Quantity and Quality of Sleep and Incidence of Type 2 Diabetes

A systematic review and meta-analysis

Francesco P. Cappuccio, frcp1
Lanfranco D’Elia, md2
Pasquale Strazzullo, md2
Michelle A. Miller, phd3

OBJECTIVE — To assess the relationship between habitual sleep disturbances and the incidence of type 2 diabetes and to obtain an estimate of the risk.

RESEARCH DESIGN AND METHODS — We conducted a systematic search of publications using MEDLINE (1955–April 2009), EMBASE, and the Cochrane Library and manual searches without language restrictions. We included studies if they were prospective with follow-up >3 years and had an assessment of sleep disturbances at baseline and incidence of type 2 diabetes. We recorded several characteristics for each study. We extracted quantity and quality of sleep, how they were assessed, and incident cases defined with different validated methods. We extracted relative risks (RRs) and 95% CI and pooled them using random-effects models. We performed sensitivity analysis and assessed heterogeneity and publication bias.

RESULTS — We included 10 studies (13 independent cohort samples; 107,756 male and female participants, follow-up range 4.2–32 years, and 3,586 incident cases of type 2 diabetes). In pooled analyses, quantity and quality of sleep predicted the risk of development of type 2 diabetes. For short duration of sleep (≤5–6 h/night), the RR was 1.28 (95% CI 1.03–1.60, P = 0.024, heterogeneity P = 0.015); for long duration of sleep (>8–9 h/night), the RR was 1.48 (1.13–1.96, P = 0.005); for difficulty in initiating sleep, the RR was 1.57 (1.25–1.97, P < 0.0001); and for difficulty in maintaining sleep, the RR was 1.84 (1.39–2.43, P < 0.0001).

CONCLUSIONS — Quantity and quality of sleep consistently and significantly predict the risk of the development of type 2 diabetes. The mechanisms underlying this relation may differ between short and long sleepers.

Sleep patterns of quantity and quality are affected by a variety of cultural, social, psychological, behavioral, pathophysiological, and environmental influences. Changes in modern society include longer working hours, more shift-work, and 24–7 availability of commodities. These changes are paralleled by secular trends of curtailed duration of sleep to fewer hours per day across westernized populations (1). These trends have led to increased reporting of fatigue and tiredness and excessive daytime sleepiness (2). Lack of sleep exerts deleterious effects on a variety of systems with detectable changes in metabolic (3,4), endocrine (5,6), and immune pathways (7). Short-term, acute, laboratory, and cross-sectional observational studies indicate that disturbed or reduced sleep is associated with glucose intolerance, insulin resistance, reduced acute insulin response to glucose, and a reduction in the disposition index (4), thus predisposing individuals to type 2 diabetes. The causality of the association and the generalizability of the results to longer-term effects of sustained sleep disturbances have been studied in prospective population studies to establish a temporal sequence between exposure and outcome. Because of the large differences in the types and sizes of populations examined, the duration of follow-up, and the size of the effects, it is difficult to draw immediate conclusions on the consistency of the associations and the size of the effect. Our aim was to review published prospective population-based studies to assess whether the global evidence supports the presence of a relationship between sleep disturbances (in quantity and quality) and the development of type 2 diabetes and to obtain a quantitative estimate of the risk.

RESEARCH DESIGN AND METHODS

Literature search
We developed a search strategy to identify studies that reported the association between sleep disturbances and incidence of type 2 diabetes. We searched the electronic databases MEDLINE (from 1955 to April 2009) and EMBASE (from 1980) as well as the Cochrane Library using the terms “sleep” and “diabetes” and “prospective” or “cohort” or “longitudinal.” Furthermore, we reviewed reference lists of original and review articles to search for more studies. No language restriction was applied. After electronic identification of 1,553 potentially relevant studies, 175 were identified for additional scrutiny. Final exclusions were made through perusal of abstracts (supplementary Fig. 1, available in an online appendix at http://care.diabetesjournals.org/cgi/content/full/dc09-1124/DC1).

