Received: 9 September 2020 | Revised: 12 October 2020 | Accepted: 13 October 2020

DOI: 10.1111/jsr.13228

Journal of Sleep Research ESRS

REVIEW PAPER

A systematic review and meta-analysis of wind turbine noise effects on sleep using validated objective and subjective sleep assessments

Tessa Liebich[1,2] | Leon Lack[1] | Kristy Hansen[3] | Branko Zajamšek[1] | Nicole Lovato[1] | Peter Catcheside[1] | Gorica Micic[1]

[1]Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, College of Medicine and Public Health, Flinders University, Bedford Park, SA, Australia

[2]College of Education, Psychology and Social Work, Flinders University, Adelaide, SA, Australia

[3]College of Science and Engineering, Flinders University, Adelaide, SA, Australia

Correspondence
Tessa Liebich, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, College of Medicine and Public Health, Box 6, 5 Laffer Drive, Flinders University, Bedford Park, SA, 5042, Australia.
Email: tessa.liebich@flinders.edu.au

Abstract

Little is known about the potential impacts of wind turbine noise (WTN) on sleep. Previous research is limited to cross-sectional studies reporting anecdotal impacts on sleep using inconsistent sleep metrics. This meta-analysis sought to comprehensively review studies evaluating the impact of WTN using widely accepted and validated objective and subjective sleep assessments. Search terms included: “wind farm noise”, “wind turbine noise”, “wind turbine sound”, “wind turbine noise exposure” AND “sleep”. Only original articles published in English published after the year 2000 and reporting sleep outcomes in the presence of WTN using polysomnography, actigraphy or psychometrically validated sleep questionnaires were included. Uniform outcomes of the retrieved studies were meta-analysed to examine WTN effects on objective and subjective sleep outcomes. Nine studies were eligible for review and five studies were meta-analysed. Meta-analyses (Hedges' *g*; 95% confidence interval [CI]) revealed no significant differences in objective sleep onset latency (0.03, 95% CI −0.34 to 0.41), total sleep time (−0.05, 95% CI −0.77 to 0.67), sleep efficiency (−0.25, 95% CI −0.71 to 0.22) or wake after sleep onset (1.25, 95% CI −2.00 to 4.50) in the presence versus absence of WTN (all $p > .05$). Subjective sleep estimates were not meta-analysed because measurement outcomes were not sufficiently uniform for comparisons between studies. This systematic review and meta-analysis suggests that WTN does not significantly impact key indicators of objective sleep. Cautious interpretation remains warranted given variable measurement methodologies, WTN interventions, limited sample sizes, and cross-sectional study designs, where cause-and-effect relationships are uncertain. Well-controlled experimental studies using ecologically valid WTN, objective and psychometrically validated sleep assessments are needed to provide conclusive evidence regarding WTN impacts on sleep.

KEYWORDS

objective sleep, polysomnography, psychometrically validated assessment, sleep disruption, subjective sleep, wind turbine noise

Note. In this manuscript, the terms ‘self-reported’ and ‘subjective’ are used interchangeably.

1 | INTRODUCTION

> **Statement of significance**
>
> Studies investigating the impact of wind turbine noise (WTN) on objectively measured sleep outcomes are scarce. Previous reviews and meta-analyses are limited to cross-sectional studies based largely on anecdotal impacts on sleep and reporting indirect and inconsistent sleep metrics. Without the use of objective and standardised questionnaires, only limited conclusions can be drawn. To date, several experimental studies have examined the impact of WTN on sleep using polysomnography, actigraphy and psychometrically validated questionnaires, calling for an updated review. The present review and meta-analysis show that key indicators of objective sleep outcomes do not appear to be impacted by WTN, whereas psychometrically validated subjective sleep outcomes showed more inconsistent findings.

There are many economic and eco-friendly advantages associated with wind turbines given the long-term sustainability of this clean energy source. However, adverse health effects have also been reported by residents who live near wind turbines (Thorne, 2011), with sleep disturbance one of the most prominent and commonly reported concerns (Basner et al., 2014; Crichton et al., 2014; Janssen et al., 2011; Krogh et al., 2011; Muzet, 2007; World Health Organization, 2011). However, other residents living at similar distances to wind turbines report no sleep disturbance or ill health effects (Thorne, 2011), thus the prevalence, severity, and impacts of potential sleep disturbance effects remain unclear.

Good sleep is essential for health and quality of life (QoL), as well as for achieving optimal neural development, learning, memory and emotional regulation (Frank et al., 2013). Insufficient sleep (i.e. difficulty initiating and maintaining sleep) can result in daytime alertness and functional impairments, mood disturbance and reduced QoL (Jalali et al., 2016a; Janssen et al., 2011; Micic et al., 2018). Pre-existing psychosocial stress and aversive noise (e.g. environmental noise) have the potential to impair one's ability to initiate and maintain sleep, which can over time lead to maladaptive coping strategies such as spending increased time in bed awake and ruminating on the noise keeping them up, thus developing conditioned responses to the noise, e.g. increased alertness (Perlis et al., 1997). This can contribute to the development of insomnia, which can have a severe impact on an individual's QoL via fatigue, lack of energy, decreased mood, irritability, and memory and cognitive impairments (Lovato et al., 2014; Sweetman et al., 2017). Given that environmental noise, such as wind turbine noise (WTN), has the potential to be a psychosocial stressor and thus result in poor sleep (Evandt et al., 2017; Perlis et al., 1997; Riemann et al., 2010), it is important to consider and review the available findings to date regarding whether WTN impacts individual objective and subjective sleep.

Sleep disturbance from common environmental noise sources (e.g. road traffic and aircraft noise) is well established (Eberhardt & Akselsson, 1987; Kuroiwa et al., 2002; Marks & Griefahn, 2007). For example, in the presence of traffic noise, aircraft noise, and rail noise at A-weighted sound pressure levels (SPLs) of 39, 44 and 50 dB(A), compared to control nights of 32 dB(A) background noise, total sleep time (TST) and sleep quality have been shown to be reduced and latency to slow-wave sleep has been shown to be prolonged (Griefahn et al., 2006). A-weighting is frequently applied to noise measurements and is similar to the hearing response of the human auditory system as it is most sensitive in the mid-frequency ranges (200–2000 Hz) compared to the lower (<200 Hz) and higher frequencies (>2,000 Hz) (Leventhall, 2004). Whilst WTN is another environmental noise source, limited research has examined its effects on human sleep and physiology. Furthermore, WTN has some acoustic features that could make it more problematic for sleep compared to other noise types.

Wind turbine noise occurs predominantly in low frequencies, which can propagate substantially longer distances and penetrate building structures more readily, and thus could potentially be more problematic for sleep compared to higher frequency noises. In addition, WTN can also exhibit substantial amplitude modulation (AM), where noise amplitude varies with time continuously with each turbine blade-tower passage and sometimes more sporadically depending on external factors, e.g. variations in the weather, wind speed, wind shear, the number and size of the turbines in the area, local topography, vegetation, and the distance between turbines and residences receiving the noise (Hansen et al., 2017; Hansen et al., 2019). As a result of low-frequency noise predominance, the time-varying nature of AM, and low background noise of rural areas where wind turbines are typically installed, there is the potential for sleep disruption to occur. The aim of the present review was to meta-analytically gather all recent evidence to date (i.e. papers published after the year 2000) to quantitatively assess and systematically review WTN impacts on objective and psychometrically validated subjective sleep.

Previous literature reviews have focussed on the correlates of WTN on annoyance and health effects rather than the specific impact of WTN on sleep (Basner & McGuire, 2018; Schmidt & Klokker, 2014). To our knowledge, only one systematic review has specifically investigated the impact of WTN on sleep (Onakpoya et al., 2015). That review was based on studies that used self-reported assessments of sleep alone, many of which involved researcher-developed sleep questionnaires, often consisting of limited items addressing the presence versus absence of self-reported sleep disturbance, rather than outcomes from psychometrically validated questionnaires that have undergone extensive reliability and validity testing. Without the use of standardised, psychometrically validated tools, limited conclusions can be drawn regarding the impact of WTN on subjective sleep. Psychometric validity of questionnaires in research demonstrates that the questionnaires systematically measure what they are designed to measure. Using standardised questionnaires is useful and necessary for allowing comparisons between studies. More recently, several experimental studies have examined the effects of WTN on sleep using polysomnography (PSG), the "gold-standard" measure of sleep, as well as actigraphy and validated questionnaires. The present review aimed to use systematic and meta-analytic approaches to describe and provide a quantitative summary of data on this topic. Where possible, the present review also aimed to quantify the strength of evidence around the impacts of WTN on objective (PSG and actigraphy) and psychometrically validated self-reported measures of sleep.

2 | METHODS

2.1 | Design

This systematic review and meta-analysis was written in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (Moher et al., 2010).

2.2 | Data sources and search strategies

A systematic literature search was performed between January and April 2020. Electronic searches were conducted in PubMed, Scopus, Science Direct, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsycARTICLES, Web of Science and the Medical Literature Analysis and Retrieval System Online (MEDLINE) databases. Search terms were "wind farm noise", "wind turbine noise", "wind turbine sound", "wind turbine noise exposure" AND "sleep" [Title/Abstract]. See Appendix S1 for the specific database search strategies used. The search was also expanded by manually identifying relevant publications from the reference lists of retrieved literature after discussion with co-authors.

2.3 | Study selection criteria

Duplicate articles were removed, and the rest were screened by the primary author (TL) according to the selection criteria presented in Table 1. The retrieved studies were also reviewed by GM in an unblinded manner. GM also helped manually identify any relevant publications that were not retrieved from database searching.

An initial search was implemented in each database, which involved searching for studies that had been published before 2000 and investigated the impact of WTN on sleep using objective and/or psychometrically validated subjective sleep assessments. Given previous reviews and a lack of relevant publications before 2000, the selection criteria were designed to capture more recent studies published after 2000, that used objective sleep measures (i.e. PSG or actigraphy), and/or psychometrically validated subjective sleep measures. These included, but were not limited to sleep diaries, the Pittsburgh Sleep Quality Index (PSQI), Insomnia Severity Index (ISI), and Epworth Sleepiness Scale (ESS).

TABLE 1 Study selection criteria

Article criteria
Original, full-text, peer-reviewed article Contains terms "wind farm noise" OR "wind turbine noise", OR "wind turbine noise exposure", OR "wind turbine sound" and "sleep" in the title/abstract Written in English Published after 2000
Sample characteristics criteria
Adults aged ≥18 years Reportedly living/working within 15 km from a wind farm or exposed to WTN as part of the study procedure
Primary outcome criteria
Evaluated the impact of WTN on any of the following objective and/or psychometrically validated subjective sleep parameters: • Sleep onset latency (SOL), total sleep time (TST), wake after sleep onset (WASO), sleep efficiency. • Global scores of PSQI, ISI and/or ESS.
Meta-analysis criteria
Examined the presence versus absence of WTN on any of the aforementioned objective and/or psychometrically validated subjective sleep parameters (i.e. included a control group/condition and WTN exposure condition)

ESS, Epworth Sleepiness Scale; ISI, Insomnia Severity Index; PSQI, Pittsburgh Sleep Quality Index; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; WTN, wind turbine noise.

2.4 | Objective sleep measurement

2.4.1 | PSG

Polysomnography is the current "gold-standard" used for objective sleep measurement, as it uses direct electroencephalography (EEG) measurements and widely accepted scoring criteria to comprehensively describe sleep–wake timing, sleep stages, sleep onset latency (SOL; the time taken to fall asleep, in minutes), wake after sleep onset (WASO; the total time spent awake between the first and last epoch of sleep, in minutes), TST (in minutes), sleep efficiency (the total time spent asleep expressed as a percentage of time available for sleep between lights out and arising from sleep), and brief arousals from sleep (Martin & Hakim, 2011). PSG is often scored in 30-s epochs and is used to classify cortical activity including sleep staging, arousals, awakenings, sleep spindles, and K-complexes according to standards developed and maintained by the American Academy of Sleep Medicine (Iber et al., 2007). Eligible PSG studies required the use of the international 10–20 system for electrode placement on both an experimental night (a night with WTN exposure) and a control night (a quiet, WTN-free night), and thus report the traditional

TABLE 2 Study characteristics including sample and testing characteristics

Study	Location/ environment	Sample size	Age, years, mean	Design
Ageborg Morsing et al. (2018a)	Sweden, laboratory	6	22.2	Experimental laboratory study
Ageborg Morsing et al. (2018b)	Sweden, laboratory	6	24	Experimental laboratory study
Jalali, et al. (2016a)	Canada, open flat agricultural fields	16	55.9	Pre–post field study
Lane et al. (2016)	Canada, rural matched areas	32	Exposed group: 60.4; Unexposed group: 41.4 (adjusted mean age 50.9)	Cross-sectional field study
Jalali et al. (2016b)	Canada, rural area with flat agricultural fields	37	54.25	Pre–post field study
Nissenbaum et al. (2012)	USA, tree covered island and mountainous topography	79	N/A	Cross-sectional field study
Abbasi et al. (2015)	Iran, mountainous topography	53	30.8	Cross-sectional field study
Michaud et al. (2016)	Canada	1,238	N/A	Cross-sectional field study
Smith et al. (2020)	Sweden, laboratory	50	51.2	Experimental laboratory study

AM, amplitude modulation; dB L_{Aeq}, equivalent continuous sound pressure level; dB(A), A-weighted decibel; ISO; International Organisation for Standardisation; SOL, sleep onset latency; SPLs, sound pressure levels; TST, total sleep time; WASO, wake after sleep onset; WTN, wind turbine noise.

