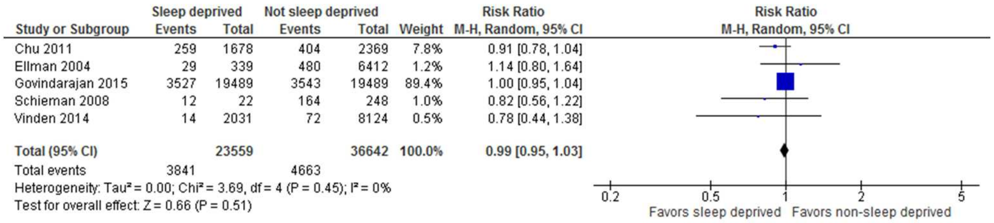


Figure 3. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

Legend: Vinden 2013 reported iatrogenic injuries; Schieman 2008, Govindarajan 2015, and Chu 2011 reported post-operative complication rate; Ellman 2004 reported post-operative complications (other types of complications reported not included in the analysis)

381x101mm (300 x 300 DPI)

Figure 4. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

381x101mm (300 x 300 DPI)

Supplementary file 1. Search Strategy

Database: In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

Date searched: 13 April 2016, updated 7 November 2017

Records retrieved: 5068 and 1442 in the update (removed duplicates retrieved in previous search)

1. Medical Staff, Hospital/
2. Physician Impairment/
3. exp Physicians/
4. allergist*.ti.
5. (an?esthetist* or an?esthesiologist*).ti.
6. cardiologist*.ti.
7. clinician*.ti.
8. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
9. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
10. dermatologist*.ti.
11. endocrinologist*.ti.
12. doctor*.ti.
13. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
14. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
15. family practitioner*.ti.
16. gastroenterologist*.ti.
17. (general practitioner* or GP*).ti.
18. (general adj2 physician*).ti.
19. geriatrician*.ti.
20. gyn?ecologist*.ti.

21. h?ematologist*.ti.

22. (health* adj2 (professional* or provider*)).ti.

23. hospitalist*.ti.

24. (house staff* or housestaff*).ti.

25. intensivist*.ti.

26. internist*.ti.

27. medical professional*.ti.

28. obstetrician*.ti.

29. oncologist*.ti.

30. ophthalmologist*.ti.

31. orthop?edist*.ti.

32. (otolaryngologist* or otorhinolaryngologist*).ti.

33. neonatologist*.ti.

34. nephrologist*.ti.

35. neurologist*.ti.

36. neuropsychiatrist*.ti.

37. neurosurgeon*.ti.

38. p?ediatrician*.ti.

39. perinatologist*.ti.

40. physician*.ti.

41. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

42. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.

43. primary care practitioner*.ti.

44. psychiatrist*.ti.

45. pulmonologist*.ti.

46. rheumatologist*.ti.

47. surgeon*.ti.

48. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

49. traumatologist*.ti.

50. urologist*.ti.

51. or/1-50 [Combined MeSH, title, and text word searches for physicians]

52. Burnout, Professional/

53. exp Circadian Rhythm/

54. exp Fatigue/

55. Occupational Health/

56. Rest/ph, px [Physiology, Psychology]

57. Sleep Deprivation/

58. Sleep Disorders, Circadian Rhythm/

59. Sleep Wake Disorders/

60. exp Stress, Psychological/

61. Workload/px [Psychology]

62. Work Schedule Tolerance/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw,kf.

64. biological rhythm*.tw,kf.

65. (burn out* or burned out* or burnt out* or burnout*).tw,kf.

66. circadian misalignment.tw,kf.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw,kf.

68. exhaust*.tw,kf.

69. fatigu*.tw,kf.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw,kf.

71. tired*.tw,kf.

72. weariness.tw,kf.

73. or/52-72 [Combined MeSH and text words for fatigue]

74. and/51,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (comment or editorial or letter).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. remove duplicates from 80

Database: Ovid Embase 1996 to 2016 Week 15

Date searched: 13 April 2016

Records retrieved: 8859

1. medical staff/
2. exp physician/
3. allergist*.ti.
4. (an?esthetist* or an?esthesiologist*).ti.
5. cardiologist*.ti.
6. clinician*.ti.
7. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
8. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
9. dermatologist*.ti.
10. endocrinologist*.ti.
11. doctor*.ti.
12. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
13. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
14. family practitioner*.ti.
15. gastroenterologist*.ti.
16. (general practitioner* or GP*).ti.
17. (general adj2 physician*).ti.

18. geriatrician*.ti.

19. gyn?ecologist*.ti.

20. h?ematologist*.ti.

21. (health* adj2 (professional* or provider*)).ti.

22. hospitalist*.ti.

23. (house staff* or housestaff*).ti.

24. intensivist*.ti.

25. internist*.ti.

26. medical professional*.ti.

27. obstetrician*.ti.

28. oncologist*.ti.

29. ophthalmologist*.ti.

30. orthop?edist*.ti.

31. (otolaryngologist* or otorhinolaryngologist*).ti.

32. neonatologist*.ti.

33. nephrologist*.ti.

34. neurologist*.ti.

35. neuropsychiatrist*.ti.

36. neurosurgeon*.ti.

37. p?ediatrician*.ti.

38. perinatologist*.ti.

39. physician*.ti.

40. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

41. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

42. primary care practitioner*.ti.

43. psychiatrist*.ti.

44. pulmonologist*.ti.

45. rheumatologist*.ti.

46. surgeon*.ti.

47. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

48. traumatologist*.ti.

49. urologist*.ti.

50. or/1-49 [Combined Emtree, title, and text word searches for physicians]

51. burnout/

52. circadian rhythm/

53. circadian rhythm sleep disorder/

54. fatigue/

55. mental stress/

56. occupational health/

57. sleep deprivation/

58. sleep waking cycle/

59. work capacity/

60. work schedule/

61. working time/

62. workload/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

64. biological rhythm*.tw.

65. (burn out* or burned out* or burnt out* or burnout*).tw.

66. circadian misalignment.tw.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw.

68. exhaust*.tw.

69. fatigu*.tw.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

71. tired*.tw.

72. weariness.tw.

73. or/51-72 [Combined Emtree and text words for fatigue]

74. and/50,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (conference* or editorial or letter or proceeding).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. limit 80 to embase

Database: Ovid PsycINFO 1987 to April Week 1 2016

Date searched: 13 April 2016

Records retrieved: 2094

1. exp Physicians/
2. allergist*.ti.
3. (an?esthetist* or an?esthesiologist*).ti.
4. cardiologist*.ti.
5. clinician*.ti.
6. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
7. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
8. dermatologist*.ti.
9. endocrinologist*.ti.
10. doctor*.ti.
11. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
12. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
13. family practitioner*.ti.
14. gastroenterologist*.ti.
15. (general practitioner* or GP*).ti.
16. (general adj2 physician*).ti.

17. geriatrician*.ti.

18. gyn?ecologist*.ti.

19. h?ematologist*.ti.

20. (health* adj2 (professional* or provider*)).ti.

21. hospitalist*.ti.

22. intensivist*.ti.

23. internist*.ti.

24. medical professional*.ti.

25. obstetrician*.ti.

26. oncologist*.ti.

27. ophthalmologist*.ti.

28. orthop?edist*.ti.

29. (otolaryngologist* or otorhinolaryngologist*).ti.

30. neonatologist*.ti.

31. nephrologist*.ti.

32. neurologist*.ti.

33. neuropsychiatrist*.ti.

34. neurosurgeon*.ti.

35. p?ediatrician*.ti.

36. perinatologist*.ti.

37. physician*.ti.

38. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

39. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

40. primary care practitioner*.ti.

41. psychiatrist*.ti.

42. pulmonologist*.ti.

43. rheumatologist*.ti.

44. surgeon*.ti.

45. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

46. traumatologist*.ti.

47. urologist*.ti.

48. or/1-47 [Combined thesaurus, title, and text word searches for physicians]

49. Compassion Fatigue/

50. Fatigue/

51. Human Biological Rhythms/

52. Occupational Health/

53. Occupational Stress/

54. Sleep/

55. Sleepiness/

56. Working Conditions/

57. Work Rest Cycles/

58. Work Week Length/

59. Work Scheduling/

60. Workday Shifts/

61. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

62. biological rhythm*.tw.

63. (burn out* or burned out* or burnt out* or burnout*).tw.

64. circadian misalignment.tw.

65. ((circadian or diurnam or ultradian) adj rhythm*).tw.

66. exhaust*.tw.

67. fatigu*.tw.

68. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

69. tired*.tw.

70. weariness.tw.

71. or/49-70 [Combined thesaurus and text words for fatigue]

72. and/48,71 [Combined concepts for physicians and fatigue]

73. limit 72 to yr="2000-Current"

74. limit 73 to (english or french)

Database: CINAHL Plus with Full Text (1937 to the present) via EBSCOhost

Date searched: 14 April 2016

Records retrieved: 3378

S1. (MH "Medical Staff, Hospital+")

S2. (MH "Physicians+")

S3. TI allertist*

S4. TI (anesthetist* or anaesthetist* or anesthesiologist* or anaesthesiologist*)

S5. TI cardiologist*

S6. TI clinician*

S7. clinician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S8. clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S9. TI dermatologist*

S10. TI endocrinologist*

S11. TI doctor*

S12. doctor* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S13. doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S14. TI "family practitioner*"

S15. TI gastroenterologist*

S16. TI ("general practitioner*" or GP*)

S17. TI (general N2 physician*)

S18. TI geriatrician*

S19. TI (gynaecologist* or gynecologist*)

S20. TI (haematologist* or hematologist*)

S21. TI hospitalist*

S22. TI ("house staff*" or housestaff*)

S23. TI intensivist*

S24. TI internist*

S25. TI obstetrician*

S26. TI oncologist*

S27. TI ophthalmologist*

S28. TI (orthopaedist* or orthopedist*)

S29. TI (otolaryngologist* or otorhinolaryngologist*)

S30. TI neonatologist*

S31. TI nephrologist*

S32. TI neurologist*

S33. TI neuropsychiatrist*

S34. TI neurosurgeon*

S35. TI (paediatrician* OR pediatrician*)

S36. TI perinatologist*

S37. TI physician*

S38. physician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S39. physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S40. TI "primary care practitioner*"

S41. TI psychiatrist*

S42. TI pulmonologist*

S43. TI rheumatologist*

S44. TI surgeon*

S45. surgeon* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S46. TI traumatologist*

S47. TI urologist*

S48. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28

OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47

S49. (MH "Circadian Rhythm")

S50. (MH "Fatigue")

S51. (MH "Impairment, Health Professional")

S52. (MH "Mental Fatigue")

S53. (MH "Occupational Health")

S54. (MH "Shiftwork")

S55. (MH "Sleep Deprivation")

S56. (MH "Sleep Disorders, Circadian Rhythm")

S57. (MH "Sleep-Wake Transition Disorders")

S58. (MH "Stress, Occupational+")

S59. (MH "Stress, Psychological")

S60. ("24 hour*" or "24 hr*" or "twenty four hour*" or "twentyfour hour*") N1 rhythm*

S61. "biological rhythm*"

S62. "burn out*" or "burned out*" or "burnt out*" or burnout*

S63. "circadian misalignment"

S64. (circadian or diurnam or ultradian) N1 rhythm*

S65. exhaust*

S66. fatigu*

S67. sleep* N3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)

S68. tired*

S69. weariness

S70. S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69

S71. S48 AND S70

S72. S48 AND S70 Limiters - Published Date: 20000101-20161231; Publication Type: Clinical Trial, Journal Article, Meta Analysis, Meta Synthesis, Practice Guidelines, Randomized Controlled Trial, Research, Review, Systematic Review; Language: English, French

Database: PubMed via NCBI Entrez

Date searched: 14 April 2016

Records retrieved: 92

(((("Medical Staff, Hospital"[mh:noexp] OR "Physician Impairment"[mh:noexp] OR "Physicians"[mh] OR allergist[ti] OR allergists[ti] OR anaesthetist[ti] OR anaesthetists[ti] OR anaesthesiologist[ti] OR anaesthesiologists[ti] OR anesthetist[ti] OR anesthetists[ti] OR anesthesiologist[ti] OR anesthesiologists[ti] OR cardiologist[ti] OR cardiologists[ti] OR clinician[ti] OR clinicians[ti] OR ((clinician[tiab] OR clinicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((clinician[tiab] OR clinicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR dermatologist[ti] OR dermatologists[ti] OR endocrinologist[ti] OR endocrinologists[ti] OR doctor[ti] OR doctors[ti] OR ((doctor[tiab] OR doctors[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((doctor[tiab] OR doctors[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "family practitioner"[ti] OR

"family practitioners"[ti] OR gastroenterologist[ti] OR gastroenterologists[ti] OR "general practice physician"[ti] OR "general practice physicians"[ti] OR "general practitioner"[ti] OR "general practitioners"[ti] OR geriatrician[ti] OR geriatricians[ti] OR gynaecologist[ti] OR gynaecologists[ti] OR gynecologist[ti] OR gynecologists[ti] OR haematologist[ti] OR haematologists[ti] OR hematologist[ti] OR hematologists[ti] OR "health care professional"[ti] OR "health care professionals"[ti] AND "health care provider"[ti] OR "health care providers" OR "health professional"[ti] OR "health professionals"[ti] OR "health provider"[ti] OR "health providers"[ti] OR "healthcare professional"[ti] OR "healthcare professionals"[ti] OR "healthcare provider"[ti] OR "healthcare providers"[ti] OR hospitalist[ti] OR hospitalists[ti] OR "house staff"[ti] OR "house staffs"[ti] OR housestaff[ti] OR housestaffs[ti] OR intensivist[ti] OR intensivists[ti] OR internist[ti] OR internists[ti] OR "medical professional"[ti] OR "medical professionals"[ti] OR obstetrician[ti] OR obstetricians[ti] OR oncologist[ti] OR oncologists[ti] OR ophthalmologist[ti] OR ophthalmologists[ti] OR orthopaedist[ti] OR orthopaedists[ti] OR orthopedist[ti] OR orthopedists[ti] OR otolaryngologist[ti] OR otolaryngologists[ti] OR otorhinolaryngologist[ti] OR otorhinolaryngologists[ti] OR neonatologist[ti] OR neonatologists[ti] OR nephrologist[ti] OR nephrologists[ti] OR neurologist[ti] OR neurologists[ti] OR neuropsychiatrist[ti] OR neuropsychiatrists[ti] OR neurosurgeon[ti] OR neurosurgeons[ti] OR paediatrician[ti] OR paediatricians[ti] OR pediatrician[ti] OR pediatricians[ti] OR perinatologist[ti] OR perinatologists[ti] OR physician[ti] OR physicians[ti] OR ((physician[tiab] OR physicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((physician[tiab] OR physicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "primary care practitioner"[ti] OR "primary care practitioners"[ti] OR psychiatrist[ti] OR psychiatrists[ti] OR pulmonologist[ti] OR pulmonologists[ti] OR rheumatologist[ti] OR rheumatologists[ti] OR surgeon[ti] OR surgeons[ti] OR ((surgeon[tiab] OR surgeons[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR

"burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR traumatologist[ti] OR traumatologists[ti] OR urologist[ti] OR urologists[ti]) AND ("Burnout, Professional"[mh:noexp] OR "Circadian Rhythm"[mh] OR "Fatigue"[mh] OR "Occupational Health"[mh:noexp] OR "Rest/physiology"[mh:noexp] OR "Rest/psychology"[mh:noexp] OR "Sleep Deprivation"[mh:noexp] OR "Sleep Disorders, Circadian Rhythm"[mh:noexp] OR "Stress, Psychological"[mh] OR "Workload/psychology"[mh] OR "Work Schedule Tolerance"[mh:noexp] OR "24 hour rhythm"[tiab] OR "24 hour rhythms"[tiab] OR "24 hr rhythm"[tiab] OR "24 hr rhythms"[tiab] OR alertness[tiab] OR "biological rhythm"[tiab] OR "biological rhythms"[tiab] OR "burn out"[tiab] OR "burned out"[tiab] OR "burnt out"[tiab] OR burnout[tiab] OR "circadian misalignment"[tiab] OR "circadian rhythm"[tiab] OR "circadian rhythms"[tiab] OR "diurnal rhythm"[tiab] OR "diurnal rhythms"[tiab] OR exhausted[tiab] OR exhaustion[tiab] OR exhausting[tiab] OR exhausts[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR (("Sleep"[mh:noexp] OR sleep[tiab] OR sleeping[tiab]) AND (deprivation[tiab] OR deprive[tiab] OR deprived[tiab] OR deprives[tiab] OR depriving[tiab] OR disorder[tiab] OR disorders[tiab] OR lack[tiab] OR lacked[tiab] OR lacking[tiab] OR lacks[tiab] OR loss[tiab] AND insufficient[tiab] OR problem[tiab] OR problems[tiab])) OR tired[tiab] OR tiredness[tiab] OR "twenty four hour rhythm"[tiab] OR "twenty four hour rhythms"[tiab] OR weariness[tiab] OR "ultradian rhythm"[tiab] OR "ultradian rhythms"[tiab])) NOT (((Animals[MESH] OR Animal Experimentation[MESH] OR "Models, Animal"[MESH] OR Vertebrates[MESH]) NOT (Humans[MESH] OR Human experimentation[MESH])) OR (((animals[tiab] OR animal model[tiab] OR rat[tiab] OR rats[tiab] OR mouse[tiab] OR mice[tiab] OR rabbit[tiab] OR rabbits[tiab] OR pig[tiab] OR pigs[tiab] OR porcine[tiab] OR swine[tiab] OR dog[tiab] OR dogs[tiab] OR hamster[tiab] OR hamsters[tiab] OR chicken[tiab] OR chickens[tiab] OR sheep[tiab]) AND (publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb])) NOT (human[ti] OR humans[ti] OR people[ti] OR children[ti] OR adults[ti] OR seniors[ti] OR patient[ti] OR patients[ti])))) NOT (editorial[pt] OR comment[pt] OR letter[pt] OR newspaper article[pt])) AND ((publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR (pubstatUSheadofprint))

Filters activated: Publication date from 2000/01/01 to 2016/12/31, English, French.

Supplementary table 1. Descriptive characteristics of the included studies

Study	Physician and patient characteristics				Setting		Interventions or	Outcomes
Country	**Type**	**n=**	**Sex (% male)**	**Age**	**Location**	**Urban or rural**	**exposures**	
Observational (exposure) studies (n=45)								
Cohort design								
Chu, 2011 [32] Canada	Surgeons	6	NR	Range: 32-55y	Tertiary care academic hospital	Urban	Insufficient sleep due to work on the night preceding surgery	Length of surgery; patient postoperative mortality, complications, length of stay
	Patients: cardiac surgery cases	4,047	NR	NR				
Ellman, 2004 [41] US	Surgeons	NR	NR	NR	University hospitals	Urban	Insufficient sleep due to work on the night preceding surgery	Length of surgery; patient complications, in-hospital mortality, length of stay, need for blood products
	Patients: adult cardiac surgery cases	6,751	70%	S: 63.4±0.7y C: 63.5±0.1y				
Govindarajan, 2015 [31] Canada	Surgeons	1,448	NR	46.3±8.7	Academic and non-academic hospitals	Mixed	Sleep deprivation due to work on the night preceding a daytime surgery	Length of surgery; Patient complications, mortality, readmissions, length of stay
	Patients: surgical cases	38,978	NR	56.4±16.6y				
Rothschild, 2009 [58] US	Surgeons Obstetrician/gynecologists	220	Surgeons: 84% OB/GYNs: 28%	Surgeons: 42.0±7.6y OB/GYNs: 42.0±9.0y	Tertiary care academic trauma centre/referral centre for high-risk obstetrics	Urban	Sleep deprivation due to work on the night preceding a daytime procedure	Patient complications, preventable complications
	Patients: surgical and obstetrics cases	Surg.: 4,471 Obst.: 4,902	Surg: S: 25% C: 28% Obst.: S: 0% C: 0%	Surg: S: 49.1±16.3y C: 50.0±16.3y Obst.: S: 32.9±5.2y C: 33.5±5.0y				
Schieman, 2007 [63] Canada	Colorectal surgeons	NR	NR	NR	University teaching hospitals	NR	Fatigue due to work on the night preceding surgery	Length of surgery; patient operative complications, length of stay, mortality, cancer recurrence
	Patients: undergoing anterior resection for rectal cancer	270	NR	S: 64.5y C: 64.4y				

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Vinden, 2014 [77] Canada	General surgeons	331	83%	48±10y	Community hospitals	Mixed	Sleep deprivation due to overnight work preceding daytime surgery	Patient mortality, operative complications
	Patients: Elective cholecystectomies	10,390	S: 27% C: 26%	S: 49±16y C: 49±16y				
Before-after design								
Amirian, 2014 [34] Denmark	Surgeons	29	55%	Median: 35y Range: 27-49y	Academic hospital	Urban	17-h night shift with sleep deprivation	Cognitive and psychomotor abilities on a laparoscopic simulation
Gerdes, 2008 [45] US	Surgeons	9	NR	NR	University Hospital	Urban	Fatigue; sleep deprivation from overnight call shift	Cognitive and psychomotor abilities
Lederer, 2006 [50] Austria	Senior anesthetists	11	82%	49.0±2.0y	Hospital	Urban	Sleep deprivation from 24-h call shift	Concentration ability; reaction time; performance on psychometric tasks
Time series design								
Leichtfried, 2011 [51] Austria	Anesthetists	10	100%	Mean: 32y Range: 29-35y	University Hospital	Urban	Sleep deprivation from 24-h shift; sleepiness, sleep hours	Melatonin metabolite profile
Cross-sectional design								
Aziz, 2004 [35] US	Family medicine physicians Various specialties	153	NR	NR	Hospitals	NR	Fatigue	Stress
Beaujouan, 2005 [36] France	Anesthesiologists	3,476	64%	≤35y: 9% 36-45y: 28% 46-55y: 49% 56-65y: 13%	Public sector General hospitals University hospitals Private hospitals	NR	Sleep deprivation	Substance abuse
Chang, 2013 [37] US	Anesthesiologists	11	64%	Mean: 38y IQR: 34-48y	Level 1 trauma centre	NR	Sleep deprivation due to 15-h overnight call shift; sleepiness	Cognitive performance; reaction time

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Chen, 2008 [38] US	Psychiatrists Internists General practitioners Surgeons Obstetrician-gynecologists Radiologists Pediatricians Other	180	77%	Academic: 79% 36-55y Private practice: 73% 36-65y	Medical school Private practices	Urban	Sleep deprivation or sleepiness	Impact on personal and professional life; perceived risk of errors
Doppia, 2011 [39] France	Anesthesiologists	565	64%	<35y: 11% 35-54y: 63% >55y: 25%	Public hospitals Private hospitals Work-health environments Public health units	NR	Sleep deprivation	Burnout
Elovaino, 2015 [42] Finland	Physicians in various specialties	1,524	40%	Median: 49.7y Range: 24-69y	Hospitals Primary care Private practice Other unspecified	NR	Sleep difficulties	Job demands and control
Gander, 2000 [43] New Zealand	Anesthetists	183	NR	Mean: 46y	Combined public/private practice Other unspecified	NR	Work-related sleep disturbance	Risk of fatigue-related errors
Harbeck, 2015 [46] Germany	Internists	20	45%	Median: 32y Range: 26-42y	Hospital	NR	Sleep disturbance due to a 24-call shift	Biochemical and physiological parameters; neurocognitive function
Heponiemi, 2014 [47] Finland	Physicians in various specialties Non-specialized physicians	1,541	40%	49.80±9.49y, Range: 24-67y	Hospitals Primary care clinic Private practice Other unspecified	NR	Sleep difficulties	Job satisfaction; work ability; psychological distress

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Jackson, 2017 [48] US	Surgeons in various subspecialties	993	61%	More; less satisfied: 30-39y: 23%;24% 40-49y: 32%;36% 50-59y: 23%;27% ≥60y: 23%;14%	Academic practice Non-academic practice	NR	Not feeling well rested	Job satisfaction
Kanieta, 2011 [49] Japan	Internists Surgeons Orthopedics Pediatricians Obstetrician-gynecologists Psychiatrists Dermatologists Urologists Opthalmologists Otorhinolaryngologists Other	3,486	66%	20-39y: 11% 40-49y: 25% 50-59y: 28% 60-69y: 16% ≥70y: 21%	Hospitals Clinics Other unspecified	NR	Sleep deprivation and difficulties; insomnia	Medical incidents
Lindfors, 2006 [52] Finland	Anesthetists	328	53%	47±7.8y Range: 32-69y	University hospitals Central and district hospitals Private sector	NR	Sleep disturbances; sleepiness	Stress; suicidal tendencies
Mahmood, 2016 [53] Norway	Generalists Internists Pediatricians Surgical specialties Anesthesiologists	450 (all time points)	41%	43y±2.8y	Public health system Private practice	NR	Sleep deprivation due to on-call shifts	Alcohol misuse

Study	Physician and patient characteristics				Setting		Interventions or	Outcomes
Country	Type	n=	Sex (% male)	Age	Location	Urban or rural	exposures	
Nishimura, 2014 [54] Japan	Neurosurgeons and neurologists	2,564	NR	NR	Stroke care centres Teaching hospitals	NR	Sleep deprivation	Burnout
Pit, 2014 [55] Australia	General practitioners	92	60%	50±10.7y	NR	Rural	Work-related sleep disturbance	Early retirement intentions
Pit, 2016 [56] Australia	General practitioners	92	60%	50±10.7y	Private (solo) practice Group practice	Rural	Work-related sleep disturbance	Sickness presenteeism
Roberts, 2014 [57] US	General internists Internal medicine hospitalists	578	58%	Hospitalists: 46.9±12.4y Generalists: 53.6±10.2y	Private practice Academic medical centre Veterans hospital Military practice Other	NR	Fatigue	Falling asleep while driving
Saadat, 2016 [60] US	Anesthesiologists	21	71%	30-40y: 57% 41-50y: 19% 51-55y: 24% Range: 32-56y	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Mood disturbances
Saadat, 2017 [59] US	Anesthesiologists	21	65%	Range: 32-56 years	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Reaction time
Sanches, 2015 [61] Spain	Emergency medicine physicians	18	28%	29.2±2.6y	Central hospital	NR	Sleep deprivation	Cognitive and psychomotor abilities
Sargent, 2009 [62] US	Orthopedic surgeons	264	92%	NR	Orthopedic surgery training programs	NR	Sleep deprivation	Burnout; psychological distress; marital satisfaction
Sende, 2012 [64] France	Emergency physicians	318	62%	39±8y	Hospitals Mobile emergency services Other unspecified	NR	Fatigue; sleep deprivation	Stress

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Sexton, 2001 [65] US	Consulting physicians: Surgeons Anesthesiologists Pulmonary physicians Cardiologists Pediatricians	271	NR	NR	Teaching and non-teaching hospitals	Urban	Fatigue	Perceived performance effectiveness
Shanafelt, 2005 [67] US, Canada, Mexico	Oncologists	241	85%	>50y: 51%	Community clinics Hospitals Private practice Academic medical centres	NR	Fatigue; sleep deprivation	Quality of life/well-being
Shanafelt, 2010 [66] US	Surgeons	7,905	87%	Median: 51y Q1: 43y Q2: 59y	Private practice Academic medical centres Veterans hospital Active military practice Retired or not in practice Other	NR	Fatigue	Perceived major medical errors
Shanafelt, 2014 [68] US	Oncologists	1,117	52%	Median: 52y	Private practice Academic practice Veteran's hospital Industry, other	NR	Fatigue	Satisfaction with work-life balance
Shirom, 2006 [69] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Perception of quality of patient care

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Shirom, 2010 [70] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Burnout
Smith, 2017 [71] UK	General practitioners Surgeons Other unspecified specialties	3,550	63%	NR	NR (varied)	NR	Perceived fatigue; sleep deprivation	Physical and mental health; competence
Starmer, 2016 [72] US	General pediatricians Pediatric surgeons Pediatric hospitalists Pediatric specialists (unspecified)	840	40%	NR	NR (some in private practice)	NR	Sleep deprivation	Burnout; balanced personal and professional commitments; life and career satisfaction
Tanti, 2017 [73] Malta	Physicians (unspecified)	204	62%	Median: 41y	Hospitals Community Office-based	NR	Fatigue	Prescribing errors
Tokuda, 2009 [74] Japan	Hospital physicians: Generalists Other unspecified specialties	236	75%	40.9±7.8y Range: 26-76y	Hospitals with ≥20 inpatient beds	NR	Sleep deprivation	Burnout; job satisfaction
Vela-Bueno, 2008 [76] Spain	Primary care physicians	113	27%	41.4±8.0y	Primary care centres	Urban	Sleep problems, insomnia	Burnout
Wada, 2010 [78] Japan	Physicians (unspecified)	3,862	78%	M: 75% 30-59y F: 85% 30-59y	Hospitals	NR	Sleep deprivation	Depressive symptoms

Study Country	Physician and patient characteristics Type	 n=	 Sex (% male)	 Age	Setting Location	 Urban or rural	Interventions or exposures	Outcomes
Non-comparative design								
Gander, 2008 [43] New Zealand	Anesthetists	20	85%	Median: 44y	Hospitals	Urban	Sleep disturbance from consecutive working days of on-call work	Psychomotor performance
Intervention studies (n=2)								
Randomized controlled trials								
Dutheil, 2013 [40] France	Emergency physicians	17	35%	39.1y±6.9y	University hospital	Urban	Fatigue related to 14-h and 24-h shifts; sleep deprivation; low sleep quality;	Perceived stress; urine interleukine-8
Uchal, 2005 [75] Norway	Surgeons Gynecologists Orthopedic surgeons Urologists Vascular surgeons	64	67%	Median: Post-call: 33.0y Post-work: 38.0y	Government hospitals	NR	Sleep deprivation due to 24-h call shift	Product quality, procedure effectiveness of a surgical simulation

