Sleep duration and health in adults: an overview of systematic reviews

Jean-Philippe Chaput, Caroline Dutil, Ryan Featherstone, Robert Ross, Lora Giangregorio, Travis J. Saunders, Ian Janssen, Veronica J. Poitras, Michelle E. Kho, Amanda Ross-White, and Julie Carrier

Abstract: The objective of this overview of systematic reviews was to examine the associations between sleep duration and health outcomes in adults. Four electronic databases were searched in December 2018 for systematic reviews published in the previous 10 years. Included reviews met the a priori determined population (community-dwelling adults aged 18 years and older), intervention/exposure/comparator (various levels of sleep duration), and outcome criteria (14 outcomes examined). To avoid overlap in primary studies, we used a priority list to choose a single review per outcome; reviews that examined the effect of age and those that looked at dose–response were prioritized. A total of 36 systematic reviews were eligible and 11 were included. Reviews included comprised 4,437,101 unique participants from 30 countries. Sleep duration was assessed subjectively in 96% of studies and 78% of studies in the reviews were prospective cohort studies. The dose–response curves showed that the sleep duration that was most favourably associated with health was 7–8 h per day. Modification of the effect by age was not apparent. The quality of the evidence ranged from low to high across health outcomes. In conclusion, the available evidence suggests that a sleep duration of 7–8 h per day is the one most favourably associated with health among adults and older adults. (PROSPERO registration no.: CRD42019119529.)

Novelty

• This is the first overview of reviews that examines the influence of sleep duration on a wide range of health outcomes in adults.
• Seven to 8 h of sleep per day was most favourably associated with health.
• Effect modification by age was not evident.

Key words: sleep duration, guidelines, public health, adults, dose–response, recommendations.

Résumé : L’objectif de cet aperçu des revues systématiques est d’examiner les associations entre la durée du sommeil et l’état de santé des adultes. En décembre 2018, quatre bases de données électroniques sont consultées pour relever des revues systématiques publiées au cours des 10 années précédentes. Les examens inclus concernent la population déterminée a priori (adultes vivant dans la communauté âgés de 18 ans et plus), l’intervention/l’exposition/comparaison (divers niveaux de durée du sommeil) et les critères de résultat (14 résultats examinés). Pour éviter le chevauchement dans les études primaires, nous utilisons une liste de priorités pour choisir une seule revue par résultat; les revues qui examinent l’effet de l’âge et celles qui examinent la dose-réponse sont priorisées. Au total, 36 revues systématiques sont qualifiées et 11 sont inclues. Les revues inclues comprennent 4,437,101 participants différents de 30 pays. La durée du sommeil est évaluée subjectivement dans 96 % des études et 78 % des études dans les revues sont des études prospectives de cohorte. Les courbes dose-réponse révèlent que la durée du sommeil la plus favorablement associée à la santé est de 7 à 8 heures par jour. La modification de l’effet selon l’âge ne ressort pas. La qualité des données probantes varie de faible à élevée selon les résultats de santé. En conclusion, les données disponibles suggèrent qu’une durée de sommeil de 7 à 8 heures par jour est la plus favorablement associée à la santé chez les adultes et les personnes âgées. (Numéro d’enregistrement PROSPERO : CRD42019119529.) [Traduit par la Rédaction]

Les nouveautés

• Il s’agit du premier aperçu des revues qui examinent l’influence de la durée du sommeil sur un large éventail de résultats de santé chez les adultes.
Introduction

Sleep is increasingly recognized as a critical component of cognitive, emotional, and physical health. Healthy sleep is characterized by adequate duration, good quality, appropriate timing, and the absence of sleep disorders (Buysse 2014; Chaput and Shiau 2019). Insufficient sleep has become a public health concern in many countries given its high prevalence and association with mortality and morbidity (Liu et al. 2016; Chaput et al. 2017). Not only can insufficient sleep adversely impact health, it can also lead to mistakes in the workplace, lower psychomotor performance, decreased work productivity, and increased risk of car accidents (Institute of Medicine (US) Committee on Sleep Medicine and Research 2006). Thus, insufficient sleep poses a substantial burden to our health and economic sectors in disability and injury each year.

Sleep duration varies across the lifespan and shows an inverse association with age (Chaput et al. 2018). Sleep duration recommendations issued by public health bodies are important for surveillance, help inform policies and interventions, and can be used to educate the general public about healthy sleep (Chaput 2019). For example, the National Sleep Foundation recommends 7–9 h of sleep per day for adults (aged 18–64 years) and 7–8 h for older adults (aged ≥65 years) (Hirshkowitz et al. 2015) while the American Academy of Sleep Medicine and Sleep Research Society recommends ≥7 h per night on a regular basis for adults aged 18–60 years (Watson et al. 2015). These American organizations used very similar guideline development processes but had different experts at the table, which resulted in slightly different sleep duration recommendations.

Given the large body of evidence linking sleep duration and health outcomes in the adult population, an overview of reviews that examine the amount of sleep most favourably associated with overall health as evidenced with dose–response curves is needed to better inform public health guidelines. An overview of reviews (rather than a review of primary studies) is needed at this stage because it allows to leverage existing research, reduces redundancy in research, and enhances efficiency in the review process considering limited timelines and resources. Furthermore, such an overview of reviews needs to examine the effect of age on the associations between sleep duration and health outcomes to determine if recommendations should be different for older compared with younger adults.

Therefore, the present work is the first to provide an overview of systematic reviews on sleep duration and health outcomes in adults aged 18 years and older. The main objective is to determine the sleep duration associated with overall health in adults and examine the impact of age. Findings from this review will inform public health guidelines around sleep duration and identify future research needs.

Materials and methods

Protocol and registration

The present overview of reviews was registered a priori with the International Prospective Register of Systematic Reviews (PROSPERO; registration no. CRD42019119529; available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?id=CRD42019119529), and was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for reporting systematic reviews and meta-analyses (Moher et al. 2009).

Eligibility criteria

The Participants, Interventions, Comparisons, Outcomes, and Study design (PICOS) framework (Schardt et al. 2007) was followed to identify key study concepts in the research question a priori and to facilitate the search process.

Population

The population of interest was community-dwelling adults aged 18 years and older, including apparently healthy adults, adults with obesity, adults with metabolic syndrome, or adults who have had one or more falls in the past year. This also included studies that, among their participant pool, included adults with a chronic condition (e.g., heart disease, diabetes, cancer). Reviews including mixed populations, that is, comprising studies with both individuals who met and those who did not meet the eligibility criteria, were included if the results pertaining to the population of interest were reported separately. If results for the population of interest were not reported separately, studies with a mixed population were included if 80% or more of the study population met the inclusion criteria (or if the sample average fit within the criteria). For example, a systematic review with no subgroup analyses that included some studies from the general population, and some from disease-specific populations, would be included if 80% or more of the participants in the systematic review were from those studies performed in the general population. Exclusion criteria included pregnant women, residents in long-term care, patients in acute care or a hospital setting, people who were unable to move under their own power, and elite athletes (e.g., varsity/ provincial level athletes or Masters’ athletes). We also excluded studies that targeted exclusively shiftworkers and individuals with a sleep disorder or diagnosed disease (e.g., insomnia, type 2 diabetes, depression) to keep the focus on the general adult population and not specific clinical populations.