"gold-standard" metrics of sleep quality described above. Studies using PSG under these conditions were considered for eligibility and no other factors, such as sampling and filter frequencies or maximum impedance values, impacted study eligibility.

2.4.2 | Actigraphy

Actigraphy is a wrist-worn motion sensor device that algorithmically infers sleep and wakefulness from gross body movements, often across 1-min epochs (Smith et al., 2018). Actigraphy provides information on sleep patterns including estimates of the timing and duration of sleep and awakenings from which SOL, TST, WASO, and sleep efficiency is inferred. Actigraphy is minimally intrusive and thus enables longer-term inferences of sleep patterns not practical via PSG (Martin & Hakim, 2011). In addition, actigraphy provides an objective marker of sleep that can be easily used in an individual's home and does not need trained personnel to set up and implement. Whilst it does require some manual scoring, actigraphy does not require rigorous and time-consuming scoring after an overnight recording unlike PSG. Actigraphy is also less impacted by recall bias, sleep misperception or misattribution of awakenings than subjective self-report measures (Martin & Hakim, 2011). However, actigraphy relies on motion without directly assessing sleep via cortical activity. This approach has high sensitivity, but low specificity for detecting sleep, with frequent misclassification of inactivity as sleep when EEG demonstrates wake. This can result in an overestimation of sleep and an underestimation of wakefulness during the night (Marino

SPL of WTN exposure	Method of SPL measurement	Distance measurements
29.5 dB L_{Aeq}, 34.1 dB L_{Aeq}, 33.7 dB L_{Aeq} indoor WTN (with varying frequencies and AM characteristics)	Three 8-hr night-time synthesised WTN exposures with varying filtering, frequency bands and AM beats	N/A
32.8 dB L_{Aeq}, 32.8 dB L_{Aeq}, 30.4 dB L_{Aeq} indoor WTN (with varying frequencies and AM characteristics)	Three 8-hr night-time synthesised WTN exposures with varying filtering, frequency bands and AM beats	N/A
Time 1: 36.55 dB(A); Time 2: 36.50 dB(A)	10-hr noise measurements at two participant's residences for 16 nights before and 16 nights after wind turbine operation	10 individuals <1,000 m from a turbine and 6 individuals >1,000 m from a turbine
N/A	8-hr equivalent A-weighted sound level L_{Aeq} from 23:00 and 7:00 hours in one participant per group for 5 nights	Exposed group mean (*SD*) distance of 794.6 (264.1) m from a turbine. Unexposed group mean (*SD*) distance of 2,931 (1,015.6) m from a turbine
Time 1: 31.52 dB(A); Time 2: 31.23 dB(A)	10-hr noise measurements at two participant's residences for 16 nights before and 16 nights after wind turbine operation	22 individuals <1,000 m from a turbine and 15 individuals >1,000 m from a turbine
WTN ranging from 32–61 dB L_{Aeq}	Predicted noise levels at various distances from both wind turbine sites	Near group: 375–1,400 m; far group: 3.3–6.6 km
83 dB(A), 66 dB(A), 60 dB(A)	8-hr equivalent sound levels (L_{Aeq}, 8 hr) according to ISO 9612:2009.	0–50 m, 50–100 m, >150 m
Calculated outdoor SPLs at dwellings reached a mean (*SD*) of 46 (7.4) dB(A) and background night-time levels ranged between 35–61 dB(A). Ontario and Prince Edward Island residents were grouped into SPL categories of <25 dB(A), 25–<30 dB(A), 30–<35 dB(A), 35–<40 dB(A) and 40–46 dB(A)	Estimation using ISO 9613-1 (ISO, 1993) and 9613-2 (ISO, 1996). Long-term 1-year A-weighted equivalent continuous outdoor SPLs (L_{Aeq})	Ontario and Prince Edward Island residents at varying distances from a wind farm (<550 m, 550 m–1 km, 1–2 km, 2–5 km, >5 km)
32 dB L_{Aeq} indoor WTN including AM	Continuous synthesised WTN based on short- and long-term recordings including AM. This was played from 22:00 to 07:00 hours. Participant's scheduled sleep opportunity was 23:00–07:00 hours and thus participants were aware of the WTN exposure. All sound was calibrated to reflect a max 45 dB L_{Aeq}	Exposed group = resided <1 km from a turbine or reported sleep disturbance or annoyance from wind turbines in the past month; unexposed group

et al., 2013; Martin & Hakim, 2011; Sivertsen et al., 2006). In addition, because the epoch length of actigraphy is 1 min, only longer duration awakenings can be captured in comparison to PSG awakenings that are much shorter in duration.

Eligible actigraphy studies required the use of actigraphy as an objective measure of sleep and thus allowed for the reporting of traditional sleep metrics (e.g. SOL, TST, WASO, sleep efficiency). Eligible actigraphy studies also needed to have a control condition (e.g. non-exposed individuals or no-WTN exposure) to be considered in the meta-analysis. Studies using actigraphy under these conditions were considered for eligibility and no other factors, e.g. manually verified scoring or specific actigraphy devices or scoring algorithms, impacted study eligibility to maximise the number of eligible studies.

2.5 | Subjective sleep assessment

Sleep perception (i.e. the individual's own account of how long it takes them to go to sleep, how many hours of sleep they received, how much time they spent awake etc.) is important, particularly when assessing the possibility of insomnia (Maich et al., 2018; Morgenthaler et al., 2007). Self-reported sleep quality assessment using sleep diaries and sleep questionnaires is central to an insomnia diagnosis and treatment, and requires psychometrically validated instruments for meaningful between-group comparisons and for tracking improvements and recovery (American Psychiatric Association, 2013). For instance, using psychometrically validated questionnaires makes it possible to combine studies that have used the same questionnaires and thus strengthen and broaden research findings.

TABLE 3 Study outcomes and the tools used to assess these outcomes and the main findings of included studies

Study	Outcomes	Tools used to assess outcomes	Study findings
Ageborg Morsing et al. (2018a)	Objective SOL, sleep efficiency, TST, subjective sleep outcomes	PSG, morning questionnaire	No significant effect of SOL between control night [mean (*SD*) 23.3 (20.6) min], 29.5 dB L_{Aeq} WTN [mean (*SD*) 20.4 (13.2) min], 34.1 dB L_{Aeq} WTN [mean (*SD*) 16.0 (7.2) min] or 33.7 dB L_{Aeq} WTN [mean (*SD*) 13.7 (8) min], $p > .01$. No significant effect of TST between control night [mean (*SD*) 425.9 (32.5) min], 29.5 dB L_{Aeq} WTN [mean (*SD*) 444.9 (13.8) min], 34.1 dB L_{Aeq} WTN [mean (*SD*) 429.2 (32.4) min] or 33.7 dB L_{Aeq} WTN [mean (*SD*) 448.6 (8.4) min], $p > .01$. No significant effect of sleep efficiency between control night [mean (*SD*) 90.0 (6.8)%], 29.5 dB L_{Aeq} WTN [mean (*SD*) 93.2 (2.5)%], 34.1 dB L_{Aeq} WTN [mean (*SD*) 90.3 (6.4)%] or 33.7 dB L_{Aeq} WTN [mean (*SD*) 93.6 (1.6)%], $p > .01$. WASO data were not analysed in this study[a]. No significant effect of subjective SOL or number of perceived awakenings
Ageborg Morsing et al. (2018b)	Objective SOL, sleep efficiency, TST, subjective sleep outcomes	PSG, morning questionnaire	No significant effect of SOL between control night [mean (*SD*) 10.3 (8.4) min], 32.8 dB L_{Aeq} WTN with window gap filtering and high-frequency AM beats [mean (*SD*) 17.5 (10.6) min], 32.8 dB L_{Aeq} WTN with window gap filtering and low-frequency AM beats [mean (*SD*) 17.0 (11.4) min] or 30.4 dB L_{Aeq} WTN with window closed filtering and low-frequency AM beats [mean (*SD*) 21.3 (25.5) min], $p > .01$. No significant effect of TST between control night [mean (*SD*) 455.2 (9.2) min], 32.8 dB L_{Aeq} WTN with window gap filtering and high-frequency AM beats [mean (*SD*) 447.5 (14.7) min], 32.8 dB L_{Aeq} WTN with window gap filtering and low-frequency AM beats [mean (*SD*) 442.7 (9.9) min] or 30.4 dB L_{Aeq} WTN with window closed filtering and low-frequency AM beats [mean (*SD*) 440.8 (34.4) min], $p > .01$. No significant effect of sleep efficiency between control night [mean (*SD*) 94.8 (1.9)%], 32.8 dB L_{Aeq} WTN with window gap filtering and high-frequency AM beats [mean (*SD*) 93.2 (3.1)%], 32.8 dB L_{Aeq} WTN with window gap filtering and low-frequency AM beats [mean (*SD*) 92.2 (2.1)%] or 30.4 dB L_{Aeq} WTN with window closed filtering and low-frequency AM beats [mean (*SD*) 91.8 (7.2)%], $p > .01$. WASO data were not analysed in this study[a]. No significant effect of subjective SOL or number of perceived awakenings
Jalali et al. (2016a)	Objective SOL, sleep efficiency, TST, WASO, subjective sleep outcomes	Ambulatory PSG, sleep diary	No significant difference between SOL at Time 1 [mean (*SD*) 14.9 (17.7) min] and Time 2 [mean (*SD*) 11.1 (16.9) min], $p = .371$. No significant difference between TST at Time 1 [mean (*SD*) 380.3 (68.8) min] and Time 2 [mean (*SD*) 402.1 (36.4) min], $p = .226$. No significant difference between WASO at Time 1 [mean (*SD*) 34.8 (26.0) min] and Time 2 [mean (*SD*) 34.4 (26.9) min], $p = .950$. No significant difference between sleep efficiency at Time 1 [mean (*SD*) 88.5 (7.1)%] and Time 2 [mean (*SD*) 89.4 (6.9)%], $p = .634$. No significant differences in subjective TST, number and length of awakenings or sleep latency at Time 1 compared with Time 2 (all $p > .05$)

(Continues)

TABLE 3 (Continued)

Study	Outcomes	Tools used to assess outcomes	Study findings
Lane et al. (2016)	Objective SOL, sleep efficiency, TST, WASO, subjective sleep outcomes	Actigraphy, sleep diary	No significant differences in SOL for exposed individuals [mean (*SD*) 6.8 (1.8) min] and unexposed individuals [mean (*SD*) 7.3 (2.3) min], $p = .22$. No significant differences in sleep efficiency for exposed individuals [mean (*SD*) 88.5 (5.4)%] and unexposed individuals [mean (*SD*) 91.0 (4.1)%], $p = .17$. No significant differences in TST for exposed individuals [mean (*SD*) 436.7 (53.6) min] and unexposed individuals [mean (*SD*) 413.7 (47.7) min], $p = .34$. No significant differences in WASO for exposed individuals [mean (*SD*) 44.0 (1.7) min] and unexposed individuals [mean (*SD*) 30.6 (1.9) min], $p = .16$
Jalali et al. (2016b)	Subjective sleep quality	PSQI, ESS ISI	PSQI scores increased from Time 1 [mean (*SD*) 4.1 (2.1)] to Time 2 [mean (*SD*) 6.2 (3.9)], $p = .006$. ESS scores also significantly increased from Time 1 [mean (*SD*) 4.7 (3.2)], to Time 2 [mean (*SD*) 7.1 (5.3)], $p = .002$. ISI scores also significantly increased from Time 1 [mean (*SD*) 3.1 (3.6)], to Time 2 [mean (*SD*) 6.4 (6.7)], $p = .005$
Nissenbaum et al. (2012)	Subjective sleep quality	PSQI, ESS	Mean PSQI scores were significantly greater in the near group (7.8) than the far group (6.0), $p = .0461$. Mean ESS scores were also significantly greater in the near group (7.8), than the far group (5.7), $p = .0322$
Abbasi et al. (2015)	Daytime sleepiness	ESS	Significant differences between ESS and occupational group, where maintenance/repair workers had the greater ESS scores [mean (*SD*) 10.5 (1.7)] than security [mean (*SD*) 6.0 (1.4)], and office administration staff [mean (*SD*) 4.0 (0.9)], $p < .001$
Michaud et al. (2016)	Objective SOL, sleep efficiency, TST, WASO and self-reported sleep quality	Actigraphy, PSQI	No significant difference between SOL ($p = .02$), sleep efficiency ($p = .05$), WASO ($p = .36$) or TST ($p = .74$), across the different exposure levels. No significant differences between mean PSQI scores across different exposure levels ($p = .75$) (mean + *SD* not reported here)
Smith et al. (2020)	Objective SOL, sleep efficiency, TST, WASO and self-reported sleep quality	PSG, morning questionnaire	No significant effect of SOL between control night [mean (*SE*) 21.3 (3.5) min], and WTN night [mean (*SE*) 25.3 (3.7) min], $p = .165$. No significant effect of TST between control night [mean (*SE*) 415.6 (5.5) min] and WTN night [mean (*SE*) 402.9 (8.6) min], $p = .543$. No significant effect of sleep efficiency between control night [mean (*SE*) 86.6 (1.2)%], and WTN night [mean (*SE*) 84.2 (1.7)%], $p = .483$. No significant effect of WASO between control night [mean (*SE*) 45.2 (5.3) min] and WTN night [mean (*SE*) 52.3 (7.5) min], $p = .50$

AM, amplitude modulation; dB L_{Aeq}, equivalent continuous sound pressure level; dB(A), A-weighted decibel; ESS, Epworth Sleepiness Scale; ISI, Insomnia Severity Index; PSG, polysomnography; PSQI, Pittsburgh Sleep Quality Index; SOL, sleep onset latency; SPLs, sound pressure levels; TST, total sleep time; WASO, wake after sleep onset; WTN, wind turbine noise.