C: control group; F: female; h: hour(s); IQR: interquartile range; M: male; NR: not reported; S: study group; SD: standard deviation; Surg: surgical; Obst: obstetric; Q: quartile; UK: United Kingdom; US: United States of America; y: year(s)

Supplementary file 3. Risk of bias assessments

Summary of risk of bias assessments for randomized controlled trials (n=2)[a]

First Author, Year	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias	Overall risk of bias[b]
Dutheil, 2013	Low	Unclear	High	High	Low	Low	High	High
Uchal, 2005	Low	Low	Unclear	Low	Low	Low	Low	Unclear

[a]Assessed using the Cochrane Collaboration's Risk of Bias Tool

[b]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cohort studies (n=6)[a]

First Author, Year	Selection					Comparability		Outcome				Total
	Representa-tiveness of exposed cohort /1	Selection of non-exposed cohort /1	Ascertain-ment of exposure /1	Outcome not present at start /1	Total /4	Compara-bility of cohorts /2	Total /2	Assess-ment of outcome /1	Adequate length of follow-up /1	Adequate follow-up of cohorts /1	Total /1	Score[b] /9
Chu, 2011	1	1	0	1	3	2	2	1	1	1	3	8
Ellman, 2004	1	1	1	1	4	1	1	1	1	1	3	8
Govindarajan, 2015	1	1	1	1	4	2	2	1	1	1	3	9
Rothschild, 2009	1	1	1	1	4	2	2	1	1	1	3	9
Schieman, 2008	1	1	1	1	4	1	1	1	1	1	3	8
Vinden, 2014	1	1	1	1	4	1	1	1	1	1	3	8

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale

[b]An overall score of 7 to 9 stars is considered as low risk of bias, 4 to 6 as unclear risk of bias, and 3 or less as high risk of bias

Summary of risk of bias assessments for before-after studies (n=3)[a]

First Author, Year	Random sequence generation[b]	Allocation concealment[b]	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Amirian, 2014	NA	NA	High	High	Low	Low	High	High
Gerdes, 2008	NA	NA	High	High	Low	Low	High	High
Lederer, 2006	NA	NA	High	High	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for before-after studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as 'not applicable' (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of risk of bias assessments for time series studies (n=1)[a]

First Author, Year	Intervention independent of other changes	Intervention effect pre-specified	Intervention unlikely to affect data collection	Allocation concealment[a]	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Leitchfried, 2011	Low	High	Low	NA	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for interrupted time series studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as not applicable (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cross-sectional studies (n=34)[a]

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Aziz, 2004	0	0	0	0	1	0	1	1
Beaujouan, 2005	1	0	1	0	1	0	1	2
Chang, 2013	1	0	1	0	1	1	2	3
Chen, 2008	1	0	1	0	1	0	1	2
Doppia, 2011	1	1	2	0	1	1	2	4
Elovaino, 2015	1	1	2	0	1	1	2	4
Gander, 2000	1	1	2	0	1	1	2	4
Harbeck, 2015	1	0	1	0	1	1	2	3
Heponiemi, 2014	1	1	2	0	1	1	2	4
Jackson, 2017	0	0	0	0	1	0	1	1
Kanieta, 2011	1	0	1	0	1	1	2	3
Lindfors, 2006	1	1	2	0	1	1	2	4
Mahmood, 2017	1	0	1	0	1	0	1	2
Nishimura, 2014	1	1	2	0	1	0	1	3
Pit, 2014	1	0	1	0	1	1	2	3
Pit, 2016	1	0	1	0	1	1	2	3
Roberts, 2014	1	1	2	0	1	0	1	3
Saadat, 2016	1	1	2	0	1	1	2	4
Saadat, 2017	1	1	2	0	1	1	2	4
Sanches, 2015	1	0	1	0	1	0	1	2
Sargent, 2009	1	0	1	0	1	0	1	2

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Sende, 2010	1	0	1	0	1	0	1	2
Sexton, 2001	1	0	1	0	1	0	1	2
Shanafelt, 2005	1	0	1	0	1	1	2	3
Shanafelt, 2010	1	1	2	0	1	0	1	3
Shanafelt, 2014	1	0	1	0	1	1	2	3
Shirom, 2006	1	1	2	0	1	1	2	4
Shirom, 2010	1	1	2	0	1	1	2	4
Smith, 2016	1	0	1	0	1	1	2	3
Starmer, 2016	1	1	2	0	1	1	2	4
Tanti, 2017	1	0	1	0	1	0	1	2
Tokuda, 2009	1	1	2	0	1	1	2	4
Vela-Bueno, 2008	1	1	2	0	1	1	2	4
Wada, 2010	1	1	2	0	1	0	1	3

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted for cross-sectional studies

[b]An overall score of 4 to 5 stars is considered as low risk of bias, 3 as unclear risk of bias, and 2 or less as high risk of bias. For response rate, ≥50% was used as the criterion to be awarded a star

Summary of quality assessments for non-comparative studies (n=1)[a]

First Author, Year	Selection			Exposure		Outcome				Total
	Adequacy of case definition /1	Representativeness of the sample /1	Total /2	Ascertainment of exposure	Total /1	Assessment of outcome /1	Same method of assessment for entire sample /1	Loss to follow-up /1	Total /3	Score[b] /6
Gander, 2008	1	1	2	0	0	0	1	1	2	4

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted by the authors to be suitable to the non-comparative design

[b]An overall score of 5 to 6 stars is considered as low risk of bias, 3 to 4 as unclear risk of bias, and 2 or less as high risk of bias

Supplementary file 4. Detailed study outcomes

Physician health and wellness outcomes and associations with fatigue

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Jackson, 2017 RoB: high	CS	Not feeling well rested: self-reported as 'unhealthy' Time points NR	71% healthy, 28% unhealthy in terms of being well rested	Job satisfaction: Abridged Job in General Scale; grouped into more or less satisfied using the median Time points NR	Job satisfaction in those more vs. less satisfied: Healthy (well rested): 85% vs. 58%, p<0001; Unhealthy (not well rested): 15% vs. 42%, p<0.001.
Nishimura, 2014 RoB: unclear	CS	Sleep hours/night: self-reported (continuous) Time points NR	Mean±SD sleep: 5.94±1.08h	Burnout: Japanese MBI (severe: EE >4.0 and either DP >2.6 or PE <4.17) Time points NR	1) Mean±SD sleep for not burned out vs. mild to moderate vs. severe: 6.07±1.15 vs. 5.88±0.94 vs. 5.63±0.94, p<0.05. 2) Association between sleep and burnout (OR (95% CI)): bivariate 0.67 (0.61-0.73), p<0.001; multivariate including work characteristics and mental health: 0.84 (0.75-0.94), p=0.002.
Sargent, 2009 RoB: high	CS	Sleep deprivation: self-reported on a 4-point scale (none, a little, quite a bit, a lot) Time points NR	21% none, 48% a little, 23% quite a bit, 8% a lot	Burnout: MBI (norms NR); Marital satisfaction: RDAS; Psychological morbidity: GHQ-12 score ≥4 Time points NR	1) Positive correlation between sleep deprivation and EE, DP, psychological distress, lower marital satisfaction, all p<0.001. No relationship with PA.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift with on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Stress during duty: 4-point scale from 'calm' to 'very demanding' Assessed post-duty	1) Mean stress score during duty: 2.1.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Leitchtfried, 2011 RoB: high	TS	24-h shift; Sleepiness: ESS (range: 0-24); Sleep hours: self-reported (continuous) Sleepiness assessed pre-shit, sleep hours pre, during and post-shift	ESS (mean (range)): 7.4 (4-12); Mean±SD sleep hours: 1) pre-study: 7.74±1.35h; 2) Pre-24-h shift (11h00 on day 1: 0.13±0.35h, 19:00 on day 1: 6.99±0.68h); 3) During the 24-h shift (07h00 on day 2: 0.0±0.0h, 19h00 on day 2, 5.49±1.95h); 4) Post-24-h shift (11h00 on day 3: 0.5±0.71h, 19h00 on day 3: 7.06±1.18h).	aMT6-s: urinalysis Assessed at 4-h intervals from 07:00 to 11:00	1) aMT6-s over shift, mean (95% CI): higher at 11:00AM pre (12.0 (6.3-8.1)) and post-shift (9.3 (3.7-14.9)) vs. during, p=0.016; 2) Correlations between sleep and aMT6-s (data NR): mild for sleep duration the night prior with aMT6-s at 3PM the following day; sleep on night 2 with aMT6-s at 3PM the next day; total sleep with aMT6-s at 11AM on third day; moderate for sleep on first night with aMT6-s at 7AM and 11AM pre-shift, 11PM during 24-h shift and 11AM post-shift; total sleep pre-shift and nocturnal sleep during 24-h shift with aMT6-s at 11PM during shift; total sleep with aMT6-s at 3PM on first and second day, 11PM on second day; 3) Correlations between ESS and aMT6-s: moderate for aMT6-s at 7AM during shift, 11AM on day off.
Beaujouan, 2005 RoB: high	CS	Sleep deprivation: 4-point scale (always, frequently, rarely, never) Time points NR	48.8% always or frequently feel sleep deprived	Substance abuse: 93-item addiction and substance abuse questionnaire Time points NR	1) 60.6% with drug dependence vs. 46.0% of those without reported sleep difficulties, p<0.001. 2) OR (95% CI) of addiction for frequently/always vs. rarely/never sleep deprived: tobacco 1.42 (1.04-1.94); tranquilizer/hypnotics 3.26 (2.12-5.02).
Doppia, 2011 RoB: low	CS	Insufficient sleep: 4-point scale (no, not really, sort of, yes) Time points NR	28.9% reported insufficient sleep during work time	Burnout: CBI (mild: 1-2.4, moderate: 2.5-3.5, severe: 3.6-5) Time points NR	1) Frequency of burnout by response for sleep sufficiency: 42.6% for no/not really, 16.3% for sort of/yes, p<0.001.
Lindfors, 2006 RoB: low	CS	Sleep hours/day: self-reported to the nearest 0.5h; Adequacy of sleep and rest: self-reported (yes/no)	Sleep hours (mean (range)): 7 (5-9)	Stress: MOSQ on a 3-point scale (no, to some extent, clearly); Thoughts of suicide: 4-point scale (‘never’ to ‘have tried’)	1) Sleep sufficiency predicted stress symptoms: bivariate β=-0.362, p<0.001; multivariate including gender, sick leave, suicide β=-0.269, p<0.001; 2) Sleep disturbance associated with thoughts of suicide, p=0.009.

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Exposures or interventions: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Time points NR	
Saadat, 2015 RoB: low	CS	Sleep deprivation (<7h/24-h) due to 17-h overnight shift; Sleepiness and alertness: VAS from 0 (not at all) to 100 (extremely) All assessed on a regular day and a post-call day	Mean±SD sleepiness on a regular day vs. post-call day: 2.99±2.18 vs. 6.79±2.30, p<0.001	Simple cognitive tests: VAS from 0 (not at all) to 100 (extremely); Mood disturbance: PMS (scoring NR) All assessed on a regular day and a post-call day	Regular day vs. post-call day, mean±SD scores: 1) Simple cognitive tests: energetic 6.04±2.27 vs. 2.53±1.87, confident 7.03±1.83 vs. 4.98±2.29, irritable 2.03±1.94 vs. 4.86±2.16, sleepy 2.99±2.18 vs. 6.79±2.30, talkative 4.46±1.74 vs. 2.41±1.97, all p<0.001; jittery 1.84±1.74 vs. 3.12±2.34, p=0.003; anxiousness ns; 2) PMS: tension 13.48±2.71 vs. 15.43±4.46, p=0.049; anger 15.24±4.41 vs. 18.14±5.92, p=0.005; fatigue 10.14±2.63 vs. 20.05±6.87, p<0.001; confusion 10.57±1.69 vs. 12.57±4.24, p=0.025; vigor 24.05±6.75 vs.16.67±5.70, p<0.001; depression: ns; total mood disturbance: 42.57±15.26 vs. 76.90±6.91, p<0.001.
ER or ICU physicians					
Dutheil, 2013 RoB: high	RCT	14-h or 24-h shift; Sleep hours: self-reported sleep and wake time; Sleep quality: VAS from 1 (low) to 100 (high); Mental and physical fatigue: VAS from 1 (low) to 100 (high) Assessed on day prior to shift; during shift; each day of protocol (work, off, clerical, control)	1) Sleep duration and quality lower during shifts (14h and 24h) than any other day, and lower during the 24-h vs. 14-h shift (p<0.05); 2) Mental and physical fatigue higher after 14-h and 24-h shift vs. control day (data NR).	Stress: VAS from 0 (low) to 100 (high); IL-8: urinalysis Assessed at 08:30 and 18:30 on each day of protocol	1) Stress: higher following 14-h and 24-h shifts vs. the control day, p<0.05 (data NR); 2) IL-8: higher following 24-h shift vs. control (p=0.007) and 14-h shift (p=0.015); ns difference between 14-h shift and control day; 3) Correlations with IL-8: sleep hours pre-24-h shift, r=-0.622, p=0.007; poor sleep quality during 14-h and 24-h shifts, r=0.452, p=0.031; 4) Multivariable regression: 24-h shift increased IL-8 by 1.9ng vs. control day, p=0.007; ns association with 14-h shift, mental or physical fatigue, sleep deprivation, 14-h shift.
Sende, 2012 RoB: high	CS	Fatigue and sleep deprivation as sources of stress	NR	Most important sources of stress among 4 categories (work-related, patient-	1) 78% indicated that sleep loss and fatigue were sources of stress.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		related, organizational, individual) Time points NR	
Generalists[b]					
Harbeck, 2015 RoB: unclear	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Biochemical (laboratory values) and physiological (heart rate variability, skin resistance, blood pressure) stress parameters Assessed before a normal day shift, and after a 24-h on call shift	Before a normal shift vs. after overnight call shift: 1) Biochemical parameters: no changes in any parameter except for thyroid stimulating hormone which was higher after the on-call shift (p = 0.049, data NR); 2) Physiological parameters: no significant changes in any parameter.
Pit, 2014 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day' Time points NR	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Early retirement (<65 years) intentions (yes/no) Time points NR	For sleep disturbance a few times a year to every day vs. never: 1) Intention to retire early: 74% vs. 26%, p<0.01; 2) Association with intention to retire early (OR (95% CI)): univariate 3.6 (1.47-8.80), p<0.01; multivariate including work, occupational, individual factors 2.91 (1.11-7.6), p<0.05; 4) RR (95% CI) for intention to retire early: 2.0 (1.18-3.49); attributable fraction: 50.0%; population attributable fraction: 37.1%.
Pit, 2016 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day'	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Sickness presenteeism: 'yes' response indicated 1 or more days	For sleep disturbance a few times a year to every day vs. never: 1) Sickness presenteeism: 32% vs. 68%, p=0.018;

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Assessed for the past 12 months	2) Association with sickness presenteeism (OR (95% CI)): 2.92 (1.19-7.16), p=0.02.
Roberts, 2014 RoB: unclear	CS	Fatigue: LAS from 0 (low) to 10 (high) Assessed for the past week	Mean (SD) score: 5.8 (2.4) for hospitalists; 5.9 (2.4) for general internists	Impact of fatigue on daily activities (falling asleep while driving) (yes/no) Time points NR	1) 8.7% of hospitalists and 4.3% of outpatient general internists had fallen asleep while driving due to fatigue.
Vela-Bueno, 2008 RoB: low	CS	Sleep Quality: PSQI (Spanish): score ≥5 indicates low quality (range; 0 to 21); Insomnia: DSM-IV criteria Time points NR; insomnia symptoms in past month	Prevalence (% (95% CI)): 1) Sleep-onset latency >30 minutes: 8.4 (4.8-11.9); 2) Wake time after sleep onset >30 minutes: 15.4 (10.8-19.9); 3) Early morning awakening: 22.5 (19.5-30.4); 4) Nonrestorative sleep: 22.5 (17.2-27.7); 5) Daytime impairment for ≥5 days in past month: 14.2 (9.7-18.6); 6) Insomnia: 18.8 (13.8-23.7).	Burnout: PBM with a 7-point scale from 1 (never) to 7 (always) Time points NR	Low vs. high burnout, mean±SD: 1) Global PSQI: 3.72±2.22 vs. 7.24±4.17, p<0.001; 2) PSQI subscores: sleep quality: 0.54±0.57 vs. 1.40±0.83, p<0.001; sleep latency: 0.51±0.80 vs. 1.38±1.03, p=0.002; sleep duration: 0.45±0.64 vs. 1.16±0.92, p=0.003; sleep efficiency: 0.21±0.57 vs. 0.77±0.98, p=0.018; sleep disturbance: ns; use of medication: 0.14±0.49 vs. 0.57±0.83, p=0.032; daytime dysfunction: 0.52±0.73 vs. 1.57±0.88, p=0.002. 3) Prevalence (95% CI) of insomnia symptoms: sleep latency: 5.5% (2.5-11.5%) vs. 21.1% (10.5-31.6%), p=0.015; wake time >30 min after sleep onset: 9.4% (1.6-17.1%) vs. 25.5% (14.2-37.7%), p=0.029; early awakening: 14.5% (5.1-23.8%) vs. 45.6 (32.7-58.4%), p<0.001; somewhat/very dissatisfied with sleep: 5.5% (2.5-11.5%) vs. 50% (37.1-62.8%), p<0.001; day impairment: 5.5% (2.5-11.5%) vs. 38.2% (25.6-50.7%), p<0.001; insomnia: 7.3% (0.4-14%) vs. 39.7% (27.1-52.2%), p<0.001.
Oncologists					
Shanafelt, 2005 RoB: unclear	CS	Fatigue: LASA QOL ≤7; Sleep deprivation: 10-point Likert scale from 0 (not at all) to 10 (stressful as can be)	75% had a high level of fatigue; Mean±SD sleep score: 4.5±2.65.	Wellbeing: 10-item LASA QOL, high ≥8 vs. low ≤7 Time points NR	1) Sleep deprivation for high vs. low overall well-being (mean±SD): 3.9±2.57 vs. 5.1±2.60, p=0.0004; 2) Lower fatigue predicted overall wellbeing in a multivariate model including personal and professional characteristics, p=0.002.

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
		Time points NR			
Shanafelt, 2014 RoB: unclear	CS	Fatigue: 10-point LAS (lower scores indicate greater fatigue) Time points NR	Mean±SD fatigue score: 5.7±2.4	Satisfaction with WLB: 5-point Likert scale from 'strongly agree' to 'strongly disagree' Time points NR	1) OR (95%CI) of lower satisfaction predicted by high fatigue (vs. not) in multivariate model including personal and work-related factors, and burnout: 0.489 (0.337-0.710), p<0.001.
Mixed groups of physicians					
Aziz, 2004 RoB: high	CS	Working while fatigued: 5-point scale from 'extreme' to 'a little' Time points NR	NR	Stress: 47-item questionnaire with a 5-point scale from 'extreme' to 'a little' Time points NR	1) Sources of stress: working while fatigued had a mean±SD score of 2.44±1.20, factor loading: 0.653, in factor analysis; 2) Inverse correlation between stress and working while fatigued: r=-0.270 (significance level NR).
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: worried about having a car accident while driving home post-call: 5.4 vs. 7.0, p<0.001; sleep loss has a major impact on personal life: 8.4 vs. 7.0, p=0.01; 3) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Elovaino, 2015 RoB: low	CS	Sleeping problems: Jenkins Scale with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006 and 2010	Mean±SD score: 2006: 2.30 (1.00); 2010: 2.35 (1.05).	Jobs demands: 5 items scored on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree); Job control: 3 items derived from the Karasek Job Questionnaire	There was no association between sleeping problems in 2006 and job demands or control in 2010.

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Exposures or interventions: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Heponiemi, 2014 RoB: low	CS	Sleeping problems: Jenkins Scale[81] with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Psychological distress: GHQ-12 with a 4-point scale (low to high); Job satisfaction: JDS with a Likert scale from 1 (strongly disagree) to 5 (strongly agree) Assessed in 2010	1) Sleeping problems associated with job satisfaction, β=-0.12, p<0.001, psychological distress, β=0.18, p<0.001; 2) Total indirect effect of on-call duty through two mediators (sleeping problems, work interference with family) (R^2 (95% CI)): job satisfaction 0.06 (-0.059, -0.016), p<0.001; psychological distress 0.16 (0.023, 0.081), p<0.001.
Mahmood, 2016 RoB: high	CS	Sleep deprivation: self-reported mean hours of sleep when on call Assessed at 4 years, 10 years, and 15 years post-graduation	Mean±SD hours: 4 years: 4.52 (2.79); 10 years: 5.38 (6.36); 15 years: 6.41 (7.14).	Alcohol use disorders: Modified 9-item version of the Alcohol Use Disorder Identification Test (AUDIT) ≥6 for men and ≥5 for women. Assessed at 4 years, 10 years, and 15 years post-graduation	There was no association between hours of sleep when on call and hazardous drinking behaviours (p=0.732)
Shirom, 2010 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Burnout: SMBM on a 7-point scale from 1 (almost never) to 7 (always)	1) Correlation between physical fatigue subscale and overall burnout: 0.88, p<0.05; 2) In a predictive structural model for burnout, physical fatigue accounted for unique variance in the burnout items not accounted for by total burnout (R^2=0.24).
Smith, 2017 RoB: unclear	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Mental and physical illness: self-reported via open-ended comments Time points NR	Some physicians reported developing mental illness (e.g., bipolar disorder, alcohol misuse) due to tiredness and stress at work; others developed physical health problems due to sleep deprivation, poor eating habits and lack of exercise.

Study Risk of Bias (RoB)	Study design	Exposures or interventions		Outcomes	Associations between exposure and outcome
		Assessment measure and time points	Baseline	Assessment measure and time points	
Starmer, 2016 RoB: low	CS	Sleep deprivation: <7 hours sleep in a typical 24-h period (self-reported) Time points NR	27.7% sleep deprived	Burnout, satisfaction with career and life, balanced personal and professional commitments: Each on a 5-point Likert scale (strongly agree to strongly disagree) Time points NR	≥7-h vs. <7-h sleep: 1) Burnout (% strongly agree/agree): 26.4% vs. 39.6%, p<0.05; career satisfaction (% strongly agree/agree) ns; life satisfaction (% completely/very satisfied): 76.4% vs. 55.9%, p<0.05; balanced personal and professional commitments (% completely/very satisfied): 49.7% vs. 26.1%. 2) <7-h sleep (vs ≥7-h) (OR, 95% CI) associated with life satisfaction 0.44 (0.29-0.67), p<0.05; balanced personal/professional commitments 0.46 (0.31-0.71), p<0.05, in a model including work and personal factors.
Tokuda, 2009 RoB: low	CS	Sleep hours/day: self-reported (continuous) Time points NR (included weekday and weekends)	Mean±SD (range) sleep hours/day: 6±0.9 (3-8)	Burnout: MBI (Japanese) with a 7-point Likert scale: 0 (none) to 6 (every day); Job satisfaction: JHPSS with a 5-point Likert scale: 1 (strongly disagree) to 5 (strongly agree) Time points NR	Maximum likelihood estimates±SE: 1) Sleeping time to job satisfaction: group 0.990±0.458, p=0.031; ns for men; women 1.711±0.805, p=0.034; 2) Sleeping time to EE: group -0.219 ±0.070, p=0.002; men, -0.215±0.082, p=0.009; ns for women.
Wada, 2010 RoB: unclear	CS	Sleep hours/day: Self-reported (continuous) Assessed for past month when not completing overnight work	<5 hours: 8.7% men, 9.9% women; 5 to <6 hours: 32.3% men, 34.6% women; 6 to <7 hours: 46.0% men, 43.7% women; ≥7 hours: 13.0% men, 11.8% women.	Depression: QIDS-SR; Japanese score <5 (no symptoms) to >20 (very severe symptoms) Assessed for past 7 days	1) Sleep hours for those with vs. without depressive symptoms: <5: 18.7% vs. 7.7% men, 20.5% vs. 8.7% women; 5 to <6: 33.7% vs. 32.2% men, 38.6% vs. 34.2% women; 6 to <7: 35.1% vs. 46.9% men; 31.8% vs. 45.1% women; 2) Association between <5h sleep (vs. 6-7h) and depressive symptoms (OR (95% CI)): univariate 2.79 (1.96-3.95) for men, 2.65 (1.47-4.78) for women; multivariate (including age and workload

Study Risk of Bias (RoB)	Study design	Exposures or interventions		Outcomes	Associations between exposure and outcome
		Assessment measure and time points	Baseline	Assessment measure and time points	
					factors) 2.70 (1.82-4.03) for men, 2.38 (1.11-5.10) for women.

[a]Includes studies of anesthetists, where these were physicians.

[b]Includes primary care physicians, internal medicine physicians, and general practitioners.

AM: morning; aMT6-s: melatonin metabolite; BA: before-after; CI: confidence interval; CBI: Copenhagen Burnout Inventory; CS: cross-sectional; DP: depersonalization; DSM: Diagnostic and Statistical Manual of Mental Disorders; EE: emotional exhaustion; ER: emergency; ESS: Epworth Sleepiness Scale; GHQ: General Health Questionnaire; h: hour(s); ICU: intensive care unit; IL-8: interleukin-8; JDS: Job Diagnostic Survey; JHPSS: Japanese Hospital Physicians Satisfaction Scale; LAS: linear analog scale; LASA: linear analog assessment scales; MBI: Maslach Burnout Inventory; MOSQ: Modified Occupational Stress Questionnaire; min: minute(s); NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PA: personal achievement; PBM: Pines Burnout Measure; PE: professional efficacy; PM: afternoon; POMS: Profile of Mood States; PSQI: Pittsburgh Sleep Quality Index; QIDS-SR: Quick Inventory Depressive Scale – Self-Reported; QOL: Quality of Life; RCT: randomized controlled trial; RDAS: Revised Dyadic Adjustment Scale; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SMBM: Shirom-Melamed Burnout Measure; TS: time series; USA: United States of America; VAS: visual analog scale; vs.: versus; WLB: work-life balance

Performance and safety outcomes related to fatigue or sleep loss among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Exposures or intervention: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Uchal, 2005 RoB: unclear	RCT	Sleep deprivation from a 24-h call shift vs. 8-h work; Sleep hours: self-reported (continuous); Sleepiness: ESS (moderate: 10-15, severe: ≥16) Assessed post-call and post-work	Median (range) sleep hours: 1.5 (0-3) post-call vs. 6.5 (5-9) post-work, p<0.05; Median ESS score: 7.0 post-call vs. 5.5 post-work, ns.	Surgical performance: laparoscopic surgical simulator(Minimally Invasivs Surgical Trainer-Virtual Reality) for product quality, procedure effectiveness Assessed post-call and post-work	Post call vs. post-work: 1) Product quality: no difference in accuracy error, tissue damage, leak rate; 2) Procedure effectiveness: no difference in goal-directed actions, non-goal directed actions, operating time.
Chu, 2011 RoB: low	CO	Sleep deprivation: self-reported hours, moderate (3-6h) or severe (<3h) Assessed the night before surgery	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived and 1,595 (39.4%) moderately sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	For 0-3 vs. 3-6 vs. >6 hours of sleep: no difference in CABG or ACC.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and another case in the next 24-h	Of 6,751 procedures, 339 (5%) performed by sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	Sleep deprived vs. non-sleep deprived: no difference in CABG or ACC.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Surgical performance: duration of surgery	Sleep deprived vs. non-sleep deprived: no difference in duration of surgery, even after stratification by type of procedure.
Amirian, 2014 RoB: high	BA	17-h night call shift; Sleep hours during the shift: Wrist-mounted Micro-Mini-Motionlogger; Sleepiness: KSS	Naps pre-call: 11 (37%) napped for median (IQR) 90 (58-128) min; Median (IQR) sleep: 91 (62-123) min on the pre-call night vs. 430 (329-449) on	Surgical performance: LapSimGyn laparoscopic simulation for time, blood loss, instrument path; D2 test of attention and concentration	Pre- vs. post-call: 1) LapSimGyn: no difference in total time, blood loss, instrument path length, instrument angular path; napping did not affect performance;

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Assessed on pre-call and on-call day; sleepiness assessed during shift	the on-call night, p<0.001; Sleep on-call: 12 (40%) slept for median (IQR) 98 (39-135) min; Significant development of sleepiness during shift (p<0.001), plateau score of 7 at 04:00 to 08:00.	Assessed on pre-call and on-call day	2) D2 test: improvement in concentration, p<0.05. No changes in any other parameters; 3) ns difference in laparoscopic simulation time in those who slept during the shift vs. not.
Gerdes, 2008 RoB: high	BA	On-call shift; Fatigue: questionnaire designed by Behrenz & Monga, 1999; Sleep hours: self-reported (continuous) Assessed in 3 sessions pre- and post-call	Fatigue differential from pre- to post-call (range): 1-7 (units unclear); Sleep during call (range): 1-5h	Psychomotor performance: virtual ring transfer task for gesture-level proficiency, hand movement smoothness, tool movement smoothness, elapsed time Assessed in 3 sessions pre- and post-call	1) Pre- to post-call: decrease in all measures of psychomotor proficiency (p<0.05, data NR) except elapsed time; no change in number of psychomotor errors; increase cognitive errors (p<0.05, data NR); 2) Cognitive errors increased exponentially as fatigue ratings increased (R^2=0.9219) and as hours of sleep declined (R^2=0.933).
Shanafelt, 2010 RoB: unclear	CS	Degree of fatigue as a contributor to errors (self-reported) Assessed for the past 3 months	NR	Perceived recent major medical errors (self-reported) Assessed for the past 3 months	1) Prevalence of perceived recent major medical error: 8.9%; 2) Of those reporting an error, 6.9% listed degree of fatigue as the greatest contributing factor.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift, on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Psychomotor performance: reaction time, critical flicker fusion, response measure, peripheral awareness; Concentration ability: scale of 0 (low tiredness) to 100 (maximum tiredness) Assessed pre- and post-duty	Pre- vs. post-duty, mean±SD: 1) Psychometric testing: recognition reaction time (ms): 439.8±50.8 vs. 480.3±58.9; motor reaction time (ms): 252.8±39.3 vs. 465.4±65.0; total reaction time (ms): 690.8±73.4 vs. 746.5±113.7; critical flicker fusion (Hz): 29.0±2.3 vs. 28.7±3.7; response measure (pixels): 647.8±126.7 vs. 598.3±138.1,

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					peripheral awareness task recognition time: 58.9±59.2 vs. 5[illegible].6±47.5; 2) Concentration ability: 26.4±23.5 vs. 56.3±23.0, p=0.007.
Chang, 2013 RoB: unclear	CS	15-h in-house overnight call; Sleepiness pre-call: ESS ≥9; Sleep hours: self-reported (continuous) Sleepiness assessed pre-call, sleep hours during call	Median (IQR) ESS: 9 (9), 64% scored ≥9; Median (IQR) hours slept during shift: 1 (0-3).	Psychomotor performance: reaction time; CCPT II; N-back; HVLT (3 trials of 12 words) Assessed at baseline and pre- and post-call	1) Afternoon baseline vs. pre-call: no difference in reaction time, CCPT, N-back, of HVLT; Morning baseline vs. post-call: 1) No change in auditory or visual reaction time; 2) CCPT (t-scores): No change in detectability, response style, hit reaction time, omissions, commissions; 3) N-back % accuracy: no change for auditory, visual, or mean N-value; 4) HVLT (t-score): mean for trials 1-3: 48.6±7.6 vs. 41.5±9.9 (p=0.04); delayed recall: ns; 5) No correlation between ESS scores pre-call or sleep during shift and any measure of psychomotor performance.
Gander, 2000 RoB: low	CS	Nights of work-related sleep disturbance: self-reported (continuous) Assessed for the past 6 months	NR	Risk of fatigue-related errors: questionnaire modelled after Gravenstein et al., 1990 Assessed for the past 6 months	1) Risk of fatigue-related errors increased with increasing nights of work-related sleep disturbance: RR: 1.25, 95% CI: 1.06-1.49.
Saadat, 2017 RoB: low	CS	Sleep deprivation due to an overnight call shift	NR	Reaction time: PVT Assessed after an overnight call shift and the morning of a regular (non-call) day	Mean (SD) reaction time was slower post-call (297.76 (83.75)) vs. on a regular day (266.58 (38.35)), p=0.0[illegible]7.