Intervention (exposure)

The intervention or exposure was sleep duration. Sleep duration could be reported in different ways in the studies and could include total sleep duration (i.e., per 24-h period, including naps, or nighttime sleep duration only). Reviews were eligible if they included primary studies that used objective (e.g., polysomnography, actigraphy/accelerometry) or subjective (e.g., self-report) measures of sleep duration (or both). For systematic reviews that included primary experimental studies, the interventions must have targeted sleep duration exclusively and not multiple health behaviours (e.g., both sleep and diet).

Comparison

Various levels of sleep duration were used for comparison. However, a comparator or control group was not required for inclusion.

Outcomes

A total of 14 health outcomes were chosen based on the literature, expert input and consensus, and recognition of the importance of including a broad range of outcomes. Eight outcomes were identified as critical (primary outcomes) by expert agreement (Ross et al. 2020): (i) mortality; (ii) incident cardiovascular disease (e.g., coronary artery disease, myocardial infarction, stroke); (iii) incident type 2 diabetes; (iv) mental health (e.g., incident depression); (v) brain health (e.g., incident neurodegenerative disease such as Alzheimer’s disease or Parkinson’s disease); (vi) cognitive function (e.g., attention, concentration, executive control, learning).
ing, memory, reaction time); (vii) falls; and (viii) accidents/injuries. Six outcomes were identified as important (secondary outcomes) by expert agreement: (i) incident obesity; (ii) biomarkers of cardiometabolic risk (e.g., insulin sensitivity, glucose tolerance, blood pressure, triglycerides); (iii) bone health (e.g., osteoporosis, bone mineral density, fractures); (iv) health-related quality of life; (v) work productivity (e.g., absenteeism, presenteeism, tests of productivity); and (vi) physical activity and sedentary behaviour.

Study designs
Published or in-press peer-reviewed systematic reviews (as defined by The Cochrane Collaboration (Higgins and Green 2011)) with or without meta-analyses were eligible for inclusion. No grey literature was eligible for inclusion. Systematic reviews that did not receive a “yes” or “partial yes” for items 4 (adequacy of literature search) and 9 (risk of bias from individual studies being included in the review) on A MeaSurement Tool to Assess systematic Reviews (AMSTAR 2) assessments were excluded as these characteristics were considered critical flaws (Shea et al. 2017). For the purpose of this overview, systematic reviews must have searched at least 2 relevant databases and provided a key word and/or search strategy. Justifying language restrictions was not required, as long as the review was published in English or French.

Information sources and search strategy
A research librarian with expertise in systematic review searching created the electronic search strategy and a second research librarian reviewed it. The complete search strategies are available in Supplement S1. The following databases were searched using the Ovid interface: MEDLINE, EMBASE, and PsycINFO. CINAHL was also searched using the EBSCO platform. Searches were conducted the week of December 18, 2018. Studies published in the previous 10 years only (December 18, 2008, until December 18, 2018) were searched to manage scope, reduce overlap, and with a goal to include the most recent body of evidence to inform current public health recommendations around sleep duration in Canada. The search was rerun on October 14, 2019, in case relevant reviews may have been published since then. Reference lists of included studies were also checked.

Study selection
Bibliographic records were extracted and imported into the Reference Manager Software (Thompson Reuters, San Francisco, Calif, USA) for removal of duplicate references. In level 1 screening, titles and abstracts of potentially relevant articles were screened by 2 independent reviewers using Covidence (Veritas Health Innovation, Melbourne, Australia). In level 2 screening, full-text copies of articles were obtained for those meeting the initial screening criteria. If an article was included by 1 reviewer and not the other at the title and abstract stage, the article was obtained for further review. Two independent reviewers examined all full-text articles. Any discrepancies were resolved with a discussion and consensus between the 2 reviewers or by a third reviewer if required.

After having the pool of eligible systematic reviews, we aimed to retain 1 review per outcome to reduce overlap in primary studies. The review that was retained was selected using a priority list that was designed to select the review that would provide the best evidence to inform sleep duration recommendations. Prioritization of systematic reviews to identify the best review for each outcome was as follows. First, studies that reported direct outcome measures were prioritized over studies that reported indirect markers of the outcome measures. Second, we prioritized reviews that examined the effect of age (e.g., if the effects were different in adults aged 18–64 years vs. ≥65 years) and dose-response (i.e., optimal sleep duration from dose–response curves). Third, if there was more than 1 review that addressed these criteria, we selected the review that was of highest quality based on full AMSTAR 2 assessment. Finally, if there were multiple reviews of high-quality we prioritized the most recent review. If a review did not address, or addressed only 1 of “age” or “dose–response”, we considered whether estimates of effect from separate reviews could be included to address these components. For example, if no reviews assessed the effect of age for the mortality outcome, we considered including estimates of effect from 1 systematic review that included adults aged 18–64 years, and estimates of effect from a second systematic review that included adults aged 65 years or older. The same strategy was applied for “dose–response”.

If estimates of effect from more than 1 systematic review needed to be included for a given outcome (i.e., to be able to evaluate the effect of age and/or dose–response), we assessed and reported on the degree of overlap in primary studies between systematic reviews using the corrected covered area (CCA) (Pieper et al. 2014). The extent of primary study overlap among the systematic reviews was interpreted as either slight (0%–5%), moderate (6%–10%), high (11%–15%), or very high (>15%) (Pieper et al. 2014). The degree of overlap was reported but not used as an exclusion criterion if papers analyzed the data differently (e.g., by age in 1 review and by dose–response in another review).

If there were no systematic reviews for a critical outcome, a de novo review was planned. De novo reviews were not planned if there were no systematic reviews for outcomes deemed important but not critical.

Detailed information about the methodology used to prepare this overview of reviews is provided in another paper published in this issue of the journal (Kho et al. 2020).

Data extraction
Microsoft Excel was used for data extraction. Data extraction was completed by 1 reviewer and verified by another reviewer. Where multiple models were reported, results from the most fully adjusted models were extracted. Important study features (e.g., author, publication year, country, study design, sample size, age, exposure, outcome, results, covariates, and quality of the evidence) were extracted. The optimal sleep duration was also extracted from dose–response curves, where available. We also extracted differences in effect by age, sex, race/ethnicity, socioeconomic status, weight status, and/or chronic disease status, where available. Reviewers were not blinded to the authors or journals when extracting data.

Risk of bias and study quality assessment
Quality assessment of included systematic reviews
Two reviewers independently assessed the methodological quality of each systematic review using the AMSTAR 2 rating scale (Shea et al. 2017). AMSTAR 2 contains 16 items to appraise the methodological aspects of reviews. All assessments were discussed and agreed upon based on discussion among the 2 reviewers (or in consultation with a third reviewer, if required). AMSTAR 2 ratings are “high”, “moderate”, “low”, and “critically low” quality. For the purposes of this review, we considered the following items noncritical: 2 (pre-registration of protocol), 3 (explanation of included study designs), 7 (justification for exclusion of individual studies), and 10 (reporting sources of funding for individual studies). The item on conflict of interest (COI) requires that COI for the systematic review and all primary studies be assessed. We modified this item to assess whether potential COI was documented only for the review itself.

*Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/apnm-2020-0034.
Quality assessment of primary studies within included systematic reviews

The quality assessment of primary studies, performed by the authors of the systematic reviews, was extracted and reported. The information was reported as indicated by the systematic review authors (e.g., “the authors concluded that the evidence was low to moderate quality”). In the event of a de novo search for a particular outcome, quality assessment using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) framework was planned.

Synthesis of results

Results were summarized via narrative synthesis, grouped by outcome. Results were described as reported by the systematic review authors, such as reporting available summary estimates and confidence intervals as well as the number of primary studies and participants that contributed to each available estimate. Tables were used to ensure consistency of data presentation across studies. A narrative description of any subgroup analyses by age, sex, race/ethnicity, socioeconomic status, weight status, and/or chronic disease status was also summarized.

Results

Description of studies

As reported in Fig. 1, a total of 2315 records were identified through database searches and 1257 records remained after removing duplicates. After titles and abstracts were screened, 130 full-text articles were obtained for further review and 36 articles met the inclusion criteria and were thus eligible for inclusion. Reasons for exclusion were as follows: wrong outcome (n = 34), wrong study design (n = 23), duplicate publication (n = 18), wrong exposure (n = 15), and wrong population (n = 4). A total of 11 articles were included in this overview of reviews after using the priority list to avoid overlap and with a goal to keep the best review per outcome. Among the 36 articles that were eligible for inclusion, reasons for excluding articles were the absence of dose-response relationships reported (n = 12), the lower priority of outcomes reported (i.e., directedness) (n = 8), and the lower quality reviews using AMSTAR 2 (n = 5). See Supplement S2 for the complete list of full-text articles excluded and the list of 25 reviews that were not chosen for this overview of reviews. Of note, the updated search strategy conducted in October 2019 did not result in relevant systematic reviews that would have been included in this overview of reviews.

Characteristics of studies sorted by outcome are summarized in Table 1 (critical outcomes) and Table 2 (important outcomes). Overall, the 11 reviews included 4,437,101 unique participants from 30 countries. The study designs used in the reviews were prospective cohort studies (78% of studies), cross-sectional studies (19% of studies), and case-control studies (3% of studies). Sleep was assessed subjectively in 96% of studies and objectively in 4% of studies. Sleep duration (hours) was reported “per night” or “per day” in the studies. The quality of the evidence, as reported in the reviews by the authors, ranged from low to high across health...
The pooled RR of the shortest and longest sleep duration vs. reference sleep duration (between 7 and 8 h for the majority of the S222 Appl. Physiol. Nutr. Metab. Vol. 45, 2020