[a]Denotes no WASO data were analysed in the study. The primary author, TL contacted the authors of these studies to obtain mean (*SD*) values to be included in the meta-analysis.

2.5.1 | Sleep diary

Sleep diaries are psychometrically validated for measuring sleep perception night-to-night (Carney et al., 2012; Maich et al., 2018). Individual questions are used to calculate common sleep parameters including time in bed, SOL, number of perceived awakenings, WASO, time of final awakening, and time out of bed. More comprehensive versions may also assess day-by-day sleep medication use, naps, caffeine, and alcohol use (Maich et al., 2018).

2.5.2 | ISI

The ISI is a seven-item self-report assessment of difficulty initiating and maintaining sleep, sleep satisfaction, and daily functioning (Morin, 1993). The total score ranges from 0 to 28, whereby higher scores indicate greater insomnia severity. Clinical score cut-offs are 0–7 = absence of insomnia, 8–14 = subthreshold insomnia, 15–21 = moderate insomnia, and 22–28 = severe insomnia. The ISI demonstrates adequate internal consistency for identifying both

TABLE 4 Reporting quality and risk of bias within identified studies using an adapted version of the STROBE checklist (Onakpoya et al., 2015)

Study	Country	Study design	Appropriate recruitment strategy?	Appropriate sampling technique?	Response rate if applicable	Representative sample?
Ageborg Morsing et al. (2018a)	Sweden	Experimental laboratory study	Somewhat - Advertising and detailed exclusion criteria	Yes - Participants were counterbalanced to receive all conditions (within-subjects cross-over design)	N/A	Yes - Noise-sensitive individuals
Ageborg Morsing et al. (2018b)	Sweden	Experimental laboratory study	Somewhat - Advertising and detailed exclusion criteria	Yes - Participants were counterbalanced to receive all conditions (within-subjects cross-over design)	N/A	Yes - Noise-sensitive individuals
Jalali et al. (2016a)	Canada	Pre–post field study	Uncertain - Inclusion criteria for home sleep assessment	Unclear - Residents who lived within 2,000 m from a proposed wind farm	N/A	Yes - Residents living within 2,000 m of a post-turbine erection site but in the pre-operational stage
Lane et al. (2016)	Canada	Cross-sectional field study	Yes - Door-to-door recruitment	Yes - Randomly sampled	50%	Yes - Individuals living near wind farm areas and a demographically matched rural control area
Jalali et al. (2016b)	Canada	Pre–post field study	Uncertain - Letters of advance notice delivered to door and door-to-door recruitment	No - Residents who lived within 2,000 m from a proposed wind farm	30%	Yes - Residents living within 2,000 m of a post-turbine erection site but in the pre-operational stage
Nissenbaum et al. (2012)	USA	Cross-sectional field study	Yes - Questionnaire face-to-face or telephone interview	Yes - Random sampling	59% for the near group - no response rate for far group	Yes - Residents living in close proximity to a wind turbine (375–1,400m) and far from a wind turbine (3,000–6,600 m)
Abbasi et al. (2015)	Iran	Cross-sectional field study	Uncertain - Based on job type, questionnaire sent	Unclear - Census	N/A	Individuals working on a wind farm (no control group as the individuals furthest away was still >150 m)

Relevant outcome measures?	Power calculation?	Appropriate statistical analysis?	Limitations/biases	Risk of bias judgement
Yes - Objective SOL, sleep efficiency, TST, WASO[a]	No	Yes - Non-parametric tests – Friedman tests, and Wilcoxon signed-rank tests.	Low sample size and representativeness of the sample, WTN was above recommended outdoor levels for Sweden. Significance levels were $p < .01$ rather than $p < .05$. Individual non-significance levels were not reported for mean (*SD*) across nights (= risk reporting bias) Some counterbalancing was used (Nights 3–5), but control night was always on night 2 No reports on blinding of participants or researchers mentioned (although a blind sleep scorer was used)	Some concerns
Yes - Objective SOL, sleep efficiency, TST, WASO[a]	No	Yes - Non-parametric tests - Friedman tests, and Wilcoxon signed-rank tests	Low sample size and representativeness of the sample, WTN was above recommended outdoor levels for Sweden. Significance levels were $p < .01$ rather than $p < .05$. Individual non-significance levels were not reported for mean (*SD*) across nights (= risk of reporting bias) Some counterbalancing was used (Nights 3–5), but control night was always on night 2 No reports on blinding of participants or researchers mentioned (although a blind sleep scorer was used)	Some concerns
Yes - Objective SOL, sleep efficiency, TST, WASO	No	Yes - Paired sample *t* test, McNemar tests, Spearman's rank correlations	Identifies lack of control in field designs (WTN exposure levels, wind speed variation), order effects and general issues with WTN exposure Participants not blinded to study aims (although a blind sleep scorer was used) Unclear whether random sampling was used (=risk of selection bias) No indication of attrition	High risk
Yes - Objective SOL, sleep efficiency, TST, WASO	No	Yes - *t* tests and Wilcoxon-Mann–Whitney tests	Notes the limitations of low statistical power and low estimates of exposure due to calm weather Random sampling used (=less risk of selection bias) Response rate stated Actigraphy scored based on algorithm (=less risk of detection bias)	Some concerns
Yes - PSQI, ISI and ESS	No	Yes - Wilcoxon signed-rank tests, Mann–Whitney tests, independent *t* tests, chi-square tests, and Spearman's rank correlations	Identifies lack of control in field designs (WTN exposure levels, wind speed variation), order effects and non-response biases Participants not blinded to study aims (= risk of selection bias) No random sampling was used (=risk of selection bias) Low response rate (30% = risk of selection bias)	High risk
Yes - PSQI, ESS	No	Yes - Descriptive and multivariate analyses	Reporting and selection biases due to both areas involving residents that benefit financially from wind turbines. Reducing property value fears, visual impacts and attitudes impacting results. No *SD*/variability measures reported Lack of variability estimates (= risk of reporting bias) Response rate only provided for near group (= risk of reporting bias) Participants not blinded to study aims Principle investigator was blind to outcome assessment	High risk
Unclear - ESS is not used to diagnose sleep disorders	No	Yes - MANOVA, Pillai's Trace test, Scheffe's post hoc test, multivariate regression	Used ESS to identify sleep disorder, fear of responding truthfully due to job Unclear in terms of whether sampling was random (=risk of selection bias) No response rate indicated (= risk of selection bias) Participants unlikely blinded to study aims. No indication of blind outcome assessment/data handling at any stage (= risk of selection bias and detection bias)	High risk

(Continues)

TABLE 4 (Continued)

Study	Country	Study design	Appropriate recruitment strategy?	Appropriate sampling technique?	Response rate if applicable	Representative sample?
Michaud et al. (2016)	Canada	Cross-sectional field study	Yes - Computer-assisted personal interviewing technique	Yes - Computer-assisted random selection method	78.9%	Yes - Individuals at varying distances from a wind farm (<550 m, 550 m–1 km, 1–2 km, 2–5 km, >5 km)
Smith et al. (2020)	Sweden	Experimental laboratory study	Yes - Postal mailings, phone calls, advertising, experimental exclusion criteria considered	Yes - Participants were counterbalanced to receive all conditions (within-subjects cross-over design)	N/A	Yes - Individuals living <1,000 m from a turbine and those not living near a turbine

ESS, Epworth Sleepiness Scale; ISI, Insomnia Severity Index; MANOVA, multivariate analysis of variance; N/A, not available; PSQI, Pittsburgh Sleep Quality Index; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; WTN, wind turbine noise.
[a]WASO data were requested by TL.

clinical (Cronbach's α = .91) and community samples (Cronbach's α = .90); hence, is considered to be a reliable tool for assessing insomnia severity (Morin et al., 2011).

2.5.3 | PSQI

The PSQI is a 19-item questionnaire that assesses sleep duration, sleep latency and the frequency/severity of specific sleep-related problems (Buysse et al., 1989). Individual items are scored into seven main components that are then summed to provide an aggregate global score. Global PSQI scores range from 0 to 21, where higher scores represent worse sleep quality and PSQI scores of >5 indicate poor sleep quality. The PSQI has good internal consistency (Cronbach's α = .83), good test–retest reliability (r = .85), adequate validity to distinguish between poor and healthy sleepers (89.6% sensitivity and 86.5% specificity), and good construct validity (r = .69) (Buysse et al., 1989).

2.5.4 | ESS

The ESS is an eight-item scale that assesses habitual daytime sleepiness or the likelihood of sleeping in particular situations (Johns, 1991). It has high test–retest reliability (r = .82) and internal consistency (Cronbach's α = .88) (Johns, 1992). The total ESS score ranges from 0 to 24, with higher scores indicating higher daytime sleepiness. Clinical cut-offs of ≥10 indicate excessive daytime sleepiness. However, worthy of note is that the ESS does not capture momentary sleepiness, where instead momentary sleepiness would be captured by the Karolinska Sleepiness Scale (KSS) (Åkerstedt and Gillberg, 1990).

2.6 | Statistical analyses

2.6.1 | Data extraction and quality assessment

Relevant data fields for extraction were identified by TL and are shown in Tables 2 and 3. In the case that data were not available for analysis in the retrieved studies, TL contacted the appropriate authors for such data. The statistics reported in the retrieved articles included mean (*SD*) or mean (*SE*) from which pooled variances were determined where possible.

The reporting quality of the included studies was assessed via the adapted Strengthening the Reporting of Observational studies in Epidemiology (STROBE) Checklist (Von Elm et al., 2007) at the study level. The adapted STROBE checklist was chosen to measure study quality because this is the checklist that has been used by the only other systematic review and meta-analysis that has investigated the impact of WTN on sleep (Onakpoya et al., 2015). Therefore, it was assumed that a larger proportion of the identified studies in the present review and meta-analysis would also be observational, and thus, it was considered appropriate to still use the adapted STROBE checklist. This involved an assessment of the recruitment and sampling technique (e.g. did they detail their techniques and was the recruited sample representative of the

Relevant outcome measures?	Power calculation?	Appropriate statistical analysis?	Limitations/biases	Risk of bias judgement
Yes - PSQI, objective SOL, sleep efficiency, WASO, TST, number of awakenings, time in bed, rate of awakenings per 1 hr in bed	Yes	Yes – Cochran-Mantel-Haenszel chi-square test, univariate logistic regression models, multiple regression models, stepwise regression analyses, generalised estimating equation methods, Poisson distributions	Describes the use of actigraphy as an objective measure of sleep, as well as the timing of objective versus subjective measures of sleep (7-day actigraphy versus PSQI over the year and 30 days). Also identifies night-to-night variation in outdoor WTN levels and the possibility that wind turbine operators altered the output of their turbines to produce desirable effects. Considered the difference in objective sleep variables from weekdays to weekends Masked the true aim of the study (= less risk of selection bias) Actigraphy scored based on algorithm (= less risk of detection bias) Random sampling (= less risk of selection bias) Adequate response rate	Low risk
Yes - Objective SOL, sleep efficiency, TST, WASO and subjective morning questionnaire	No	Yes - Mixed-effects regression models.	Acknowledges self-selection bias, self-report habitual sleep times, lower ecological validity due to being in a laboratory Participants not blinded (although a blind sleep scorer was used) (=risk of selection bias) Counterbalanced WTN and control night Reported outcome variables	Some concerns

interested population and sampled in an adequate way?), response rate, relevant outcome measures, appropriate statistical analyses, and any limitations and biases. Based on the identified limitations and risk of selection bias, reporting bias, detection bias, and attrition/response rate if applicable; a risk of bias judgement (i.e. low risk, some concerns, and high risk) was also made and is shown in Table 4. The reporting quality and risk of bias judgement were assessed independently by TL in an unblinded manner and reviewed by GM. Differential judgements by TL and GM were resolved by a third author (PC).