Study Risk of Bias (RoB)	Study design	Exposures or intervention		Outcomes	Associations between exposure and outcome
		Assessment measure and time points	**Baseline**	**Assessment measure and time points**	
Gander, 2008 RoB: unclear	NC	Sleep loss across consecutive working days or on-call work: Wrist-mounted Actiwatch (Mini Mitter, Bend, Oregon, US), sleep and duty diary Assessed over a 2-week period including a weekend of rostered shifts or on-call	≥2 hours sleep <baseline: 8% of 24-h periods that included day work vs. 14% that included day + call; Sleep hours: mean 0.6h less sleep when working day shifts (p=0.014) and 0.8h less sleep when working day shifts + call (p=0.013) vs. off.	Psychomotor performance: PVT Assessed within 2 hours pre- and post-call	1) In fixed model analysis for reaction time including sleep, time since waking, work hours: acute sleep loss associated with slower median reaction time, $F_{(1,184)}$=5.70, p<0.05; longer time since waking associated with poorer performance on the slowest 10%, $F_{(1,185)}$=5.13, p<0.05; 2) Reaction time across 12 consecutive work days: no change in pre-duty reaction times but post-duty reaction times slowed linearly, median -3.38, p<0.001; decline in performance across 10 minutes became progressively steeper both pre- and post-duty, p=0.020.
ER or ICU physicians					
Sanches, 2015 RoB: high	CS	Acute sleep deprivation (<5h of night sleep after a night shift of 12h) Sleep hours: 7-day Actigraphy via SenseWear® Pro2 Armband; Sleepiness: ESS; Sleep quality: PSQI Assessed the week and night before the psychomotor tests	Non-sleep deprived vs. sleep deprived: PSQI >5: 0% vs. 33%, ns; ESS≥10: 11% vs. 67% Sleep time (mean±SD) in week before tests: duration and number of naps higher in sleep deprived group, but diurnal sleep hours lower, 428.6±30.1 vs. 375.8±55.9, p=0.038; Sleep quality (mean±SD): week before tests: 3.3±0.7 vs. 2.6±0.3, p=0.013; night before tests: 3.1±0.8 vs. 1.9±1.0, p=0.020.	Psychomotor performance via Battery Test Reaction 5 (v1): StimulTest, InstrucTest, MovemTest; TP test of visual attention Assessed on morning after night shift 8	Sleep deprived group vs. non-sleep deprived, mean±SD: 1) InstrucTest: correct answers: 169.4 (16.0) vs. 148.3 (28.[illegible]), p=0.070; wrong answers: ns; perfection index (%): 99.6 (0.3) vs. 98.9 (1.3), p=0.021; response latency (sec/click): ns; 2) StimulTest: correct answers: 170.7 (21.9) vs. 145.1 (17.[illegible]), p=0.022; wrong answers: ns; perfection index (%): ns; response latency (sec/click): 1.06 (0.1) vs. 1.24 (0.1), p=0.022; 3) MovemTest: ns for any parameter; 4) TP: omitted symbols: 34.2±18.4 vs. 62.7±44.0, p=0.034; concentration index (%): 14.1±8.9 vs. 30.0±25.9, p=0.019; quality index (%): 13.8±8.6 vs. 29.2±26.4, p=0.031; correct/wrong symbols: ns; Correlations between sleep and tests: 1) TP for sleep hours nights 1-6: omitted symbols: r=-0.686, p=0.011 for non-sleep-

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					deprived, ns for sleep-deprived; concentration index (%): r=-0.359, p=0.037 for sleep-deprived, ns for non-sleep deprived; r=-0.359, p=0.037 for the group; no other significant correlations. 2) No correlation between PSQI, ESS and any of the psychomotor tests.
Generalists[b]					
Harbeck, 2015	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Neurocognitive parameters: computerized attentional test (vigilance, alertness); D2 letter cancellation test (divided attention); Trail Making Test (visual attention, task switching); Digit Span, Digit Symbol Substitution Test, Weschler Memory Scale (memory functions) Assessed before a normal day shift, and after a 24-h on call shift	Intrinsic alertness, focused attention and vigilance were similar on both occasions; Phasic alertness improved following the on-call shift: mean (SD) 24.8 (15.6) vs. 38.3 (21.5), p = 0.022.
Mixed specialties or undefined populations					
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: written an incorrect order: 8.8 vs. 7.3, p=0.02; might fall asleep while examining a patient: 13.2 vs. 7.7, p=0.001; look forward to sleeping at grand rounds: 10.4 vs. 7.4, p=0.002;

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					3) No difference in ESS score for those who agree/strongly agree vs. other response: work is unaffected by sleep loss and fatigue, thinking is unaffected by sleep loss, sleep loss and fatigue affect my medical decisions, have heard of others making medical errors due to sleep loss and fatigue, never make errors in prescriptions on post-call days, have made medical errors because of sleep loss and fatigue; 4) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Heponiemi, 2014 RoB: low	CS	Sleeping problems: 4-item Jenkins Scale on 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Work ability: Work Ability Index on scale from 1 (could not work at all) to 10 (best work ability) Assessed in 2010	1) On-call duty had an indirect effect on work ability (R^2=0.11, 95% CI: -0.122, -0.031, p<0.001) through two mediators (work interference with family, sleeping problems); 2) Sleeping problems inversely associated with work ability, β=-0.29, p<0.001.
Kanieta, 2011 RoB: unclear	CS	Sleep hours: self-reported (continuous) Sleepiness and sleep difficulties: 5-point scale from 1 (never) to 5 (always); Insomnia: ≥3 sleep difficulties Assessed for the past month	Insufficient rest: 32.5%; Daytime sleepiness: 3.5%; Insomnia: 20.0%; Sleep time (mean±SD min): 279.8±60.9	Self-reported medical incidents: 4-point scale from 1 (never) to 4 (often) Assessed for the past month	1) Prevalence of medical incidents (% (95% CI)): sleep deprived (26.8% (24.2, 29.4)) vs. not (15.2% (13.7, 16.7)), p<0.01; insomnia (24.8% (21.6, 28.0)) insomnia vs. not (17.6% (16.2, 19.0)), p<0.01; ≥6h sleep (18.3% (16.8, 19.8)) vs. <6h (21.7% (18.8, 24.6)), p=0.03; 2) Predictors of medical incidents in multivariate model including personal and work-related factors (OR (95% CI)): lacking rest due to sleep deprivation vs. not (1.65 (1.33-2.04)), p<0.01); insomnia vs. not (1.45 (1.16-1.82), p<0.01); ns for sleep hours.

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Sexton, 2001 RoB: high	CS	Fatigue as a factor impacting performance Time points NR	NR	Performance effectiveness measured by 1 question: agree, neutral, disagree Time points NR	1) "When fatigued, I perform effectively during critical phases of operations/patient care": Anesthetic: 47% agree; 15% neutral; 38% disagree; Surgical: 70% agree; 12% neutral; 18% disagree.
Shirom, 2006 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Quality of care: Adapted 15-item SERVQUAL with a 5-point Likert scale from 1 (very small extent) to 5 (very large extent) Time points NR	1) Quality of care positively predicted by fatigue in a model incorporating several other components of burnout, β=0.17, p<0.05.
Smith, 2017 RoB: moderate	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Perceived competence: self-reported via open-ended comments Time points NR	Some physicians indicated that continual tiredness and exhaustion led to concerns that it would affect their competence; some felt that professional performance was compromised at times of physical and mental fatigue.
Tanti, 2017 RoB: high	CS	Fatique: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	NR	Prescribing errors: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	Perception of the contribution of fatigue to prescribing errors differed by physician type (p<0.05): 54% of community doctors, 96% hospital doctors, 8% of office-working doctors perceived a very high or high association between fatigue and prescribing errors.

[a]Includes studies of anesthetists, where these were physicians.
[b]Includes primary care physicians, internal medicine physicians, and general practitioners.
ACC: aortic cross-clamp time; BA: before-after; CABG: cardiopulmonary bypass time; CCPT II: Connor's Continuous Performance Test II; CI: confidence interval; CO: cohort; CS: cross-sectional; ER: emergency; ESS: Epworth Sleepiness Scale; h: hour(s); HVLT: Hopkin's Verbal Learning Task; Hz: Hertz; ICU: intensive care unit; IQR: interquartile range; KSS: Karolinska Sleep Scale; min: minutes; ms: millisecond(s); N-back: Dual N-back test; NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; PVT: Psychomotor vigilance Performance Task; RR: risk ratio; RCT: randomized controlled trial; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SERVQUAL: Service Quality Measure; SMBM: Shirom-Melamed Burnout Measure; TP: Toulouse-Piéron test; TS: time series; US: United States of America; vs.: versus

Patient outcomes related to fatigue or sleep restriction among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Surgeons					
Chu, 2011 RoB: low	CO	Sleep deprivation: moderate (3-6 h) or severe (<3-h) sleep deprivation the night before surgery (self-reported hours)	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived, 1,595 (39.4%) by moderately sleep-deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) 0-3 vs. 3-6 vs. >6 hours of sleep: No difference in incidence of mortality, incidence of 10 major complications (except septicemia, 3.6% vs. 0.9% vs. 0.8%, p=0.03), ICU length of stay; in-hospital length of stay (days): 7.0 vs. 6.0 vs. 7.0, p<0.001.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and performed a subsequent case in the next 24-h	Of 6,751 procedures, 339 (5%) were performed by sleep deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, need for blood products, complications (operative, neurologic, renal, infectious, pulmonary), in-hospital length of stay.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Chart review: mortality, surgical complications, readmission, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, surgical complications, readmissions within 30 days, or length of stay.
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse surgical complications Assessed during and post-surgery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total number of complications, preventable complications, type of complications; 2) Operating room procedures with complications, OR (95% CI): 8.5% for 0-6h sleep vs. 1.1% for >6h sleep, 2.70 (1.13-6.48), p=0.03; 3) All procedures with complications, OR (95% CI): 6.2% for 0-6h sleep vs. 3.4% for >6h sleep, 1.72 (1.02-2.89), p=0.04.

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Schieman, 2007 RoB: low	CO	Fatigue: surgeon billed for clinical work after 22:00 the night before surgery	Of 270 procedures, 22 (8%) were performed by fatigued surgeons	Chart review: surgical complications, length of stay, mortality, cancer recurrence Assessed during and post-surgery	1) Fatigued vs. non-fatigued surgeons: no difference in intra- or post-operative complication rate, length of stay, in-hospital length of stay, cancer recurrence.
Vinden, 2014 RoB: low	CO	Sleep deprivation (at risk): surgeon worked 00:00 to 07:00 and performed surgery 07:00 to 18:00	Of 94,183 surgeries, 2,078 (2.2%) were performed by surgeons who were 'at risk'	Chart review: conversion to open procedure (from laparoscopic), iatrogenic injuries, mortality Assessed during and post-surgery	1) At risk vs. not at risk surgeon: no difference in incidence of conversion to open procedure, iatrogenic injuries, mortality, in either univariate or multivariate analyses.
Obstetricians					
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse obstetric complications Assessed during and post-delivery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total complications, preventable complications, type of complications; 2) No association between sleep deprivation and proportion of procedures with complications, nor difference for 0-6h vs. >6h of sleep opportunity.

CI: confidence interval; CO: cohort; h: hours; NR: not reported; OR: odds ratio; RoB: Risk of Bias; SD: standard deviation; US: United States of America; vs.: versus

Supplementary file 5. Statistical analyses

Dichotomous outcomes

Outcome or subgroup	Number of studies	Number of participants	Pooled risk ratio (95% CI)	Heterogeneity	
				P	I^2
1.1 Patient mortality	5	60,436	0.98 (0.84, 1.15)	0.73	0%
1.2 Intra-operative complications	3	19,798	suppressed	0.007	82%
1.2.1 Surgical procedure	3[a]	14,896	suppressed	<0.001	88%
1.2.2 Obstetric procedure	1[a]	4,902	suppressed	NA	NA
1.3 Post-operative complications	5	60,201	0.99 (0.95, 1.03)	0.45	0%

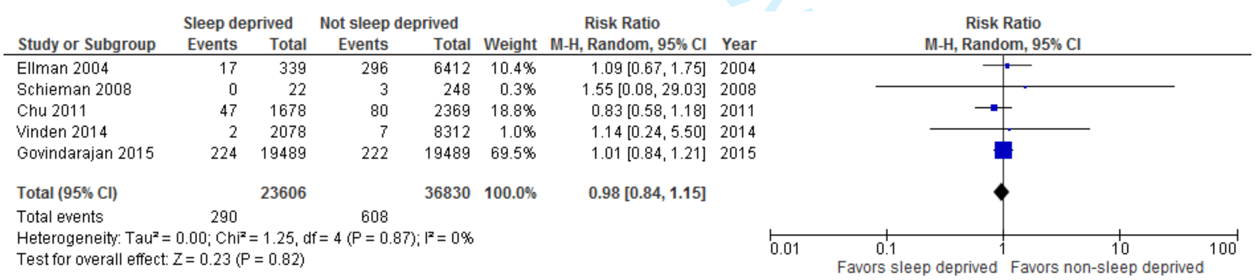
[a] Rothschild, 2009 is represented in both analyses

Continuous outcomes

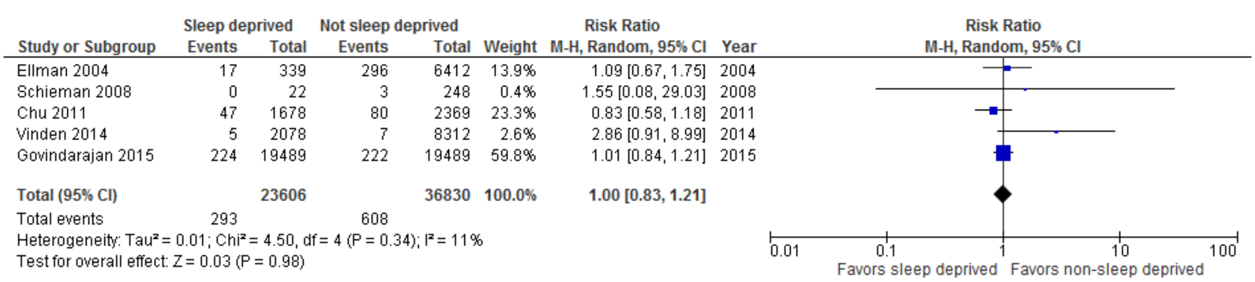
Outcome or subgroup	Number of studies	Number of participants	Pooled mean difference (95% CI)	Heterogeneity	
				P	I^2
1.4 Operating time (minutes)	4	50,046	-0.14 (-1.60, 1.33)	0.70	0%
1.5 Length of hospital stay (days)	4	50,046	suppressed	<0.001	86%
1.5.1 Cardiac surgeries	2	10,798	suppressed	0.01	84%
1.5.2 Elective surgeries	1	38,978	suppressed	NA	NA
1.5.3 Anterior resection for anal cancer	1	270	suppressed	NA	NA

CPBT: cardiopulmonary bypass time; NA: not applicable

1.1 Patient mortality



Sensitivity analysis using highest possible number of events for Vinden 2014

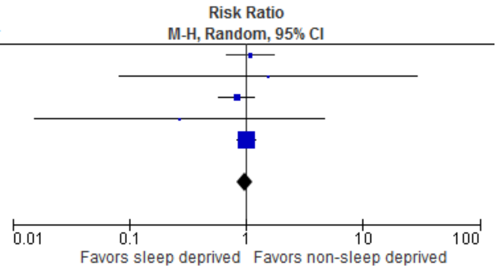


Sensitivity analysis using lowest possible number of events for Vinden 2014

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Ellman 2004	17	339	296	6412	10.5%	1.09 [0.67, 1.75]	2004
Schieman 2008	0	22	3	248	0.3%	1.55 [0.08, 29.03]	2008
Chu 2011	47	1678	80	2369	18.9%	0.83 [0.58, 1.18]	2011
Vinden 2014	0	2078	7	8312	0.3%	0.27 [0.02, 4.67]	2014
Govindarajan 2015	224	19489	222	19489	70.0%	1.01 [0.84, 1.21]	2015
Total (95% CI)		**23606**		**36830**	**100.0%**	**0.98 [0.84, 1.14]**	
Total events	288		608				

Heterogeneity: Tau² = 0.00; Chi² = 2.01, df = 4 (P = 0.73); I² = 0%

Test for overall effect: Z = 0.29 (P = 0.77)

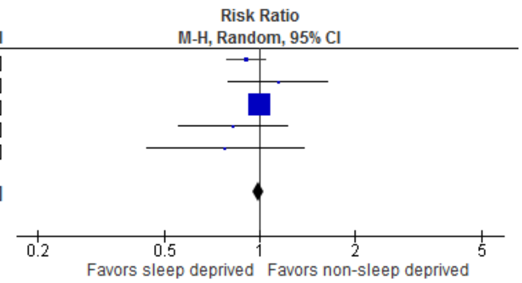


1.3 Post-operative complications

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI
Chu 2011	259	1678	404	2369	7.8%	0.91 [0.78, 1.04]
Ellman 2004	29	339	480	6412	1.2%	1.14 [0.80, 1.64]
Govindarajan 2015	3527	19489	3543	19489	89.4%	1.00 [0.95, 1.04]
Schieman 2008	12	22	164	248	1.0%	0.82 [0.56, 1.22]
Vinden 2014	14	2031	72	8124	0.5%	0.78 [0.44, 1.38]
Total (95% CI)		**23559**		**36642**	**100.0%**	**0.99 [0.95, 1.03]**
Total events	3841		4663			

Heterogeneity: Tau² = 0.00; Chi² = 3.69, df = 4 (P = 0.45); I² = 0%

Test for overall effect: Z = 0.66 (P = 0.51)

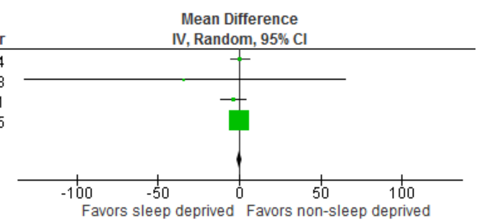


1.4 Operating time (minutes)

Study or Subgroup	Sleep deprived Mean	Sleep deprived SD	Sleep deprived Total	Not sleep deprived Mean	Not sleep deprived SD	Not sleep deprived Total	Weight	Mean Difference IV, Random, 95% CI	Year
Ellman 2004	107.7	55.24	339	107.4	56.05	6412	5.9%	0.30 [-5.74, 6.34]	2004
Schieman 2008	114	226.6	22	148	226.6	248	0.0%	-34.00 [-132.80, 64.80]	2008
Chu 2011	110.2	119.13	1678	114	120.88	2369	3.8%	-3.80 [-11.30, 3.70]	2011
Govindarajan 2015	156	75.56	19489	156	80	19489	90.2%	0.00 [-1.54, 1.54]	2015
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.14 [-1.60, 1.33]**	

Heterogeneity: Tau² = 0.00; Chi² = 1.42, df = 3 (P = 0.70); I² = 0%

Test for overall effect: Z = 0.18 (P = 0.86)



Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8

Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-11, Table 1, Supplementary file 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	11, Supplementary file 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	p. 12-18; Supplementary file 4; figures 2-4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	p. 12-18, figures 2-4
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Supplementary file 5
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-19
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	19-20
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	20
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	21

Appendix 1. PRISMA checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**. Page 2 of 2

BMJ Open

The impact of fatigue and insufficient sleep on physician and patient outcomes: A systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021967.R3
Article Type:	Research
Date Submitted by the Author:	07-Aug-2018
Complete List of Authors:	Gates, Michelle; University of Alberta, Pediatrics Wingert, Aireen; University of Alberta, Pediatrics Featherstone, Robin; University of Alberta, Pediatrics Samuels, Charles; Centre for Sleep and Human Performance Simon, Christopher; Canadian Medical Association Dyson, Michele; University of Alberta, Pediatrics
Primary Subject Heading:	Occupational and environmental medicine
Secondary Subject Heading:	Evidence based practice, Health services research, Medical education and training, Occupational and environmental medicine
Keywords:	Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, SLEEP MEDICINE

The impact of fatigue and insufficient sleep on physician and patient outcomes: A systematic review

Michelle Gates, Aireen Wingert, Robin Featherstone, Charles Samuels, Christopher Simon, Michele P Dyson

Michelle Gates, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486C, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Aireen Wingert, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-490B, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Project coordinator

Robin Featherstone, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-486D, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Research librarian

Charles Samuels, Centre for Sleep and Human Performance, Suite 106, 51 Sunpark Drive SE, Calgary, Alberta, Canada, T2X 3V4
Physician and medical director

Christopher Simon, Canadian Medical Association, 1867 Alta Vista Drive, Ottawa, Ontario, Canada, K1G 5W8
Advisor, physician health

Michele P Dyson, Alberta Research Centre for Health Evidence (ARCHE), Edmonton Clinic Health Academy 4-474, University of Alberta, 11405-87 Avenue NW, Edmonton, Alberta, Canada T6G 1C9
Assistant professor

Correspondence to: MP Dyson mdyson@ualberta.ca

Word count (main text): 5,929

ABSTRACT

Objectives: For physicians in independent practice, we synthesized evidence on the (a) impacts of insufficient sleep and fatigue on health and performance, and patient safety; (b) effectiveness of interventions targeting insufficient sleep and fatigue.

Design: We systematically reviewed online literature. After piloting, one reviewer selected studies by title and abstract; full texts were then reviewed in duplicate. One reviewer extracted data; another verified a random 10% sample. Two reviewers assessed risk of bias. We pooled findings via meta-analysis when appropriate, or narratively.

Data sources: We searched Medline, Embase, PsycINFO, CINAHL and PubMed for published studies in April 2016; Medline was updated in November 2017. We searched Embase for conference proceedings, and hand-searched meeting abstracts, association and foundation websites.

Eligibility criteria for selecting studies: English or French language primary research studies published from 2000-2017 examining the effect of fatigue or sleep-related exposures or interventions on any outcome among physicians in independent practice and their patients.

Results: Of 16,154 records identified, we included 47 quantitative studies of variable quality. 28 studies showed associations between fatigue or insufficient sleep and physician health and well-being outcomes. 21 studies showed no association with surgical performance, and mixed findings for psychomotor performance, work performance, and medical errors. We pooled data from six cohort studies for patient outcomes. For sleep deprived versus non-sleep deprived surgeons, we found no difference in patient mortality (n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, I^2 = 0% (P = 0.73)) nor postoperative complications (n = 60,201, RR 0.99, 95% CI 0.95 to 1.03, I^2 = 0% (P = 0.45)). The findings for intraoperative complications and length of stay were considerably heterogeneous.

Conclusions: Fatigue and insufficient sleep may be associated with negative physician health outcomes. Current evidence is inadequate to inform practice recommendations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The review was informed by the methods outlined by Cochrane and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.
- The review was limited by the quality of the included studies, which was often poor. Confidence in our conclusions may be weakened due to multiple comparisons.
- We have focused on evidence from high income countries; our findings may not be generalizable to other settings.

BACKGROUND

The working hours of physicians have been a topic of debate for many years.[1] Beginning in the late 1980s, evidence indicating that medical resident fatigue could negatively impact their cognitive functioning and performance, resulting in an increased risk of medical error, began to accumulate.[2] In response, by the early 2000s physicians' regulatory bodies worldwide began to take action toward restricting the work hours of medical residents and ensuring adequate time for recovery between shifts.[3-5] Since their implementation in the United States by the Accreditation Council for Graduate Medical Education (ACGME), the impact of work hour regulations has been widely researched. Still, evidence for impacts on patient care, resident training and wellbeing remains equivocal.[6-9] This is likely because work hours are only one of many contributors to fatigue and physician wellbeing. In fact, the ACGME has recently reversed the 2011 changes that limited resident work hours to 16 hours per shift and the requirement for 8 hours of time off between shifts. This decision was made in favour of promoting "flexibility" for residency training program work hours and scheduling.

The focus on medical trainees has left physicians in independent practice as a relatively neglected group in research and policy. In Canada, there is no concrete regulation on the hours or patterns in which physicians choose to work.[10] In the absence of clear policies, physicians trained under traditional systems may find it difficult to work shorter hours or take more frequent breaks.[1] Indeed, more than 40% of practicing physicians in the United States work in excess of 80 hours per week.[11] While long work hours remain a cultural norm in medicine, in comparable high-risk industries (e.g., aviation), work patterns and work hours are tightly regulated.[12] The need for similar evidence-based policies in medicine has become a topic of increased interest. Exemplar of this, an evidence-based guideline for fatigue risk management in emergency medical services,[13] informed by a comprehensive set of systematic reviews, has recently been published. For physicians, it has been argued that there is a need to adapt healthcare systems and provide support in identifying the signs of fatigue and mitigating its risks.[1]

Besides potentially affecting patient outcomes, fatigue can impact the health and wellbeing of physicians themselves. Burnout, just one outcome related to fatigue, has been described as epidemic among physicians[14-16] and ultimately affects recruitment and retention of physicians both in community and acute care settings. While the effect of physician wellbeing on the sustainability of healthcare systems has recently received increased attention,[17] evidence-based solutions to burnout

remain relatively elusive.[18] What is clear, is that comprehensive organisational-level efforts are necessary to fully address the issue.[19] Research addressing the factors that influence burnout and overall physician wellness is needed to inform system- and individual-level strategies.[20, 21] To date, evidence of the effects of fatigue and the role of chronic insufficient sleep on physicians in independent practice has not been synthesized, making it unclear what gaps in knowledge remain unaddressed.

Given this void, we undertook a systematic review focusing broadly on primary research relevant to the Canadian context as a fundamental starting point to examine the effects of fatigue and chronic insufficient sleep on physicians in independent practice, and on interventions to combat these effects. Our review was guided by the following research questions: Among physicians in independent practice, (1) what are the impacts of fatigue and chronic insufficient sleep on physician health, physician performance, and patient safety; and (2) what is the effectiveness of interventions that target fatigue and chronic insufficient sleep, in terms of improving physician and patient outcomes?