Inclusion and exclusion criteria
Studies had to fulfill the following criteria for inclusion: original article, prospective design, assessment of sleep disturbances (short or long duration as well as difficulty in initiating or maintaining sleep) as baseline exposure, incident cases of type 2 diabetes as outcome, follow-up of at least 3 years, adult population, and indication of the number of subjects exposed and of the rate or number of incident cases in different sleep disturbance categories. No sample size restriction was applied. Stud-
Statistical analysis

Definition of sleep disturbance

Duration of sleep was assessed by self-reported habitual sleep duration using the Sleep Duration Index (SDI) (8,11–13). Sleep disturbance was defined as duration of sleep less than 6 or 7 h/night (10,15). Long sleep was defined as 8 h/night (8,11,13,14). Difficulty in initiating or maintaining sleep was assessed by questionnaire (Table 1). The latter measures are components of sleep quality. Sleep maintenance reflects sleep consolidation.

Statistical analysis

The quality of the studies included in the meta-analysis was evaluated by the Downs and Black Quality Index score system (18), a validated checklist for assessing the quality of both randomized and nonrandomized studies. It consists of several items distributed among five subscales: reporting, power, and external validity, subscales: randomization and power. For the assessment of nonrandomized studies, the maximum score is 10.09. The maximum score of randomized studies is 19. Relevant studies were included only if the participants were randomized. Data were extracted independently by two investigators (F.P.C. and L.D.), and differences were resolved by discussion and consensus with either P.S. or M.A.M.

Table 1—Description of the studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Cohort</th>
<th>Sex</th>
<th>Follow-up</th>
<th>Incidence assessment</th>
<th>Outcome assessment</th>
<th>Exposure</th>
<th>Score</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayas et al. (8)</td>
<td>NHS Women</td>
<td></td>
<td>10.0</td>
<td>40–65</td>
<td>Questionnaire</td>
<td>SD</td>
<td>15</td>
<td>0.9-69</td>
</tr>
<tr>
<td>Kawakami et al. (16)</td>
<td>Electrical Co. Men</td>
<td>8.0</td>
<td>NA</td>
<td>8</td>
<td>DIS, DMS</td>
<td>Questionnaire</td>
<td>WHO criteria</td>
<td>1.68</td>
</tr>
<tr>
<td>Nilsson et al. (9)</td>
<td>MPP Men</td>
<td></td>
<td>15.2</td>
<td>Mean 46.2</td>
<td>Questionnaire</td>
<td>Questionnaire</td>
<td>FBG</td>
<td>4.3%</td>
</tr>
<tr>
<td>Björkelund et al. (10)</td>
<td>Gothenburg Women</td>
<td>32.0</td>
<td>38–60</td>
<td>15</td>
<td>SD</td>
<td>Interview</td>
<td>Multiple diagnoses</td>
<td>8.7%</td>
</tr>
<tr>
<td>Mallon et al. (11)</td>
<td>Combined</td>
<td></td>
<td>10</td>
<td>40–70</td>
<td>SD, DIS, DMS</td>
<td>Questionnaire</td>
<td>Questionnaire</td>
<td>9.1%</td>
</tr>
<tr>
<td>Meisinger et al. (17)</td>
<td>Combined</td>
<td></td>
<td>7.5</td>
<td>25–74</td>
<td>DIS, DMS</td>
<td>Questionnaire</td>
<td>Self-reported validated</td>
<td>3.85, 2.18</td>
</tr>
<tr>
<td>Yaggi et al. (12)</td>
<td>MMAS Men</td>
<td></td>
<td>13–17</td>
<td>40–70</td>
<td>SD</td>
<td>Questionnaire</td>
<td>Evidence of diagnosis</td>
<td>6.11</td>
</tr>
<tr>
<td>Gangwisch et al. (13)</td>
<td>NHANES I</td>
<td></td>
<td>8–10</td>
<td>32–86</td>
<td>SD</td>
<td>Questionnaire</td>
<td>Multiple methods</td>
<td>4.8%</td>
</tr>
<tr>
<td>Hayashino et al. (15)</td>
<td>HIPOP-OHP</td>
<td></td>
<td>4.2</td>
<td>19–69</td>
<td>SD, DIS, DMS</td>
<td>Questionnaire</td>
<td>Multiple methods</td>
<td>3.5%</td>
</tr>
<tr>
<td>Beihl et al. (14)</td>
<td>IRAS Combined</td>
<td></td>
<td>5.0</td>
<td>40–69</td>
<td>SD</td>
<td>Questionnaire</td>
<td>OGTT</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