2.6.2 | Meta-analyses

All analyses were conducted using *Meta-Essentials: Workbooks for meta-analysis, version 1.5* (Hak et al., 2016) for estimation of pooled mean effects and 95% confidence intervals (95% CIs) using random-effects models. Hedges' *g* is the appropriate effect size to use when analysing group differences (i.e. between an experimental versus control group) and when sample sizes are small (Borenstein et al., 2011). The present review also reported on prediction intervals (PI), which involve the range in which 95% of future studies are predicted to fall and the assessment of heterogeneity and potential for publication biases. Meta-analyses were conducted on all eligible retrieved studies that used uniform objective or self-reported measures of sleep to investigate the impact of the presence versus absence of WTN exposure on sleep outcomes (SOL, TST, WASO, sleep efficiency, PSQI, ISI and/or ESS scores).

2.6.3 | Heterogeneity and risk of biases

The *Q*-statistic (Cochrane's *Q*) was also reported to indicate the average variability of the effect size for each sleep parameter. A significant *Q*-statistic suggests that the variability in the effect size is greater than expected by chance (Hak et al., 2016). The *Q*-statistic is limited as it can be impacted by sample size biases between studies and thus should be interpreted with the I^2 statistic, which indicates the proportion of variance of real differences in effect sizes (Hak et al., 2016). In the present meta-analyses, I^2, the *Q*-statistic, and the significance level are reported. In the event of a high I^2 (>50%), a subgroup analysis will be sought, as this indicates that the meta-analysed studies are less likely to be of the same population.

In addition, funnel plots of the effect size in comparison to the *SE* for each sleep parameter were used to assess the potential for publication bias. Symmetrical funnel plots are strongly indicative of minimal bias. A further assessment of bias in individual studies is provided in Table 4.

3 | RESULTS

3.1 | Selection of studies

Figure 1 illustrates the PRISMA flow diagram, outlining the study selection process at each stage of screening. The database search strategy identified 451 records and seven additional records through consultation with co-authors to identify pertinent articles not captured by the search strategy and screening reference lists of included articles. In all, 324 records remained after removing duplicates and 49 remained after abstract screening. Full-text screening excluded 41 studies, mainly due to absence of key outcomes, leaving eight studies that met the inclusion criteria (Figure 1). One of the eligible studies reported on two separate pilot studies and, therefore, this was treated as two separate records, making nine eligible studies for qualitative synthesis after abstract screening. For quantitative synthesis, four studies were excluded for reasons detailed in Figure 1. Studies that did not uniformly or comparably measure objective and subjective outcomes and thus could not be meta-analysed were discussed separately. Ageborg Morsing et al. (2018a, 2018b) had three different WTN exposure nights [at outdoor SPLs of 40, 45, 50 dB equivalent continuous SPL (L_{Aeq})]. In this case, the 45 dB L_{Aeq} conditions were chosen as the WTN condition to be included in the meta-analysis. This was due to the World Health Organization stating that at night, outdoor SPLs should not exceed 45 dB L_{Aeq} (Bergland & Lindvall, 1995). Additionally, Ageborg Morsing et al. (2018a, 2018b) did not report WASO in their studies and therefore, TL contacted the primary authors of these studies to obtain WASO data for inclusion in this review and meta-analysis.

3.2 | Study demographics

Table 2 summarises the sample and testing characteristics and Table 3 summarises the outcomes, tools used to assess each outcome, and the study findings. In all, 1,517 participants were assessed in the nine included studies. Three experimental laboratory studies were conducted in Sweden (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Smith et al., 2020), two cross-sectional and two longitudinal studies in Canada (Jalali et al., 2016a; Jalali et al., 2016b; Lane et al., 2016; Michaud et al., 2016) one cross-sectional study in Iran (Abbasi et al., 2015) and one cross-sectional study in the USA (Nissenbaum et al., 2012). Topography varied between the non-laboratory studies, with two study locations being in mountainous areas, (Abbasi et al., 2015; Nissenbaum et al., 2012) and four Canadian studies in rural areas with flat, open fields (Jalali et al., 2016a; Jalali et al., 2016b; Lane et al., 2016; Michaud et al., 2016). The mean (*SD*, range) age across all studies was 41.3 (15.0, 22–56) years. The distance from wind turbines ranged between <50 m and 11.2 km and outdoor SPLs ranged from <25 to 83 dB(A). It is worth noting that studies included not only individuals who lived near wind turbines, but also individuals with no prior exposure to WTN and those who worked on wind farms. Three studies used synthesised WTN recordings, (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Smith et al., 2020), four studies used 8–10 hr recordings of WTN measured inside participants' homes (Abbasi et al., 2015; Jalali et al., 2016a; Jalali et al., 2016b; Lane et al., 2016), and two studies used estimations/predictions of WTN using International Organisation for Standardisation (ISO) models (Michaud et al., 2016; Nissenbaum et al., 2012). Five of the nine studies assessed SOL, TST, WASO and sleep efficiency using objective measures of sleep, which included one actigraphy-based study and four PSG studies that were included in the meta-analysis (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Jalali et al., 2016a; Lane et al., 2016; Smith et al., 2020).

3.3 | Reporting quality

Despite three of the retrieved studies being experimental studies, the STROBE Checklist was still used as a measure of reporting quality and bias given the larger proportion of studies still being cross-sectional in nature. Table 4 summarises the reporting quality of all nine included studies. As shown in Table 4, all studies used appropriate statistical methods to compare groups and associations, and used relevant and appropriate, objective or psychologically validated self-report outcome measures, as per the study selection criteria (Table 1). Four of the studies used the "gold-standard" PSG to assess sleep outcomes objectively (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Jalali et al., 2016a; Smith et al., 2020), two used actigraphy to assess sleep objectively (Lane et al., 2016; Michaud et al., 2016), and the remaining three used psychometrically validated subjective sleep questionnaires including the PSQI, ESS and ISI (Abbasi et al., 2015; Jalali et al., 2016b; Nissenbaum et al., 2012). One of the actigraphy-based studies also used the PSQI to assess self-reported sleep quality (Michaud et al., 2016) and the other actigraphy study also used a sleep diary to assess self-reported perception of sleep in addition to objective sleep outcomes (Lane et al., 2016). The four remaining objective studies (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Jalali et al., 2016a; Smith et al., 2020) also used a sleep diary or morning questionnaire to assess self-reported sleep outcomes.

Recruitment and sampling strategies varied from appropriate to low quality. For cross-sectional studies, recruitment and sampling strategies included questionnaires, door-to-door recruitment, face-to-face/telephone interviews, random sampling, a computer-assisted personal interviewing technique and the use of census data (Onakpoya et al., 2015; Von Elm et al., 2007). For the two longitudinal studies (Jalali et al., 2016a; Jalali et al., 2016b), recruitment involved door-to-door recruitment for those meeting specified criteria including being aged ≥18 years, healthy, good sleepers, no sleep medication, no hearing loss, and no other significant sources of noise disruption (such as traffic or rail noise). Sampling strategies for the longitudinal studies involved selecting residents living within 2 km of a pre-operational wind farm to reflect a baseline control condition (Jalali et al., 2016a; Jalali et al. 2016b). The three experimental studies (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Smith et al., 2020) also utilised advertising and detailed exclusion/inclusion criteria, and all adopted a counterbalanced design. Smith et al. (2020) in particular, provided detailed information regarding their recruitment and sampling strategies in their supplementary analyses.

However, some of these criteria/strategies have the potential to introduce bias, particularly without random sampling to minimise

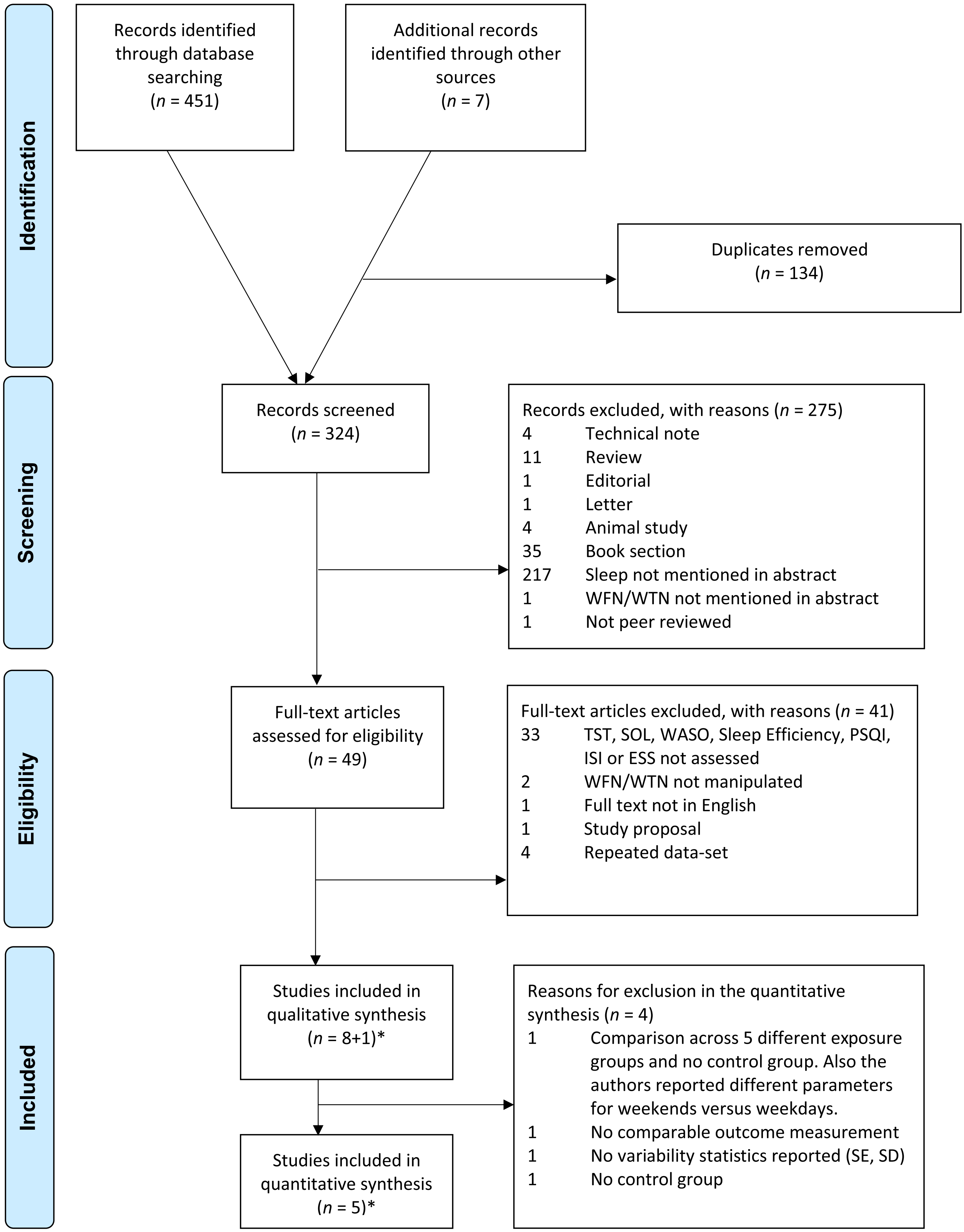

FIGURE 1 PRISMA flow diagram showing the process for inclusion. Note. *1 additional study was included as one record conducted and analysed two separate studies

potential attitudinal biases around perceived annoyance and sleep impacts. Multiple additional factors could also confound WTN effects on sleep, e.g. hearing loss with ageing populations or industrial noise exposure, and common pre-existing sleep problems. Excluding participants with hearing loss or sleep apnea could help to avoid confounding, but might not adequately represent rural residents surrounding wind farms or wind turbines. For example, heightened low-frequency hearing acuity, increased wake across the night, conscious noise exposure or pre-existing sleep problems, can all impact sleep quality and by excluding participants that do not experience these factors, may impact the generalisability of study findings.

Furthermore, only one study (Michaud et al., 2016) reported sample size (power) calculations and only four of the nine studies provided response rates, from which the mean (*SD*) response rate across the studies was 54.5 (20.3)% (Jalali et al., 2016b; Lane et al., 2016; Michaud et al., 2016; Nissenbaum et al., 2012).

Based on the biases summarised in Table 4, the overall reporting quality was classed as "low" according to the STROBE checklist and identified limitations and biases. In terms of the risk of bias judgements for each study, four of the nine studies identified had a high risk of bias, and another four had some concerns of bias, with only one having a low risk of bias.