METHODS

Review conduct

The conduct of this systematic review was guided by Cochrane standards.[22] The research team convened to plan the key research questions and methodology but did not register a formal protocol. The findings are reported in adherence with the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement.[23] Ethical approval was not required for this study.

Patient involvement

Patients were not involved.

Literature search

An information specialist developed a search strategy that included concepts related to physicians, fatigue and sleep. On 13 April 2016 we searched the following online databases with coverage in the biomedical sciences and psychology: Medline, Embase, PsycINFO, CINAHL and PubMed, limited to English and French language articles published from 2000 to 2016. We updated the Medline search in November 2017, as this database offered the highest precision. Though fatigue among physicians is not a new phenomenon,[2] we limited our search to articles published post-2000 to include studies relevant to current physician practice. Work hour limitations have existed in European countries since 1993, but

implementation in the United States (2003)[5] and Canada (2013) for residents is more recent.[24] We aimed to include studies published in this era of increased awareness about the potential impacts of long work hours. To locate unpublished studies, we searched Embase for conference proceedings since 2000 and hand-searched meeting abstracts of the Canadian Conference on Physician Health and the International Conference on Physician Health (2012 to 2016). We also searched the following association and foundation websites: American Medical Association, Australian Medical Association, British Medical Association, Canadian Medical Association, European Medical Association, National Sleep Foundation, Ontario Medical Association and the World Medical Association. The complete search strategy undertaken is reported in Supplementary file 1.

Inclusion criteria

Primary studies (quantitative or qualitative) of fatigue- or sleep-related exposures or interventions among physicians in independent practice were eligible for inclusion. We included physicians practicing in any medical specialty and in any healthcare setting within a high income country,[25] to identify practices comparable to the Canadian setting. Studies including physicians-in-training were included only if data for physicians in independent practice could be isolated. Exposures of interest included fatigue, insufficient sleep, or sleepiness. We also included studies of any intervention that aimed to reduce fatigue or sleep loss with any comparator (or no comparator). All reported outcomes, measured at any time, were eligible for inclusion.

We excluded commentaries, letters, editorials and dissertations. Systematic reviews, health technology assessments, economic evaluations and practice guidelines were excluded, although the reference lists of these as well as the included studies were scanned for potential primary studies. Studies that focused solely on physicians-in-training (e.g., trainees, residents, fellows, interns, medical students, junior doctors, registrars) were ineligible. To maintain the focused scope of the review, we excluded work hours, work load, and any other exposure or intervention that was indirectly related to fatigue or sleep.

Study selection

The study team piloted the selection criteria, which were then applied by two independent reviewers following a two-phase process. We first screened titles and abstracts for potential relevance. Then, we retrieved all records classified as “include” or “unsure” and reviewed their full text for eligibility. Any

disagreements between reviewers were resolved by discussion or third-reviewer consultation when necessary.

Data extraction

Reviewers used a standardized form to extract data in Microsoft Office Excel (*v.* 2016, Microsoft Corporation, Redmond, WA). One reviewer independently extracted data from each included study and a second reviewer verified a random 10% sample. Since no major errors or omissions were noted, we did not undertake further verification.

We extracted the following data: country of publication; funding source; study design; inclusion and exclusion criteria; population characteristics (i.e., sample size, age and gender distribution, physician specialty); setting (i.e., physician workplace, urban or rural); exposure or intervention; definition of fatigue or insufficient sleep; sleep and fatigue scales used and timing of measurement; comparators (if applicable); and outcomes.

Risk of bias appraisal

Two reviewers independently assessed the risk of bias in each included study using standard tools. Disagreements were resolved via discussion or by consulting a third reviewer. We used the Cochrane Risk of Bias tool[22] to assess randomised controlled trials. Adapted versions of the tool developed by the Effective Practice and Organization of Care group[26] were used to assess before-after and time series studies. We used the Newcastle-Ottawa Quality Assessment Scale[27] to appraise cohort studies. We adapted the scale to assess cross-sectional studies and the one non-comparative study.

Evidence synthesis

We considered clinical and methodological heterogeneity in our decision on whether to proceed with meta-analysis for the outcomes identified. For most outcomes, we found high levels of heterogeneity in study design, populations, exposures or interventions, and outcome measures and chose not pool the data via meta-analysis. Thus, we have presented the findings for most outcomes narratively and in summary tables.

When statistical pooling was appropriate, this was undertaken using Review Manager (RevMan v.5.3, Copenhagen: The Nordic Cochrane Centre, the Cochrane Collaboration, 2014) via pairwise meta-analysis

using the DerSimonian and Laird random effects model (given expected heterogeneity).[28] We pooled dichotomous outcomes using the relative risk (95% confidence interval (CI)) and continuous outcomes using the mean difference (95% CI) since the units across studies were consistent (i.e., minutes). When meta-analysis was conducted, we assessed statistical heterogeneity using the chi-square test (using P = 0.10 as the threshold for significance), and quantified the extent of heterogeneity using the I^2 statistic.[29] We considered an I^2 value of 0% to 40% to be low (potentially unimportant), 30% to 60% to be moderate, 50% to 90% to be substantial, and 75% to 100% to be considerable heterogeneity.[22] Subgroup and sensitivity analyses were conducted when appropriate to explore heterogeneity. We intended to assess small study bias visually by inspecting funnel plots and statistically using Egger's regression test, but did not due to the small number (i.e., less than 8) of studies included in the meta-analyses.[30]

When data were not presented in the format required for meta-analysis, we estimated means or standard deviations (SDs) using standard equations. We used the median instead of the mean for one study[31] for the outcomes of length of stay and operating time. Additionally, for one study[32] in the length of stay analysis where the SD could not be estimated, we substituted the mean variance of other studies within the meta-analysis.[33]

RESULTS

We identified 16,083 unique records via the database searches, 56 grey literature sources, and 14 additional records in reference lists of systematic reviews. We excluded 15,016 citations by title and abstract, and another 1,090 by full text. Forty-seven studies[31, 32, 34-78] were eligible for inclusion, and 6[31, 32, 41, 58, 63, 77] were included in meta-analysis for the outcomes of operating time, intra- and post-operative complications, patient mortality and length of hospital stay. Figure 1 shows the flow of studies through the selection process.

Included study characteristics

A summary of the study characteristics is provided in Table 1. Supplementary file 2 presents descriptive information for each included study. There were 45 observational studies[31, 32, 34-39, 41-74, 76-78] and two intervention studies.[40, 75] All studies were quantitative. Nearly half (n = 20/47, 43%) of the studies took place in North America,[31, 32, 35, 37, 38, 41, 45, 48, 57-60, 62, 63, 65-68, 72, 77] and slightly more than one-third (n = 16/47, 34%) in Europe.[34, 36, 39, 40, 42, 46, 47, 50-53, 61, 64, 73, 75, 76]

Table 1. Summary characteristics of the included studies

Study characteristics	**n**	**%**	**Physician characteristics**	**n**	**%**	**Exposures, interventions and outcomes**	**n**	**%**
Study design			**Gender**			**Exposures (observational)**[a]	**45**	**96**
Cross-sectional	34	72	Reported[b]	38	81	Fatigue-related	15	32
Cohort	6	13	>50% male	30	79	Sleep-related	37	79
Before-after	3	6	**Age**			Overnight or extended shifts	18	38
RCT	2	4	Reported[b]	38	81	**Interventions (experimental)**	**2**	**4**
Time series	1	2	Range (years)	20 to >70		**Outcomes**		
Non-comparative	1	2	**Specialty area**[c]			Physician health and wellbeing	28	60
Region and country			Surgeons	13	28	Work and life satisfaction	9	19
North America	20	43	Anesthesiologists	10	21	Burnout	7	15
US	15	32	Generalists	7	15	Stress	8	17
Canada	4	9	ED or ICU physicians	3	6	Mental health and wellbeing	7	15
Canada, US & Mexico	1	2	Oncologists	2	4	Other health-related outcomes	5	11
Europe	16	34	Obstetrician-gynecologists	1	2	Physician performance, risk of error	21	45
France	4	9	Mixed groups	14	30	Psychomotor performance	7	15
Finland	3	6	**Work setting**[d]			Work ability and quality of care	5	11
Spain	2	4	Hospitals	37	79	Incidence of medical errors	5	11
Austria	2	4	Private practice	13	28	Surgical efficiency, effectiveness	6	13
Norway	2	4	Primary care centres, outpatient clinics	7	15	Patient outcomes	6	13
Denmark	1	2	Academic practice, training programs	5	11			
Germany	1	2	Other (e.g., industry, military)	11	23			
Malta	1	2	Not reported	3	6			
Japan	4	9	**Urban or rural**					
Australia	2	4	Reported[b]	16	34			
Israel	2	4	Urban	12	75			
New Zealand	2	4	Rural	2	13			
United Kingdom	1	2	Mixed	2	13			

ED: emergency department; ICU: intensive care unit; RCT: randomised controlled trial; US: United States of America

[a]Exposures that have been directly related to an outcome. Some studies included multiple exposures.

[b]Percentages presented using the total number of studies where the outcome was reported as the denominator.
[c]Anesthesiologists include physician anesthetists; generalists include primary care physicians, internists, and general practitioners; mixed groups refers to studies including more than one physician group or specialty (usually large-scale surveys). In some studies, multiple distinct groups were represented.
[d]As defined by the authors. Values for the settings will exceed 100% because studies may occur in more than one setting.

The 47 studies reported outcomes for 36,190 (range = 6 to 7,905) physicians and 69,809 (range = 270 to 38,978) adult patients. About half reported on surgeons (n = 13/47, 28%),[31, 32, 34, 41, 45, 48, 54, 58, 62, 63, 66, 75, 77] or anesthesiologists/physician anesthetists (n = 10/47, 21%).[36, 37, 39, 43, 44, 50-52, 59, 60] Where it was reported, the samples tended to be predominantly male (n = 30/38, 79%) and physician age varied widely. Hospitals were the most common setting (n = 37/47, 79%).[31, 32, 34-37, 39-43, 45-47, 49-52, 54, 57-61, 63-70, 73-75, 77, 78] In the studies where it was reported (n = 16/47, 34%),[31, 32, 34, 38, 40, 41, 43, 45, 50, 51, 55, 56, 58, 65, 76, 77] all but four studies[31, 55, 56, 77] took place in solely an urban setting.

Fifteen (32%) studies reported on fatigue-related exposures (e.g., as a source of stress, exhaustion, physical fatigue; hereafter referred to as ‘fatigue’),[35, 40, 45, 48, 57, 63-71, 73] while others (n = 37/47, 79%) reported on sleep-related exposures (e.g., sleep hours, insufficient sleep, sleep deprivation, sleep disruption, sleepiness; hereafter referred to as ‘insufficient sleep’).[31, 32, 34, 36-47, 49-56, 58-62, 64, 67, 71, 72, 74-78] A few (n = 5/47, 11%) reported on both.[40, 45, 64, 67, 71] In some cases (n = 18/47, 38%), fatigue or insufficient sleep were related to overnight work or long on-call shifts.[31, 32, 34, 37, 40, 41, 43, 45, 46, 50, 51, 53, 58-60, 63, 75, 77] Measured outcomes varied widely and were ultimately organised into physician physical and mental health, physician performance and risk of error, and patient outcomes.

Risk of bias appraisal

The overall quality of the body of research was poor; 62% (n = 29/47) of studies were rated at unclear or high risk of bias. Of the two randomised controlled trials, one was rated as unclear overall risk of bias[75] and one as high risk.[40] All cohort studies were at low risk of bias (mean score: 8.4/9, range: 8-9).[31, 32, 41, 58, 63, 77] All of the before-after studies were rated as high risk of bias.[34, 45, 50] The single time series study was assessed at high risk of bias.[51] The cross-sectional studies varied in performance (mean score: 3.0/5, range: 1-4); only one-third (n = 12/34, 35%) were at low risk of bias.[39, 42, 44, 47, 52, 59, 60, 69, 70, 72, 74, 76] The one non-comparative study was at unclear risk of bias.[43] Detailed assessments of the sources of bias per study are shown in Supplementary file 3.

Physician health and wellbeing outcomes

Twenty-eight studies reported on physician health and wellbeing-related outcomes,[35, 36, 38-40, 42, 46-48, 50-57, 60, 62, 64, 67, 68, 70-72, 74, 76, 78] including burnout (n = 7), stress (n = 8), mental health and wellbeing (n = 7), life and job satisfaction (n = 9) and other markers of health (n = 5) (Supplementary file 4).

Seven cross-sectional studies reported on burnout (5 low[39, 70, 72, 74, 76], 1 unclear[54], 1 high risk of bias[62]) among surgeons,[54, 62] anesthesiologists,[39] generalists,[76] and other mixed groups.[70, 72, 74] Two studies reported on surgeons; the larger (n = 2,564, low risk of bias) study of neurosurgeons showed increased odds of burnout with sleep deprivation (hours of sleep per night; OR 0.84, 95% CI 0.75 to 0.94, P = 0.002).[54] Among anesthesiologists one study (n = 565, low risk of bias) indicated that burnout (measured via Maslach Burnout Inventory) was more prevalent among the sleep-deprived ('lack of sleep' on one question; 47.6% vs. 16.3%, P < 0.001).[39] In one small (n = 11) study of generalists, those with burnout (measured via Pines Burnout Measure) had poorer Pittsburgh Sleep Quality Index scores (7.24±4.17 vs. 2.72±2.22, P < 0.001).[76] In the two larger studies of mixed physician groups (low risk of bias), burnout (measured via 5-point scale) was more prevalent among those who were sleep deprived (<7 hours of sleep per 24 hours; 39.6% vs. 26.4%, P < 0.05),[72] and physical fatigue ('feeling tired' on a 7-point scale) was correlated with burnout (Shirom-Melamed Burnout Measure; r = 0.88, P < 0.05).[70] In summary, evidence from 7 cross-sectional studies (71% at low risk of bias), showed associations between insufficient sleep and burnout.

Six cross-sectional studies (2 low[47, 52], 1 unclear[46], 3 high risk of bias[35, 62, 64]), one uncontrolled before-after study (high risk of bias[50]), and one intervention study (high risk of bias[40]) reported on stress outcomes among surgeons,[62] anesthesiologists,[50, 52] emergency physicians,[40, 64] internal medicine physicians,[46] and mixed groups.[35, 47] In a small sample (n = 20) of internal medicine physicians, insufficient sleep related to a 24-hour call shift showed no association with biochemical or physiological stress parameters, except levels of thyroid stimulating hormone, which was higher post-shift (P = 0.049, data not reported).[46] The remaining observational studies suggested that there was an association between insufficient sleep or fatigue and stress. The one study of orthopedic surgeons (n = 264, high risk of bias) showed that insufficient sleep (measured on a 3-point scale) and psychological distress (measured via General Health Questionnaire-12) were correlated (data not reported, P < 0.001).[62] The two reports on anesthesiologists were of varied quality; the larger (n = 328, low risk of bias) study showed that stress symptoms (measured via Modified Occupational Stress Questionnaire) were predicted by sleep sufficiency (self-reported on one question, β = -0.269, P < 0.001).[52] Among the two studies reporting on mixed groups of physicians, the larger (n = 1,541, low risk of bias) study showed an association between sleep problems (4 questions derived from Jenkins scale) and psychological distress (General Health Questionnaire-12; β = 0.18, P < 0.001).[47] One RCT assessed the impact of insufficient sleep from shift work (14-hour or 24-hour shifts), showing that stress (on a visual analog scale) among emergency physicians (n = 17) was higher following the shift as compared to a control day (data not reported, P < 0.05).[40] In summary, evidence from one intervention study at high risk of bias and all but one of the 7 observational studies (29% at low risk of bias) identified supported an inverse association between fatigue or sleep deprivation and stress.

Seven cross-sectional studies (2 low,[52, 60] 3 unclear,[67, 71, 78] 2 high risk of bias[36, 53]) reported on aspects of mental health including addiction or substance misuse,[36, 53, 71] depression,[78] thoughts of suicide,[52] mood disturbance,[60, 71] and overall wellbeing.[67] One study,[53] which was at high risk of bias, showed no association between hours of sleep when on call and hazardous drinking behaviours (via Alcohol Use Disorder Identification Test). Meanwhile, the six other studies all showed associations between insufficient sleep and fatigue and reduced mental health. Three studies reported on anesthetists,[36, 52, 60] with two large surveys showing increased odds of tobacco (OR 1.42, 95% CI 1.04 to 1.94) and tranquilizer/hypnotics (OR 3.26, 95% CI 2.12 to 5.02) dependency being predicted by sleep deprivation (measured by one question),[36] and sleep disturbance being associated with thoughts of suicide (using a 4-point scale; P = 0.009).[52] A small study (n = 21) showed greater mood disturbance following a 17-hour night shift than a usual day (Profile of Mood States score 42.57±15.26 vs. 70.90±6.91, P < 0.001).[60] Among oncologists (n = 241), overall wellbeing was predicted by lower levels of fatigue after controlling for personal and professional characteristics (assessed via linear analog scale quality of life survey, P = 0.002).[67] A large (n = 3,862, unclear risk of bias) study of physicians showed that insufficient sleep (lower sleep hours when not at work in the past month) was associated with increased odds of depression (Quick Inventory Depressive Scale; OR 2.70, 95% CI 1.82 to 4.03 for men; OR 2.38, 95% CI 1.11 to 5.10 for women).[78] In open-ended questions, senior physicians in one study (unclear risk of bias) attributed the development of mental illness to tiredness and stress at work.[71] In summary, 7 cross-sectional studies (29% at low risk of bias) were identified, and of these 6 supported an association between insufficient sleep or fatigue and negative mental health outcomes.

Nine cross-sectional studies (4 low,[42, 47, 72, 74] 2 unclear,[55, 68] 3 high risk of bias[38, 48, 62]) reported on outcomes related to job satisfaction,[42, 47, 48, 55, 72, 74] life satisfaction,[38, 62, 72] or work-life balance.[68, 72] The six studies that investigated job satisfaction were all at low risk of bias and generally included mixed groups of physicians;[47, 72, 74] one study reported on general practitioners,[55], another on surgeons,[48] and one on mixed specialties.[42] Three studies showed that reductions in sleep duration and/or quality[47, 48, 74] were associated with reduced job satisfaction. Meanwhile one showed no association between insufficient sleep (<7 hours per 24-hour period) and career satisfaction (measured on a 5-point Likert scale),[72] and another showed no relationship between earlier sleep disturbance (Jenkins Scale) and later job demands or job control (measured via 5-point scale).[42] A single study (n = 92) reporting on rural general practitioners indicated that frequent sleep disturbance (measured on a 7-point scale) predicted the intention to retire early (OR 2.91, 95% CI 1.11 to 7.6, P < 0.05).[55] In summary, 6 cross-sectional studies (all at low risk of bias) were identified, and all but two[42, 72] of these studies showed that insufficient sleep and fatigue were associated with reductions in satisfaction.

The three studies reported on life satisfaction.[38, 62, 72] Of two studies among mixed physician groups,[38, 72] the one larger (n = 840) study showed that insufficient sleep (< 7 hours per day) was a predictor of reduced life satisfaction (measured on a 5-point Likert scale; OR 0.44, 95% CI 0.29 to 0.67, P ≤ 0.05).[72] One study at high risk of bias reported on orthopedic surgeons (n = 264), showing that sleep deprivation (measured via 3-point scale) was correlated with lower marital satisfaction (Revised Dyadic Adjustment Scale; data not reported, P < 0.001).[62] Two large studies at low or unclear risk of bias reported on work-life balance.[68, 72] Among oncologists (n = 1,117), reduced satisfaction with work-life balance (measured on a 5-point Likert scale) was predicted by high levels of fatigue (measured via 10-point visual analog scale), even when controlling for personal and work-related factors and burnout (OR 0.489, 95% CI 0.337 to 0.710, P < 0.001).[68] Among a mixed group of physicians (n = 840, low risk of bias), insufficient sleep (<7 hours in a typical 24-hour period) predicted a reduced perception of having balanced personal and professional commitments (5-point Likert scale; OR 0.46, 95% CI 0.31 to 0.71, P ≤ 0.05).[72] In summary, 3 cross-sectional studies (all unclear or high risk of bias) supported an association between insufficient sleep or fatigue and reduced life satisfaction, and 2 cross-sectional studies (50% low risk of bias) supported an association with reduced work-life balance.

Four cross sectional studies (3 unclear,[56, 57, 71] 1 high risk of bias[38]) and one time series study (high risk of bias[51]) reported on other health-related outcomes. Among a mixed group of physicians (n = 180), one study at high risk of bias showed that Epworth Sleepiness Scale scores were higher among physicians who worried about having a car accident while driving home (7.0 vs. 5.4, P < 0.001).[38] Among generalists (n = 578), almost 1 in 10 (8.7%) admitted to falling asleep while driving due to fatigue.[57] Also among generalists (n = 92), those with frequent work-related sleep disturbance (measured on a 7-point scale) were at increased odds of sickness presenteeism (OR 2.92, 95% CI 1.19 to 7.16, P = 0.02).[56] The one time series study concluded that a single 24-h shift did not cause major chronodisruption (based on serum melatonin measurement) among anesthetists (n = 10).[51] Meanwhile, open-ended comments from a large sample (n = 3,550) of senior physicians suggests that they attributed the development of physical health problems to a lifestyle of insufficient sleep, poor eating habits and lack of exercise imposed by their jobs.[71] In summary, 5 cross sectional studies (0% at low risk of bias) supported associations between insufficient sleep and fatigue and varied deleterious health outcomes (i.e., car accidents, sickness presenteeism, physical health problems). One time series study at high risk of bias did not support such a relationship.

Physician performance and risk of errors

Twenty-one studies reported on physician performance and safety-related outcomes,[31, 32, 34, 37, 38, 41, 43-47, 49, 50, 59, 61, 65, 66, 69, 71, 73, 75] including surgical efficiency and effectiveness (n = 6), psychomotor performance (n = 7), work ability and quality of care (n = 5) and medical errors (n = 5) (Supplementary file 4).

Four cohort studies (all low risk of bias[31, 32, 41, 63]), one before-after study (high risk of bias[34]) and one randomized controlled trial (high risk of bias[75]) examined the effects of insufficient sleep from overnight work or extended shifts, during surgeries[31, 32, 41] or laparoscopic simulations.[34, 75] We pooled the data from these studies[31, 32, 41, 63] via meta-analysis, which showed no difference in operating time (sometimes referred to as surgeon efficiency) between sleep deprived and non-sleep deprived surgeons (Figure 2; n = 50,046, MD -0.14, 95% CI -1.60 to 1.33, I^2 = 0% (P = 0.70)). Of studies not meta-analysed, the small (n = 29) before-after study showed no impact of sleep deprivation from shift-work nor of sleep hours on performance on a laparoscopic simulation (LapSimGyn).[34] One small (n = 64) intervention study compared a 24-hour shift to a usual work day, also finding no detriment to performance on a laparoscopic simulation (Minimally Invasive Surgical Trainer-Virtual Reality) despite diminished sleep hours while working on-call.[75] In summary, pooled data from 4 cohort studies (100% low risk of bias) showed no effect of insufficient sleep on surgical efficiency. Additional data from one RCT (high risk of bias) and one before-after study (high risk of bias) also showed no association between insufficient sleep and performance on laparoscopic simulations.

Two before-after studies (high risk of bias[45, 50]) and five cross-sectional studies (2 low,[43, 59] 3 unclear,[37, 46] 1 high risk of bias [61]) reported on psychomotor performance outcomes among surgeons,[45] anesthesiologists,[37, 43, 50, 59] emergency physicians,[61] and internal medicine physicians.[46] Among a small group of surgeons (n = 9), performance on a virtual ring transfer task deteriorated after an on-call shift (data not reported, P < 0.05).[45] The four studies among anesthetists reported mixed findings. One small (n = 11) before-after study showed longer reaction times (690.8±73.4 vs. 746.5±113.7 milliseconds) and reduced concentration ability (26.4±23.5 vs. 56.3±23.0 on a 100-point scale, P = 0.007) following a 24-hour shift with insufficient sleep;[50] Two others found that insufficient sleep due to overnight shifts was associated with slower reaction times.[43, 59] Conversely, a small study (n = 11) found no effect of overnight shiftwork with insufficient sleep on any measure of psychomotor performance except Hopkin's Verbal Learning Test (t-score of 48.6±7.6 vs. 41.5±9.9, P = 0.04).[37] Among emergency physicians (n = 18), one study (high risk of bias) showed that those who were sleep deprived (<5 hours sleep after a 24-hour shift) had a reduced performance on most but not all psychomotor tests (Battery Test Reaction 5),[61] while among internal medicine physicians (n = 20, low risk of bias), neurocognitive parameters did not seem to worsen post-call.[46] In summary, two before-after (0% low risk of bias) and 5 cross-sectional studies (40% low risk of bias) showed mixed results for the association between fatigue or insufficient sleep and psychomotor performance.

Five cross-sectional studies (2 low,[47, 69] 1 unclear,[71] 2 high risk of bias[38, 65]) reported on associations between sleep deprivation or fatigue and work ability or perceived performance, all among mixed groups of physicians.[38, 47, 65, 69, 71] The two large studies at low risk of bias showed that sleep problems and fatigue were inversely associated with physicians' perceived quality of work.[47, 69] Among 1,541 physicians in Finland, sleeping problems (measured by 4 questions from the Jenkins Scale) were inversely associated with scores on the Work Ability Index (β = -0.29, P < 0.001),[47] while a study of 890 physicians from Israel demonstrated that perceived quality of care was predicted by fatigue (1 item on the Shirom-Melamed Burnout Measure) even after controlling for components of burnout (β = 0.17, P < 0.05).[69] Similarly, in one study, comments from senior physicians suggested that continual

tiredness and exhaustion negatively affected their perceived competence.[71] The two studies[38, 65] that were at high risk of bias had conflicting findings. In summary, 5 cross-sectional studies (40% at low risk of bias) reported on perceived work performance; those that were at low risk of bias supported an association between fatigue or insufficient sleep and reduced performance.

Five cross-sectional studies (1 low,[44] 2 unclear,[49, 66] 2 high risk of bias[38, 73]) reported on associations between insufficient sleep or fatigue and self-reported medical errors among surgeons,[66] anesthesiologists[44] and mixed groups of physicians.[38, 49, 73] A large (n = 7,905) study at unclear risk of bias showed that only 6.9% of surgeons reported fatigue as the most important contributor to medical errors.[66] Among anesthesiologists, a smaller study (n = 183) at low risk of bias showed that the risk of self-reported fatigue-related errors increased with more nights of work-related sleep disturbance (RR 1.25, 95% CI 1.06 to 1.49).[44] Two of the studies reporting on mixed groups of physicians had conflicting results,[38, 49] while another reported that physicians' opinions on the association between fatigue and prescribing errors differed by work setting.[73] One-third (34%) of community-based, 96% of hospital-based, and 8% of office-based physicians believed that there was a high or very high association between fatigue and prescribing errors ($P < 0.05$).[73] In summary, 5 cross-sectional studies (20% at low risk of bias) reported on self-reported errors, and these showed mixed findings for associations with fatigue or insufficient sleep.

Patient Outcomes

Six large (n = 270 to 38,978) cohort studies at low risk of bias reported on patient outcomes, all related to surgical[31, 32, 41, 58, 63, 77] or obstetric[58] procedures (Supplementary file 4). In these studies, insufficient sleep or fatigue were typically defined as overnight work prior to a daytime procedure[31, 41, 58, 63, 77]; though two studies measured sleep hours[32] or 'sleep opportunity'.[58] We pooled data for procedures performed by sleep deprived versus non-sleep deprived surgeons (or obstetrician-gynecologists in one case[58]). Analyses showed no difference in the rate of post-operative complications (Figure 3; 5 studies,[31, 32, 41, 63, 77] n = 60,201, RR 0.99, 95% CI 0.95 to 1.03, $I^2 = 0\%$ ($P = 0.45$) nor patient mortality (Figure 4; 5 studies,[31, 32, 41, 63, 77] n = 60,436, RR 0.98, 95% CI 0.84 to 1.15, $I^2 = 0\%$ ($P = 0.73$)). One study[77] in the mortality analysis reported the number of deaths only as ≤5. We assumed 2 events for this study (midpoint between 0 and 5); sensitivity analysis using the lowest (i.e., 0) and highest (i.e., 5) possible number of events did not change the overall result (Supplementary file 5). We found considerable between-study heterogeneity in the analyses for intraoperative complications (I^2 = 82%) and length of stay (I^2 = 86%), which could not be explained via subgroup analyses by procedure type, thus we have suppressed the average estimates of effect. For length of stay, the results of one study on cardiac surgeries favoured sleep deprived surgeons,[32] while the others[31,41,63] had null results. For intraoperative complications, the findings of one study[63] favoured non-sleep deprived surgeons, but the others[58,77] had null results.

DISCUSSION

Fatigue and chronic insufficient sleep are two potential drivers of reduced physician wellbeing[17, 19] that have thus far been understudied in physicians in independent practice. Burnout is becoming increasingly prevalent among physicians,[14-16] and recent research indicates that comprehensive individual- and system-level strategies are needed to address the problem.[6-9, 19, 21] We have systematically reviewed evidence from a heterogeneous array of available studies reporting on diverse outcomes related to physicians in independent practice and their patients. The included studies were often at high or unclear risk of bias, included small samples of physicians, and inconsistently measured and reported exposures and outcomes. The key message gleaned from this review is that despite growing interest in the topic of physician wellness, the robust evidence needed to inform individual and systems-level fatigue management strategies is lacking.