*Data from ref. 18. †1, age, body mass, and sex (when combined); 2, leisure-time physical activity; 3, smoking; 4, alcohol consumption; 5, family history of diabetes, hypertension, and high cholesterol; 6, hypertension; 7, depression; 8, socioeconomic variables; 9, snoring. ‡Per 1,000 person-years. §Cumulative incidence. DIS, difficulty in initiating sleep; DMS, difficulty in maintaining sleep; FBG, fasting blood glucose; NHANES I, National Health and Nutrition Examination Survey I; NHS, Nurses Health Study; HIPOP-OHP, High-Risk and Population-Oriented Hypertension Study; NA, not applicable; OGGT, oral glucose tolerance test.
Sleep disturbances and type 2 diabetes

phenotypes to stabilize the variances and to normalize their distribution. We estimated the pooled RR (and 95% CI) using a random-effects model. By comparison with the reference category of sleep disturbance, we estimated the pooled risk and 95% CI of developing type 2 diabetes for the short and the long sleep category and for difficulty in initiating or maintaining sleep separately. Heterogeneity among studies was tested by Q statistics and quantified by I² and H statistics (19). We also performed meta-regression using a random-effects model (20). Funnel plot asymmetry was used to detect publication bias, with the application of Egger’s regression test (21). When indicated, we recalculated the combined estimate after imputation from the asymmetry of the funnel plot of the number of “missing” studies and their effect sizes and SEMs, a method known as “trim and fill” (21). The influence of individual studies was examined by omitting one or more study at a time to see the extent to which inferences depended on a particular study or group of studies (sensitivity analysis). Subgroup analysis was performed to assess possible sources of statistical heterogeneity and to check for the potential impact of sex and duration of follow-up on the relationship between sleep disturbances and incidence of type 2 diabetes. All statistical analyses were performed using MIX software (version 1.7) (22).

RESULTS

Characteristics

Ten studies (reporting on 13 cohorts) were included in the meta-analysis (8–17) (supplementary Fig. 1). When results were reported for men and women separately, they were entered into the analyses as separate cohorts. Table 1 summarizes the characteristics of the studies. Overall, the systematic review included 107,756 participants. Five studies recruited both men and women (11,13–15,17), two studies recruited only women (8,10), and three studies recruited only men (9,12,16). Four studies were from Europe, four were from the U.S., and two were from Japan. One study reported results by ethnicity (14). The majority of studies were population-based cohorts (9–12,17), two had multicenter recruitment (8,14), one was a national survey (13), and two were from occupational cohorts (15,16). The median Quality Score Index was 16 (range 8–17). Median follow-up was 9.5 years (4.2–32). All studies assessed sleep disturbances by questionnaire. The methods to ascertain new cases of type 2 diabetes varied among studies: in five studies, questionnaires were used (8,9,11,15,17), with additional validation (8,9); in the other five studies, more direct diagnostic criteria were used (10,12–14). The total number of incident cases of type 2 diabetes was 3,586. Of the five studies that included both men and women, two reported outcomes separately for men and women (11,17). Overall, nine cohorts reported data on the relationship between type 2 diabetes and short sleep, seven on long sleep, six on difficulty in initiating sleep, and six on difficulty in maintaining sleep (Table 1).

Short duration of sleep

Short duration of sleep was associated with a greater risk of developing type 2 diabetes (Fig. 1A) with no evidence of publication bias (supplementary Fig. 2A). There was statistical heterogeneity among studies. The effect in men (RR 2.07 [95% CI 1.16–3.72]) tended to be larger than that in women (1.07 [0.90–1.28], heterogeneity test P = 0.04). The effect was not affected by the restriction of the analysis to studies defining short sleep as ≤5 h, in which incident cases were assessed by questionnaire (n = 5; 1.36 [1.10–1.68], P = 0.004).