3.4 | Meta-analysis of objectively measured sleep parameters

Whilst six studies used objective measures of sleep (four PSG and two actigraphy), only five used uniform outcomes (four PSG and one actigraphy) and thus were included in the meta-analysis. The actigraphy study by Michaud et al. (2016) compared five different exposure groups in the field, the lowest exposure being <25 dB(A) and thus did not have a control no-WTN exposure condition. Whilst 25 dB(A) could be argued to reflect a control condition, participants were still exposed to WTN and thus could invalidate participant responses who are exposed to this level of WTN. Four objective sleep parameters were comparable across the five objective studies that assessed the impact of WTN exposure on sleep relative to control no-WTN exposure. These included SOL, TST, WASO, and sleep efficiency.

As there are known limitations of actigraphy versus PSG measures, meta-analyses were initially run without the actigraphy study (Lane et al., 2016) to minimise the potential for biases associated with actigraphy compared to PSG. However, the overall results remained unchanged with versus without this study included (all $p > .05$), and thus all five studies that used objective measures of sleep (PSG and actigraphy) were meta-analysed together.

3.4.1 | SOL, TST, sleep efficiency and WASO meta-analytic results

Figure 2 shows the mean differences between the presence and absence of WTN exposure in SOL, TST, sleep efficiency, and WASO of the five included studies. The Hedges' g (95% CI) and associated meta-analytic statistics are shown in Table 5. Individual study mean (*SD*) values are displayed in Table 3. When all available studies were combined, there were no statistically significant effects of WTN exposure on SOL, TST, sleep efficiency or WASO compared to no-WTN exposure. As shown in Table 5, heterogeneity between studies was low and not statistically significant for SOL, TST and sleep efficiency, but was high for WASO, suggesting that WASO effects cannot be considered to be generalisable across studies. A meaningful subgroup analysis was not possible with only five studies, but when the actigraphy study was removed from the meta-analysis, heterogeneity in WASO decreased from 89.77% ($p < .001$) to 12.82% ($p = .328$), whereby the heterogeneity was no longer significant. Overall, this suggests that for WASO, the meta-analysed studies are likely not considered to be of the same population.

With only five included studies, evaluating the risk of bias across studies was difficult to assess and thus these results should be interpreted with caution. Figure 3 shows the funnel plots that were constructed for each sleep parameter in the meta-analysis. Upon visual inspection of each funnel plot, SOL, TST, sleep efficiency and WASO appeared symmetrical, indicating minimal publication bias across studies. The Duval and Tweedie "Trim and Fill" method was used to determine the presence of any missing unpublished studies and where they would likely fall within the funnel plot as well as calculating an adjusted, combined effect size after including any missing studies in the analysis (Duval & Tweedie, 2000). This method was used as it allows for filling each plot by including any trimmed studies on the right-hand side and the imputed studies on the left side of the mean. By using this method, no studies were deemed missing in any of the funnel plots (a–d) and thus no data points were imputed into Figure 3 and all adjusted combined effect sizes remained identical to the unadjusted combined effect sizes.

3.5 | Systematic narrative review of objectively and subjectively measured sleep parameters

3.5.1 | Actigraphy

Two studies used actigraphy to assess the impact of WTN on sleep (Lane et al., 2016; Michaud et al., 2016). One of these cross-sectional studies was initially based on weekdays versus weekend sleep data, but was then adjusted using least squares mean (95% CI) to account for province and day of the week (Michaud et al., 2016). Table 6 shows the adjusted results of this actigraphy study of 1,238 participants, which found no significant differences between WTN exposure levels in SOL, TST, sleep efficiency or WASO (Michaud et al., 2016). Lane et al. (2016) assessed sleep using actigraphy in 12 WTN exposed individuals and 10 WTN non-exposed individuals and also found no evidence to support that WTN significantly impacted objectively assessed sleep parameters including sleep efficiency, SOL, WASO, TST, time in bed or number of awakenings.

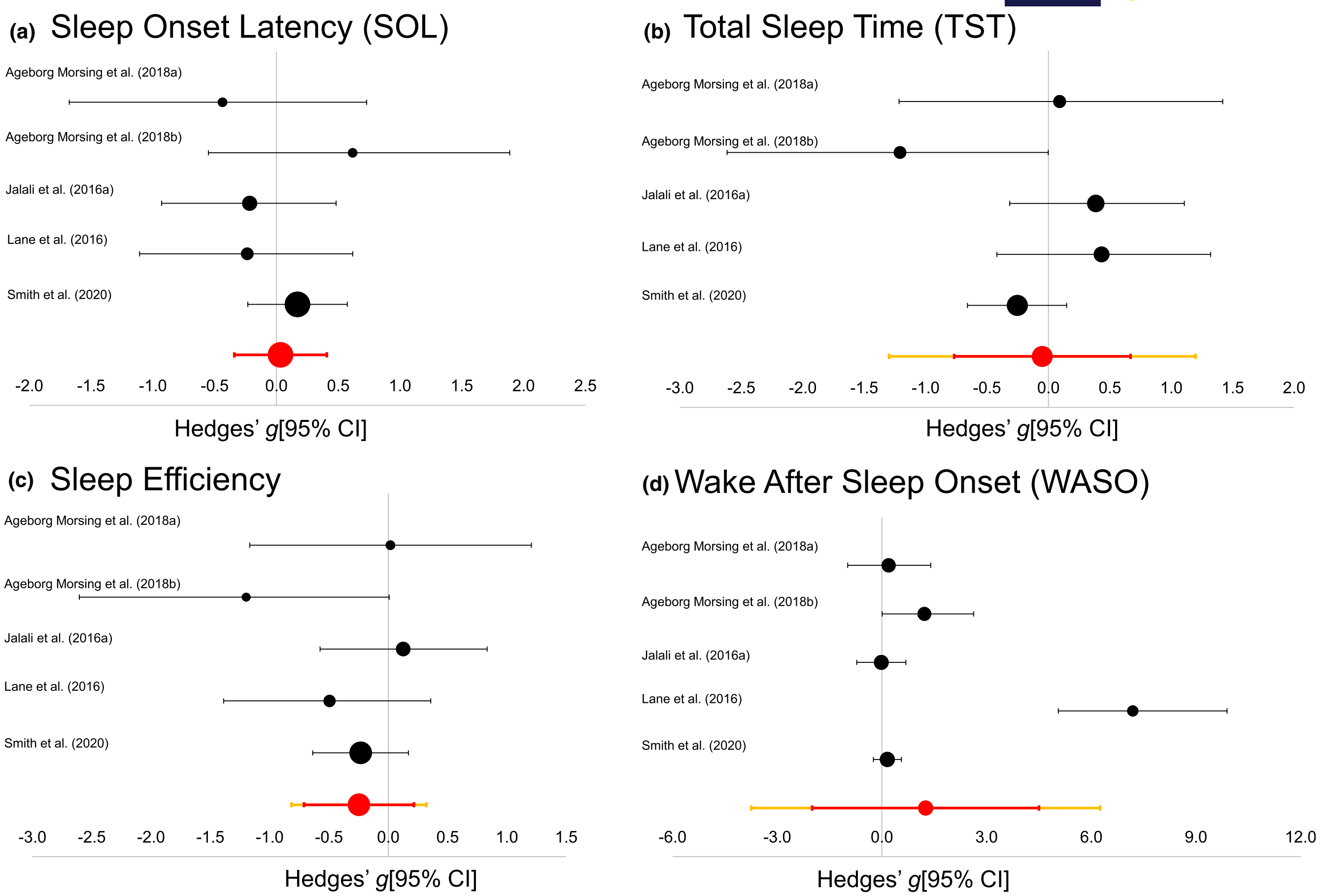

FIGURE 2 Graphical representation of pooled mean effects (effect sizes) for SOL (a), TST (b), sleep efficiency (c) and WASO (d) in the presence and absence of WTN exposure. Negative values on the *x* axis indicate a shorter SOL, less TST, lower sleep efficiency, and a lower amount of WASO in the presence of WTN exposure, while positive values indicate a longer SOL, greater TST, greater sleep efficiency, and a higher amount of WASO in the presence of WTN exposure, compared to control, no WTN exposure. The relative size of the point estimates indicates the study's weighting in the generation of the meta-analytic result. Red error bars represent 95% confidence intervals (CI). The orange error bars indicate 95% predicted interval estimates of where 95% of future studies are predicted to lie. In (a), no orange error bars are present as the 95% prediction intervals are identical to the 95% CI. All studies which evaluated SOL, TST, sleep efficiency and WASO were included in these figures.

TABLE 5 Hedges' *g* (95% CI) and associated meta-analytic statistics

Objective sleep parameter	Range of Hedges' *g* between studies	Combined Hedges' *g* (95% CI)[a]	*p*	95% prediction interval (PI)[b]	Heterogeneity (Q, I^2, *p* value)
SOL, min	−0.44–0.62	0.03(−0.34 to 0.41)	.806	−0.34 to 0.41	Q = 3.29, I^2 = 0%, *p* = .510
TST, min	−1.21–0.43	−0.05(−0.77 to 0.67)	.849	−1.30 to 1.20	Q = 7.81, I^2 = 48.8%, *p* = .099
Sleep efficiency, %	−1.20–0.13	−0.25(−0.71 to 0.22)	.139	−0.82 to 0.32	Q = 4.4, I^2 = 9.16%, *p* = .354
WASO, min	−0.02–7.19	1.25(−2.00 to 4.50)	.284	−3.48 to 5.99	Q = 39.09, I^2 = 89.77%, *p* <.001

[a]95% CI = 95% confidence interval.

[b]PI = 95% prediction interval; 95% of future studies effects are predicted to fall within this range.

3.5.2 | PSG in field versus laboratory settings

The four PSG studies included in the present review and meta-analysis involved both ambulatory PSG in the field and three PSG laboratory studies, for which there were varying results. Jalali et al. (2016a) found no significant differences between objective sleep parameters (including but not limited to SOL, sleep stage distribution, WASO, and TST) from the pre- to post-operational stage of a wind farm. However, average A-weighted WTN exposures were also not significantly different between pre- versus post-operational stages [mean (*SD*) Time 1: 36.5 (4.2) dB(A) versus Time 2: 36.5 (4.2) dB(A), *p* = .959].

The two experimental pilot studies by Ageborg Morsing et al. (2018a, 2018b) used PSG and a morning questionnaire to

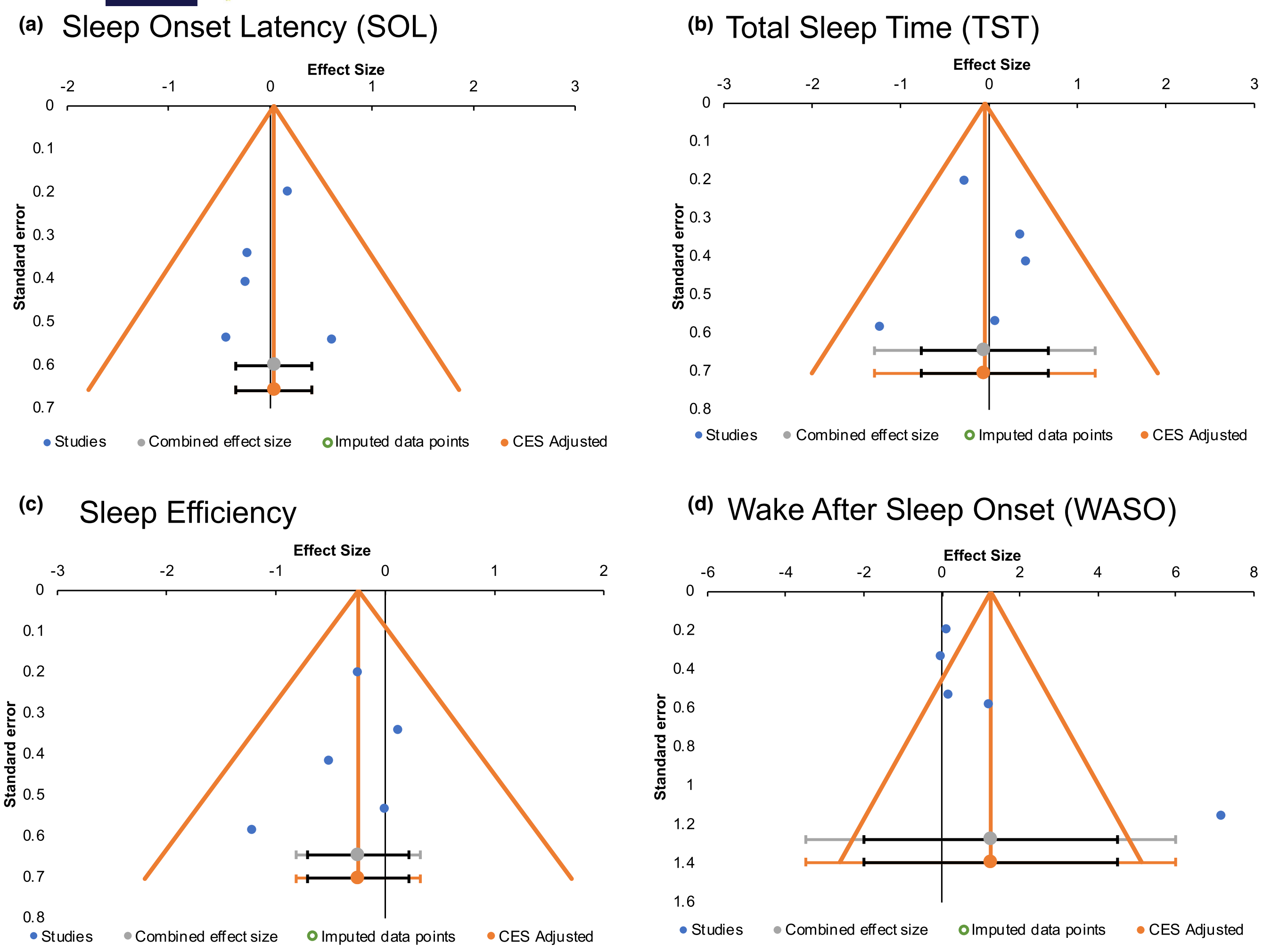

FIGURE 3 Graphical representation of each funnel plot for SOL (a), TST (b), sleep efficiency (c) and WASO (d). These four plots indicate that SOL, TST, sleep efficiency and WASO appear symmetrical, suggesting minimal publication bias. CES: combined effect size.