Overview of key systematic reviews that examined the relationship between sleep duration and critical health outcomes in adults.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design and number of primary studies included</th>
<th>Population</th>
<th>Intervention/exposure and comparator</th>
<th>Outcome measurement</th>
<th>Main findings</th>
<th>Quality of the evidence</th>
<th>AMSTAR 2 rating and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yin et al. (2017)</td>
<td>Systematic review and dose–response meta-analysis of prospective cohort studies (n = 43) with 2.3 to 34 y of follow-up</td>
<td>N = 241 107 in 43 articles from 14 countries: USA (n = 10), Japan (n = 8), Taiwan (n = 4), UK (n = 4), China (n = 3), Finland (n = 2), Israel (n = 2), Korea (n = 2), Spain (n = 2), Sweden (n = 2), Australia (n = 1), Brazil (n = 1), Denmark (n = 1), and Netherlands (n = 1), Age: 18-98 y</td>
<td>Short and long sleep duration (in most studies the reference was between 7 and 8 h/d)</td>
<td>Subjective measurement: self-reported questionnaire (n = 26) and interview (n = 17)</td>
<td>The pooled RR of the shortest and longest sleep duration vs. reference sleep duration (between 7 and 8 h for the majority of the studies) was 1.13 (95% CI: 1.09–1.17), with low to moderate heterogeneity (I² = 27.0%, p &lt; 0.01), and 1.35 (95% CI: 1.29–1.40), with high heterogeneity (I² = 76.2%, p &lt; 0.01), respectively. U-shaped curvilinear association with the lowest risk of all-cause mortality at a sleep duration of about 7 h/d. The pooled RR for all-cause mortality was 1.06 (95% CI: 1.04–1.07) per 1-h reduction of sleep duration, with moderate to high heterogeneity (I² = 55.5%, p &lt; 0.01). For long sleep, a nonlinear association between long sleep duration and all-cause mortality was found (p = 0.02), and the pooled RR for all-cause mortality was 1.13 (95% CI: 1.10–1.15) per 1-h increase of sleep duration, with high heterogeneity (I² = 76.5%, p &lt; 0.01). Overall, this review demonstrated a U-shaped association between sleep duration and risk of all-cause mortality, with the lowest risk observed with approximately 7 h of sleep duration. Sleep durations that were too short or too long were significantly associated with elevated risks of all-cause mortality. Compared with 7 h/d, a 1-h decrease was associated with a 5% increased risk of all-cause mortality and a 1-h increase in sleep duration was associated with a 3% increased risk of all-cause mortality.</td>
<td>Quality scores of included studies ranged from 5/9 to 7/9 on the Newcastle–Ottawa scale (4–6 = moderate quality and 7–8 = high quality). Mean score = 6.9</td>
<td>Moderate</td>
</tr>
<tr>
<td>Itani et al. (2017)</td>
<td>Systematic review and meta-analysis of prospective cohort studies (n = 38) with 2.8 to 34 y of follow-up</td>
<td>N = 1 298 970 in 38 articles from 14 countries: Japan (n = 10), USA (n = 10), Sweden (n = 3), Taiwan (n = 3), UK (n = 3), Australia (n = 1), Brazil (n = 1), China (n = 1), Denmark (n = 1), Finland (n = 1), Israel (n = 1), Korea (n = 1), Netherlands (n = 1), and Spain (n = 1), Age: 20–98 y</td>
<td>Short sleep duration (in most cases &lt;5 or 6 h/d)</td>
<td>Objective measurement: polysomnography (n = 1) and actigraphy (n = 2). Subjective measurement: self-reported questionnaire (n = 23) and interview (n = 12)</td>
<td>Mortality Measurement: death certificate and medical record/registry</td>
<td>Meta-analysis contained 36 articles and indicated that compared with normal sleep (as reported in the original studies, short sleep showed a statistically significant increase in mortality due to all causes at an RR of 1.12 (95% CI: 1.10–1.15) per 1-h increment of sleep duration). No variation of effect by age or sex. Variation of effect by other variables not reported</td>
<td>Quality scores of included studies were all considered &quot;high&quot; quality on the Newcastle–Ottawa scale as they ranged from 7/9 to 9/9 (7–9 = high quality). Mean score = 8.0</td>
</tr>
<tr>
<td>Jike et al. (2018)</td>
<td>Systematic review and meta-analysis of prospective cohort studies (n = 37) with 2.8 to 34 y of follow-up</td>
<td>N = 1 296 470 in 37 articles from 14 countries: Japan (n = 10), USA (n = 8), Taiwan (n = 3), UK (n = 3), China (n = 2), Brazil (n = 2), Sweden (n = 2), Australia (n = 1), Denmark (n = 1), Finland (n = 1), Korea (n = 1), Netherlands (n = 1), and Spain (n = 1), Age: 19–98 y</td>
<td>Long sleep duration (in most cases &gt;9 h/d)</td>
<td>Objective measurement: actigraphy (n = 1). Subjective measurement: self-reported questionnaire (n = 24) and interview (n = 12)</td>
<td>Mortality Measurement: death certificate and medical record/registry</td>
<td>Meta-analysis contained 36 articles that indicated that compared with normal sleep duration (as reported in the original studies), long sleep duration was associated with a statistically significant increase in all-cause mortality, with an RR of 1.39 (95% CI: 1.33–1.47, p &lt; 0.001, F = 83%). Overall, long sleep duration (greater than 8 or 9 h) was associated with a 39% absolute increase in mortality compared with normal sleep. Variation of effect: No variation of effect by age or sex. Variation of effect by other variables not reported</td>
<td>Quality scores of included studies were all considered &quot;high&quot; quality on the Newcastle–Ottawa scale as they ranged from 7/9 to 9/9 (7–9 = high quality). Mean score = 8.0</td>
</tr>
</tbody>
</table>
Table 1 (continued).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design and number of primary studies included</th>
<th>Population</th>
<th>Intervention/exposure and comparator</th>
<th>Outcome measurement</th>
<th>Main findings</th>
<th>Quality of the evidence</th>
<th>AMSTAR 2 rating and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome: Incident cardiovascualr disease</strong></td>
<td>Wang et al. (2016) Systematic review and meta-analysis of prospective cohort studies (n = 17) with 5 to 31 y of follow-up</td>
<td>N = 527,440 in 17 articles from 11 countries: USA (n = 4), Japan (n = 3), Sweden (n = 2), Canada (n = 1), China (n = 1), Denmark (n = 1), England (n = 1), Europe (n = 1), Germany (n = 1), Netherlands (n = 1), and Singapore (n = 1) Age: 18–90 y</td>
<td>Short and long sleep (in most studies the reference was between 7 to 8 h/d) Subjective measurement: self-reported by questionnaire for all studies</td>
<td>Incidence of cardiovascular disease Measurement: confirmed by hospitalization/death certificates, autopsy, information from coroner, medical registry, National Medical Registry withdrawal history files</td>
<td>The meta-analysis contained 14 samples from 10 articles for the short sleep duration and 11 samples from 8 articles for the long sleep duration, which indicated that compared with normal sleep of 7 h/d, both short and long sleep duration showed a statistically significant increase in the incidence of cardiovascular heart diseases at a RR of 1.11 (95% CI: 1.05–1.16) and 1.07 (95% CI: 1.00–1.15), respectively. These findings showed a U-shaped relationship between sleep duration and risk of cardiovascular heart disease, with the lowest risk at 7–8 h/d. Overall, this review demonstrated a U-shaped association between sleep duration and cardiovascular heart disease, with the lowest risk observed with approximately 7 h of sleep duration. Sleep duration that were too short or too long were significantly associated with elevated risks of cardiovascular heart disease. Compared with 7 h of sleep per day, a 1-h decrease in sleep duration was associated with 11% increased risk of cardiovascular heart disease and a 1-h increase in sleep duration was associated with 7% increased risk of cardiovascular heart disease. Variation of effect: No variation of effect by sex. Variation of effect by other variables not reported.</td>
<td>Quality scores of included studies ranged from 6/9 to 9/9 on the Newcastle-Ottawa scale (6 = moderate quality and 7–9 = high quality). Mean score = 7.2</td>
<td>Moderate More than 1 noncritical weakness (authors did not report if selection was performed in duplicate and did not report on the sources of funding for the studies included)</td>
</tr>
<tr>
<td><strong>Outcome: Incident type 2 diabetes</strong></td>
<td>Shan et al. (2015) Systematic review and meta-analysis of prospective cohort studies (n = 10) with 2 to 17 y of follow-up</td>
<td>N = 482,502 in 10 articles from 4 regions: USA (n = 5), Asia (n = 2), Europe (n = 2), and Australia (n = 1) Age: 19 to 86 y</td>
<td>Short and long sleep duration (in most studies the reference was between 7 to 8 h/d) Subjective measurement: self-reported by questionnaire for all studies</td>
<td>Incidence of type 2 diabetes Measurement: hospital record, self-report of physician diagnosis, fasting plasma glucose or prescribed medications, WHO criteria, multiple methods</td>
<td>A meta-analysis of 9 samples from 8 articles for short sleep duration and 6 samples from 6 articles for long sleep duration revealed that compared with normal sleep of 7 h of sleep per day, both short and long sleep duration showed a statistically significant increase in the incidence of type 2 diabetes, with RR of 1.09 (95% CI: 1.04–1.15) for each one hour decrease and 1.14 (95% CI: 1.03–1.26) for each 1 h increase. This review reported a U-shaped dose–response relationship between sleep duration and risk of type 2 diabetes, with the lowest risk observed at a sleep duration of 7 to 8 h/d. Overall, this review demonstrated a U-shaped association between sleep duration and the incidence of type 2 diabetes, with the lowest risk observed with approximately 7 to 8 h of sleep duration. Sleep durations that were too short or too long were significantly associated with elevated risks of type 2 diabetes. Compared with 7 h/d, a 1-h decrease was associated with 9% increased risk of type 2 diabetes and a 1-h increase in sleep duration was associated with 14% increased risk of type 2 diabetes. Variation of effect: Variation of effect by sex, race/ethnicity, socioeconomic status, weight status, or chronic disease status not reported</td>
<td>Quality scores of included studies ranged from 6/9 to 8/9 on the Newcastle-Ottawa scale (6 = moderate quality and 7–9 = high quality). Mean score = 6.7</td>
<td>Moderate More than 1 noncritical weakness (authors did not report all the potentially relevant information concerning the PICOS, whether data extraction was performed in duplicate, and did not report on the sources of funding for the studies included)</td>
</tr>
</tbody>
</table>
Table 1 (continued).