Long duration of sleep

Long duration of sleep was associated with a greater risk of developing type 2 diabetes (Fig. 1B) with no evidence of publication bias (supplementary Fig. 2B) and no statistically significant heterogeneity. The effect was not altered by the restriction of the analysis to studies defining long sleep as ≥9 h (n = 5; RR 1.38 [95% CI 1.15–1.65], P = 0.0006) or to those in which incident cases were assessed by questionnaire (n = 4; 1.59 [1.15–2.21], P = 0.0053).

Difficulty in initiating sleep

Difficulty in initiating sleep was associated with a greater risk of type 2 diabetes (Fig. 2A) with no evidence of publication bias (supplementary Fig. 2C) and no statistical heterogeneity. The effect was not altered by the restriction of the analysis to studies in which incident cases were assessed by direct clinical assessments (n = 4; RR 1.58 [95% CI 1.13–2.21], P = 0.0082). The trim and fill method imputed two studies with a revised estimate of 1.45 [1.13–1.86].

Difficulty in maintaining sleep

Difficulty in maintaining sleep was associated with a greater risk of type 2 diabetes (Fig. 2B) with no evidence of publication bias (supplementary Fig. 2D) and no statistical heterogeneity. The effect was not altered by the restriction of the analysis to studies in which incident cases were assessed by direct clinical assessments (n = 4; RR 1.67 [95% CI 1.30–2.14], P < 0.0001). The effect estimates were comparable in men (n = 3,207 incident cases; 2.29 [1.28–4.10], P = 0.005) and in women (n = 2,107 incident cases; 1.95 [1.22–3.12], P = 0.005, heterogeneity test P = 0.68). The trim and fill method imputed two studies with a revised estimate of 1.62 (1.18–2.24).

Duration of follow-up

The risk of developing type 2 diabetes showed an overall trend to increase with the duration of follow-up (Fig. 3). The RR increments per year of follow-up were estimated at 2% for short sleep (1.02 [95% CI 0.93–1.12]) after the exclusion of an outlier (10) (coefficient before exclusion 0.98 [0.96–1.01]), 7% for long sleep (1.07 [0.99–1.16]), no increments for difficulty in initiating sleep (1.00 [0.95–1.05]), and 12% (1.12 [1.00–1.24], P = 0.04) for difficulty in maintaining sleep.

CONCLUSIONS — This study provides for the first time quantitative pooled estimates of the associations between measures of quantity and quality of habitual sleep and the incidence of type 2 diabetes in studies around the world. It shows an unambiguous and consistent pattern of increased risk of developing type 2 diabetes at either end of the distribution of sleep duration and with qualitative disturbances of sleep. The risk varies between 28% in people who report habitual sleep of <5–6 h/night and 84% in those with difficulties in maintaining their sleep. The presence of little or no statistical heterogeneity among studies, the absence of publication bias, and the high statistical power confer further strength to our results. The effects were, by and large, comparable in men and women (with the exception of short sleep) and did not depend on the type of assessment of exposure and outcome or on the definitions of short or long sleep. A large number of potential confounders, particularly age and BMI, were considered in the primary analyses. The effect tended to increase with the duration of follow-up.
These results are of interest for several reasons. First, the association is consistent in different populations. Although the meta-analysis detected some statistical heterogeneity among studies (in particular, in the short sleep category), further sensitivity analysis and the absence of publication bias are in favor of similar effects across populations. Second, they indicate an effect size of potential public health relevance, consistent across the sexes and depending on the duration of follow-up.