Traditionally, much of the fatigue-related research has focused on hazards to patients. The current review included six cohort studies showing that insufficient sleep and/or fatigue did not seem to result in increased rates of patient morality or post-operative complications; findings for length of stay and intra-operative complications were inconclusive. Evidence for psychomotor performance, surgical skills and errors suggest that there is indeed a potential for negative outcomes. The included studies, like many of the others in this and other systematic reviews,[79] employed indirect definitions that make it difficult to classify sleep deprived physicians with certainty. In recent years there has been a shift away from the singular focus on patient safety toward a more comprehensive view that also considers the detrimental effects of fatigue, sleep loss and other occupational hazards on physician wellness.[80] Evidence from this review supports that fatigue and insufficient sleep may be negatively associated with physician health and wellbeing. It is now recognized that health systems cannot be sustained by a workforce that is facing an epidemic of burnout.[19, 81, 82]

In light of high rates of burnout, the ongoing dialogue about the need for a cultural shift in the practice of medicine[83, 84] is now more important than ever. Recognition of the potential effects of physician fatigue on patients, physicians, and healthcare systems as a whole must be emphasized at a systemic level, encouraging a shift in which the risks are viewed as unacceptable.[1, 20, 80] Likewise, although research to date has focused largely on individual-level approaches to address burnout, it is now clear that placing the burden of a system-level problem solely on the individual is unlikely to bring about significant and lasting change.[85] Recent research has highlighted physician burnout as a system-driven issue that will require corresponding national-scale multicomponent solutions.[1, 19, 81, 82] As such, in the past several years both the American and Canadian Medical Associations have developed policies and programs that address physician health.[81, 86] The Canadian Medical Association's new policy on physician health calls on broad stakeholder groups (e.g., policymakers, regional health authorities, governments) to take shared responsibility for the health of physicians and to make meaningful and concerted efforts towards promoting a healthy and sustainable workforce.[81]

The most salient finding of this review is that the current evidence is insufficient to inform policy and practice. Correspondingly, a 2016 research summit on physician wellness and burnout outlined the need for timely, relevant and methodologically robust research to inform practice and policy.[21] The findings herein may be used as motivation for researchers and practitioners to develop and design methodologically strong research programs related to physician fatigue, inform successful research grant proposals, and lobby healthcare organizations to increase the focus on physician fatigue management programs. It will be important to make use of existing validated measures[87-89] consistently in future research. Identifying outcomes of importance to physicians and their patients should be prioritized, such that these may be collected within intervention studies. Reporting these consistently will allow for the effective synthesis of findings and reduce research waste.[90] Integrated knowledge translation strategies involving multiple stakeholder groups (e.g., physicians, patients, medical schools, physicians' associations and governing bodies, policymakers) may help to ensure that the research is relevant and facilitates decision-making.[91]

Strengths and Limitations

Our systematic review is the first to synthesize evidence on the effects of fatigue and insufficient sleep on physicians in independent practice. The review is timely, given recent calls for research into individual and organisational solutions for burnout,[20, 21] and an increased focus on physician health.[80, 81] While we have identified a diverse body of evidence, we could not draw definitive conclusions due to methodological weaknesses (e.g., 62% at high risk of bias, reliance primarily on cross-sectional designs and uncontrolled studies, subjective measurement of exposures and outcomes, small sample sizes, inclusion of predominantly male physicians within urban settings) and heterogeneous outcome measures in the included studies. Given that the 2017 update search was limited to one database, it is possible that a small number of relevant studies could have been missed. We believe that the likelihood that these might alter the conclusions of the review is low. The findings may have been influenced by publication bias, and may not be generalized to all settings, given our restriction to high income countries. Confidence in the conclusions is limited due to multiple comparisons.

CONCLUSION

The evidence synthesized in this review suggests that fatigue and insufficient sleep are associated with some detrimental physician health and wellbeing outcomes; the evidence for potential associations with performance and safety outcomes was mixed. Meta-analyses for patient outcomes demonstrated that in many cases, potential relationships with physician sleep deprivation remain unclear. Our overall confidence in the findings is low, owing to multiple comparisons and a body of research that is hindered by methodological weaknesses. Further methodologically robust research that includes consistent outcomes that are of interest to physicians and their patients is needed to inform strong practice recommendations and policy decisions.

ACKNOWLEDGMENTS

We would like to thank Dr. Meghan Sebastianski and Andrea Milne for assistance with screening, and Dr. Allison Gates for screening, data verification, and quality assessment of the French studies.

COPYRIGHT/LICENSE FOR PUBLICATION

Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) license any third party to do any or all of the above.

COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: Dr. Christopher Simon is employed by the Canadian Medical Association, who provided financial support for the research; there are no other relationships or activities that could appear to have influenced the submitted work.

CONTRIBUTOR STATEMENT

All authors contributed to the conception and design of the project. MG and AW contributed to the acquisition, analysis and interpretation of the data, and drafted the manuscript. RF contributed to acquisition of data. CSa, CSi and MPD contributed to interpretation of data and revised the manuscript for important intellectual content. All authors approved the final version of the manuscript as submitted.

FUNDING STATEMENT

This research was supported by the Canadian Medical Association. The funder had no role in the study design; collection, analysis and interpretation of data; the writing of the report; and the decision to submit the article for publication.

ROLE OF FUNDERS

Dr. Christopher Simon is employed by the Canadian Medical Association. The remaining authors are independent from the funders. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

TRANSPARENCY DECLARATION

The lead author (MG) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted; and all discrepancies from the study as planned have been explained.

DATA ACCESS STATEMENT

All authors, external and internal, had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the interpretation.

DATA SHARING STATEMENT

The data pertaining to this systematic review are available from the corresponding author upon reasonable request.

FIGURE CAPTIONS

Figure 1. Flow of records through the selection process

Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

Figure 3. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

Legend: Vinden 2013 reported iatrogenic injuries; Schieman 2008, Govindarajan 2015, and Chu 2011 reported post-operative complication rate; Ellman 2004 reported post-operative complications (other types of complications reported not included in the analysis)

Figure 4. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

REFERENCES

1. Greig P, Snow R. Fatigue and risk: are train drivers safer than doctors? *BMJ* 2017;359. doi: https://doi.org/10.1136/bmj.j5107
2. Parker JBR. The effects of fatigue on physician performance - an underestimated cause of physician impairment and increased patient risk. *Can J Anaesth* 1987;34(5):489-95.
3. Temple J. Resident duty hours around the globe: where are we now? BMC Med Educ 2014;14(Suppl 1):S8. https://doi.org/10.1186/1472-6920-14-S1-S8
4. Imrie K, Frank J, The National Steering Committee on Resient Duty Hours. Fatigue, risk, & excellence: towards a pan-Canadian consensus on resident duty hours. Ottawa, Ontario: The Royal College of Physicians and Surgeons of Canada, 2013. http://www.residentdutyhours.ca/final_report.php (accessed 12 Jan 2018).
5. Accreditation Council for Graduate Medical Education. History of duty hours. 2017. http://www.acgme.org/What-We-Do/Accreditation/Clinical-Experience-and-Education-formerly-Duty-Hours/History-of-Duty-Hours (accessed 12 Jan 2018).
6. Bolster L, Rourke L. The effect of restricting residents' duty hours on patient safety, resident well-being, and resident education: an updated systematic review. *J Grad Med Educ* 2015;7(3):349-63.
7. Harris JD, Staheli G, LeClere L, Andersone D, McCormick F. What effects have resident work-hour changes had on education, quality of life, and safety? A systematic review. *Clin Orthop* 2015;473(5):1600-8.
8. Ahmed N, Devitt KS, Keshet I, et al. A systematic review of the effects of resident duty hour restrictions in surgery: impact on resident wellness, training, and patient outcomes. *Ann Surg* 2014;259(6):1041-53.
9. Peets A, Ayas NT. Restricting resident work hours: the good, the bad, and the ugly. *Crit Care Med* 2012;40(3):960-6.
10. Canadian Medical Association. CMA Policy: Management of physician fatigue. Ottawa, Canada: Canadian Medical Association; 2014.
11. Anim M, Markert RJ, Wood VC, Schuster BL. Physician practice patterns resemble ACGME duty hours. *Am J Med* 2009;122:587-93.
12. Transport Canada. Fatigue risk management system for the canadian aviation industry: policies and procedures development guidelines. Ottawa, Canada: Her Majesty the Queen in Right of Canada, as

represented by the Minister of Transport, 2008. https://www.tc.gc.ca/eng/civilaviation/publications/TP14576-6042.htm (accessed 12 Jan 2018).

13. Patterson PD, Higgins JS, Lang ES, et al. Evidence-based guidelines for fatigue risk management in ems: formulating research questions and selecting outcomes. *Prehosp Emerg Care* 2017;21(2):149-56.
14. Boudreau RA, Grieco RL, Cahoon SL, Robertson RC, Wedel RJ. The pandemic from within: two surveys of physician burnout in Canada. *Can J Commun Ment Health* 2006;25(2):71-88.
15. Kumar S. Burnout and doctors: prevalence, prevention and intervention. *Healthcare* 2016;4(3):37. doi: 10.3390/healthcare4030037
16. Shanafelt TD, Boone S, Tan L, et al. Burnout and satisfaction with work-life balance among us physicians relative to the general us population. *Arch Intern Med* 2012;172(18):1377-85.
17. Brady KJS, Trockel MT, Khan CT, et al. What do we mean by physician wellness? A systematic review of its definition and measurement. *Acad Psychiatry* 2017. doi: 10.1007/s40596-017-0781-6
18. Fralick M, Flegel K. Physician burnout: who will protect us from ourselves? *CMAJ* 2014;186(10):731. doi: 10.1503/cmaj.140588
19. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. *Mayo Clin Proc* 2017;92(1):129-46.
20. Lemaire JB, Wallace JE. Burnout among doctors. BMJ 2017;358:j3360. doi: 10.1136/bmj.j3360
21. Dyrbye LN, Trockel M, Frank E, et al. Development of a research agenda to identify evidence-based strategies to improve physician wellness and reduce burnout. *Ann Intern Med* 2017;166(10):743-4.
22. Higgins JPT, Green S (editors). The Cochrane handbook for systematic reviews of interventions, version 5.1.0. London, UK: The Cochrane Collaboration, 2011. http://www.handbook.cochrane.org (accessed 12 Jan 2018).
23. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;6(7):e1000100. https://doi.org/10.1371/journal.pmed.1000100
24. Pattani R, Wu PE, Dhalla IA. Resident duty hours in Canada: past, present and future. *CMAJ* 2014;186(10):761-5.
25. International Monetary Fund. World economic outlook: too slow for too long. Washington, DC: International Monetary Fund, 2016. http://www.imf.org/external/pubs/ft/weo/2016/01/ (accessed 12 Jan 2018).

26. Cochrane Effective Practice and Organisation of Care. Suggested risk of bias criteria for EPOC reviews. Oslo, Norway: Norwegian Knowledge Centre for the Health Services, 2016. http://epoc.cochrane.org/epoc-specific-resources-review-authors (accessed 12 Jan 2018).

27. Wells GA SB, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa, Canada: The Ottawa Hospital Research Institute, 2014. http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (accessed 12 Jan 2018).

28. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.

29. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-60.

30. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.

31. Govindarajan A, Urbach DR, Kumar M, et al. Outcomes of daytime procedures performed by attending surgeons after night work. *N Engl J Med* 2015;373(9):845-53.

32. Chu MW, Stitt LW, Fox SA, Kiaii B, Quantz M, Guo L, et al. Prospective evaluation of consultant surgeon sleep deprivation and outcomes in more than 4000 consecutive cardiac surgical procedures. *Arch Surg* 2011;146(9):1080-5.

33. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol* 59(1):7-10.

34. Amirian I, Andersen LT, Rosenberg J, Gogenur I. Laparoscopic skills and cognitive function are not affected in surgeons during a night shift. *J Surg Educ* 2014;71(4):543-50.

35. Aziz A. Sources of perceived stress among american medical doctors: a cross-cultural perspective. Cross Cultural Management 2004;11(4):28-39.

36. Beaujouan L, Czernichow S, Pourriat JL, Bonnet F. Prevalence and risk factors for substance abuse and dependence among anaesthetists: a national survey. *Ann Fr Anesth Réanim* 2005;24(5):471-9.

37. Chang LC, Mahoney JJ, 3rd, Raty SR, Ortiz J, Apodaca S, De La Garza R 2nd. Neurocognitive effects following an overnight call shift on faculty anesthesiologists. *Acta Anaesthesiol Scand* 2013;57(8):1051-7.

38. Chen I, Vorona R, Chiu R, Ware JC. A survey of subjective sleepiness and consequences in attending physicians. *Behav Sleep Med* 2008;6(1):1-15.

39. Doppia MA, Estryn-Béhar M, Fry C, Guetarni K, Lieutaud T, comité de pilotage de l'enquête SESMAT. Burnout in French doctors: a comparative study among anaesthesiologists and other specialists in French hospitals (SESMAT study). *Ann Fr Anesth Réanim* 2011;30(11):782-94.
40. Dutheil F, Trousselard M, Perrier C, et al. Urinary interleukin-8 is a biomarker of stress in emergency physicians, especially with advancing age--the JOBSTRESS randomized trial. *PLoS One* 2013;8(8):e71658. https://doi.org/10.1371/journal.pone.0071658
41. Ellman PI, Law MG, Tache-Leon C, et al. Sleep deprivation does not affect operative results in cardiac surgery. *Ann Thorac Surg* 2004;78(3):906-11.
42. Elovainio M, Heponiemi T, Jokela M, et al. Stressful work environment and wellbeing: what comes first? *J Occup Health Psychol* 2015;20(3):289-300.
43. Gander P, Millar M, Webster C, Merry A. Sleep loss and performance of anaesthesia trainees and specialists. *Chronobiol Int* 2008;25(6):1077-91.
44. Gander PH, Merry A, Millar MM, Weller J. Hours of work and fatigue-related error: a survey of New Zealand anaesthetists. *Anaesth Intensive Care* 2000;28(2):178-83.
45. Gerdes J, Kahol K, Smith M, Leyba MJ, Ferrara JJ. Jack Barney award: the effect of fatigue on cognitive and psychomotor skills of trauma residents and attending surgeons. *Am J Surg* 2008;196(6):813-9.
46. Harbeck B, Suefke S, Haas CS, Lehnert H, Kropp P, Moenig H. No stress after 24-hour on-call shifts? *J Occup Health* 2015;57(5):438-47.
47. Heponiemi T, Puttonen S, Elovainio M. On-call work and physicians' well-being: testing the potential mediators. *Occup Med* 2014;64(5):352-7.
48. Jackson TN, Pearcy CP, Khorgami Z, Agrawal V, Taubman KE, Truitt MS. The physician attrition crisis: a cross-sectional survey of the risk factors for reduced job satisfaction among US surgeons. *World J Surg* 2017;24. doi: 10.1007/s00268-017-4286-y
49. Kaneita Y, Ohida T. Association of current work and sleep situations with excessive daytime sleepiness and medical incidents among Japanese physicians. *J Clin Sleep Med* 2011;7(5):512-22.
50. Lederer W, Kopp M, Hahn O, et al. Post-duty psychomotor performance in young and senior anaesthetists. *Eur J Anaesthesiol* 2006;23(3):251-6.
51. Leichtfried V, Putzer G, Perkhofer D, Schobersberger W, Benzer A. Circadian melatonin profiles during single 24-h shifts in anesthetists. *Sleep Breath* 2011;15(3):503-12.
52. Lindfors PM, Nurmi KE, Meretoja OA, et al. On-call stress among Finnish anaesthetists. *Anaesthesia* 2006;61(9):856-66.

53. Mahmood JI, Stoen Grotmol K, Tesli M, Vaglum P, Tyssen R. Contextual factors and mental distress as possible predictors of hazardous drinking in Norwegian medical doctors: a 15-year longitudinal, nationwide study. *Eur Addict Res* 2017;23(1):19-27.
54. Nishimura K, Nakamura F, Takegami M, et al. Cross-sectional survey of workload and burnout among Japanese physicians working in stroke care: the nationwide survey of acute stroke care capacity for proper designation of comprehensive stroke center in Japan (J-ASPECT) study. *Circ Cardiovasc Qual Outcomes* 2014;7(3):414-22.
55. Pit SW, Hansen V. Factors influencing early retirement intentions in Australian rural general practitioners. *Occup Med* 2014;64(4):297-304.
56. Pit SW, Hansen V. The relationship between lifestyle, occupational health, and work-related factors with presenteeism amongst general practitioners. *Arch Environ Occup Health* 2016;71(1):49-56.
57. Roberts DL, Shanafelt TD, Dyrbye LN, West CP. A national comparison of burnout and work-life balance among internal medicine hospitalists and outpatient general internists. *J Hosp Med* 2014;9(3):176-81.
58. Rothschild JM, Keohane CA, Rogers S, et al. Risks of complications by attending physicians after performing nighttime procedures. *JAMA* 2009;302(14):1565-72.
59. Saadat H, Bissonnette B, Tumin D, et al. Effects of partial sleep deprivation on reaction time in anesthesiologists. *Paediatr Anaesth* 2017;27(4):358-62.
60. Saadat H, Bissonnette B, Tumin D, et al. Time to talk about work-hour impact on anesthesiologists: The effects of sleep deprivation on Profile of Mood States and cognitive tasks. *Paediatr Anaesth* 2016;26(1):66-71.
61. Sanches I, Teixeira F, dos Santos JM, Ferreira AJ. Effects of acute sleep deprivation resulting from night shift work on young doctors. *Acta Med Port* 2015;28(4):457-62.
62. Sargent MC, Sotile W, Sotile MO, Rubash H, Barrack RL. Quality of life during orthopaedic training and academic practice. Part 1: orthopaedic surgery residents and faculty. *J Bone Joint Surg Am* 2009;91(10):2395-405.
63. Schieman C, MacLean AR, Buie WD, Rudmik LR, Ghali WA, Dixon E. Does surgeon fatigue influence outcomes after anterior resection for rectal cancer? *Am J Surg* 2008;195(5):684-7.
64. Sende J, Jbeili C, Schvahn S, Khalid M, Asaph J, Romano H, et al. Stress factors and stress consequences among emergency physicians: vational survey. *Annales Francaises de Médecine d'Urgence* 2012;2(4):224-31.

65. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. *Hum Perform Extrem Environ* 2001;6(1):6-11.

66. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and medical errors among American surgeons. *Ann Surg* 2010;251(6):995-1000.

67. Shanafelt TD, Novotny P, Johnson ME, et al. The well-being and personal wellness promotion strategies of medical oncologists in the North Central Cancer Treatment Group. *Oncology* 2005;68(1):23-32.

68. Shanafelt TD, Raymond M, Kosty M, et al. Satisfaction with work-life balance and the career and retirement plans of US oncologists. *J Clin Oncol* 2014;32(11):1127-35.

69. Shirom A, Nirel N, Vinokur AD. Overload, autonomy, and burnout as predictors of physicians' quality of care. *J Occup Health Psychol* 2006;11(4):328-42.

70. Shirom A, Nirel N, Vinokur AD. Work hours and caseload as predictors of physician burnout: the mediating effects by perceived workload and by autonomy. *J Appl Psychol* 2010;59(4):539-65.

71. Smith F, Goldacre MJ, Lambert TW. Adverse effects on health and wellbeing of working as a doctor: views of the UK medical graduates of 1974 and 1977 surveyed in 2014. *J R Soc Med* 2017;110(5):198-207.

72. Starmer AJ, Frintner MP, Freed GL. Work-life balance, burnout, and satisfaction of early career pediatricians. *Pediatrics* 2016;e20153183. doi: 10.1542/peds.2015-3183

73. Tanti A, Camilleri M, Borg AA, et al. Opinions of Maltese doctors and pharmacists on medication errors. *Int J Saf Med* 2017;29(1-2):81-99.

74. Tokuda Y, Hayano K, Ozaki M, Bito S, Yanai H, Koizumi S. The interrelationships between working conditions, job satisfaction, burnout and mental health among hospital physicians in Japan: a path analysis. *Ind Health* 2009;47(2):166-72.

75. Uchal M, Tjugum J, Martinsen E, Qiu X, Bergamaschi R. The impact of sleep deprivation on product quality and procedure effectiveness in a laparoscopic physical simulator: a randomized controlled trial. *Am J Surg* 2005;189(6):753-7.

76. Vela-Bueno A, Moreno-Jimenez B, Rodriguez-Munoz A, et al. Insomnia and sleep quality among primary care physicians with low and high burnout levels. *J Psychosom Res* 2008;64(4):435-42.

77. Vinden C, Nash DM, Rangrej J, et al. Complications of daytime elective laparoscopic cholecystectomies performed by surgeons who operated the night before. *Obstet Gynecol Surv* 2014;69(2):71-3.

78. Wada K, Yoshikawa T, Goto T, et al. National survey of the association of depressive symptoms with the number of off duty and on-call, and sleep hours among physicians working in Japanese hospitals: a cross sectional study. *BMC Public Health* 2010;10:127. doi: 10.1186/1471-2458-10-127
79. Sturm L, Dawson D, Vaughan R, et al. Effects of fatigue on surgeon performance and surgical outcomes: a systematic review. *ANZ J Surg* 2011;81(7-8):502-9.
80. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet* 2009;374(9702):1714-21.
81. Canadian Medical Association. CMA policy: physician health. Ottawa, Canada: Canadian Medical Assocation, 2017. http://policybase.cma.ca/dbtw-wpd/Policypdf/PD18-01.pdf (accessed 12 Jan 2018).
82. Shanafelt T, Goh J, Sinsky C. The business case for investing in physician well-being. *JAMA Intern Med* 2017;177(2):195-205.
83. Gaba DM, Howard SK. Fatigue among clinicians and the safety of patients. *N Engl J Med* 2002;347(16):1249-55.
84. Ward S, Outram S. Medicine: in need of culture change. *Itern Med J* 2016;46(1):112-6.
85. Panagioti M, Panagopoulou E, Bower P, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. *JAMA Intern Med* 2017;177(2):195-205.
86. American Medical Association. STEPSforward™. Chicago, IL: American Medical Assocation, 2017. https://www.stepsforward.org/modules/joy-in-medicine (accessed 12 Jan 2018).
87. Jackson C. The Chalder Fatigue scale (CFQ 11). Occup Med 2015;65(1):86. doi: 10.1093/occmed/kqu168
88. Chalder T, Berelowitz G, Pawlikowska T, et al. Development of a fatigue scale. *J Psychosom Res* 1993;37(2):147-53.
89. Smets E, Garssen B, Bonke Bd, De Haes J. The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *J Psychosom Res* 1995;39(3):315-25.
90. Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383(9912):166-75.
91. Gagliardi AR, Berta W, Kothari A, Boyko J, Urquhart R. Integrated knowledge translation (IKT) in health care: a scoping review. *Implement Sci* 2016;11:38. https://doi.org/10.1186/s13012-016-0399-1

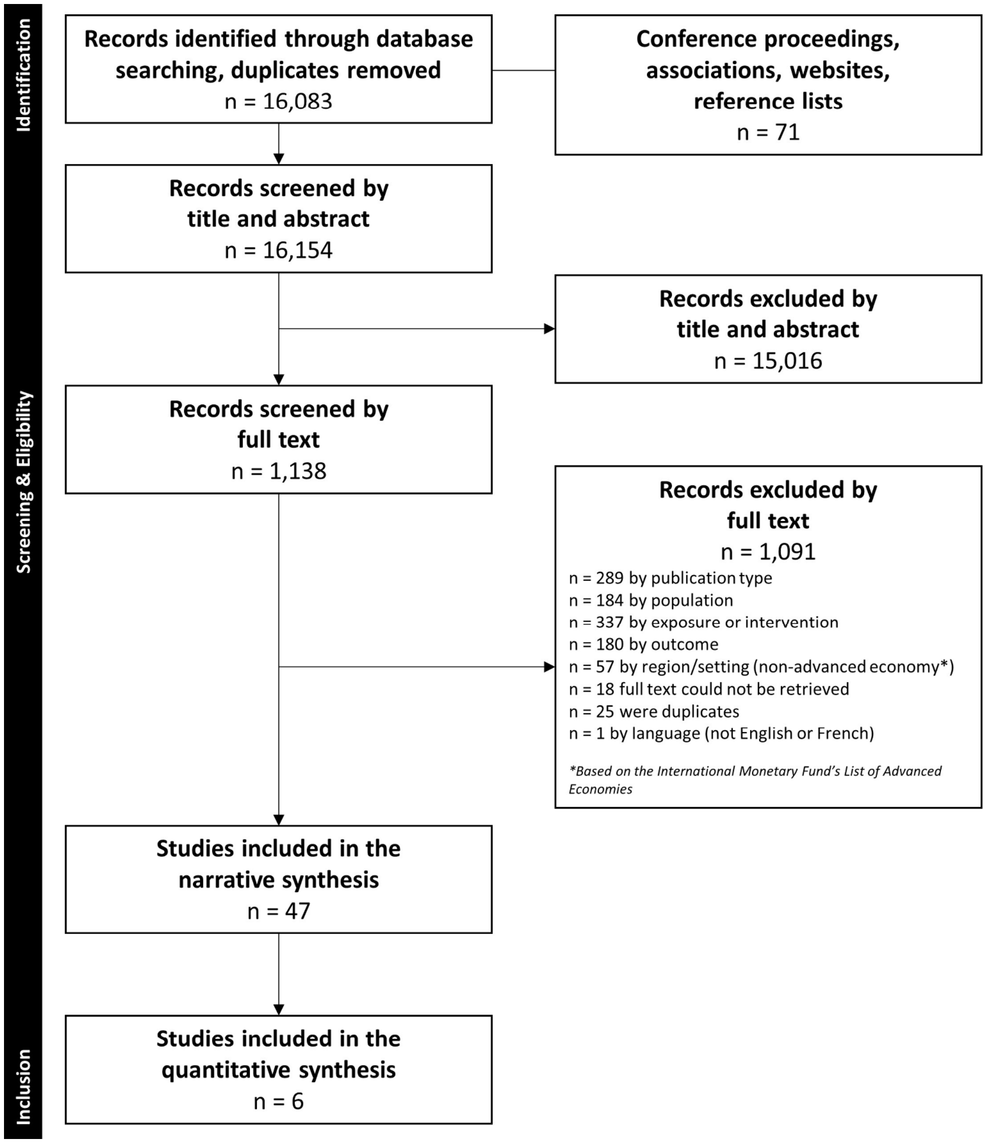


Figure 1. Flow of records through the selection process

190x215mm (300 x 300 DPI)

	Sleep deprived			Not sleep deprived				Mean Difference	
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI	Year
Ellman 2004	107.7	55.24	339	107.4	56.05	6412	5.9%	0.30 [-5.74, 6.34]	2004
Schieman 2008	114	226.6	22	148	226.6	248	0.0%	-34.00 [-132.80, 64.80]	2008
Chu 2011	110.2	119.13	1678	114	120.88	2369	3.8%	-3.80 [-11.30, 3.70]	2011
Govindarajan 2015	156	75.56	19489	156	80	19489	90.2%	0.00 [-1.54, 1.54]	2015
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.14 [-1.60, 1.33]**	

Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.42$, df = 3 (P = 0.70); $I^2 = 0\%$
Test for overall effect: Z = 0.18 (P = 0.86)

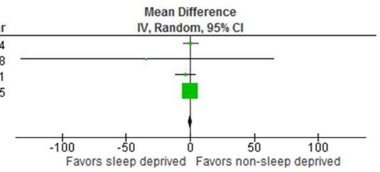


Figure 2. Forest plot for operating time among sleep deprived and non-sleep deprived surgeons

381x101mm (300 x 300 DPI)

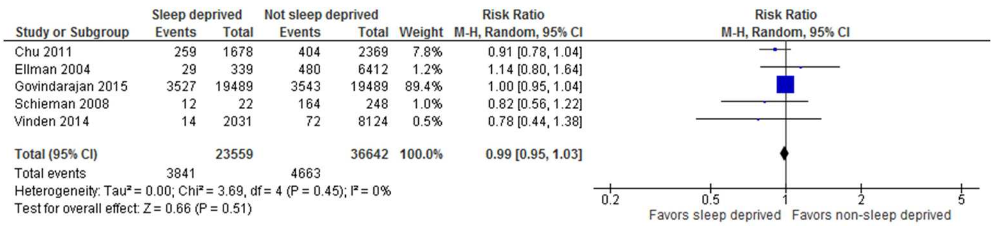


Figure 3. Forest plot for post-operative complications among surgeries performed by sleep deprived and non-sleep deprived surgeons

Legend: Vinden 2013 reported iatrogenic injuries; Schieman 2008, Govindarajan 2015, and Chu 2011 reported post-operative complication rate; Ellman 2004 reported post-operative complications (other types of complications reported not included in the analysis)

381x101mm (300 x 300 DPI)

Figure 4. Forest plot for patient mortality among surgeries performed by sleep deprived and non-sleep deprived surgeons

381x101mm (300 x 300 DPI)

Supplementary file 1. Search Strategy

Database: In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present

Date searched: 13 April 2016, updated 7 November 2017

Records retrieved: 5068 and 1442 in the update (removed duplicates retrieved in previous search)

1. Medical Staff, Hospital/
2. Physician Impairment/
3. exp Physicians/
4. allergist*.ti.
5. (an?esthetist* or an?esthesiologist*).ti.
6. cardiologist*.ti.
7. clinician*.ti.
8. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
9. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
10. dermatologist*.ti.
11. endocrinologist*.ti.
12. doctor*.ti.
13. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.
14. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.
15. family practitioner*.ti.
16. gastroenterologist*.ti.
17. (general practitioner* or GP*).ti.
18. (general adj2 physician*).ti.
19. geriatrician*.ti.
20. gyn?ecologist*.ti.