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design and number of primary studies included</th>
<th>Population</th>
<th>Intervention/exposure and comparator</th>
<th>Outcome measurement</th>
<th>Main findings</th>
<th>Quality of the evidence</th>
<th>AMSTAR 2 rating and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome: Mental health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jike et al. (2018)</td>
<td>Systematic review of prospective cohort study (n = 2) with 5 y of follow-up</td>
<td>N = 15 204 in 1 article (~40% male) from the USA</td>
<td>Long sleep duration (&gt;8 h/d) Subjective measurement: self-reported by questionnaire (n = 2)</td>
<td>Incidence of depression Measurement: total PHQ-9 score ≥ 5</td>
<td>Short sleep duration was not significantly associated with the incidence of depression compared with normal sleep as defined in the original studies (RR: 1.34, 95% CI: 1.15–1.56) and long sleep duration (RR: 1.21, 95% CI: 1.06–1.39) were associated with a higher incidence of cognitive disorders. The reference category was the middle category of sleep duration, closest to 7 h/d or night. A U-shaped dose-response relationship was observed with 7–8 h of sleep as the optimal duration. Stratified analyses showed that the association between short sleep duration and cognitive disorders was more robust with self-reported sleep than objectively measured sleep. Overall, the lowest incident risk of cognitive disorders was found at the sleep duration of 7 to 8 h/d</td>
<td>Quality scores of the included studies ranged between 6/9 and 8/9 on the Newcastle–Ottawa scale (6 = moderate quality and 7–8 = high quality). Mean score = 6.9</td>
<td>Low One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search and did not provide a list of the full texts that were excluded)</td>
</tr>
<tr>
<td>Outcome: Brain health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lo et al. (2015)</td>
<td>Systematic review and meta-analysis of observational studies (n = 18); this included cross-sectional (n = 11) and prospective cohort studies (n = 7) with up to 22.1 y of follow-up</td>
<td>N = 97 264 in 18 articles from 14 countries: USA (n = 5), China (n = 2), Finland (n = 2), Spain (n = 2), Canada (n = 1), UK (n = 1), France (n = 1), Germany (n = 1), Singapore (n = 1), and 1 multi-country (China, Ghana, India, Mexico, Russia, South Africa, and UK)</td>
<td>Sleep duration (reference was 7–8 h/d of sleep across most studies) Subjective measurement: self-reported by questionnaire (n = 7)</td>
<td>Cognitive function including complex cognition, executive functions, verbal memory, working memory, and processing speed Measurement: assessed objectively using different cognitive performance tasks</td>
<td>Compared with the reference category, both short sleep duration (RR: 1.34, 95% CI: 1.15–1.56) and long sleep duration (RR: 1.21, 95% CI: 1.06–1.39) were associated with a higher incidence of cognitive disorders. The reference category was the middle category of sleep duration, closest to 7 h/d or night. A U-shaped dose-response relationship was observed with 7–8 h of sleep as the optimal duration. Stratified analyses showed that the association between short sleep duration and cognitive disorders was more robust with self-reported sleep than objectively measured sleep. Overall, the lowest incident risk of cognitive disorders was found at the sleep duration of 7 to 8 h/d</td>
<td>Quality scores of the included studies ranged between 6/9 and 8/9 on the Newcastle–Ottawa scale (6 = moderate quality and 7–8 = high quality). Mean score = 6.9</td>
<td>Low One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search and did not provide a list of the full texts that were excluded)</td>
</tr>
<tr>
<td>Outcome: Cognitive functions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al. (2018)</td>
<td>Systematic review and dose–response meta-analysis of prospective cohort studies (n = 6) with 5 y of follow-up</td>
<td>N = 22 187 in 9 articles from 6 countries: USA (n = 4), Canada (n = 1), and 1 multi-country (China, Ghana, India, Mexico, Russia, South Africa, and UK)</td>
<td>Sleep duration in most studies the reference was around 7 h/d Objective measurement: actigraphy (n = 2) Subjective measurement: self-reported by questionnaire (n = 7)</td>
<td>Measurement: diagnostics from the DSMIV for dementia; screen to identify potential dementia cases, TICS, or NINCDS-ADRDA for Alzheimer’s disease; Chinese MMSE score, SMS score, or visit 3 sleep Visit change score for cognitive decline; and poor performance on CERAD test, MMSE score, or SMS score for cognitive impairment</td>
<td>Compared with the reference category, both short sleep duration (RR: 1.34, 95% CI: 1.15–1.56) and long sleep duration (RR: 1.21, 95% CI: 1.06–1.39) were associated with a higher incidence of cognitive disorders. The reference category was the middle category of sleep duration, closest to 7 h/d or night. A U-shaped dose-response relationship was observed with 7–8 h of sleep as the optimal duration. Stratified analyses showed that the association between short sleep duration and cognitive disorders was more robust with self-reported sleep than objectively measured sleep. Overall, the lowest incident risk of cognitive disorders was found at the sleep duration of 7 to 8 h/d</td>
<td>Quality scores of the included studies ranged between 6/9 and 8/9 on the Newcastle–Ottawa scale (6 = moderate quality and 7–8 = high quality). Mean score = 6.9</td>
<td>Low One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search and did not provide a list of the full texts that were excluded)</td>
</tr>
</tbody>
</table>
One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search strategy and did not provide a list of the full texts that were excluded).

Overall, both short and long sleep duration are associated with falls, with the lowest risk of falls at 7–8 h/d. Variation of effect:

- No variation of effect by age.
- Stratified analyses by sex and race/ethnicity did not significantly affect the associations between short sleep duration and falls. Subgroup analysis by race/ethnicity significantly affected the association between long sleep duration and falls, with a stronger association among Caucasians (OR: 1.69, 95% CI: 1.36–2.08) compared with Asians (OR: 1.23, 95% CI: 1.04–1.45).

**Outcome: Falls**

- **Study design and number of primary studies included:** Systematic review and meta-analysis of observational studies (n = 7) including cross-sectional (n = 6) and cohort (n = 1) study designs.
- **Population:** N = 232,829 in 7 articles from 6 countries: South Korea (n = 2), Australia (n = 1), Japan (n = 1), Spain (n = 1), Taiwan (n = 1), and USA (n = 1).
- **Age:** 59–80 y.
- **Sleep duration (reference category was around 7–8 h/d of sleep across most studies):** Falls defined as an unintentional change in position resulting in coming to rest at a lower level or on the ground. Studies reporting that the participants experienced at least 1 fall according to the above definition during the follow-up period were included in the meta-analysis, regardless of the location where falls occur.
- **Outcome measurement:** self-reported by questionnaire for all studies.
- **Main findings:** Compared with the reference category, both short (OR: 1.32, 95% CI: 1.21–1.46) and long sleep duration (OR: 1.35, 95% CI: 1.17–1.56) were associated with falls. A U-shaped curve was observed in the dose-response analysis, and the lowest risk of falls was 7–8 h of sleep per day.
- **Quality of the evidence:** Using the guidelines of Methodological Evaluation of Observational Research and the quality assessment tool of observational cohort and cross-sectional studies from the National Heart, Lung, and Blood Institute, the quality scores of studies ranged between 10 and 14. The scale ranges from 0 to 17. Mean score = 11.9.
- **AMSTAR 2 rating and rationale:** low One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search strategy and did not provide a list of the full texts that were excluded).

**Outcome: Accidents/injuries**

- **Study design and number of primary studies included:** Systematic review and meta-analysis of observational studies (n = 27) including cross-sectional (n = 16), case-control (n = 7), and prospective cohort studies (n = 4) with 3 to 10 y of follow-up.
- **Population:** N = 268,332 in 27 articles from 15 countries: USA (n = 15), France (n = 4), Sweden (n = 3), Canada (n = 2), Israel (n = 2), Japan (n = 2), Australia (n = 1), Ethiopia (n = 1), Finland (n = 1), India (n = 1), Italy (n = 1), Malaysia (n = 1), New Zealand (n = 1), Pakistan (n = 1), and UK (n = 1).
- **Age:** 15–97 y.
- **Sleep duration (short sleep duration defined as <6 or <7 h/d in most cases):** Sleep duration defined as <6 or <7 h/d in most cases.
- **Objective measurement:** Polysomnography (n = 2).
- **Subjective measurement:** self-reported by questionnaire (n = 19) or interview (n = 6).
- **Work injury of any severity (minor, major, or fatal):** Work injuries were registered by the company’s administration or an official body, self-reported, or diagnosed by a physician.
- **Outcome measurement:** self-reported.
- **Main findings:** This meta-analysis of 27 studies revealed that workers with short sleep duration were at a significantly higher risk of being injured than workers without short sleep duration (RR: 1.35, 95% CI: 1.16–1.58, p = 0.001). Approximately 13% of work injuries could be attributed to sleep problems (including short sleep duration).
- **Quality of the evidence:** Quality scores of included studies were considered high in 20 studies (74%) and in 7 studies (26%) were considered of poor quality based on a modified version of the Newcastle–Ottawa scale.
- **AMSTAR 2 rating and rationale:** low One critical flaw with noncritical weaknesses (authors did not report some of the information needed in the literature search strategy and did not provide a list of the full texts that were excluded).