**Study limitations**

First, with the exception of one study (16), all had a Downs and Black score between 15 and 18 of 19, indicating high quality. Second, a meta-analysis of observational data cannot directly control for confounding. We made an attempt to allow for multiple confounding by including adjusted estimates from multivariate models from each contributing study. However, residual confounding and bias remain a possibility. For instance, low levels of physical activity or poor diet that are causally related to type 2 diabetes may have also influenced sleep patterns. Third, there was no evidence of publication bias. However, the results can only be representative of the studies included and may not be easily extrapolated to other settings. For instance, most studies were carried out in Europe, the U.S., and Japan. They cannot therefore represent the wider populations across the globe, particularly those from the Indian subcontinent (where type 2 diabetes is highly prevalent) or from Africa. All studies used self-reported sleep disturbances (either as quantity or as quality of sleep). These methods have limitations in that they often may not allow (unless explicitly built as additional questions) differentiation of time asleep from time in bed or estimation of the number and duration of naps when assessing duration of sleep. On the other hand, sleep studies using objective measures of sleep are not practical and often not feasible in large prospective population studies. Sleep diaries, actigraphy, and polysomnography from some large population and small-scale investigations have shown good correlations between subjective estimates of sleep duration and more direct assessments (23–25). Furthermore, assessments of sleep durations in the primary health care setting, when collected, rely exclusively on self-reported data from patients.

Quantity and quality of sleep were assessed at one point in all studies. A single measure of exposure may not fully capture the sustained effects of sleep disruption over time on long-term morbidity. The studies analyzed did not always exclude subjects with obstructive sleep apnea-hypopnea syndrome (OSAS). These represent ~4% of middle-aged men and ~2% of middle-aged women (26, 27). OSAS is associated with obesity, short and disrupted sleep, excessive daytime sleepiness, and high rates of morbidity and mortality, predominantly due to cardiovascular disease (28). Although it is possible that the presence of patients with OSAS may have contributed to the risk of type 2 diabetes, the adjustment for obesity or BMI in almost every study would have, at least in part, corrected for this.
Sex differences in the risk associated with duration of sleep have been reported (29–33). Our analysis was repeated after stratification by sex, wherever possible. No differences were detected between short duration of sleep or difficulty in maintaining sleep and the development of type 2 diabetes. Ideally long duration of follow-up would be needed to assess the influence of sleep on health over the life course (34). We excluded a priori studies with short follow-up (<3 years) to avoid the possibility that measurements of sleep quantity and quality would be too close to the diagnosis of type 2 diabetes. We included studies with follow-up ranging from 4.2 to 32 years. The effect size was directly related to the duration of follow-up for some measures of sleep disturbance, suggesting the possibility of a time-dependent cumulative effect. We were unable to stratify studies by age-groups because of the inconsistent reporting of age in the original studies.

Potential mechanisms

Causative mechanisms relating sleep problems to adverse health outcomes include reciprocal changes in circulating levels of leptin and ghrelin (6,35). These in turn would increase appetite and caloric intake, reduce energy expenditure (3,4), and facilitate the development of obesity (4,6) and impaired glycemic control (36), increasing cardiovascular risk. Increased cortisol secretion and altered growth hormone metabolism have also been implicated (37). Low-grade inflammation is activated during short sleep, with possible implications not only for cardiovascular disease (7) but
also for other chronic conditions including cancer. The association of difficulty of initiating or maintaining sleep could be related to the same mechanisms, as an expression of reduced total sleep duration. Finally, elevated levels of dopamine and symptoms of gastroesophageal reflux have recently been described as important contributors to difficulties in maintaining sleep and insomnia (38, 39). Conversely, there is a less clear indication of possible mechanisms mediating the effect of long duration of sleep as a cause of type 2 diabetes. Depressive symptoms, low socioeconomic status, unemployment, a low level of physical activity, undiagnosed health conditions, and poor general health have all been shown to be associated with long duration of sleep and to confound the association with morbidity as well as mortality (40).

Disrupted sleep, both in quantity and quality, should be regarded as a behavioral risk factor for the development of type 2 diabetes, heavily determined by the environment and possibly amenable to modification through both education and counseling as well as through favorable modifications of physical and working environments to allow sufficient sleep and avoid habitual and sustained sleep deprivation and disruption.

Acknowledgments — This work is part of the Programme “Sleep, Health & Society” of the University of Warwick and was supported in part by a European Commission grant (FP7-Health-2007-201550).

No potential conflicts of interest relevant to this article were reported.

We thank Chen Ji for assistance with meta-regressions.

References
33. Suarez EC. Gender-specific associations between disturbed sleep and biomarkers of inflammation, coagulation and insulin resistance. Brain Behav Immunity 2008;22(Suppl. 1):29–30