TABLE 6 Results of a cross-sectional study by Michaud et al. (2016) depicting the measured sleep outcome in comparison to each WTN exposure [dB(A)]

	Sleep parameter, least squares mean (95% CI)			
SPL, dB(A)	SOL, min	TST, min	Sleep efficiency, %	WASO, min
<25 dB(A)	9.9 (6.2–13.6)	447.9 (422.6–473.2)	84.0 (81.9–86.0)	60.9 (54.2–67.6)
25–30 dB(A)	4.4 (1.4–7.5)	442.7 (412.8–472.6)	86.0 (84.1–88.0)	58.6 (50.6–66.6)
30–35 dB(A)	8.1 (5.3–11.0)	438.5 (416.4–460.6)	82.8 (80.8–84.8)	62.7 (57.1–68.2)
35–40 dB(A)	8.5 (6.2–10.8)	444.4 (423.1–465.7)	83.9 (82.2–85.6)	60.8 (55.6–66.0)
40–46 dB(A)	9.9 (7.4–12.4)	438.5 (416.1–460.9)	83.5 (81.7–85.3)	64.1 (57.8–70.3)
p	.1783	.7348	.0519	.3596

Permission for reproduction of this table has been approved by Michaud et al. (2016).

examine objective and self-reported sleep parameters in six participants who had not had prior exposure to WTN. Participants spent five nights in the sleep laboratory and were exposed to various types of WTN and 18 dB L_{Aeq} control background noise. Results showed some significant impacts of WTN on sleep, which are summarised in Table 7. Overall, these two studies found some evidence that wakefulness increases with strong AM and lower frequencies, that deep sleep is reduced in the presence of higher frequencies and stronger AM, and that light sleep increases with higher frequencies and stronger AM. No other significant effects were found in terms of objective sleep parameters in either study. It is worth noting that these two studies used a WTN level that represented worst-case weather conditions designed to increase the likelihood of showing noise effects compared to control.

Lastly, in an experimental study of 50 individuals living within 1 km of a wind turbine and/or reporting annoyance or sleep disturbance by WTN over the past month compared to individuals living further away from wind turbines, Smith et al. (2020) found

TABLE 7 Two experimental pilot studies by Ageborg Morsing et al. (2018a, 2018b) depicting mean (*SD*), statistical significance and Cohen's *d* (95% CI) for each experimental night

PILOT STUDY A					
	Control 18 dB L_{Aeq}	50 dB L_{Aeq} outdoors; 33.7 dB L_{Aeq} indoors (window closed filtering)	40 dB L_{Aeq} outdoors; 29.5 dB L_{Aeq} indoors (window open filtering)	45 dB L_{Aeq} outdoors; 34.1 dB L_{Aeq} indoors (window open filtering)	Post hoc comparisons (*p*, Cohen's *d* (95% CI)
Awakenings, *n*/hr	1.75 (0.63)	2.47 (0.66)	1.69 (0.63)	1.58 (0.75)	[a]$p = .046$, $d = 1.12$ (−0.10 to 2.33) [b]n.s. [c]n.s. [d]$p = .028$, $d = 1.21$ (−0.02 to 2.44) [e]n.s. [f]$p = .028$, $d = 1.26$ (0.02 to 2.50)
PILOT STUDY B					
	Control 18 dB L_{Aeq}	40 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, high AM)	50 dB L_{Aeq} outdoors; 30.4 dB L_{Aeq} indoors (window closed filtering, low AM)	45 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, low AM)	Post hoc comparisons (*p*, Cohen's *d* (95% CI)
N3%	22.8 (4.9)	21.7 (5.3)	22.0 (4.0)	18.0 (3.7)	[a]n.s. [b]$p = .043$, $d = -0.22$ (−1.35 to 0.92) [c]n.s. [d]$p = .046$, $d = -0.06$ (−1.20 to 1.07) [e]n.s. [f]n.s.
	Control 18 dB L_{Aeq}	40 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, high AM)	50 dB L_{Aeq} outdoors; 30.4 dB L_{Aeq} indoors (window closed filtering, low AM)	45 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, low AM)	Post hoc comparisons (*p*, Cohen's *d* (95% CI)
First awakening, min	39.8 (30.0)	58.8 (51.4)	57.3 (59.6)	26.3 (34.7)	[a]n.s. [b]n.s. [c]n.s. [d]$p = .028$, $d = -0.64$ (−1.80 to 0.52); [e]$p = .028$, $d = -0.74$ (−1.91 to 0.43) [f]n.s. [g]n.s.
	Control 18 dB L_{Aeq}	50 dB L_{Aeq} outdoors; 30.4 dB L_{Aeq} indoors (window closed filtering, low AM)	45 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, low AM)	40 dB L_{Aeq} outdoors; 32.8 dB L_{Aeq} indoors (window open filtering, high AM)	Post hoc comparisons (*p*, Cohen's *d* (95% CI)
Maximum continuous N2, min	38.3 (8.0)	26.9 (5.7)	36.1 (9.0)	27.7 (6.6)	[a]$p = .027$, $d = -1.64$ (−2.95 to −0.33) [b]$p = .027$, $d = -1.45$ (−2.72 to −0.18) [c]n.s. [d]n.s. [e]$p = .046$, $d = -1.06$ (−2.27 to 0.15) [f]$p = .028$, $d = -1.22$ (−2.45 to 0.01)

AM, amplitude modulation; dB L_{Aeq}, equivalent continuous sound pressure level,N2, Stage 2 sleep; N3, Stage 3 sleep; n.s., non-significant comparison ($p > .05$). Superscript letters indicate the following paired comparisons: [a]control versus 50 dB L_{Aeq} outdoors; [b]control versus 40 dB L_{Aeq} outdoors; [c]control versus 45 dB L_{Aeq} outdoors; [d]40 dB L_{Aeq} outdoors versus 50 dB L_{Aeq} outdoors; [e]40 dB L_{Aeq} outdoors versus 45 dB L_{Aeq} outdoors; [f]50 dB L_{Aeq} outdoors versus 45 dB L_{Aeq} outdoors;. Permission for reproduction of this table has been approved by Ageborg Morsing et al. (2018a, 2016b).

a significant difference in the percentage of Stage 3 (N3) sleep (p = .034), where the maximum continuous N3 duration in the exposed group was 6.8 min (estimated marginal mean) longer than in the reference group. Smith et al. (2020) also found a significant main effect of study night on the latency to rapid eye movement (REM) sleep and percentage of REM sleep, with an 11.1-min reduction in REM sleep time and a 16.8-min extension of REM latency in the WTN exposure night compared to the control night. No other significant interactions between study night (WTN; 32 dB L_{Aeq} with varying filtering and AM depth, control; 13 dB L_{Aeq} background noise) and study group (reference, exposed) were found for the remaining PSG outcomes investigated.

3.5.3 | Sleep diaries

All five objective studies used some form of sleep diary to assess subjective sleep parameters, (Ageborg Morsing et al., 2018a; Ageborg Morsing et al., 2018b; Jalali et al., 2016a; Lane et al., 2016; Smith et al., 2020), but only Jalali et al. (2016a) reported quantitative sleep diary-based subjective sleep parameters. Jalali et al. (2016a) showed no significant impacts of WTN on TST (p = .472), number of awakenings (p = .126), length of awakenings (p = .062) or SOL (p = .942) from pre- to post-wind farm operation. Lane et al. (2016) used an adapted version of the Pittsburgh Sleep Diary to assess the impact of WTN on exposed versus non-exposed individuals' self-reported sleep. This involved asking participants what time they got into bed, time they fell asleep, their wake-up time, and a sleep quality on a 6-point rating scale. Lane et al. (2016) did not specifically report subjective SOL, TST, number of awakenings, length of awakenings, or WASO, but reported that noise-exposed participants went to bed significantly earlier than the non-exposed participants (p = .02) and went to sleep significantly earlier than the unexposed group (p = .03). No other significant differences in subjective sleep quality were reported.

Ageborg Morsing et al. (2018a, 2018b) used morning questionnaires to assess subjective sleep parameters in the presence versus absence of WTN exposure in a controlled sleep laboratory, in addition to their objective measures of sleep (PSG) on noise-sensitive individuals. The sleep items involved 11-point numerical scales and 5-point descriptive scales (i.e. "very good" to "very bad"). It was also reported that the questionnaire asked about perceived sleep latency and number of awakenings. Ageborg Morsing et al. (2018a) found no significant differences in any of the subjective sleep variables, whereas Ageborg Morsing et al. (2018b) found greater difficulty falling asleep with 32.8 dB(A) indoor WTN exposure (window gap filtering and high AM frequency) and with 30.4 dB(A) indoor WTN exposure (closed window, low AM frequency) compared to a control night (p = .032). No other significant effects were found between the 3 WTN exposure nights and the control night including SOL and number of awakenings.

In addition to objective sleep measures, Smith et al. (2020) also used morning questionnaires to assess subjective sleep parameters in the presence and absence of WTN exposure on 50 individuals. Smith et al. (2020) found no significant interactions between study night and study group but found significantly lower sleep quality, greater difficulty falling back to sleep after an awakening, increased difficulty sleeping, sleeping worse than usual, and waking more frequently after the WTN exposure night compared to the control night. Similarly, Smith et al. (2020) also found that noise-exposed participants rated their sleep quality as being more negative than the control group after both nights.

3.5.4 | ISI

The pre–post field study on 37 individuals before and after the operation of wind turbines was the only study that reported on insomnia severity scores and found that self-reported insomnia symptoms were significantly higher from pre- to post-operational wind turbines [mean (SD) score 3.1 (3.6) versus 6.4 (6.7), p = .005], with a moderate effect size (Cohen's d = 0.62) (Jalali et al., 2016b). Whilst there was a notable increase in ISI scores, it is important to note that these scores are below levels considered to reflect subthreshold insomnia (ISI >7). These findings also showed that 45.9% of the 37 participants had a negative attitude, 18.9% had a neutral attitude, and 32.4% had a positive attitude towards wind turbines. Jalali et al. (2016b) further reported that changes in ISI scores from Time 1 to Time 2 were strongly associated with negative attitudes to WTN (p = .003).

3.5.5 | PSQI

Three studies used the PSQI to assess the impact of WTN on perceived sleep quality (Jalali et al., 2016b; Michaud et al., 2016; Nissenbaum et al., 2012). Jalali et al. (2016b) showed that self-reported sleep quality was significantly poorer following compared to prior to WTN exposure [mean (SD) score 6.2 (3.9) versus 4.1 (2.1), p = .006] with a moderate effect size (Cohen's d = 0.67). PSQI scores of >5 are considered to indicate poor sleep, so these results support a shift from good to poor sleep with WTN exposure. Jalali et al. (2016b) also found that almost 50% of participants had a negative attitude towards wind turbines and that changes in PSQI scores from Time 1 to Time 2 were strongly associated with negative attitudes (p = .002). Nissenbaum et al. (2012) conducted a cross-sectional field study in 79 individuals showing similar results, whereby participants living near a wind turbine (375–1,400 m) showed poorer sleep quality than participants living further away (3,000–6,600 m) from a wind turbine (mean score 7.8 versus 6.0, p = .046). However, variance was not reported so effect sizes could not be calculated, and A-weighted noise levels were variable ranging from 32–61 dB L_{Aeq}. Lastly, Michaud et al. (2016) reported no significant relationships between PSQI scores and model estimated WTN exposure levels.