**Note:** 3MS, Modified Mini-Mental State; CERAD, Consortium to Establish a Registry for Alzheimer’s Disease; CI, confidence interval; DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, 4th edition; MMSE, Mini-Mental State Examination; NINCDS-ADRDA, National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association; OR, odds ratio; PHQ-9, patient health questionnaire with 9 items; PICOS, Participants, Interventions, Comparisons, Outcomes, and Study design; RR, relative risk; TICS, telephone interview for cognitive status; WHO, World Health Organization.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Study design and number of primary studies included</th>
<th>Population</th>
<th>Intervention/exposure and comparator</th>
<th>Outcome measurement</th>
<th>Main findings</th>
<th>Quality of the evidence</th>
<th>AMSTAR 2 rating and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome: Incident obesity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Jike et al. (2018)</strong></td>
<td>Systematic review and dose-response meta-analysis of prospective cohort studies (n = 15) with 1 to 16 y of follow-up</td>
<td>N = 318 437 in 15 articles from 8 countries: Japan (n = 6), Spain (n = 2), USA (n = 2), UK (n = 1), Canada (n = 1), Finland (n = 1), Korea (n = 1), and Thailand (n = 1) Age: 18–90 y</td>
<td>Long sleep duration (in most cases &gt;9 h/d)</td>
<td>Incidence of obesity (mean weight gain relative to reference category, BMI ≥25 kg/m² and ≥30 kg/m², body weight increase &gt;5 kg, and WC of ≥90 cm for men and ≥85 cm for women)</td>
<td>A meta-analysis of 13 articles revealed that compared with normal sleep, long sleep duration was significantly associated with increased incidence of obesity among female (RR = 1.10, 95% CI: 1.00–1.26 for female vs. RR = 1.10, 95% CI: 1.00–1.26 for male). Variation of effect by other variables not reported</td>
<td>Quality scores of included studies ranged from 5/9 to 9/9 on the Newcastle-Ottawa scale (4–6 = moderate quality, 7–9 = high quality). Mean score = 6.1</td>
<td>High One noncritical weakness (authors did not report on the sources of funding for the studies included)</td>
</tr>
<tr>
<td><strong>Outcome: Biomarkers of cardiometabolic risk</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Li et al. (2019)</strong></td>
<td>Systematic review and dose-response meta-analysis of prospective cohort studies (n = 9) with 2 to 10 y of follow-up</td>
<td>N = 48 525 in 9 articles from 4 countries: China (n = 3), South Korea (n = 2), USA (n = 3), and Spain (n = 1) Age: 18–46 y</td>
<td>Short and long sleep duration (reference category was around 7–8 h/d of sleep across most studies)</td>
<td>Incidence of hypertension was defined as a blood pressure ≥140/90 mm Hg, taking anti-hypertensive medications, or self-reported</td>
<td>This meta-analysis containing 9 articles revealed that compared with 7 h of sleep (reference category), 6 h and ≤5 h of sleep per night were associated with higher incidence of hypertension, while 9 h of sleep per night was associated with a decreased risk of incidence of hypertension. There was no significant difference between the 7 h and &gt;9 h groups. Overall, this dose–response meta-analysis showed that hypertension incidence was higher with shorter sleep (≤6 h/d) durations than 7 to 9 h of sleep.</td>
<td>Quality scores of included studies ranged between 6/9 and 8/9 on the Newcastle–Ottawa scale (4–6 = moderate quality, 7–9 = high quality). Mean score = 7.1</td>
<td>Low One critical flaw with noncritical weaknesses (authors did not report on some of the information needed in the literature search strategy and did not provide a list of the full texts that were excluded along with the justifications)</td>
</tr>
</tbody>
</table>
Overall, this meta-analysis suggests that long sleep duration (8 h or more per day) is associated with a higher risk of osteoporosis in middle-aged and elderly women. Variation of effect: No variation of effect. Quality score of included studies ranged from 7/9 to 8/9 on the Newcastle–Ottawa scale (defined as medium quality by the authors). Mean score = 7.2.

## Data synthesis

### Critical outcomes

#### Mortality

We included 3 reviews examining the association between sleep duration and mortality in adults (Table 1). Yin et al. (2017) showed that compared with 7 h of sleep per day, a 1-h decrease in sleep duration was associated with 6% increased risk of all-cause mortality and a 1-h increase in sleep duration was associated with a 13% increased risk of all-cause mortality (N = 241 107 adults in 43 articles). Itani et al. (2017) showed that short sleep duration (<6 h/day) was associated with a 12% absolute increase in mortality risk compared with normal sleep (N = 1 298 970 adults in 38 articles) while Jike et al. (2018) showed that long sleep duration (>8 or 9 h/day) was associated with a 39% absolute increase in mortality compared with normal sleep (N = 1 296 470 adults in 37 articles). These last 2 reviews were published by the same authors; they decided to publish 1 systematic review on “short sleep” and another one on “long sleep” in 2 different publications instead of 1. The extent of primary study overlap was thus very high (CCA: 65.7%). The quality of the evidence and AMSTAR 2 ratings were moderate-to-high.

#### Incident cardiovascular disease

Wang et al. (2016) reported that compared with 7 h of sleep per day, a 1-h decrease in sleep duration was associated with an 11% increased risk of incident cardiovascular heart disease and a 1-h increase in sleep duration was associated with 7% increased risk of incident cardiovascular heart disease (N = 517 440 adults in 17 articles) (Table 1). The quality of the evidence was moderate-to-high and the AMSTAR 2 rating was moderate.

#### Incident type 2 diabetes

Shan et al. (2015) showed that compared with 7 h per day, a 1-h decrease in sleep duration was associated with 9% increased risk of incident type 2 diabetes and a 1-h increase in sleep duration was associated with 14% increased risk of incident type 2 diabetes (N = 482 502 adults in 10 articles) (Table 1). The quality of the evidence was moderate-to-high and the AMSTAR 2 rating was moderate.

#### Mental health (incident depression)

Itani et al. (2017) reported that short sleep duration (<6 or ≤7 h/day) was not significantly associated with the incidence of depression compared with normal sleep duration (N = 16 257 adults in 2 articles) (Table 1). Jike et al. (2018) reported that long sleep duration (>8 h/day) was not significantly associated with the incidence of depression compared with normal sleep duration (N = 15 204 adults in 1 article). These 2 reviews were published by the same authors but were split into 2 publications (1 focusing on short sleep and 1 focusing on long sleep). The extent of primary study overlap was thus very high (CCA: 50%). The quality of the evidence was moderate-to-high and the AMSTAR 2 rating was high.

#### Brain health (incident cognitive disorders)

Wu et al. (2018) showed that the lowest incident risk of cognitive disorders was found at a sleep duration of 7 to 8 h per day (N = 22 187 adults in 9 articles) (Table 1). The quality of the evidence was moderate-to-high and the AMSTAR 2 rating was low.