21. h?ematologist*.ti.

22. (health* adj2 (professional* or provider*)).ti.

23. hospitalist*.ti.

24. (house staff* or housestaff*).ti.

25. intensivist*.ti.

26. internist*.ti.

27. medical professional*.ti.

28. obstetrician*.ti.

29. oncologist*.ti.

30. ophthalmologist*.ti.

31. orthop?edist*.ti.

32. (otolaryngologist* or otorhinolaryngologist*).ti.

33. neonatologist*.ti.

34. nephrologist*.ti.

35. neurologist*.ti.

36. neuropsychiatrist*.ti.

37. neurosurgeon*.ti.

38. p?ediatrician*.ti.

39. perinatologist*.ti.

40. physician*.ti.

41. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

42. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw,kf.

43. primary care practitioner*.ti.

44. psychiatrist*.ti.

45. pulmonologist*.ti.

46. rheumatologist*.ti.

47. surgeon*.ti.

48. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw,kf.

49. traumatologist*.ti.

50. urologist*.ti.

51. or/1-50 [Combined MeSH, title, and text word searches for physicians]

52. Burnout, Professional/

53. exp Circadian Rhythm/

54. exp Fatigue/

55. Occupational Health/

56. Rest/ph, px [Physiology, Psychology]

57. Sleep Deprivation/

58. Sleep Disorders, Circadian Rhythm/

59. Sleep Wake Disorders/

60. exp Stress, Psychological/

61. Workload/px [Psychology]

62. Work Schedule Tolerance/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw,kf.

64. biological rhythm*.tw,kf.

65. (burn out* or burned out* or burnt out* or burnout*).tw,kf.

66. circadian misalignment.tw,kf.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw,kf.

68. exhaust*.tw,kf.

69. fatigu*.tw,kf.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw,kf.

71. tired*.tw,kf.

72. weariness.tw,kf.

73. or/52-72 [Combined MeSH and text words for fatigue]

74. and/51,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (comment or editorial or letter).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. remove duplicates from 80

Database: Ovid Embase 1996 to 2016 Week 15

Date searched: 13 April 2016

Records retrieved: 8859

1. medical staff/
2. exp physician/
3. allergist*.ti.
4. (an?esthetist* or an?esthesiologist*).ti.
5. cardiologist*.ti.
6. clinician*.ti.
7. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
8. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
9. dermatologist*.ti.
10. endocrinologist*.ti.
11. doctor*.ti.
12. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
13. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
14. family practitioner*.ti.
15. gastroenterologist*.ti.
16. (general practitioner* or GP*).ti.
17. (general adj2 physician*).ti.

18. geriatrician*.ti.

19. gyn?ecologist*.ti.

20. h?ematologist*.ti.

21. (health* adj2 (professional* or provider*)).ti.

22. hospitalist*.ti.

23. (house staff* or housestaff*).ti.

24. intensivist*.ti.

25. internist*.ti.

26. medical professional*.ti.

27. obstetrician*.ti.

28. oncologist*.ti.

29. ophthalmologist*.ti.

30. orthop?edist*.ti.

31. (otolaryngologist* or otorhinolaryngologist*).ti.

32. neonatologist*.ti.

33. nephrologist*.ti.

34. neurologist*.ti.

35. neuropsychiatrist*.ti.

36. neurosurgeon*.ti.

37. p?ediatrician*.ti.

38. perinatologist*.ti.

39. physician*.ti.

40. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

41. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

42. primary care practitioner*.ti.

43. psychiatrist*.ti.

44. pulmonologist*.ti.

45. rheumatologist*.ti.

46. surgeon*.ti.

47. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

48. traumatologist*.ti.

49. urologist*.ti.

50. or/1-49 [Combined Emtree, title, and text word searches for physicians]

51. burnout/

52. circadian rhythm/

53. circadian rhythm sleep disorder/

54. fatigue/

55. mental stress/

56. occupational health/

57. sleep deprivation/

58. sleep waking cycle/

59. work capacity/

60. work schedule/

61. working time/

62. workload/

63. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

64. biological rhythm*.tw.

65. (burn out* or burned out* or burnt out* or burnout*).tw.

66. circadian misalignment.tw.

67. ((circadian or diurnam or ultradian) adj rhythm*).tw.

68. exhaust*.tw.

69. fatigu*.tw.

70. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

71. tired*.tw.

72. weariness.tw.

73. or/51-72 [Combined Emtree and text words for fatigue]

74. and/50,73 [Combined concepts for physicians and fatigue]

75. animals/ not (animals/ and humans/)

76. 74 not 75

77. (conference* or editorial or letter or proceeding).pt.

78. 76 not 77

79. limit 78 to yr="2000-Current"

80. limit 79 to (english or french)

81. limit 80 to embase

Database: Ovid PsycINFO 1987 to April Week 1 2016

Date searched: 13 April 2016

Records retrieved: 2094

1. exp Physicians/
2. allergist*.ti.
3. (an?esthetist* or an?esthesiologist*).ti.
4. cardiologist*.ti.
5. clinician*.ti.
6. (clinician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
7. (clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
8. dermatologist*.ti.
9. endocrinologist*.ti.
10. doctor*.ti.
11. (doctor* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.
12. (doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.
13. family practitioner*.ti.
14. gastroenterologist*.ti.
15. (general practitioner* or GP*).ti.
16. (general adj2 physician*).ti.

17. geriatrician*.ti.

18. gyn?ecologist*.ti.

19. h?ematologist*.ti.

20. (health* adj2 (professional* or provider*)).ti.

21. hospitalist*.ti.

22. intensivist*.ti.

23. internist*.ti.

24. medical professional*.ti.

25. obstetrician*.ti.

26. oncologist*.ti.

27. ophthalmologist*.ti.

28. orthop?edist*.ti.

29. (otolaryngologist* or otorhinolaryngologist*).ti.

30. neonatologist*.ti.

31. nephrologist*.ti.

32. neurologist*.ti.

33. neuropsychiatrist*.ti.

34. neurosurgeon*.ti.

35. p?ediatrician*.ti.

36. perinatologist*.ti.

37. physician*.ti.

38. (physician* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

39. (physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)).tw.

40. primary care practitioner*.ti.

41. psychiatrist*.ti.

42. pulmonologist*.ti.

43. rheumatologist*.ti.

44. surgeon*.ti.

45. (surgeon* adj2 (absent* or burn out* or burnout* or coping or distress* or duty hour* or fatigue or health* or impair* or resilien* or satisfaction or sleep* or well being* or wellbeing* or wellness* or work* hour* or work life balance)).tw.

46. traumatologist*.ti.

47. urologist*.ti.

48. or/1-47 [Combined thesaurus, title, and text word searches for physicians]

49. Compassion Fatigue/

50. Fatigue/

51. Human Biological Rhythms/

52. Occupational Health/

53. Occupational Stress/

54. Sleep/

55. Sleepiness/

56. Working Conditions/

57. Work Rest Cycles/

58. Work Week Length/

59. Work Scheduling/

60. Workday Shifts/

61. ((24 hour* or 24 hr* or twenty four hour* or twentyfour hour*) adj rhythm*).tw.

62. biological rhythm*.tw.

63. (burn out* or burned out* or burnt out* or burnout*).tw.

64. circadian misalignment.tw.

65. ((circadian or diurnam or ultradian) adj rhythm*).tw.

66. exhaust*.tw.

67. fatigu*.tw.

68. (sleep* adj3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)).tw.

69. tired*.tw.

70. weariness.tw.

71. or/49-70 [Combined thesaurus and text words for fatigue]

72. and/48,71 [Combined concepts for physicians and fatigue]

73. limit 72 to yr="2000-Current"

74. limit 73 to (english or french)

Database: CINAHL Plus with Full Text (1937 to the present) via EBSCOhost

Date searched: 14 April 2016

Records retrieved: 3378

S1. (MH "Medical Staff, Hospital+")

S2. (MH "Physicians+")

S3. TI allertist*

S4. TI (anesthetist* or anaesthetist* or anesthesiologist* or anaesthesiologist*)

S5. TI cardiologist*

S6. TI clinician*

S7. clinician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S8. clinician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S9. TI dermatologist*

S10. TI endocrinologist*

S11. TI doctor*

S12. doctor* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S13. doctor* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S14. TI "family practitioner*"

S15. TI gastroenterologist*

S16. TI ("general practitioner*" or GP*)

S17. TI (general N2 physician*)

S18. TI geriatrician*

S19. TI (gynaecologist* or gynecologist*)

S20. TI (haematologist* or hematologist*)

S21. TI hospitalist*

S22. TI ("house staff*" or housestaff*)

S23. TI intensivist*

S24. TI internist*

S25. TI obstetrician*

S26. TI oncologist*

S27. TI ophthalmologist*

S28. TI (orthopaedist* or orthopedist*)

S29. TI (otolaryngologist* or otorhinolaryngologist*)

S30. TI neonatologist*

S31. TI nephrologist*

S32. TI neurologist*

S33. TI neuropsychiatrist*

S34. TI neurosurgeon*

S35. TI (paediatrician* OR pediatrician*)

S36. TI perinatologist*

S37. TI physician*

S38. physician* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or "work* hour*" or "work life balance")

S39. physician* and (cities or city or communit* or country* or frontier* or north* or remote or rural* or suburb* or urban*)

S40. TI "primary care practitioner*"

S41. TI psychiatrist*

S42. TI pulmonologist*

S43. TI rheumatologist*

S44. TI surgeon*

S45. surgeon* N2 (absent* or "burn out*" or burnout* or coping or distress* or "duty hour*" or fatigue or health* or impair* or resilien* or satisfaction or sleep* or "well being*" or wellbeing* or wellness* or work* hour* or "work life balance")

S46. TI traumatologist*

S47. TI urologist*

S48. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28

OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46 OR S47

S49. (MH "Circadian Rhythm")

S50. (MH "Fatigue")

S51. (MH "Impairment, Health Professional")

S52. (MH "Mental Fatigue")

S53. (MH "Occupational Health")

S54. (MH "Shiftwork")

S55. (MH "Sleep Deprivation")

S56. (MH "Sleep Disorders, Circadian Rhythm")

S57. (MH "Sleep-Wake Transition Disorders")

S58. (MH "Stress, Occupational+")

S59. (MH "Stress, Psychological")

S60. ("24 hour*" or "24 hr*" or "twenty four hour*" or "twentyfour hour*") N1 rhythm*

S61. "biological rhythm*"

S62. "burn out*" or "burned out*" or "burnt out*" or burnout*

S63. "circadian misalignment"

S64. (circadian or diurnam or ultradian) N1 rhythm*

S65. exhaust*

S66. fatigu*

S67. sleep* N3 (depriv* or disorder* or disrupt* or lack* or loss or insufficien* or problem*)

S68. tired*

S69. weariness

S70. S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65 OR S66 OR S67 OR S68 OR S69

S71. S48 AND S70

S72. S48 AND S70 Limiters - Published Date: 20000101-20161231; Publication Type: Clinical Trial, Journal Article, Meta Analysis, Meta Synthesis, Practice Guidelines, Randomized Controlled Trial, Research, Review, Systematic Review; Language: English, French

Database: PubMed via NCBI Entrez

Date searched: 14 April 2016

Records retrieved: 92

(((("Medical Staff, Hospital"[mh:noexp] OR "Physician Impairment"[mh:noexp] OR "Physicians"[mh] OR allergist[ti] OR allergists[ti] OR anaesthetist[ti] OR anaesthetists[ti] OR anaesthesiologist[ti] OR anaesthesiologists[ti] OR anesthetist[ti] OR anesthetists[ti] OR anesthesiologist[ti] OR anesthesiologists[ti] OR cardiologist[ti] OR cardiologists[ti] OR clinician[ti] OR clinicians[ti] OR ((clinician[tiab] OR clinicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((clinician[tiab] OR clinicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR dermatologist[ti] OR dermatologists[ti] OR endocrinologist[ti] OR endocrinologists[ti] OR doctor[ti] OR doctors[ti] OR ((doctor[tiab] OR doctors[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((doctor[tiab] OR doctors[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "family practitioner"[ti] OR

"family practitioners"[ti] OR gastroenterologist[ti] OR gastroenterologists[ti] OR "general practice physician"[ti] OR "general practice physicians"[ti] OR "general practitioner"[ti] OR "general practitioners"[ti] OR geriatrician[ti] OR geriatricians[ti] OR gynaecologist[ti] OR gynaecologists[ti] OR gynecologist[ti] OR gynecologists[ti] OR haematologist[ti] OR haematologists[ti] OR hematologist[ti] OR hematologists[ti] OR "health care professional"[ti] OR "health care professionals"[ti] AND "health care provider"[ti] OR "health care providers" OR "health professional"[ti] OR "health professionals"[ti] OR "health provider"[ti] OR "health providers"[ti] OR "healthcare professional"[ti] OR "healthcare professionals"[ti] OR "healthcare provider"[ti] OR "healthcare providers"[ti] OR hospitalist[ti] OR hospitalists[ti] OR "house staff"[ti] OR "house staffs"[ti] OR housestaff[ti] OR housestaffs[ti] OR intensivist[ti] OR intensivists[ti] OR internist[ti] OR internists[ti] OR "medical professional"[ti] OR "medical professionals"[ti] OR obstetrician[ti] OR obstetricians[ti] OR oncologist[ti] OR oncologists[ti] OR ophthalmologist[ti] OR ophthalmologists[ti] OR orthopaedist[ti] OR orthopaedists[ti] OR orthopedist[ti] OR orthopedists[ti] OR otolaryngologist[ti] OR otolaryngologists[ti] OR otorhinolaryngologist[ti] OR otorhinolaryngologists[ti] OR neonatologist[ti] OR neonatologists[ti] OR nephrologist[ti] OR nephrologists[ti] OR neurologist[ti] OR neurologists[ti] OR neuropsychiatrist[ti] OR neuropsychiatrists[ti] OR neurosurgeon[ti] OR neurosurgeons[ti] OR paediatrician[ti] OR paediatricians[ti] OR pediatrician[ti] OR pediatricians[ti] OR perinatologist[ti] OR perinatologists[ti] OR physician[ti] OR physicians[ti] OR ((physician[tiab] OR physicians[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR "burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR ((physician[tiab] OR physicians[tiab]) AND (cities[tiab] OR city[tiab] OR community[tiab] OR communities[tiab] OR country[tiab] OR countryside[tiab] OR frontier[tiab] OR north[tiab] OR northern[tiab] OR remote[tiab] OR rural[tiab] OR suburb[tiab] OR suburbs[tiab] OR suburban[tiab] OR urban[tiab] OR urbanite[tiab])) OR "primary care practitioner"[ti] OR "primary care practitioners"[ti] OR psychiatrist[ti] OR psychiatrists[ti] OR pulmonologist[ti] OR pulmonologists[ti] OR rheumatologist[ti] OR rheumatologists[ti] OR surgeon[ti] OR surgeons[ti] OR ((surgeon[tiab] OR surgeons[tiab]) AND (absent[tiab] OR absentee[tiab] OR absenteeism[tiab] OR absenteeisms[tiab] OR

"burned out"[tiab] OR "burn out"[tiab] OR "burn outs"[tiab] OR burnout[tiab] OR burnouts[tiab] OR "burnt out"[tiab] OR coping[tiab] OR distress[tiab] OR distressed[tiab] OR distresses[tiab] OR distressing[tiab] OR "duty hour"[tiab] OR "duty hours"[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR health[tiab] OR healthy[tiab] OR impair[tiab] OR impaired[tiab] OR impairment[tiab] OR resilience[tiab] OR resiliency[tiab] OR resilient[tiab] OR satisfaction[tiab] OR sleep[tiab] OR sleepiness[tiab] OR sleeps[tiab] OR "well being"[tiab] OR wellbeing[tiab] OR wellness[tiab] OR "work hours"[tiab] OR "working hours"[tiab] OR "work life balance"[tiab])) OR traumatologist[ti] OR traumatologists[ti] OR urologist[ti] OR urologists[ti]) AND ("Burnout, Professional"[mh:noexp] OR "Circadian Rhythm"[mh] OR "Fatigue"[mh] OR "Occupational Health"[mh:noexp] OR "Rest/physiology"[mh:noexp] OR "Rest/psychology"[mh:noexp] OR "Sleep Deprivation"[mh:noexp] OR "Sleep Disorders, Circadian Rhythm"[mh:noexp] OR "Stress, Psychological"[mh] OR "Workload/psychology"[mh] OR "Work Schedule Tolerance"[mh:noexp] OR "24 hour rhythm"[tiab] OR "24 hour rhythms"[tiab] OR "24 hr rhythm"[tiab] OR "24 hr rhythms"[tiab] OR alertness[tiab] OR "biological rhythm"[tiab] OR "biological rhythms"[tiab] OR "burn out"[tiab] OR "burned out"[tiab] OR "burnt out"[tiab] OR burnout[tiab] OR "circadian misalignment"[tiab] OR "circadian rhythm"[tiab] OR "circadian rhythms"[tiab] OR "diurnal rhythm"[tiab] OR "diurnal rhythms"[tiab] OR exhausted[tiab] OR exhaustion[tiab] OR exhausting[tiab] OR exhausts[tiab] OR fatigue[tiab] OR fatigued[tiab] OR fatigues[tiab] OR fatiguing[tiab] OR (("Sleep"[mh:noexp] OR sleep[tiab] OR sleeping[tiab]) AND (deprivation[tiab] OR deprive[tiab] OR deprived[tiab] OR deprives[tiab] OR depriving[tiab] OR disorder[tiab] OR disorders[tiab] OR lack[tiab] OR lacked[tiab] OR lacking[tiab] OR lacks[tiab] OR loss[tiab] AND insufficient[tiab] OR problem[tiab] OR problems[tiab])) OR tired[tiab] OR tiredness[tiab] OR "twenty four hour rhythm"[tiab] OR "twenty four hour rhythms"[tiab] OR weariness[tiab] OR "ultradian rhythm"[tiab] OR "ultradian rhythms"[tiab])) NOT (((Animals[MESH] OR Animal Experimentation[MESH] OR "Models, Animal"[MESH] OR Vertebrates[MESH]) NOT (Humans[MESH] OR Human experimentation[MESH])) OR (((animals[tiab] OR animal model[tiab] OR rat[tiab] OR rats[tiab] OR mouse[tiab] OR mice[tiab] OR rabbit[tiab] OR rabbits[tiab] OR pig[tiab] OR pigs[tiab] OR porcine[tiab] OR swine[tiab] OR dog[tiab] OR dogs[tiab] OR hamster[tiab] OR hamsters[tiab] OR chicken[tiab] OR chickens[tiab] OR sheep[tiab]) AND (publisher[sb] OR inprocess[sb] OR pubmednotmedline[sb])) NOT (human[ti] OR humans[ti] OR people[ti] OR children[ti] OR adults[ti] OR seniors[ti] OR patient[ti] OR patients[ti])))) NOT (editorial[pt] OR comment[pt] OR letter[pt] OR newspaper article[pt])) AND ((publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR (pubstatUSheadofprint))

Filters activated: Publication date from 2000/01/01 to 2016/12/31, English, French.

Supplementary table 1. Descriptive characteristics of the included studies

Study	Physician and patient characteristics				Setting		Interventions or	Outcomes
Country	Type	n=	Sex (% male)	Age	Location	Urban or rural	exposures	
Observational (exposure) studies (n=45)								
Cohort design								
Chu, 2011 [32] Canada	Surgeons	6	NR	Range: 32-55y	Tertiary care academic hospital	Urban	Insufficient sleep due to work on the night preceding surgery	Length of surgery; patient postoperative mortality, complications, length of stay
	Patients: cardiac surgery cases	4,047	NR	NR				
Ellman, 2004 [41] US	Surgeons	NR	NR	NR	University hospitals	Urban	Insufficient sleep due to work on the night preceding surgery	Length of surgery; patient complications, in-hospital mortality, length of stay, need for blood products
	Patients: adult cardiac surgery cases	6,751	70%	S: 63.4±0.7y C: 63.5±0.1y				
Govindarajan, 2015 [31] Canada	Surgeons	1,448	NR	46.3±8.7	Academic and non-academic hospitals	Mixed	Sleep deprivation due to work on the night preceding a daytime surgery	Length of surgery; Patient complications, mortality, readmissions, length of stay
	Patients: surgical cases	38,978	NR	56.4±16.6y				
Rothschild, 2009 [58] US	Surgeons Obstetrician/gynecologists	220	Surgeons: 84% OB/GYNs: 28%	Surgeons: 42.0±7.6y OB/GYNs: 42.0±9.0y	Tertiary care academic trauma centre/referral centre for high-risk obstetrics	Urban	Sleep deprivation due to work on the night preceding a daytime procedure	Patient complications, preventable complications
	Patients: surgical and obstetrics cases	Surg.: 4,471 Obst.: 4,902	Surg: S: 25% C: 28% Obst.: S: 0% C: 0%	Surg: S: 49.1±16.3y C: 50.0±16.3y Obst.: S: 32.9±5.2y C: 33.5±5.0y				
Schieman, 2007 [63] Canada	Colorectal surgeons	NR	NR	NR	University teaching hospitals	NR	Fatigue due to work on the night preceding surgery	Length of surgery; patient operative complications, length of stay, mortality, cancer recurrence
	Patients: undergoing anterior resection for rectal cancer	270	NR	S: 64.5y C: 64.4y				

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Vinden, 2014 [77] Canada	General surgeons	331	83%	48±10y	Community hospitals	Mixed	Sleep deprivation due to overnight work preceding daytime surgery	Patient mortality, operative complications
	Patients: Elective cholecystectomies	10,390	S: 27% C: 26%	S: 49±16y C: 49±16y				
Before-after design								
Amirian, 2014 [34] Denmark	Surgeons	29	55%	Median: 35y Range: 27-49y	Academic hospital	Urban	17-h night shift with sleep deprivation	Cognitive and psychomotor abilities on a laparoscopic simulation
Gerdes, 2008 [45] US	Surgeons	9	NR	NR	University Hospital	Urban	Fatigue; sleep deprivation from overnight call shift	Cognitive and psychomotor abilities
Lederer, 2006 [50] Austria	Senior anesthetists	11	82%	49.0±2.0y	Hospital	Urban	Sleep deprivation from 24-h call shift	Concentration ability; reaction time; performance on psychometric tasks
Time series design								
Leichtfried, 2011 [51] Austria	Anesthetists	10	100%	Mean: 32y Range: 29-35y	University Hospital	Urban	Sleep deprivation from 24-h shift; sleepiness, sleep hours	Melatonin metabolite profile
Cross-sectional design								
Aziz, 2004 [35] US	Family medicine physicians Various specialties	153	NR	NR	Hospitals	NR	Fatigue	Stress
Beaujouan, 2005 [36] France	Anesthesiologists	3,476	64%	≤35y: 9% 36-45y: 28% 46-55y: 49% 56-65y: 13%	Public sector General hospitals University hospitals Private hospitals	NR	Sleep deprivation	Substance abuse
Chang, 2013 [37] US	Anesthesiologists	11	64%	Mean: 38y IQR: 34-48y	Level 1 trauma centre	NR	Sleep deprivation due to 15-h overnight call shift; sleepiness	Cognitive performance; reaction time

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Chen, 2008 [38] US	Psychiatrists Internists General practitioners Surgeons Obstetrician-gynecologists Radiologists Pediatricians Other	180	77%	Academic: 79% 36-55y Private practice: 73% 36-65y	Medical school Private practices	Urban	Sleep deprivation or sleepiness	Impact on personal and professional life; perceived risk of errors
Doppia, 2011 [39] France	Anesthesiologists	565	64%	<35y: 11% 35-54y: 63% >55y: 25%	Public hospitals Private hospitals Work-health environments Public health units	NR	Sleep deprivation	Burnout
Elovaino, 2015 [42] Finland	Physicians in various specialties	1,524	40%	Median: 49.7y Range: 24-69y	Hospitals Primary care Private practice Other unspecified	NR	Sleep difficulties	Job demands and control
Gander, 2000 [43] New Zealand	Anesthetists	183	NR	Mean: 46y	Combined public/private practice Other unspecified	NR	Work-related sleep disturbance	Risk of fatigue-related errors
Harbeck, 2015 [46] Germany	Internists	20	45%	Median: 32y Range: 26-42y	Hospital	NR	Sleep disturbance due to a 24-call shift	Biochemical and physiological parameters; neurocognitive function
Heponiemi, 2014 [47] Finland	Physicians in various specialties Non-specialized physicians	1,541	40%	49.80±9.49y, Range: 24-67y	Hospitals Primary care clinic Private practice Other unspecified	NR	Sleep difficulties	Job satisfaction; work ability; psychological distress

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Jackson, 2017 [48] US	Surgeons in various subspecialties	993	61%	More; less satisfied: 30-39y: 23%;24% 40-49y: 32%;36% 50-59y: 23%;27% ≥60y: 23%;14%	Academic practice Non-academic practice	NR	Not feeling well rested	Job satisfaction
Kanieta, 2011 [49] Japan	Internists Surgeons Orthopedics Pediatricians Obstetrician-gynecologists Psychiatrists Dermatologists Urologists Opthalmologists Otorhinolaryngologists Other	3,486	66%	20-39y: 11% 40-49y: 25% 50-59y: 28% 60-69y: 16% ≥70y: 21%	Hospitals Clinics Other unspecified	NR	Sleep deprivation and difficulties; insomnia	Medical incidents
Lindfors, 2006 [52] Finland	Anesthetists	328	53%	47±7.8y Range: 32-69y	University hospitals Central and district hospitals Private sector	NR	Sleep disturbances; sleepiness	Stress; suicidal tendencies
Mahmood, 2016 [53] Norway	Generalists Internists Pediatricians Surgical specialties Anesthesiologists	450 (all time points)	41%	43y±2.8y	Public health system Private practice	NR	Sleep deprivation due to on-call shifts	Alcohol misuse

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Nishimura, 2014 [54] Japan	Neurosurgeons and neurologists	2,564	NR	NR	Stroke care centres Teaching hospitals	NR	Sleep deprivation	Burnout
Pit, 2014 [55] Australia	General practitioners	92	60%	50±10.7y	NR	Rural	Work-related sleep disturbance	Early retirement intentions
Pit, 2016 [56] Australia	General practitioners	92	60%	50±10.7y	Private (solo) practice Group practice	Rural	Work-related sleep disturbance	Sickness presenteeism
Roberts, 2014 [57] US	General internists Internal medicine hospitalists	578	58%	Hospitalists: 46.9±12.4y Generalists: 53.6±10.2y	Private practice Academic medical centre Veterans hospital Military practice Other	NR	Fatigue	Falling asleep while driving
Saadat, 2016 [60] US	Anesthesiologists	21	71%	30-40y: 57% 41-50y: 19% 51-55y: 24% Range: 32-56y	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Mood disturbances
Saadat, 2017 [59] US	Anesthesiologists	21	65%	Range: 32-56 years	Tertiary care academic children's hospital	NR	Sleep deprivation due to 17-h night call shift	Reaction time
Sanches, 2015 [61] Spain	Emergency medicine physicians	18	28%	29.2±2.6y	Central hospital	NR	Sleep deprivation	Cognitive and psychomotor abilities
Sargent, 2009 [62] US	Orthopedic surgeons	264	92%	NR	Orthopedic surgery training programs	NR	Sleep deprivation	Burnout; psychological distress; marital satisfaction
Sende, 2012 [64] France	Emergency physicians	318	62%	39±8y	Hospitals Mobile emergency services Other unspecified	NR	Fatigue; sleep deprivation	Stress

Study Country	Physician and patient characteristics Type	n=	Sex (% male)	Age	Setting Location	Urban or rural	Interventions or exposures	Outcomes
Sexton, 2001 [65] US	Consulting physicians: Surgeons Anesthesiologists Pulmonary physicians Cardiologists Pediatricians	271	NR	NR	Teaching and non-teaching hospitals	Urban	Fatigue	Perceived performance effectiveness
Shanafelt, 2005 [67] US, Canada, Mexico	Oncologists	241	85%	>50y: 51%	Community clinics Hospitals Private practice Academic medical centres	NR	Fatigue; sleep deprivation	Quality of life/well-being
Shanafelt, 2010 [66] US	Surgeons	7,905	87%	Median: 51y Q1: 43y Q2: 59y	Private practice Academic medical centres Veterans hospital Active military practice Retired or not in practice Other	NR	Fatigue	Perceived major medical errors
Shanafelt, 2014 [68] US	Oncologists	1,117	52%	Median: 52y	Private practice Academic practice Veteran's hospital Industry, other	NR	Fatigue	Satisfaction with work-life balance
Shirom, 2006 [69] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Perception of quality of patient care