3.5.6 | ESS

Three studies used the ESS and consistently reported significant associations between daytime sleepiness and WTN exposure (Abbasi et al., 2015; Jalali et al., 2016b; Nissenbaum et al., 2012). Jalali et al. (2016b) showed that self-reported daytime sleepiness of residents was significantly greater following the post-operation of wind turbines compared to pre-operational wind turbines [mean (*SD*) score 7.1 (5.3) versus 4.7 (3.2), $p = .002$], with a moderate effect (Cohen's $d = 0.56$). However, daytime sleepiness did not reach the clinical cut-off indicative of excessive daytime sleepiness. Nissenbaum et al. (2012) showed similar results, whereby participants living 375–1,400 m from a wind turbine showed greater ESS scores (mean 7.8) than participants living 3,000–6,600 m away from a wind turbine (mean 5.7; $p = .032$). Again, effect sizes could not be calculated as variance was not reported. In addition, Abbasi et al. (2015) showed significantly greater ESS scores for wind farm maintenance staff than security staff and administrative staff. In addition, the maintenance staff showed clinically relevant ESS scores (>10), indicating significant daytime sleepiness. However, while ESS scores of >10 indicate excessive daytime sleepiness, attribution to necessarily indicate the presence of a sleep disorder and/or sleep disturbance is problematic.

4 | DISCUSSION

We examined existing literature to evaluate and meta-analyse the potential impact of WTN on sleep using objective and/or psychometrically validated subjective measures of sleep. To our knowledge, only one systematic review and meta-analysis has previously examined this question, and was limited to self-reported, cross-sectional study outcomes available at that time (Onakpoya et al., 2015). Several more recently published studies have included objective measures and more validated questionnaires widely used in sleep research to assess sleep outcomes in the presence versus absence of WTN. Nine studies met eligibility criteria and of those, six used objective sleep measurement (PSG or actigraphy) and three used psychometrically validated questionnaires. Included objective studies varied in methodologies and outcome measures (field, laboratory, PSG, actigraphy); however, five of the six objective studies uniformly reported key sleep outcomes including SOL, TST, WASO, and sleep efficiency. The meta-analysis of five studies found no evidence to support that objectively measured sleep latency, sleep efficiency, time spent asleep and awake during the night are significantly different in the presence versus absence of WTN exposure.

However, it is worth noting that Jalali et al. (2016a) and Jalali et al. (2016b) reported that average A-weighted WTN exposure was not significantly different between pre- versus post-operational stages and thus it is perhaps not surprising that objective sleep outcomes were not impacted. Furthermore, findings by Smith et al. (2020) were also not surprising, given they assessed perceived sleep disturbance in a group who were already self-reporting sleep disturbance or presenting with annoyance towards WTN in comparison to a general sample of unexposed individuals.

Field studies are clearly the most ecologically valid and most representative of real-world WTN conditions in comparison to in-laboratory studies. However, field studies lack control over extraneous variables such as changes in wind speed, wind direction, atmospheric turbulence, topography, study blinding, nocebo effects, trial design quality, and other environmental factors including visual impacts that also have the potential to impact objective and subjective sleep disturbance (Aziz, 2017; Micic et al., 2018). For example, many of these factors can influence the airflow, turbulence and propagation of WTN leading to variability in AM, infrasound, tonality and swish components, and thus could play a part in reports of sleep disturbance. Study design differences and the way in which noise exposure is conducted could importantly influence different findings across studies (Micic et al., 2018).

Whilst actigraphy is an objective measure, unlike PSG it does not directly monitor cortical activity so relies on sleep–wake inferences based on pre-defined activity thresholds. Thus, actigraphy has poor specificity for discriminating wake from sleep when activity is low (Marino et al., 2013). Actigraphy is also unreliable for detecting micro-arousals, which may or may not be associated with gross body movements. In addition, whilst actigraphy is able to record data across the day, the algorithms that are used during sleep periods at night may not directly translate to detecting sleep during the day. Further shortcomings of actigraphy involve the fact that manual scoring is at times still used, which can introduce human error, inter- and intra-scorer variability (Driller et al., 2016). Whilst, computerised scoring algorithms help to reduce human error, automatic scoring algorithms are still faced with limitations, due to heavily relying on the estimation of sleep parameters rather than the actual activity measurement (de Souza et al., 2003) and the possibility of the off-wrist detection being mis-scored as sleep (Grandner & Rosenberger, 2019).

In contrast, whilst PSG is technical, intrusive, expensive and still subject to inter- and intra-scorer variability given the need for study setup, supervision and manual scoring by skilled sleep technicians with extensive training in scoring EEG activity (Van De Water et al., 2011) experimental laboratory-based studies using PSG do allow for substantially superior control of most extraneous variables that may confound sleep outcomes (Aziz, 2017). PSG also allows for the measurement of more fine-grain microstructural changes and sleep stage changes that extend beyond basic sleep architecture (Aziz, 2017). Ultimately, carefully controlled experimental laboratory studies are needed to definitively establish the impact of WTN on sleep. More controlled experimental studies evaluating WTN exposure effects compared to quiet control conditions using PSG sleep assessment and a repeated measures design have shown some significant impacts of WTN on the timing of the first awakening, frequency of awakenings per hour, reductions in deep sleep, less continuous time spent in N2 sleep, prolonged REM latency, and decreased REM sleep percentage. Whilst these repeated measures designs have not shown significant effects on the standard sleep metrics (SOL, WASO, TST and sleep efficiency), these results do suggest that some more detailed changes in cortical activity, sleep stage changes, and physiology in sleep can become impacted by WTN. Whilst two of the experimental laboratory-based studies had limited samples (e.g. six each), finer-grained analyses of sleep outcomes beyond basic sleep architecture are warranted in future PSG studies to investigate impacts of WTN on sleep using larger sample sizes. On the other hand, participants in these studies were also likely aware of the WTN exposure before falling asleep and during night-time awakenings as the noise exposure was present from lights out time and played until lights on in the morning. This could have potentially biased not only participant's self-reported responses, but also their objective sleep quality. Given the difficulties in controlling extraneous factors in field studies, between-subjects designs, and with participant awareness and potential attitudinal biases; the interpretation of sleep findings from field studies is particularly problematic. Thus, future repeated measures, laboratory-based PSG experimental studies, using study protocols designed to compare the presence versus absence of psychological awareness of noise exposure are clearly needed. For example, methodologies including WTN exposure only during sleep versus wake versus continuously throughout the night could allow for a deeper exploration of both psychological and physiological factors that may influence WTN noise effects on objective and subjective sleep measures.

Three of the nine studies only used psychometrically validated sleep questionnaires, and the six objective studies also used psychometrically validated sleep questionnaires in addition to either PSG or actigraphy. Findings based on self-reported sleep perception were mixed, likely partly reflecting the use of different assessment tools assessing somewhat different sleep outcomes. Studies including items that assessed self-reported sleep parameters including SOL, WASO, TST, and number of awakenings found no significant impacts of WTN in comparison to control background noise without WTN. Jalali et al. (2016b), Michaud et al. (2016) and Nissenbaum et al. (2012) all used the PSQI to assess sleep quality in the presence of WTN and produced mixed findings. Jalali et al. (2016b) found poorer PSQI and ISI scores post- compared to pre-operational WTN exposure. These results could have been impacted by the absence of study blinding, as participants were fully aware of the impending turbine presence and noise exposure, for which participant attitude and expectation bias risks are high, particularly for self-report outcomes (Jalali et al., 2016a; Jalali et al., 2016b). For example, given no significant differences in pre–post WTN exposure levels, these results suggest that being aware of a wind farm beginning operation may have contributed to increased ISI scores. Visual impacts and awareness of wind turbine existence and attitudes towards wind farms, instead of the WTN itself could also play a role. Furthermore, the study by Jalali et al. (2016b) also showed that almost 50% of participants had a negative attitude towards wind turbines, thus, attitudinal effects appear likely to help explain why participants self-reported poorer sleep quality and insomnia symptoms following the operation of the wind turbines. The between-groups study by Nissenbaum et al. (2012) was based on a combination of both predicted and measured WTN and found significantly poorer PSQI scores in participants who lived near wind turbines, compared to unexposed residents. Michaud et al. (2016) also assessed PSQI scores based on five WTN exposure levels and found no significant relationship between PSQI scores and WTN exposure. However, given the large-scale cross-sectional study design, these authors were reliant on WTN exposure estimates from sound propagation models rather than direct noise measurements, which may not necessarily adequately capture difference in noise exposures between regions and groups.

Abbasi et al. (2015), Jalali et al. (2016b) and Nissenbaum et al. (2012) assessed daytime sleepiness using the ESS. Although methods varied, consistent associations between WTN exposure and daytime sleepiness were found. Jalali et al. (2016b) showed significantly stronger associations in daytime sleepiness after wind farm operation. Likewise, Nissenbaum et al. (2012) showed that exposed individuals living close to wind turbines showed significantly greater daytime sleepiness than unexposed people living further away from wind turbines. Abbasi et al. (2015) assessed three wind farm worker groups (maintenance, security, and administrative staff) during wind farm operation, which were used to manipulate the relative distance and thus SPLs of WTN exposure. Abbasi et al. (2015) used 8-hr equivalent sound levels in their study. Results showed a dose–response relationship, whereby those working closer to wind turbines showed greater daytime sleepiness than those working further away. These results do not provide support for a particular wind farm worker job type being associated with sleep disturbance/sleep disorder presence or even momentary daytime sleepiness but rather speaks to their habitual daytime sleepiness symptoms, thus may not necessarily indicate sleep disturbance or sleep disorder presence. However, cross-sectional associations of daytime exposure levels between different worker types are inherently problematic. This is because group differences could potentially be confounded by uncontrolled factors such as differential participant characteristics (age, gender) and risk factors for sleep problems unrelated to daytime noise exposure presumably away from the usual sleep environment.

4.1 | Strengths and limitations

This is the first systematic review and meta-analysis that has investigated the impact of WTN on key sleep outcomes. The review used a robust search strategy to identify pertinent articles and underwent a reporting quality assessment based on an adapted version of the STROBE checklist (Von Elm et al., 2007). A limitation of this review was that there were only a small number of identified studies included with varied methodologies and outcome variables, which prevented a more comprehensive meta-analysis of quantifiable sleep measures. In addition, although a second author did review the retrieved studies for eligibility and reporting quality, only TL initially screened abstracts for eligibility and reporting quality. Furthermore, this review and meta-analysis treated WTN exposure as a binary outcome (i.e. exposed versus unexposed to WTN), despite the differences in WTN levels and acoustical characteristics between and within the different studies.

Studies also used mixed methods with variable measurement and model-based estimates of noise exposure levels and differing WTN exposures including worst-case WTN with characteristics that could be particularly problematic (e.g. AM, infrasound, tonality) through to more typical WTN. Given inevitable variability in weather and local conditions known to influence WTN, well-controlled laboratory studies are clearly needed to more specifically determine WTN effects on sleep.

In addition, despite the use of objective and psychometrically validated subjective sleep measures; selection and response biases, as well as the absence of study blinding may importantly influence both objective and subjective sleep measures. For example, strong negative or positive attitudes towards wind turbines and awareness of study conditions or interventions appear likely to impact study participation, self-reported and potentially objective sleep parameters through expectation effects on abilities and times taken to fall asleep, remain asleep versus awake overnight, and to wake following sleep. With the exception of Smith et al. (2020) who reported that participants were not blinded in their study, all other studies did not report nor consider blinding. As blinding is inherently difficult and not considered or reported by the large majority of these studies, the results should be interpreted with caution. Therefore, based on the aforementioned limitations and risk of biases, the strength of the present evidence surrounding the impact of WTN on objective and psychometrically validated sleep is lacking.

Overall, the results of the present systematic review and meta-analysis do not support that WTN significantly impacts the main objective markers of sleep quality including SOL, TST, WASO, and sleep efficiency, but does appear to impact some subjective sleep outcomes; which supports the notion of WTN being an environmental psychosocial stressor that has the ability to contribute to self-reported sleep disturbance, and in some instances, as well as impact some sleep stage shifts and number of awakenings per hour. However, future experiments should consider including WTN exposures only during wake versus only during established sleep (e.g. ≥N2 sleep) to help separate potential psychological versus physiological influences of WTN on sleep. Only exposing participants to WTN during established sleep periods would help to avoid potential participant expectation biases to more specifically investigate the impact of WTN on objective and subjective measures of sleep. Similarly, comparisons between WTN versus quiet control exposures during wake periods may also be needed to more specifically test for conscious awareness and bias effects on sleep propensity that also strongly influence both objective and subjective sleep.