#### Cognitive function

Lo et al. (2016) reported that extreme sleep durations (both short and long) were associated with cognitive decline compared with the reference sleep of 7 to 8 h/day (N = 97 264 adults in 18 articles) (Table 1). The quality of the evidence was judged “satisfactory” with a modified version of the Downs and Black Quality Index score system and the AMSTAR 2 rating was low.
Falls
Wu and Sun (2017) showed that both short and long sleep duration were associated with falls, with the lowest risk of falls at 7–8 h per day (N = 212,829 adults in 7 articles) (Table 1). The quality scores of included studies ranged between 10 and 14 using a scale ranging from 0 to 17 and the AMSTAR 2 rating was low.

Accidents and injuries
Uehli et al. (2014) reported that workers with shorter sleep durations (<6 or <7 h/day) significantly increased the risk of work-related injuries compared with workers without short sleep durations (N = 268,332 adults in 27 articles) (Table 1). The quality of the evidence ranged from poor to high, depending on the outcome, and the AMSTAR 2 rating was low.

Important outcomes

Incident obesity
Itani et al. (2017) showed that short sleep duration (<6 h/day) was associated with a 38% absolute increase in the incidence of obesity (N = 322,842 adults in 21 articles) while Jike et al. (2018) showed that long sleep duration (>9 h/day) was associated with an 8% absolute increase in the incidence of obesity compared with normal sleep duration (N = 318,437 adults in 15 articles) (Table 1). These 2 reviews were published by the same authors but were split into 2 publications (1 focusing on short sleep and 1 focusing on long sleep). The extent of primary study overlap was thus very high (CCA: 71.4%). The quality of the evidence was moderate-to-high and the AMSTAR 2 ratings were high.

Biomarkers of cardiometabolic risk (incident hypertension)
Li et al. (2019) showed that the incidence of hypertension was higher with shorter sleep durations (<6 h/day) compared with 7 to 9 h of sleep (N = 48,525 adults in 9 articles) (Table 2). The quality of the evidence was moderate-to-high and the AMSTAR 2 rating was low.

Bone health (risk of osteoporosis)
Moradi et al. (2017) reported that long sleep duration (≥8 h/day) was associated with a higher risk of osteoporosis in middle-aged and elderly women compared with normal sleep (N = 72,326 adults in 6 articles) (Table 2). The quality of the evidence was moderate and the AMSTAR 2 rating was low.

Health-related quality of life
No systematic review was identified that examined the association between sleep duration and health-related quality of life.

Work productivity
No systematic review was identified that examined the association between sleep duration and work productivity.

Physical activity and sedentary behaviour
No systematic review examined the association between sleep duration and physical activity or sedentary behaviour.

Summary of findings
A high-level summary of findings by health outcome can be found in Table 3. This table shows the optimal sleep duration from evidence of dose–response curves as well as the possible variation of effect by age for each outcome. Overall, the optimal sleep duration from dose–response curves appears to be at 7–8 h per day. We observed that 60% of reviews (n = 9) did not report a possible variation of effect by age, 27% (n = 4) did report that there were no modification of the effect by age, and 13% (n = 2) showed a significant relationship in adults aged <65 years only (obesity outcome).

| Table 3. Optimal sleep duration from evidence of dose–response curves and variation of the effect by age for each outcome. |
|Reference | Outcome | Optimal sleep duration from dose–response curves and variation of the effect by age |
|Yin et al. (2017) | Mortality | 7 h/d | Variation of the effect by age not reported |
|Itani et al. (2017) | Mortality | No dose–response (only short sleep duration examined) | No variation of the effect by age |
|Jike et al. (2018) | Mortality | No dose–response (only long sleep duration examined) | No variation of the effect by age |
|Wang et al. (2016) | Incident cardiovascular disease | 7–8 h/d | Variation of the effect by age not reported |
|Shan et al. (2015) | Incident type 2 diabetes | 7–8 h/d | Variation of the effect by age not reported |
|Itani et al. (2017) | Incident depression | No dose–response (only short sleep duration examined) | Variation of the effect by age not reported |
|Jike et al. (2018) | Incident depression | No dose–response (only long sleep duration examined) | Variation of the effect by age not reported |
|Wu et al. (2018) | Incident cognitive disorders | 7–8 h/d | Variation of the effect by age not reported |
|Lo et al. (2016) | Cognitive functions | 7–8 h/d | No variation of the effect by age |
|Wu and Sun (2017) | Falls | 7–8 h/d | No variation of the effect by age |
|Uehli et al. (2014) | Accidents/injuries | No dose–response | Variation of the effect by age not reported |
|Itani et al. (2017) | Incident obesity | No dose–response (only short sleep duration examined) | Increased risk of obesity in short sleepers <65 y and not among those ≥65 y |
|Jike et al. (2018) | Incident obesity | No dose–response (only long sleep duration examined) | Increased risk of obesity in long sleepers <65 y and not among those ≥65 y |
|Li et al. (2019) | Incident hypertension | 7 h or more | Variation of the effect by age not reported |
|Moradi et al. (2017) | Risk of osteoporosis | | Variation of the effect by age not reported |
Variation of the effect by sex or other factors

Another objective of this review was to examine if the associations between sleep duration and outcomes differed not only by age but also by sex, race/ethnicity, socioeconomic status, weight status, and/or chronic disease status by looking at subgroup analyses and/or effect modification. No reviews reported information about a possible variation of effect by socioeconomic status, weight status, or chronic disease status. Among the reviews that reported a possible variation of effect by sex, \( n = 9 \) reported no variation of effect by sex while \( n = 1 \) (Jike et al. 2018) reported that the association between long sleep duration and incident obesity was significant among female (RR = 1.12, 95% confidence interval (CI): 1.06–1.18) but not male (RR = 0.97, 95% CI: 0.92–1.03) participants. With regard to race/ethnicity, 1 study reported no variation of effect (Wu et al. 2018) while another study (Wu and Sun 2017) found a stronger association between short sleep duration and falls among Caucasians (odds ratio (OR) = 1.69, 95% CI: 1.36–2.08) compared with Asians (OR = 1.23, 95% CI: 1.04–1.45).

Discussion

This overview of reviews synthesized peer-reviewed scientific evidence from 11 systematic reviews examining the associations between sleep duration and health outcomes in adults aged 18 years and older. The reviews included over 4 million participants from 30 countries, and the majority of studies (78%) were prospective cohort studies that relied on self-reported sleep (96% of studies) for the assessment of sleep duration. The overall quality of evidence (primary studies contained within the reviews) ranged from low to high across outcomes, and AMSTAR 2 ratings (quality of systematic reviews) also ranged from low to high. Collectively, a U-shaped association between sleep duration and health outcomes was observed. Evidence from dose–response curves showed that the sleep duration that was most favourably associated with the health outcomes examined was around 7–8 h per day. Among the critical outcomes examined, no reviews reported a possible modification of the effect by age.