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Shirom, 2010 [70] Israel	Opthalmologists Dermatologists Otolaryngologists Gynecologists General surgeons Cardiologists	890	80%	Median: 52y SD: 7.2y	Community clinics Acute care hospital outpatient clinics	NR	Physical fatigue	Burnout
Smith, 2017 [71] UK	General practitioners Surgeons Other unspecified specialties	3,550	63%	NR	NR (varied)	NR	Perceived fatigue; sleep deprivation	Physical and mental health; competence
Starmer, 2016 [72] US	General pediatricians Pediatric surgeons Pediatric hospitalists Pediatric specialists (unspecified)	840	40%	NR	NR (some in private practice)	NR	Sleep deprivation	Burnout; balanced personal and professional commitments; life and career satisfaction
Tanti, 2017 [73] Malta	Physicians (unspecified)	204	62%	Median: 41y	Hospitals Community Office-based	NR	Fatigue	Prescribing errors
Tokuda, 2009 [74] Japan	Hospital physicians: Generalists Other unspecified specialties	236	75%	40.9±7.8y Range: 26-76y	Hospitals with ≥20 inpatient beds	NR	Sleep deprivation	Burnout; job satisfaction
Vela-Bueno, 2008 [76] Spain	Primary care physicians	113	27%	41.4±8.0y	Primary care centres	Urban	Sleep problems, insomnia	Burnout
Wada, 2010 [78] Japan	Physicians (unspecified)	3,862	78%	M: 75% 30-59y F: 85% 30-59y	Hospitals	NR	Sleep deprivation	Depressive symptoms

Study Country	Physician and patient characteristics				Setting		Interventions or exposures	Outcomes
	Type	n=	Sex (% male)	Age	Location	Urban or rural		
Non-comparative design								
Gander, 2008 [43] New Zealand	Anesthetists	20	85%	Median: 44y	Hospitals	Urban	Sleep disturbance from consecutive working days on-call work	Psychomotor performance
Intervention studies (n=2)								
Randomized controlled trials								
Dutheil, 2013 [40] France	Emergency physicians	17	35%	39.1y±6.9y	University hospital	Urban	Fatigue related to 14-h and 24-h shifts; sleep deprivation, low sleep quality;	Perceived stress; urine interleukine-8
Uchal, 2005 [75] Norway	Surgeons Gynecologists Orthopedic surgeons Urologists Vascular surgeons	64	67%	Median: Post-call: 33.0y Post-work: 38.0y	Government hospitals	NR	Sleep deprivation due to 24-h call shift	Product quality, procedure effectiveness of a surgical simulation

C: control group; F: female; h: hour(s); IQR: interquartile range; M: male; NR: not reported; S: study group; SD: standard deviation; Surg: surgical; Obst: obstetric; Q: quartile; UK: United Kingdom; US: United States of America; y: year(s)

Supplementary file 3. Risk of bias assessments

Summary of risk of bias assessments for randomized controlled trials (n=2)[a]

First Author, Year	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias	Overall risk of bias[b]
Dutheil, 2013	Low	Unclear	High	High	Low	Low	High	High
Uchal, 2005	Low	Low	Unclear	Low	Low	Low	Low	Unclear

[a]Assessed using the Cochrane Collaboration's Risk of Bias Tool

[b]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cohort studies (n=6)[a]

First Author, Year	Selection					Comparability		Outcome				Total Score[b] /9
	Representa-tiveness of exposed cohort /1	Selection of non-exposed cohort /1	Ascertain-ment of exposure /1	Outcome not present at start /1	Total /4	Compara-bility of cohorts /2	Total /2	Assess-ment of outcome /1	Adequate length of follow-up /1	Adequate follow-up of cohorts /1	Total /1	
Chu, 2011	1	1	0	1	3	2	2	1	1	1	3	8
Ellman, 2004	1	1	1	1	4	1	1	1	1	1	3	8
Govindarajan, 2015	1	1	1	1	4	2	2	1	1	1	3	9
Rothschild, 2009	1	1	1	1	4	2	2	1	1	1	3	9
Schieman, 2008	1	1	1	1	4	1	1	1	1	1	3	8
Vinden, 2014	1	1	1	1	4	1	1	1	1	1	3	8

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale

[b]An overall score of 7 to 9 stars is considered as low risk of bias, 4 to 6 as unclear risk of bias, and 3 or less as high risk of bias

Summary of risk of bias assessments for before-after studies (n=3)[a]

First Author, Year	Random sequence generation[b]	Allocation concealment[b]	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Amirian, 2014	NA	NA	High	High	Low	Low	High	High
Gerdes, 2008	NA	NA	High	High	Low	Low	High	High
Lederer, 2006	NA	NA	High	High	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for before-after studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as 'not applicable' (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of risk of bias assessments for time series studies (n=1)[a]

First Author, Year	Intervention independent of other changes	Intervention effect pre-specified	Intervention unlikely to affect data collection	Allocation concealment[a]	Incomplete outcome data	Selective reporting	Other sources of bias[c]	Overall risk of bias[d]
Leitchfried, 2011	Low	High	Low	NA	Low	Low	High	High

[a]Assessed using Cochrane Effective Practice and Organization of Care (EPOC) Review Group's criteria for interrupted time series studies, adapted from the Cochrane Collaboration Risk of Bias Tool

[b]Assessed as not applicable (NA) when the studies did not include a control group

[c]Assessed as High due to lack of a control group

[d]Overall risk of bias is Low if all domains are rated as low, High if at least one domain is assessed as high, and Unclear if at least one domain is assessed as unclear and no domains are assessed as high

Summary of quality assessments for cross-sectional studies (n=34)[a]

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Aziz, 2004	0	0	0	0	1	0	1	1
Beaujouan, 2005	1	0	1	0	1	0	1	2
Chang, 2013	1	0	1	0	1	1	2	3
Chen, 2008	1	0	1	0	1	0	1	2
Doppia, 2011	1	1	2	0	1	1	2	4
Elovaino, 2015	1	1	2	0	1	1	2	4
Gander, 2000	1	1	2	0	1	1	2	4
Harbeck, 2015	1	0	1	0	1	1	2	3
Heponiemi, 2014	1	1	2	0	1	1	2	4
Jackson, 2017	0	0	0	0	1	0	1	1
Kanieta, 2011	1	0	1	0	1	1	2	3
Lindfors, 2006	1	1	2	0	1	1	2	4
Mahmood, 2017	1	0	1	0	1	0	1	2
Nishimura, 2014	1	1	2	0	1	0	1	3
Pit, 2014	1	0	1	0	1	1	2	3
Pit, 2016	1	0	1	0	1	1	2	3
Roberts, 2014	1	1	2	0	1	0	1	3
Saadat, 2016	1	1	2	0	1	1	2	4
Saadat, 2017	1	1	2	0	1	1	2	4
Sanches, 2015	1	0	1	0	1	0	1	2
Sargent, 2009	1	0	1	0	1	0	1	2

First Author, Year	Selection			Outcome				Total Score[b] /5
	Adequacy of case definition /1	Representative-ness of the sample /1	Total /2	Assessment of outcome /1	Same method of ascertainment for entire sample /1	Response rate /1	Total /3	
Sende, 2010	1	0	1	0	1	0	1	2
Sexton, 2001	1	0	1	0	1	0	1	2
Shanafelt, 2005	1	0	1	0	1	1	2	3
Shanafelt, 2010	1	1	2	0	1	0	1	3
Shanafelt, 2014	1	0	1	0	1	1	2	3
Shirom, 2006	1	1	2	0	1	1	2	4
Shirom, 2010	1	1	2	0	1	1	2	4
Smith, 2016	1	0	1	0	1	1	2	3
Starmer, 2016	1	1	2	0	1	1	2	4
Tanti, 2017	1	0	1	0	1	0	1	2
Tokuda, 2009	1	1	2	0	1	1	2	4
Vela-Bueno, 2008	1	1	2	0	1	1	2	4
Wada, 2010	1	1	2	0	1	0	1	3

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted for cross-sectional studies

[b]An overall score of 4 to 5 stars is considered as low risk of bias, 3 as unclear risk of bias, and 2 or less as high risk of bias. For response rate, ≥50% was used as the criterion to be awarded a star

Summary of quality assessments for non-comparative studies (n=1)[a]

First Author, Year	Selection			Exposure		Outcome				Total Score[b] /6
	Adequacy of case definition /1	Representat-iveness of the sample /1	Total /2	Ascertain-ment of exposure	Total /1	Assessment of outcome /1	Same method of assessment for entire sample /1	Loss to follow-up /1	Total /3	
Gander, 2008	1	1	2	0	0	0	1	1	2	4

[a]Assessed using the Newcastle-Ottawa Quality Assessment Scale, adapted by the authors to be suitable to the non-comparative design

[b]An overall score of 5 to 6 stars is considered as low risk of bias, 3 to 4 as unclear risk of bias, and 2 or less as high risk of bias

Supplementary file 4. Detailed study outcomes

Physician health and wellness outcomes and associations with fatigue

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Exposures or interventions: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Jackson, 2017 RoB: high	CS	Not feeling well rested: self-reported as 'unhealthy' Time points NR	71% healthy, 28% unhealthy in terms of being well rested	Job satisfaction: Abridged Job in General Scale; grouped into more or less satisfied using the median Time points NR	Job satisfaction in those more vs. less satisfied: Healthy (well rested): 85% vs. 58%, p<0001; Unhealthy (not well rested): 15% vs. 42%, p<0.001.
Nishimura, 2014 RoB: unclear	CS	Sleep hours/night: self-reported (continuous) Time points NR	Mean±SD sleep: 5.94±1.08h	Burnout: Japanese MBI (severe: EE >4.0 and either DP >2.6 or PE <4.17) Time points NR	1) Mean±SD sleep for not burned out vs. mild to moderate vs. severe: 6.07±1.15 vs. 5.88±0.94 vs. 5.63±0.94, p<0.05. 2) Association between sleep and burnout (OR (95% CI)): bivariate 0.67 (0.61-0.73), p<0.001; multivariate including work characteristics and mental health: 0.84 (0.75-0.94), p=0.002.
Sargent, 2009 RoB: high	CS	Sleep deprivation: self-reported on a 4-point scale (none, a little, quite a bit, a lot) Time points NR	21% none, 48% a little, 23% quite a bit, 8% a lot	Burnout: MBI (norms NR); Marital satisfaction: RDAS; Psychological morbidity: GHQ-12 score ≥4 Time points NR	1) Positive correlation between sleep deprivation and EE, DP, psychological distress, lower marital satisfaction, all p<0.001. No relationship with PA.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift with on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Stress during duty: 4-point scale from 'calm' to 'very demanding' Assessed post-duty	1) Mean stress score during duty: 2.1.

Study Risk of Bias (RoB)	Study design	Exposures or interventions		Outcomes	Associations between exposure and outcome
		Assessment measure and time points	Baseline	Assessment measure and time points	
Leitchtfried, 2011 RoB: high	TS	24-h shift; Sleepiness: ESS (range: 0-24); Sleep hours: self-reported (continuous) Sleepiness assessed pre-shit, sleep hours pre, during and post-shift	ESS (mean (range)): 7.4 (4-12); Mean±SD sleep hours: 1) pre-study: 7.74±1.35h; 2) Pre-24-h shift (11h00 on day 1: 0.13±0.35h, 19:00 on day 1: 6.99±0.68h); 3) During the 24-h shift (07h00 on day 2: 0.0±0.0h, 19h00 on day 2, 5.49±1.95h); 4) Post-24-h shift (11h00 on day 3: 0.5±0.71h, 19h00 on day 3: 7.06±1.18h).	aMT6-s: urinalysis Assessed at 4-h intervals from 07:00 to 11:00	1) aMT6-s over shift, mean (95% CI): higher at 11:00AM pre (12.2 (6.3-8.1)) and post-shift (9.3 (3.7-14.9)) vs during, p=0.016; 2) Correlations between sleep and aMT6-s (data NR): mild for sleep duration the night prior with aMT6-s at 3PM the following day; sleep on night 2 with aMT6-s at 3PM the next day; total sleep with aMT6-s at 11AM on third day; moderate for sleep on first night with aMT6-s at 7AM and 11AM pre-shift, 11PM during 24-h shift and 11AM post-shift; total sleep pre-shift and nocturnal sleep during 24-h shift with aMT6-s at 11PM during shift; total sleep with aMT6-s at 3PM on first and second day, 11PM on second day; 3) Correlations between ESS and aMT6-s: moderate for aMT6-s at 7AM during shift, 11AM on day off.
Beaujouan, 2005 RoB: high	CS	Sleep deprivation: 4-point scale (always, frequently, rarely, never) Time points NR	48.8% always or frequently feel sleep deprived	Substance abuse: 93-item addiction and substance abuse questionnaire Time points NR	1) 60.6% with drug dependence vs. 46.0% of those without reported sleep difficulties, p<0.001. 2) OR (95% CI) of addiction for frequently/always vs. rarely/never sleep deprived: tobacco 1.42 (1.04-1.94); tranquilizer/hypnotics 3.26 (2.12-5.02).
Doppia, 2011 RoB: low	CS	Insufficient sleep: 4-point scale (no, not really, sort of, yes) Time points NR	28.9% reported insufficient sleep during work time	Burnout: CBI (mild: 1-2.4, moderate: 2.5-3.5, severe: 3.6-5) Time points NR	1) Frequency of burnout by response for sleep sufficiency: 40.6% for no/not really, 16.3% for sort of/yes, p<0.001.
Lindfors, 2006 RoB: low	CS	Sleep hours/day: self-reported to the nearest 0.5h; Adequacy of sleep and rest: self-reported (yes/no)	Sleep hours (mean (range)): 7 (5-9)	Stress: MOSQ on a 3-point scale (no, to some extent, clearly); Thoughts of suicide: 4-point scale ('never' to 'have tried')	1) Sleep sufficiency predicted stress symptoms: bivariate β=-0.362, p<0.001; multivariate including gender, sick leave, suicide β=-0.269, p<0.001; 2) Sleep disturbance associated with thoughts of suicide, p=0.009.

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Exposures or interventions: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Time points NR	
Saadat, 2015 RoB: low	CS	Sleep deprivation (<7h/24-h) due to 17-h overnight shift; Sleepiness and alertness: VAS from 0 (not at all) to 100 (extremely) All assessed on a regular day and a post-call day	Mean±SD sleepiness on a regular day vs. post-call day: 2.99±2.18 vs. 6.79±2.30, p<0.001	Simple cognitive tests: VAS from 0 (not at all) to 100 (extremely); Mood disturbance: PMS (scoring NR) All assessed on a regular day and a post-call day	Regular day vs. post-call day, mean±SD scores: 1) Simple cognitive tests: energetic 6.04±2.27 vs. 2.53±1.87, confident 7.03±1.83 vs. 4.98±2.29, irritable 2.03±1.04 vs. 4.86±2.16, sleepy 2.99±2.18 vs. 6.79±2.30, talkative 4.46±1.74 vs. 2.41±1.97, all p<0.001; jittery 1.94±1.74 vs. 3.12±2.34, p=0.003; anxiousness ns; 2) PMS: tension 13.48±2.71 vs. 15.43±4.46, p=0.049; anger 15.24±4.41 vs. 18.14±5.92, p=0.005; fatigue 10.14±2.63 vs. 20.05±6.87, p<0.001; confusion 10.57±1.69 vs. 12.57±4.24, p=0.025; vigor 24.95±6.75 vs.16.67±5.70, p<0.001; depression: ns; total mood disturbance: 42.57±15.26 vs. 76.90±6.91, p<0.001.
ER or ICU physicians					
Dutheil, 2013 RoB: high	RCT	14-h or 24-h shift; Sleep hours: self-reported sleep and wake time; Sleep quality: VAS from 1 (low) to 100 (high); Mental and physical fatigue: VAS from 1 (low) to 100 (high) Assessed on day prior to shift; during shift; each day of protocol (work, off, clerical, control)	1) Sleep duration and quality lower during shifts (14h and 24h) than any other day, and lower during the 24-h vs. 14-h shift (p<0.05); 2) Mental and physical fatigue higher after 14-h and 24-h shift vs. control day (data NR).	Stress: VAS from 0 (low) to 100 (high); IL-8: urinalysis Assessed at 08:30 and 18:30 on each day of protocol	1) Stress: higher following 14-h and 24-h shifts vs. the control day, p<0.05 (data NR); 2) IL-8: higher following 24-h shift vs. control (p=0.007) and 14-h shift (p=0.015); ns difference between 14-h shift and control day; 3) Correlations with IL-8: sleep hours pre-24-h shift, r=-0.627, p=0.007; poor sleep quality during 14-h and 24-h shifts, r=0.452, p=0.031; 4) Multivariable regression: 24-h shift increased IL-8 by 1.9ng vs. control day, p=0.007; ns association with 14-h shift, mental or physical fatigue, sleep deprivation, 14-h shift.
Sende, 2012 RoB: high	CS	Fatigue and sleep deprivation as sources of stress	NR	Most important sources of stress among 4 categories (work-related, patient-	1) 78% indicated that sleep loss and fatigue were sources of stress.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR		related, organizational, individual) Time points NR	
Generalists[b]					
Harbeck, 2015 RoB: unclear	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Biochemical (laboratory values) and physiological (heart rate variability, skin resistance, blood pressure) stress parameters Assessed before a normal day shift, and after a 24-h on call shift	Before a normal shift vs. after overnight call shift: 1) Biochemical parameters: no changes in any parameter except for thyroid stimulating hormone which was higher after the on-call shift (p = 0.049, data NR); 2) Physiological parameters: no significant changes in any parameters
Pit, 2014 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day' Time points NR	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Early retirement (<65 years) intentions (yes/no) Time points NR	For sleep disturbance a few times a year to every day vs. never: 1) Intention to retire early: 74% vs. 26%, p<0.01; 2) Association with intention to retire early (OR (95% CI)): univariate 3.6 (1.47-8.80), p<0.01; multivariate including work, occupational, individual factors 2.91 (1.11-7.6), p<0.05; 4) RR (95% CI) for intention to retire early: 2.0 (1.18-3.49); attributable fraction: 50.0%; population attributable fraction: 37.1%.
Pit, 2016 RoB: unclear	CS	Work-related sleep disturbance: 7-point scale from 'never' to 'every day'	Work-related sleep disturbance: 41% never, 59% a few times a year to every day	Sickness presenteeism: 'yes' response indicated 1 or more days	For sleep disturbance a few times a year to every day vs. never: 1) Sickness presenteeism: 32% vs. 68%, p=0.018;

Study Risk of Bias (RoB)	Study design	Exposures or interventions: Assessment measure and time points	Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
		Time points NR		Assessed for the past 12 months	2) Association with sickness presenteeism (OR (95% CI)): 2.92 (1.19-7.16), p=0.02.
Roberts, 2014 RoB: unclear	CS	Fatigue: LAS from 0 (low) to 10 (high) Assessed for the past week	Mean (SD) score: 5.8 (2.4) for hospitalists; 5.9 (2.4) for general internists	Impact of fatigue on daily activities (falling asleep while driving) (yes/no) Time points NR	1) 8.7% of hospitalists and 4.3% of outpatient general internists had fallen asleep while driving due to fatigue.
Vela-Bueno, 2008 RoB: low	CS	Sleep Quality: PSQI (Spanish): score ≥5 indicates low quality (range; 0 to 21); Insomnia: DSM-IV criteria Time points NR; insomnia symptoms in past month	Prevalence (% (95% CI)): 1) Sleep-onset latency >30 minutes: 8.4 (4.8-11.9); 2) Wake time after sleep onset >30 minutes: 15.4 (10.8-19.9); 3) Early morning awakening: 22.5 (19.5-30.4); 4) Nonrestorative sleep: 22.5 (17.2-27.7); 5) Daytime impairment for ≥5 days in past month: 14.2 (9.7-18.6); 6) Insomnia: 18.8 (13.8-23.7).	Burnout: PBM with a 7-point scale from 1 (never) to 7 (always) Time points NR	Low vs. high burnout, mean±SD: 1) Global PSQI: 3.72±2.22 vs. 7.24±4.17, p<0.001; 2) PSQI subscores: sleep quality: 0.54±0.57 vs. 1.40±0.83, p<0.001; sleep latency: 0.51±0.80 vs. 1.38±1.03, p=0.002; sleep duration: 0.45±0.64 vs. 1.16±0.92, p=0.003; sleep efficiency: 0.21±0.57 vs. 0.77±0.98, p=0.018; sleep disturbance: ns; use of medication: 0.14±0.49 vs. 0.57±0.83, p=0.032; daytime dysfunction: 0.52±0.73 vs. 1.57±0.88, p=0.002. 3) Prevalence (95% CI) of insomnia symptoms: sleep latency: 5.5% (2.5-11.5%) vs. 21.1% (10.5-31.6%), p=0.015; wake time >30 min after sleep onset: 9.4% (1.6-17.1%) vs. 25.5% (14.2-37.7%), p=0.029; early awakening: 14.5% (5.1-23.8%) vs. 45.6 (32.7-58.4%), p<0.001; somewhat/very dissatisfied with sleep: 5.5% (2.5-11.5%) vs. 50% (37.1-62.8%), p<0.001; day impairment: 5.5% (2.5-11.5%) vs. 38.2% (25.6-50.7%), p<0.001; insomnia: 7.3% (0.4-14%) vs. 39.7% (27.1-52.2%), p<0.001.
Oncologists					
Shanafelt, 2005 RoB: unclear	CS	Fatigue: LASA QOL ≤7; Sleep deprivation: 10-point Likert scale from 0 (not at all) to 10 (stressful as can be)	75% had a high level of fatigue; Mean±SD sleep score: 4.5±2.65.	Wellbeing: 10-item LASA QOL, high ≥8 vs. low ≤7 Time points NR	1) Sleep deprivation for high vs. low overall well-being (mean±SD): 3.9±2.57 vs. 5.1±2.60, p=0.0004; 2) Lower fatigue predicted overall wellbeing in a multivariate model including personal and professional characteristics, p=0.002.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Time points NR			
Shanafelt, 2014 RoB: unclear	CS	Fatigue: 10-point LAS (lower scores indicate greater fatigue) Time points NR	Mean±SD fatigue score: 5.7±2.4	Satisfaction with WLB: 5-point Likert scale from 'strongly agree' to 'strongly disagree' Time points NR	1) OR (95%CI) of lower satisfaction predicted by high fatigue (as input) in multivariate model including personal and work-related factors, and burnout: 0.489 (0.337-0.710), p<0.001.
Mixed groups of physicians					
Aziz, 2004 RoB: high	CS	Working while fatigued: 5-point scale from 'extreme' to 'a little' Time points NR	NR	Stress: 47-item questionnaire with a 5-point scale from 'extreme' to 'a little' Time points NR	1) Sources of stress: working while fatigued had a mean±SD score of 2.44±1.20, factor loading: 0.653, in factor analysis; 2) Inverse correlation between stress and working while fatigued: r=-0.270 (significance level NR).
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: worried about having a car accident while driving home post-call: 5.4 vs. 7.0, p<0.001; sleep loss has a major impact on personal life: 8.4 vs. 7.0, p=0.01; 3) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Elovaino, 2015 RoB: low	CS	Sleeping problems: Jenkins Scale with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006 and 2010	Mean±SD score: 2006: 2.30 (1.00); 2010: 2.35 (1.05).	Jobs demands: 5 items scored on a 5-point scale from 1 (strongly disagree) to 5 (strongly agree); Job control: 3 items derived from the Karasek Job Questionnaire	There was no association between sleeping problems in 2006 and job demands or control in 2010.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Heponiemi, 2014 RoB: low	CS	Sleeping problems: Jenkins Scale[81] with a 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Psychological distress: GHQ-12 with a 4-point scale (low to high); Job satisfaction: JDS with a Likert scale from 1 (strongly disagree) to 5 (strongly agree) Assessed in 2010	1) Sleeping problems associated with job satisfaction, β=-0.12, $p<0.001$, psychological distress, β=0.18, $p<0.001$; 2) Total indirect effect of on-call duty through two mediators (sleeping problems, work interference with family) (R^2 (95% CI)): job satisfaction 0.06 (-0.059, -0.016), $p<0.001$; psychological distress 0.16 (0.023, 0.081), $p<0.001$.
Mahmood, 2016 RoB: high	CS	Sleep deprivation: self-reported mean hours of sleep when on call Assessed at 4 years, 10 years, and 15 years post-graduation	Mean±SD hours: 4 years: 4.52 (2.79); 10 years: 5.38 (6.36); 15 years: 6.41 (7.14).	Alcohol use disorders: Modified 9-item version of the Alcohol Use Disorder Identification Test (AUDIT) ≥6 for men and ≥5 for women. Assessed at 4 years, 10 years, and 15 years post-graduation	There was no association between hours of sleep when on call and hazardous drinking behaviours ($p=0.732$)
Shirom, 2010 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Burnout: SMBM on a 7-point scale from 1 (almost never) to 7 (always)	1) Correlation between physical fatigue subscale and overall burnout: 0.88, $p<0.05$; 2) In a predictive structural model for burnout, physical fatigue accounted for unique variance in the burnout items not accounted for by total burnout ($R^2=0.24$).
Smith, 2017 RoB: unclear	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Mental and physical illness: self-reported via open-ended comments Time points NR	Some physicians reported developing mental illness (e.g., bipolar disorder, alcohol misuse) due to tiredness and stress at work; others developed physical health problems due to sleep deprivation, poor eating habits and lack of exercise.

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
Starmer, 2016 RoB: low	CS	Sleep deprivation: <7 hours sleep in a typical 24-h period (self-reported) Time points NR	27.7% sleep deprived	Burnout, satisfaction with career and life, balanced personal and professional commitments: Each on a 5-point Likert scale (strongly agree to strongly disagree) Time points NR	≥7-h vs. <7-h sleep: 1) Burnout (% strongly agree/agree): 26.4% vs. 39.6%, p<0.05; career satisfaction (% strongly agree/agree) ns; life satisfaction (% completely/very satisfied): 76.4% vs. 55.9%, p<0.05; balanced personal and professional commitments (% completely/very satisfied): 49.7% vs. 26.1%. 2) <7-h sleep (vs. ≥7-h) (OR, 95% CI) associated with life satisfaction 0.44 (0.29-0.67), p<0.05; balanced personal/professional commitments 0.46 (0.31-0.71), p<0.05, in a model including work and personal factors.
Tokuda, 2009 RoB: low	CS	Sleep hours/day: self-reported (continuous) Time points NR (included weekday and weekends)	Mean±SD (range) sleep hours/day: 6±0.9 (3-8)	Burnout: MBI (Japanese) with a 7-point Likert scale: 0 (none) to 6 (every day); Job satisfaction: JHPSS with a 5-point Likert scale: 1 (strongly disagree) to 5 (strongly agree) Time points NR	Maximum likelihood estimates±SE: 1) Sleeping time to job satisfaction: group 0.990±0.458, p=0.031; ns for men; women 1.711±0.805, p=0.034; 2) Sleeping time to EE: group -0.219 ±0.070, p=0.002; men -0.215±0.082, p=0.009; ns for women.
Wada, 2010 RoB: unclear	CS	Sleep hours/day: Self-reported (continuous) Assessed for past month when not completing overnight work	<5 hours: 8.7% men, 9.9% women; 5 to <6 hours: 32.3% men, 34.6% women; 6 to <7 hours: 46.0% men, 43.7% women; ≥7 hours: 13.0% men, 11.8% women.	Depression: QIDS-SR; Japanese score <5 (no symptoms) to >20 (very severe symptoms) Assessed for past 7 days	1) Sleep hours for those with vs. without depressive symptoms: <5: 18.7% vs. 7.7% men, 20.5% vs. 8.7% women; 5 to <6: 33.7% vs. 32.2% men, 38.6% vs. 34.2% women; 6 to <7: 35.1% vs. 46.9% men; 31.8% vs. 45.1% women; 2) Association between <5h sleep (vs. 6-7h) and depressive symptoms (OR (95% CI)): univariate 2.79 (1.96-3.95) for men, 2.65 (1.47-4.78) for women; multivariate (including age and workload

Study Risk of Bias (RoB)	Study design	Exposures or interventions Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					factors) 2.70 (1.82-4.03) for men, 2.38 (1.11-5.10) for women.

[a]Includes studies of anesthetists, where these were physicians.

[b]Includes primary care physicians, internal medicine physicians, and general practitioners.