A key observation when looking at the available evidence was the reliance on self-reported sleep duration measures (96% of studies). Sleep questions typically used in population health surveys tend to overestimate actual sleep duration compared with objective measures and can introduce inaccuracies (Girschik et al. 2012). Using an objective measure when assessing sleep may be especially important for adults with poor sleep quality because time in bed, as assessed with sleep duration questionnaire items, is not the same as actual sleep duration as not all time in bed is spent asleep (Chaput et al. 2018). There is a growing popularity in the use of actigraphy/accelerometry in epidemiologic research, which may have implications for sleep duration recommendations in the future (e.g., lower optimal sleep duration if objectively assessed compared with self-reported time in bed for example). With advances in wearable health technologies (e.g., Fitbit, Apple Watch), measures of sleep duration can be readily derived and tracked by the general public and future studies will need to re-examine dose–response curves with the use of other and more objective sleep assessments (and actigraphy assessment is not without limitations).

A goal of this overview of reviews was to provide evidence that could be used to better inform sleep duration recommendations for public health guidance. Currently, the National Sleep Foundation recommends 7–9 h of sleep per day for adults (18–64 years) and 7–8 h for older adults (≥65 years) (Hirshkowitz et al. 2015) while the American Academy of Sleep Medicine and Sleep Research Society recommends ≥7 h per night for adults aged 18–60 years (Watson et al. 2015). Our findings are more in line with the National Sleep Foundation recommendations and suggest that 7–8 h of sleep per day is the optimal amount based on dose-response curves. While public health guidelines are informed by the best available evidence, they are also informed by expert consensus, stakeholder consultations, and consideration of values and preferences, applicability, feasibility, and equity (Tremblay et al. 2016). For example, the well-known inter-individual variability in sleep needs (Chaput et al. 2018) means that 9 h of sleep per day may well be the optimal duration for a given individual; it may thus be ill-advised to recommend less sleep to this individual even if the sleep duration recommendations say “7–8 h per day”. This highlights the potential complexity of conveying such a message in public health because it is never a one-size-fits-all approach and we need to adapt our recommendations on a case-by-case basis.

Our findings clearly show the presence of a U-shaped association between sleep duration and health outcomes, with both short and long sleep durations being associated with adverse health outcomes. While there is a large body of evidence providing biological plausibility for short sleep as causally related to adverse health, the role of long sleep is unclear (Chaput et al. 2018). Long sleep is generally associated with other health problems (e.g., depression, chronic pain, obstructive sleep apnea), which may confound the association between sleep duration and health. Thus, the observed association between long sleep and poor health may reflect reverse causation and residual confounding (Knutson and Turek 2006; Stamatakis and Punjabi 2007). The absence of plausible biological mechanisms by which long sleep per se can cause adverse health was a key factor for the American Academy of Sleep Medicine and Sleep Research Society to recommend a threshold value (≥7 h) for sleep duration rather than a range. However, long sleep duration may be a useful marker of other possible health problems and may also be indicative of poor sleep efficiency (i.e., spending a lot of time in bed but with low quality sleep).

Primary studies included in the systematic reviews synthesized herein included both nighttime sleep duration and 24-h sleep duration depending on the tool used, and all presented results in terms of “hours per day”. A meta-analysis of prospective cohort studies using both nighttime sleep duration and 24-h sleep duration reported similar findings with all-cause mortality in adults (Shen et al. 2016). However, recent systematic reviews that examined daytime napping only and mortality concluded that naps longer than 30 min per day were associated with an increased risk of mortality (da Silva et al. 2016; Yamada et al. 2015; Zhong et al. 2015). Napping is also discouraged among people with insomnia as it can reduce the homeostatic sleep drive and perpetuate nighttime insomnia (Robbins et al. 2019). However, naps are encouraged to counteract sleep loss and achieve more sleep in some populations such as shiftworkers (Faraut et al. 2017). Napping is a heterogeneous and complex sleep habit that is beyond the scope of this overview of reviews and that has been recently reviewed in detail elsewhere (Faraut et al. 2017).

There were no relevant systematic reviews identified that examined the associations between sleep duration and health-related quality of life, work productivity, or physical activity/sedentary behaviour. We also excluded reviews with clinical populations (e.g., exclusively in patients with sleep disorders) to focus on community-dwelling adults in the general population. The present overview of reviews focused on sleep duration only and did not include other important characteristics of sleep health such as sleep quality, sleep timing, sleep consistency, daytime alertness, and the absence of sleep disorders (Buysses 2014; Chaput and Shiau 2019). Future studies are thus needed to broaden the scope of this overview of reviews.

We also examined whether the associations between sleep duration and health outcomes varied as a function of age, sex, race/ethnicity, socioeconomic status, weight status, and/or chronic disease status. Unfortunately, many systematic reviews did not conduct subgroup analyses, making it difficult to know whether the findings observed can be generalized to all community-
dwellings adults. However, among the reviews that included subgroup analyses, the findings were not different based on age or sex of the participants.

A number of limitations should be highlighted. First, the available evidence largely consisted of studies relying on self-reported sleep (96% of studies), with different categorization of sleep durations across studies. Second, many reviews did not report subgroup analyses, so it is difficult to determine whether the relationships observed apply broadly to all adults in the general population. Third, this review excluded studies that targeted clinical populations exclusively as well as those that examined other characteristics of sleep health. Fourth, reviews pooled studies that assessed sleep duration “per night” and “per day” together so the influence of napping on the observed relationships is not known. Fifth, a single measure of sleep duration at 1 point in time may not fully capture the chronic effects of sleep duration over time when related to long-term disease incidence. Sixth, this review only included systematic reviews published in English or French and conducted over the past 10 years. However, excluding non-English publications from evidence syntheses does not impact conclusions according to a recent meta-epidemiologic study (Nussbaumer-Streit et al. 2020). Finally, the risk of publication bias (i.e., an overrepresentation of studies with significant findings) cannot be discarded.

Conclusion

A comprehensive body of evidence supports the presence of a U-shaped association between sleep duration and health outcomes in adults. Dose–response curves showed that the sleep duration that was most favorably associated with the health outcomes that were examined was around 7–8 h per day in adults, with no apparent modification of the effect by age in the few studies that looked at it. The present overview of reviews is important to inform public health recommendations around healthy sleep duration of adults. Given the growing interest in using actigraphy/accelerometry in population health research, future studies will need to assess the impact of this approach on the measured associations between sleep duration and health outcomes. Future work is also needed to better examine whether sleep duration recommendations should vary between adults and older adults.

Conflict of interest statement

T.J.S. reports grants from the Public Health Agency of Canada during the conduct of the study and personal fees from the Public Health Agency of Canada and the PEI Public Schools Branch. V.J.P. reports personal fees from the Canadian Society for Exercise Physiology during the conduct of the study and is a Canadian Agency for Drugs and Technologies in Health (CADTH) employee. The current work was unrelated to her employment, and CADTH had no role in the funding, design, or oversight of the work reported. M.E.K. reports personal fees and nonfinancial support from the Canadian Society for Exercise Physiology during the conduct of the study. A.R.W. reports personal fees from ProQuest LLC, outside of the submitted work. J.C. reports grants from Canopy Growth, Rana, Philips/Respironics, Merck, and Eisai, outside the submitted work. The remaining authors declare that they have no conflicts of interest.

Acknowledgements

This study has been made possible through funding from the Public Health Agency of Canada, the Canadian Society for Exercise Physiology, and Queen’s University. The funder had no role in the design, interpretation, and publication of this manuscript.

References