AM: morning; aMT6-s: melatonin metabolite; BA: before-after; CI: confidence interval; CBI: Copenhagen Burnout Inventory; CS: cross-sectional; DP: depersonalization; DSM: Diagnostic and Statistical Manual of Mental Disorders; EE: emotional exhaustion; ER: emergency; ESS: Epworth Sleepiness Scale; GHQ: General Health Questionnaire; h: hour(s); ICU: intensive care unit; IL-8: interleukin-8; JDS: Job Diagnostic Survey; JHPSS: Japanese Hospital Physicians Satisfaction Scale; LAS: linear analog scale; LASA: linear analog assessment scales; MBI: Maslach Burnout Inventory; MOSQ: Modified Occupational Stress Questionnaire; min: minute(s); NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PA: personal achievement; PBM: Pines Burnout Measure; PE: professional efficacy; PM: afternoon; POMS: Profile of Mood States; PSQI: Pittsburgh Sleep Quality Index; QIDS-SR: Quick Inventory Depressive Scale – Self-Reported; QOL: Quality of Life; RCT: randomized controlled trial; RDAS: Revised Dyadic Adjustment Scale; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SMBM: Shirom-Melamed Burnout Measure; TS: time series; US: United States of America; VAS: visual analog scale; vs.: versus; WLB: work-life balance

Performance and safety outcomes related to fatigue or sleep loss among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Exposures or intervention: Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Surgeons					
Uchal, 2005 RoB: unclear	RCT	Sleep deprivation from a 24-h call shift vs. 8-h work; Sleep hours: self-reported (continuous); Sleepiness: ESS (moderate: 10-15, severe: ≥16) Assessed post-call and post-work	Median (range) sleep hours: 1.5 (0-3) post-call vs. 6.5 (5-9) post-work, p<0.05; Median ESS score: 7.0 post-call vs. 5.5 post-work, ns.	Surgical performance: laparoscopic surgical simulator(Minimally Invasivs Surgical Trainer-Virtual Reality) for product quality, procedure effectiveness Assessed post-call and post-work	Post call vs. post-work: 1) Product quality: no difference in accuracy error, tissue damage, leak rate; 2) Procedure effectiveness: no difference in goal-directed actions, non-goal directed actions, operating time.
Chu, 2011 RoB: low	CO	Sleep deprivation: self-reported hours, moderate (3-6h) or severe (<3h) Assessed the night before surgery	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived and 1,595 (39.4%) moderately sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	For 0-3 vs. 3-6 vs. >6 hours of sleep: no difference in CABG or ACC.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and another case in the next 24-h	Of 6,751 procedures, 339 (5%) performed by sleep-deprived surgeons	Surgical performance: CABG, ACC Assessed during surgery	Sleep deprived vs. non-sleep deprived: no difference in CABG or ACC.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Surgical performance: duration of surgery	Sleep deprived vs. non-sleep deprived: no difference in duration of surgery, even after stratification by type of procedure.
Amirian, 2014 RoB: high	BA	17-h night call shift; Sleep hours during the shift: Wrist-mounted Micro-Mini-Motionlogger; Sleepiness: KSS	Naps pre-call: 11 (37%) napped for median (IQR) 90 (58-128) min; Median (IQR) sleep: 91 (62-123) min on the pre-call night vs. 430 (329-449) on	Surgical performance: LapSimGyn laparoscopic simulation for time, blood loss, instrument path; D2 test of attention and concentration	Pre- vs. post-call: 1) LapSimGyn: no difference in total time, blood loss, instrument path length, instrument angular path; napping did not affect performance;

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
		Assessed on pre-call and on-call day; sleepiness assessed during shift	the on-call night, p<0.001; Sleep on-call: 12 (40%) slept for median (IQR) 98 (39-135) min; Significant development of sleepiness during shift (p<0.001), plateau score of 7 at 04:00 to 08:00.	Assessed on pre-call and on-call day	2) D2 test: improvement in concentration, p<0.05. No changes in any other parameters; 3) ns difference in laparoscopic simulation time in those who slept during the shift vs. not.
Gerdes, 2008 RoB: high	BA	On-call shift; Fatigue: questionnaire designed by Behrenz & Monga, 1999; Sleep hours: self-reported (continuous) Assessed in 3 sessions pre- and post-call	Fatigue differential from pre- to post-call (range): 1-7 (units unclear); Sleep during call (range): 1-5h	Psychomotor performance: virtual ring transfer task for gesture-level proficiency, hand movement smoothness, tool movement smoothness, elapsed time Assessed in 3 sessions pre- and post-call	1) Pre- to post-call: decrease in all measures of psychomotor proficiency (p<0.05, data NR) except elapsed time; no change in number of psychomotor errors; increase cognitive errors (p<0.05, data NR); 2) Cognitive errors increased exponentially as fatigue ratings increased (R^2=0.9219) and as hours of sleep declined (R^2=0.933).
Shanafelt, 2010 RoB: unclear	CS	Degree of fatigue as a contributor to errors (self-reported) Assessed for the past 3 months	NR	Perceived recent major medical errors (self-reported) Assessed for the past 3 months	1) Prevalence of perceived recent major medical error: 8.9%; 2) Of those reporting an error, 6.9% listed degree of fatigue as the greatest contributing factor.
Anesthesiologists[a]					
Lederer, 2006 RoB: high	BA	24-h shift, on-call duty; Sleep hours and interruptions: self-reported; Tiredness: VAS from 0 (low) to 100 (high) Assessed pre- and post-duty	Mean±SD sleep: 4.1±1.7h; Number of interruptions: 0.8±1.1; Tiredness pre- vs. post-duty: 30.9±27.5 vs. 59.5±18.9, p=0.01.	Psychomotor performance: reaction time, critical flicker fusion, response measure, peripheral awareness; Concentration ability: scale of 0 (low tiredness) to 100 (maximum tiredness) Assessed pre- and post-duty	Pre- vs. post-duty, mean±SD: 1) Psychometric testing: recognition reaction time (ms): 439.8±50.8 vs. 480.3±58.9; motor reaction time (ms): 252.8±39.3 vs. 465.4±65.0; total reaction time (ms): 690.8±73.4 vs. 746.5±113.7; critical flicker fusion (Hz): 29.0±2.3 vs. 28.7±3.7; response measure (pixels): 647.8±126.7 vs. 598.3±138.1,

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					peripheral awareness task recognition time: 58.9±59.2 vs. 51.6±47.5; 2) Concentration ability: 26.4±23.5 vs. 56.3±23.0, p=0.007.
Chang, 2013 RoB: unclear	CS	15-h in-house overnight call; Sleepiness pre-call: ESS ≥9; Sleep hours: self-reported (continuous) Sleepiness assessed pre-call, sleep hours during call	Median (IQR) ESS: 9 (9), 64% scored ≥9; Median (IQR) hours slept during shift: 1 (0-3).	Psychomotor performance: reaction time; CCPT II; N-back; HVLT (3 trials of 12 words) Assessed at baseline and pre- and post-call	1) Afternoon baseline vs. pre-call: no difference in reaction time, CCPT, N-back, of HVLT; Morning baseline vs. post-call: 1) No change in auditory or visual reaction time; 2) CCPT (t-scores): No change in detectability, response style, hit reaction time, omissions, commissions; 3) N-back % accuracy: no change for auditory, visual, or mean N-value; 4) HVLT (t-score): mean for trials 1-3: 48.6±7.6 vs. 41.5±9.9 (p=0.04); delayed recall: ns; 5) No correlation between ESS scores pre-call or sleep during shift and any measure of psychomotor performance.
Gander, 2000 RoB: low	CS	Nights of work-related sleep disturbance: self-reported (continuous) Assessed for the past 6 months	NR	Risk of fatigue-related errors: questionnaire modelled after Gravenstein et al., 1990 Assessed for the past 6 months	1) Risk of fatigue-related errors increased with increasing nights of work-related sleep disturbance: RR: 1.25, 95% CI: 1.06-1.49.
Saadat, 2017 RoB: low	CS	Sleep deprivation due to an overnight call shift	NR	Reaction time: PVT Assessed after an overnight call shift and the morning of a regular (non-call) day	Mean (SD) reaction time was slower post-call (297.76 (83.75)) vs. on a regular day (266.58 (38.35)), p=0.047.

Study Risk of Bias (RoB)	Study design	Exposures or intervention		Outcomes	Associations between exposure and outcome
		Assessment measure and time points	Baseline	Assessment measure and time points	
Gander, 2008 RoB: unclear	NC	Sleep loss across consecutive working days or on-call work: Wrist-mounted Actiwatch (Mini Mitter, Bend, Oregon, US), sleep and duty diary Assessed over a 2-week period including a weekend of rostered shifts or on-call	≥2 hours sleep <baseline: 8% of 24-h periods that included day work vs. 14% that included day + call; Sleep hours: mean 0.6h less sleep when working day shifts (p=0.014) and 0.8h less sleep when working day shifts + call (p=0.013) vs. off.	Psychomotor performance: PVT Assessed within 2 hours pre- and post-call	1) In fixed model analysis for reaction time including sleep, time since waking, work hours: acute sleep loss associated with slower median reaction time, $F_{(1,184)}$=5.70, p<0.05; longer time since waking associated with poorer performance on the slowest 10%, $F_{(1,185)}$=5.13, p<0.05; 2) Reaction time across 12 consecutive work days: no change in pre-duty reaction times but post-duty reaction times slowed linearly, median -3.88, p<0.001; decline in performance across 10 minutes became progressively steeper both pre- and post-duty, p=0.020.
ER or ICU physicians					
Sanches, 2015 RoB: high	CS	Acute sleep deprivation (<5h of night sleep after a night shift of 12h) Sleep hours: 7-day Actigraphy via SenseWear® Pro2 Armband; Sleepiness: ESS; Sleep quality: PSQI Assessed the week and night before the psychomotor tests	Non-sleep deprived vs. sleep deprived: PSQI >5: 0% vs. 33%, ns; ESS≥10: 11% vs. 67% Sleep time (mean±SD) in week before tests: duration and number of naps higher in sleep deprived group, but diurnal sleep hours lower, 428.6±30.1 vs. 375.8±55.9, p=0.038; Sleep quality (mean±SD): week before tests: 3.3±0.7 vs. 2.6±0.3, p=0.013; night before tests: 3.1±0.8 vs. 1.9±1.0, p=0.020.	Psychomotor performance via Battery Test Reaction 5 (v1): StimulTest, InstrucTest, MovemTest; TP test of visual attention Assessed on morning after night shift 8	Sleep deprived group vs. non-sleep deprived, mean±SD: 1) InstrucTest: correct answers: 169.4 (16.0) vs. 148.3 (28.3), p=0.070; wrong answers: ns; perfection index (%): 99.6 (0.3) vs. 98.9 (1.3), p=0.021; response latency (sec/click): ns; 2) StimulTest: correct answers: 170.7 (21.9) vs. 145.1 (17.7), p=0.022; wrong answers: ns; perfection index (%): ns; response latency (sec/click): 1.06 (0.1) vs. 1.24 (0.1), p=0.022; 3) MovemTest: ns for any parameter; 4) TP: omitted symbols: 34.2±18.4 vs. 62.7±44.0, p=0.034; concentration index (%): 14.1±8.9 vs. 30.0±25.9, p=0.019; quality index (%): 13.8±8.6 vs. 29.2±26.4, p=0.031; correct/wrong symbols: ns; Correlations between sleep and tests: 1) TP for sleep hours nights 1-6: omitted symbols: r=-0.686, p=0.011 for non-sleep-

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					deprived, ns for sleep-deprived; concentration index (%): r=-0.359, p=0.037 for sleep-deprived, ns for non-sleep deprived; r=-0.359, p=0.037 for the group; no other significant correlations. 2) No correlation between PSQI, ESS and any of the psychomotor tests.
Generalists[b]					
Harbeck, 2015	CS	24-hours on-call shift with sleep disturbance: self-reported number of sleep disturbances and hours of sleep per night Assessed before a normal day shift, and after a 24-h on call shift	1) Sleep hours on a normal day vs. following a 24-h shift: <2 hours: 0 vs. 5.9%; 2-4 hours: 5.9% vs. 47.1%; 4-6 hours: 11.8% vs. 35.3%; >6 hours: 82.4% vs. 11.8% 2) Number of sleep disturbances a normal day vs. following a 24-h shift: 0: 82.4% vs. 11.8%; 1: 11.8% vs. 35.3%; 2: 5.9% vs. 47.1%; 3: 0% vs. 5.9%; 4: 0% vs. 0%; >4: 0% vs. 0%	Neurocognitive parameters: computerized attentional test (vigilance, alertness); D2 letter cancellation test (divided attention); Trail Making Test (visual attention, task switching); Digit Span, Digit Symbol Substitution Test, Weschler Memory Scale (memory functions) Assessed before a normal day shift, and after a 24-h on call shift	Intrinsic alertness, focused attention and vigilance were similar on both occasions; Phasic alertness improved following the on-call shift: mean (SD) 24.8 (15.6) vs. 38.3 (21.5), p = 0.022.
Mixed specialties or undefined populations					
Chen, 2008 RoB: high	CS	Sleepiness: ESS score ≥11 Time points NR	Mean±SD ESS score: 7.8±4.0, range: 0-20, 23% had scores ≥11.	Impact on work and personal life: Impact Questionnaire with a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree) Time points NR	1) Impact score correlated with ESS, r=0.31, p<0.05; 2) ESS score was higher among physicians who agree/strongly agree vs. other response: written an incorrect order: 8.8 vs. 7.3, p=0.02; might fall asleep while examining a patient: 13.2 vs. 7.7, p=0.001; look forward to sleeping at grand rounds: 10.4 vs. 7.4, p=0.002;

Study Risk of Bias (RoB)	Study design	Exposures or intervention Assessment measure and time points	Baseline	Outcomes Assessment measure and time points	Associations between exposure and outcome
					3) No difference in ESS score for those who agree/strongly agree vs. other response: work is unaffected by sleep loss and fatigue, thinking is unaffected by sleep loss, sleep loss and fatigue affect my medical decisions, have heard of others making medical errors due to sleep loss and fatigue, never make errors in prescriptions on post-call days, have made medical errors because of sleep loss and fatigue; 4) Higher ESS scores predicted by impact score in multivariate regression including personal and work-related factors: β=0.11, p=0.005.
Heponiemi, 2014 RoB: low	CS	Sleeping problems: 4-item Jenkins Scale on 6-point scale from 1 (never) to 6 (every night) Assessed in 2006	Mean±SD (range) score: 2.30±1.00 (1-6)	Work ability: Work Ability Index on scale from 1 (could not work at all) to 10 (best work ability) Assessed in 2010	1) On-call duty had an indirect effect on work ability (R^2=0.11, 95% CI: -0.122, -0.031, p<0.001) through two mediators (work interference with family, sleeping problems); 2) Sleeping problems inversely associated with work ability, β=-0.29, p<0.001.
Kanieta, 2011 RoB: unclear	CS	Sleep hours: self-reported (continuous) Sleepiness and sleep difficulties: 5-point scale from 1 (never) to 5 (always); Insomnia: ≥3 sleep difficulties Assessed for the past month	Insufficient rest: 32.5%; Daytime sleepiness: 3.5%; Insomnia: 20.0%; Sleep time (mean±SD min): 279.8±60.9	Self-reported medical incidents: 4-point scale from 1 (never) to 4 (often) Assessed for the past month	1) Prevalence of medical incidents (% (95% CI)): sleep deprived (26.8% (24.2, 29.4)) vs. not (15.2% (13.7, 16.7)), p<0.01; insomnia (24.8% (21.6, 28.0)) insomnia vs. not (17.6% (16.2, 19.0)), p<0.01; ≥6h sleep (18.3% (16.8, 19.8)) vs. <6h (21.7% (18.8, 24.6)), p=0.03; 2) Predictors of medical incidents in multivariate model including personal and work-related factors (OR (95% CI)): lacking rest due to sleep deprivation vs. not (1.65 (1.33-2.04)), p<0.01); insomnia vs. not (1.45 (1.16-1.82), p<0.01); ns for sleep hours.

Study Risk of Bias (RoB)	Study design	Exposures or intervention: Assessment measure and time points	Baseline	Outcomes: Assessment measure and time points	Associations between exposure and outcome
Sexton, 2001 RoB: high	CS	Fatigue as a factor impacting performance Time points NR	NR	Performance effectiveness measured by 1 question: agree, neutral, disagree Time points NR	1) "When fatigued, I perform effectively during critical phases of operations/patient care": Anesthetic: 47% agree; 15% neutral; 38% disagree; Surgical: 70% agree; 12% neutral; 18% disagree.
Shirom, 2006 RoB: low	CS	Tiredness and exhaustion: SMBM Physician Fatigue Subscale on a 7-point scale from 1 (almost never) to 7 (always) Time points NR	NR	Quality of care: Adapted 15-item SERVQUAL with a 5-point Likert scale from 1 (very small extent) to 5 (very large extent) Time points NR	1) Quality of care positively predicted by fatigue in a model incorporating several other components of burnout, β=0.17, p<0.05.
Smith, 2017 RoB: moderate	CS	Sleep deprivation: self-reported via open-ended comments Time points NR	NR	Perceived competence: self-reported via open-ended comments Time points NR	Some physicians indicated that continual tiredness and exhaustion led to concerns that it would affect their competence; some felt that professional performance was compromised at times of physical and mental fatigue.
Tanti, 2017 RoB: high	CS	Fatique: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	NR	Prescribing errors: questionnaire on contributors to prescribing errors, with a 5-point Likert scale (very high to very low association) Time points NR	Perception of the contribution of fatigue to prescribing errors differed by physician type (p<0.05): 74% of community doctors, 96% hospital doctors, 8% of office-working doctors perceived a very high or high association between fatigue and prescribing errors.

[a]Includes studies of anesthetists, where these were physicians.
[b]Includes primary care physicians, internal medicine physicians, and general practitioners.
ACC: aortic cross-clamp time; BA: before-after; CABG: cardiopulmonary bypass time; CCPT II: Connor's Continuous Performance Test II; CI: confidence interval; CO: cohort; CS: cross-sectional; ER: emergency; ESS: Epworth Sleepiness Scale; h: hour(s); HVLT: Hopkin's Verbal Learning Task; Hz: Hertz; ICU: intensive care unit; IQR: interquartile range; KSS: Karolinska Sleep Scale; min: minutes; ms: millisecond(s); N-back: Dual N-back test; NA: not applicable; NR: not reported; ns: not statistically significant; OR: odds ratio; PSQI: Pittsburgh Sleep Quality Index; PVT: Psychomotor vigilance Performance Task; RR: risk ratio; RCT: randomized controlled trial; RoB: Risk of Bias; SD: standard deviation; SE: standard error; SERVQUAL: Service Quality Measure; SMBM: Shirom-Melamed Burnout Measure; TP: Toulouse-Piéron test; TS: time series; US: United States of America; vs.: versus

Patient outcomes related to fatigue or sleep restriction among physicians in independent practice

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Surgeons					
Chu, 2011 RoB: low	CO	Sleep deprivation: moderate (3-6 h) or severe (<3-h) sleep deprivation the night before surgery (self-reported hours)	Of 4,047 procedures, 83 (2.1%) performed by severely sleep-deprived, 1,595 (39.4%) by moderately sleep-deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) 0-3 vs. 3-6 vs. >6 hours of sleep: No difference in incidence of mortality, incidence of 10 major complications (except septicemia, 3.6% vs. 0.9% vs. 0.8%, p=0.03), ICU length of stay; in-hospital length of stay (days): 7.0 vs. 6.0 vs. 7.0, p<0.001.
Ellman, 2004 RoB: low	CO	Sleep deprivation: performed a case starting 22:00 to 05:00, or ending 22:00 to 07:30 and performed a subsequent case in the next 24-h	Of 6,751 procedures, 339 (5%) were performed by sleep deprived surgeons	Chart review: mortality, surgical complications, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, need for blood products, complications (operative, neurologic, renal, infectious, pulmonary), in-hospital length of stay.
Govindarajan, 2015 RoB: low	CO	Sleep deprivation: treated patients from midnight to 07:00 and performed a subsequent case on the same day	NR	Chart review: mortality, surgical complications, readmission, length of stay Assessed during and post-surgery	1) Sleep deprived vs. non-sleep deprived: no difference in mortality, surgical complications, readmissions within 30 days, or length of stay.
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse surgical complications Assessed during and post-surgery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total number of complications, preventable complications, type of complications; 2) Operating room procedures with complications, OR (95% CI): 8.5% for 0-6h sleep vs. 3.1% for >6h sleep, 2.70 (1.13-6.48), p=0.03; 3) All procedures with complications, OR (95% CI): 6.2% for 0-6h sleep vs. 3.4% for >6h sleep, 1.72 (1.02-2.8[illegible]), p=0.04.

Study Risk of Bias (RoB)	Study design	Exposures Intervention or assessment scale and time points	Baseline	Outcome Measures Assessment scale and time points	Associations between exposure and outcome
Schieman, 2007 RoB: low	CO	Fatigue: surgeon billed for clinical work after 22:00 the night before surgery	Of 270 procedures, 22 (8%) were performed by fatigued surgeons	Chart review: surgical complications, length of stay, mortality, cancer recurrence Assessed during and post-surgery	1) Fatigued vs non-fatigued surgeons: no difference in intra- or post-operative complication rate, length of stay, in-hospital length of stay, cancer recurrence.
Vinden, 2014 RoB: low	CO	Sleep deprivation (at risk): surgeon worked 00:00 to 07:00 and performed surgery 07:00 to 18:00	Of 94,183 surgeries, 2,078 (2.2%) were performed by surgeons who were 'at risk'	Chart review: conversion to open procedure (from laparoscopic), iatrogenic injuries, mortality Assessed during and post-surgery	1) At risk vs not at risk surgeon: no difference in incidence of conversion to open procedure, iatrogenic injuries, mortality, in either univariate or multivariate analyses.
Obstetricians					
Rothschild, 2009 RoB: low	CO	Sleep deprivation: daytime procedures following an overnight procedure; Sleep opportunity: 0-6h, <6h	NR	Chart review: frequency of adverse obstetric complications Assessed during and post-delivery	1) Post-nighttime vs. control: no difference in number of procedures with complications, total complications, preventable complications, type of complications; 2) No association between sleep deprivation and proportion of procedures with complications, nor difference for 0-6h vs. >6h of sleep opportunity.

CI: confidence interval; CO: cohort; h: hours; NR: not reported; OR: odds ratio; RoB: Risk of Bias; SD: standard deviation; US: United States of America; vs.: versus

Supplementary file 5. Statistical analyses

Dichotomous outcomes

Outcome or subgroup	Number of studies	Number of participants	Pooled risk ratio (95% CI)	Heterogeneity	
				P	I^2
1.1 Patient mortality	5	60,436	0.98 (0.84, 1.15)	0.73	0%
1.2 Intra-operative complications	3	19,798	suppressed	0.007	82%
1.2.1 Surgical procedure	3[a]	14,896	suppressed	<0.001	88%
1.2.2 Obstetric procedure	1[a]	4,902	suppressed	NA	NA
1.3 Post-operative complications	5	60,201	0.99 (0.95, 1.03)	0.45	0%

[a] Rothschild, 2009 is represented in both analyses

Continuous outcomes

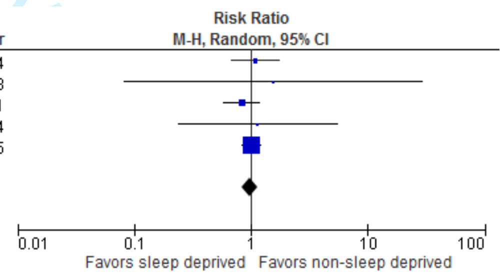
Outcome or subgroup	Number of studies	Number of participants	Pooled mean difference (95% CI)	Heterogeneity	
				P	I^2
1.4 Operating time (minutes)	4	50,046	-0.14 (-1.60, 1.33)	0.70	0%
1.5 Length of hospital stay (days)	4	50,046	suppressed	<0.001	86%
1.5.1 Cardiac surgeries	2	10,798	suppressed	0.01	84%
1.5.2 Elective surgeries	1	38,978	suppressed	NA	NA
1.5.3 Anterior resection for anal cancer	1	270	suppressed	NA	NA

CPBT: cardiopulmonary bypass time; NA: not applicable

1.1 Patient mortality

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Ellman 2004	17	339	296	6412	10.4%	1.09 [0.67, 1.75]	2004
Schieman 2008	0	22	3	248	0.3%	1.55 [0.08, 29.03]	2008
Chu 2011	47	1678	80	2369	18.8%	0.83 [0.58, 1.18]	2011
Vinden 2014	2	2078	7	8312	1.0%	1.14 [0.24, 5.50]	2014
Govindarajan 2015	224	19489	222	19489	69.5%	1.01 [0.84, 1.21]	2015
Total (95% CI)		**23606**		**36830**	**100.0%**	**0.98 [0.84, 1.15]**	
Total events	290		608				

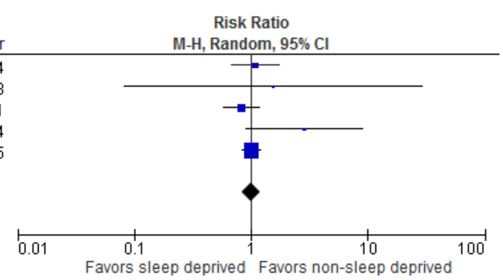
Heterogeneity: Tau² = 0.00; Chi² = 1.25, df = 4 (P = 0.87); I² = 0%
Test for overall effect: Z = 0.23 (P = 0.82)



Sensitivity analysis using highest possible number of events for Vinden 2014

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Ellman 2004	17	339	296	6412	13.9%	1.09 [0.67, 1.75]	2004
Schieman 2008	0	22	3	248	0.4%	1.55 [0.08, 29.03]	2008
Chu 2011	47	1678	80	2369	23.3%	0.83 [0.58, 1.18]	2011
Vinden 2014	5	2078	7	8312	2.6%	2.86 [0.91, 8.99]	2014
Govindarajan 2015	224	19489	222	19489	59.8%	1.01 [0.84, 1.21]	2015
Total (95% CI)		**23606**		**36830**	**100.0%**	**1.00 [0.83, 1.21]**	
Total events	293		608				

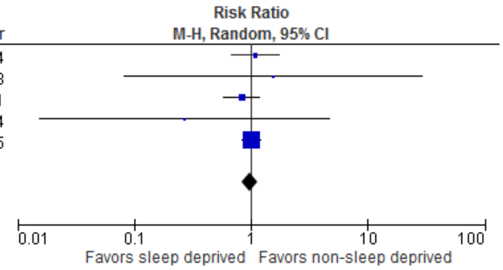
Heterogeneity: Tau² = 0.01; Chi² = 4.50, df = 4 (P = 0.34); I² = 11%
Test for overall effect: Z = 0.03 (P = 0.98)



Sensitivity analysis using lowest possible number of events for Vinden 2014

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
Ellman 2004	17	339	296	6412	10.5%	1.09 [0.67, 1.75]	2004
Schieman 2008	0	22	3	248	0.3%	1.55 [0.08, 29.03]	2008
Chu 2011	47	1678	80	2369	18.9%	0.83 [0.58, 1.18]	2011
Vinden 2014	0	2078	7	8312	0.3%	0.27 [0.02, 4.67]	2014
Govindarajan 2015	224	19489	222	19489	70.0%	1.01 [0.84, 1.21]	2015
Total (95% CI)		**23606**		**36830**	**100.0%**	**0.98 [0.84, 1.14]**	
Total events	288		608				

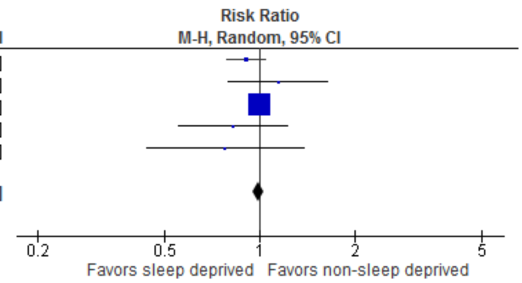
Heterogeneity: Tau² = 0.00; Chi² = 2.01, df = 4 (P = 0.73); I² = 0%
Test for overall effect: Z = 0.29 (P = 0.77)



1.3 Post-operative complications

Study or Subgroup	Sleep deprived Events	Sleep deprived Total	Not sleep deprived Events	Not sleep deprived Total	Weight	Risk Ratio M-H, Random, 95% CI
Chu 2011	259	1678	404	2369	7.8%	0.91 [0.78, 1.04]
Ellman 2004	29	339	480	6412	1.2%	1.14 [0.80, 1.64]
Govindarajan 2015	3527	19489	3543	19489	89.4%	1.00 [0.95, 1.04]
Schieman 2008	12	22	164	248	1.0%	0.82 [0.56, 1.22]
Vinden 2014	14	2031	72	8124	0.5%	0.78 [0.44, 1.38]
Total (95% CI)		**23559**		**36642**	**100.0%**	**0.99 [0.95, 1.03]**
Total events	3841		4663			

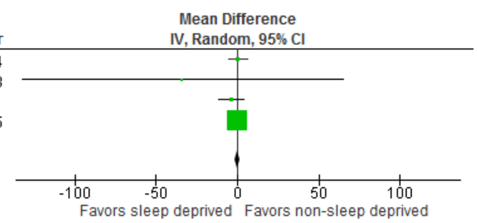
Heterogeneity: Tau² = 0.00; Chi² = 3.69, df = 4 (P = 0.45); I² = 0%
Test for overall effect: Z = 0.66 (P = 0.51)



1.4 Operating time (minutes)

Study or Subgroup	Sleep deprived Mean	Sleep deprived SD	Sleep deprived Total	Not sleep deprived Mean	Not sleep deprived SD	Not sleep deprived Total	Weight	Mean Difference IV, Random, 95% CI	Year
Ellman 2004	107.7	55.24	339	107.4	56.05	6412	5.9%	0.30 [-5.74, 6.34]	2004
Schieman 2008	114	226.6	22	148	226.6	248	0.0%	-34.00 [-132.80, 64.80]	2008
Chu 2011	110.2	119.13	1678	114	120.88	2369	3.8%	-3.80 [-11.30, 3.70]	2011
Govindarajan 2015	156	75.56	19489	156	80	19489	90.2%	0.00 [-1.54, 1.54]	2015
Total (95% CI)			**21528**			**28518**	**100.0%**	**-0.14 [-1.60, 1.33]**	

Heterogeneity: Tau² = 0.00; Chi² = 1.42, df = 3 (P = 0.70); I² = 0%
Test for overall effect: Z = 0.18 (P = 0.86)



Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4-5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Supplementary file 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6-7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	8

Appendix 1. PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8, Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8-11, Table 1, Supplementary file 2
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	11, Supplementary file 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	p. 12-18; Supplementary file 4; figures 2-4
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	p. 12-18, figures 2-4
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Not applicable
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Supplementary file 5
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	18-19
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	19-20
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	20
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	21

Appendix 1. PRISMA checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**. Page 2 of 2