5 | CONCLUSION

In summary, the present review used a systematic and meta-analytic approach to investigate WTN effects on objective and subjective sleep outcomes. Nine studies using objective and/or psychometrically validated subjective sleep measures were identified and included. To date, various methodologies, noise measurements, and outcome assessments have been used and shown mixed findings. Assessments of WTN impacts on sleep using "gold-standard" PSG assessment are starting to emerge and the present review provides an update and summary of these findings. This meta-analysis suggests that key indicators of objectively measured sleep macrostructure (i.e. sleep latency, sleep efficiency, total time spent asleep and awake) under well-controlled laboratory conditions and in the field are not significantly impacted by WTN compared to no-WTN noise control conditions. However, studies that have used a repeated measures design, under controlled laboratory conditions have shown some significant changes in more detailed measures of cortical activity and sleep stages. Whilst the "gold-standard" PSG is the most objective and most direct way to measure physiological impacts of WTN on sleep; self-report measures are also needed to assess perceived sleep quality, particularly for evaluating insomnia. Overall, few studies have used psychometrically validated subjective measures of sleep. Due to inconsistent findings and mixed methodologies, a meta-analysis of subjective sleep outcomes was not possible (e.g. sleep quality, insomnia severity, and daytime sleepiness). However, available data support that insomnia symptom severity, sleep quality, and daytime sleepiness are impacted by WTN exposure in comparison to no WTN exposure, whereas sleep diary parameters, e.g. self-reported SOL, TST, WASO, and sleep efficiency show less consistent findings. Future studies should more strongly consider potential confounding through selection and response biases and study blinding effects in their design, and also consider noise stimuli representative of typical WTN exposure, as well as the less common WTN features such as AM, infrasound and swish characteristics. Finally, methodologies that expose individuals to WTN only during sleep versus wake periods may be important to help separate subjective versus objective sleep effects and the likelihood that psychological awareness, attitudinal, and/or noise-sensitivity factors could also impact sleep.

ACKNOWLEDGEMENTS

This work was performed at Flinders University, Adelaide at the Adelaide Institute of Sleep Health. TL would also like to thank Hannah Scott and Claire Dunbar, for their advice and support throughout this process.

CONFLICT OF INTEREST

KH has research funds from the Australian Research Council. PC has research funds from the National Health and Medical Research Council. BZ has research funds from the National Health and Medical Research Council. LL is a shareholder in Re-time Pty. Ltd. and has received royalties from Retime, which is unrelated to the submitted work. TL, NL and GM have no conflicts of interest to declare.

AUTHORS CONTRIBUTIONS

All authors contributed to the manuscript as follows: TL: conceptualisation, methodology, software, formal analysis, data curation, writing – original draft preparation, writing – review and editing, visualisation. PC: writing – review and editing, supervision. LL: writing – review and editing, supervision. KH: writing – review and editing. BZ: writing – review and editing, visualisation. NL: writing – review and editing. GM: conceptualization, methodology, writing – review and editing, supervision. All authors have read and approved the final manuscript.

ORCID

Tessa Liebich https://orcid.org/0000-0001-9209-2165
Leon Lack https://orcid.org/0000-0002-8505-5873
Kristy Hansen https://orcid.org/0000-0003-2400-6207
Branko Zajamšek https://orcid.org/0000-0003-4467-8588
Nicole Lovato https://orcid.org/0000-0001-8990-6658
Peter Catcheside https://orcid.org/0000-0002-9372-6788
Gorica Micic https://orcid.org/0000-0001-5281-8603

REFERENCES

Abbasi, M., Monazzam, M. R., Akbarzadeh, A., Zakerian, S. A., & Ebrahimi, M. H. (2015). Impact of wind turbine sound on general health, sleep disturbance and annoyance of workers: A pilot-study in Manjil wind farm, Iran. *Journal of Environmental Health Science and Engineering*, *13*, 71. https://doi.org/10.1186/s40201-015-0225-8

Ageborg Morsing, J., Smith, M. G., Ögren, M., Thorsson, P., Pedersen, E., Forssén, J., & Persson Waye, K. (2018a). Wind turbine noise and sleep: Pilot studies on the influence of noise characteristics. *International Journal of Environmental Research and Public Health*, *15*, 2573. https://doi.org/10.3390/ijerph15112573

Ageborg Morsing, J., Smith, M., Ögren, M., Thorsson, P., Pedersen, E., Forssén, J., Persson Waye, K. (2018b). Wind Turbine Noise and Sleep: Pilot Studies on the Influence of Noise Characteristics. *International Journal of Environmental Research and Public Health*, *15*, (11), 2573https://doi.org/10.3390/ijerph15112573

Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1–2), 29–37. https://doi.org/10.3109/00207459008994241

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.

Aziz, H. A. (2017). Comparison between field research and controlled laboratory research. *Archives of Clinical and Biomedical Research*, *1*, 101–104. https://doi.org/10.26502/acbr.50170011

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., & Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. *The Lancet*, *383*(9925), 1325–1332. https://doi.org/10.1016/S0140-6736(13)61613-X

Basner, M., & McGuire, S. (2018). WHO environmental noise guidelines for the European region: A systematic review on environmental noise and effects on sleep. *International Journal of Environmental Research and Public Health*, *15*, 519. https://doi.org/10.3390/ijerph15030519

Bergland, B., & Lindvall, T. (1995). Community noise. *Archives of the Center for Sensory Research*, *2*(1), 1–195.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons Ltd.

Buysse, D. J., Reynolds, C. F. III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*, 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep*, *35*, 287–302. https://doi.org/10.5665/sleep.1642

Crichton, F., Dodd, G., Schmid, G., Gamble, G., Cundy, T., & Petrie, K. J. (2014). The power of positive and negative expectations to influence reported symptoms and mood during exposure to wind farm sound. *Health Psychology*, *33*, 1588. https://doi.org/10.1037/hea0000037

de Souza, L., Benedito-Silva, A. A., Pires, M. L. N., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep*, *26*(1), 81–85. https://doi.org/10.1093/sleep/26.1.81

Driller, M., McQuillan, J., & O'Donnell, S. (2016). Inter-device reliability of an automatic-scoring actigraph for measuring sleep in healthy adults. *Sleep Science*, *9*(3), 198–201. https://doi.org/10.1016/j.slsci.2016.08.003

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eberhardt, J. L., & Akselsson, K. R. (1987). The disturbance by road traffic noise of the sleep of young male adults as recorded in the home. *Journal of Sound and Vibration*, *114*, 417–434. https://doi.org/10.1016/S0022-460X(87)80014-7

Evandt, J., Oftedal, B., Hjertager Krog, N., Nafstad, P., Schwarze, P. E., Marit Aasvang, G. (2017). A Population-Based Study on Nighttime Road Traffic Noise and Insomnia. *Sleep*, *40*, (2), https://doi.org/10.1093/sleep/zsw055

Frank, E., Sidor, M. M., Gamble, K. L., Cirelli, C., Sharkey, K. M., Hoyle, N., Tikotzky, L., Talbot, L. S., McCarthy, M. J., & Hasler, B. P. (2013). Circadian clocks, brain function, and development. *Annals of the New York Academy of Sciences*, *1306*, 43–67. https://doi.org/10.1111/nyas.12335

Grandner, M. A., & Rosenberger, M. E. (2019). Actigraphic sleep tracking and wearables: Historical context, scientific applications and guidelines, limitations, and considerations for commercial sleep devices. In M. L. Grandner (Ed.), *Sleep and Health* (pp. 147–157). : Academic Press.

Griefahn, B., Marks, A., & Robens, S. (2006). Noise emitted from road, rail and air traffic and their effects on sleep. *Journal of Sound and Vibration*, *295*, 129–140. https://doi.org/10.1016/j.jsv.2005.12.052

Hak, T., Van Rhee, H. J., & Suurmond, R. (2016). *How to interpret results of meta-analysis*. (Version 1.0) : Erasmus Rotterdam Institute of Management. Retrieved from www.erim.eur.nl/researchsupport/meta-essentials/downloads

Hansen, C. H., Doolan, C. J., & Hansen, K. L. (2017). *Wind farm noise: Measurement, assessment, and control*. John Wiley & Sons Ltd.

Hansen, K. L., Nguyen, P., Zajamšek, B., Catcheside, P., & Hansen, C. H. (2019). Prevalence of wind farm amplitude modulation at long-range residential locations. *Journal of Sound and Vibration*, *455*, 136–149. https://doi.org/10.1016/j.jsv.2019.05.008

Iber, C., Ancoli-Israel, S., Chesson, A. L., & Quan, S. F. (2007). *The AASM manual for the scoring of sleep and associated events: Rules. terminology and technical specification* (1st ed.). American Academy of Sleep Medicine.

Jalali, L., Bigelow, P., Nezhad-Ahmadi, M. R., Gohari, M., Williams, D., & McColl, S. (2016a). Before–after field study of effects of wind turbine noise on polysomnographic sleep parameters. *Noise & Health*, *18*, 194. https://doi.org/10.4103/1463-1741.189242

Jalali, L., Nezhad-Ahmadi, M. R., Gohari, M., Bigelow, P., & McColl, S. (2016b). The impact of psychological factors on self-reported sleep disturbance among people living in the vicinity of wind turbines. *Environmental Research*, *148*, 401–410. https://doi.org/10.1016/j.envres.2016.04.020

Janssen, S. A., Basner, M., Griefahn, B., Miedema, H., & Kim, R. (2011). Environmental noise and sleep disturbance. In World Health Organisation (Ed.), *Burden of disease from environmental noise* (pp. 55–70). World Health Organisation.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep*, *14*, 540–545. https://doi.org/10.1093/sleep/14.6.540

Johns, M. W. (1992). Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep*, *15*, 376–381. https://doi.org/10.1093/sleep/15.4.376

Krogh, C. M. E., Gillis, L., Kouwen, N., & Aramini, J. (2011). WindVOiCe, a self-reporting survey: Adverse health effects, industrial wind turbines, and the need for vigilance monitoring. *Bulletin of Science, Technology & Society*, *31*, 334–345. https://doi.org/10.1177/0270467611412551

Kuroiwa, M., Xin, P., Suzuki, S., Sasazawa, Y., Kawada, T., & Tamura, Y. (2002). Habituation of sleep to road traffic noise observed not by polygraphy but by perception. *Journal of Sound and Vibration*, *250*, 101–106. https://doi.org/10.1006/jsvi.2001.3901

Lane, J. D., Bigelow, P. L., Majowicz, S. E., & McColl, R. S. (2016). INTERNATIONAL PERSPECTIVES: Impacts of industrial wind turbine noise on sleep quality: Results from a field study of rural residents in Ontario, Canada. *Journal of Environmental Health*, *79*, 8–13.

Leventhall, H. G. (2004). Low frequency noise and annoyance. *Noise and Health*, *6*(23), 59–72.

Lovato, N., Lack, L., Wright, H., & Kennaway, D. J. (2014). Evaluation of a brief treatment program of cognitive behavior therapy for insomnia in older adults. *Sleep*, *37*, 117–126. https://doi.org/10.5665/sleep.3320

Maich, K. H., Lachowski, A. M., & Carney, C. E. (2018). Psychometric properties of the consensus sleep diary in those with insomnia disorder. *Behavioral Sleep Medicine*, *16*, 117–134. https://doi.org/10.1080/15402002.2016.1173556

Marino, M., Li, Y. I., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M. (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*, *36*, 1747–1755. https://doi.org/10.5665/sleep.3142

Marks, A., & Griefahn, B. (2007). Associations between noise sensitivity and sleep, subjectively evaluated sleep quality, annoyance, and performance after exposure to nocturnal traffic noise. *Noise & Health*, *9*(34), 1.

Martin, J. L., & Hakim, A. D. (2011). Wrist actigraphy. *Chest*, *139*, 1514–1527. https://doi.org/10.1378/chest.10-1872

Michaud, D. S., Feder, K., Keith, S. E., Voicescu, S. A., Marro, L., Than, J., Guay, M., Denning, A., Murray, B. J., Weiss, S. K., Villeneuve, P. J., van den Berg, F., & Bower, T. (2016). Effects of wind turbine noise on self-reported and objective measures of sleep. *Sleep*, *39*, 97–109. https://doi.org/10.5665/sleep.5326

Micic, G., Zajamsek, B., Lack, L., Hansen, K., Doolan, C., Hansen, C., Vakulin, A., Lovato, N., Bruck, D., Chai-Coetzer, C. L., Mercer, J., & Catcheside, P. (2018). A review of the potential impacts of wind farm noise on sleep. *Acoustics Australia*, *46*, 87–97. https://doi.org/10.1007/s40857-017-0120-9

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*, 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson, A., Coleman, J., Lee-Chiong, T., Pancer, J., & Swick, T. J. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: An update for 2007. *Sleep*, *30*, 519–529. https://doi.org/10.1093/sleep/30.4.519

Morin, C. M. (1993). *Insomnia: Psychological Assessment and Management*. Guilford Press

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*, *34*, 601–608. https://doi.org/10.1093/sleep/34.5.601

Muzet, A. (2007). Environmental noise, sleep and health. *Sleep Medicine Reviews*, *11*, 135–142. https://doi.org/10.1016/j.smrv.2006.09.001

Nissenbaum, M. A., Aramini, J. J., & Hanning, C. D. (2012). Effects of industrial wind turbine noise on sleep and health. *Noise & Health*, *14*, 237. https://doi.org/10.4103/1463-1741.102961

Onakpoya, I. J., O'Sullivan, J., Thompson, M. J., & Heneghan, C. J. (2015). The effect of wind turbine noise on sleep and quality of life: A systematic review and meta-analysis of observational studies. *Environment International*, *82*, 1–9. https://doi.org/10.1016/j.envint.2015.04.014

Perlis, M. L., Giles, D. E., Mendelson, W. B., Bootzin, R. R., & Wyatt, J. K. (1997). Psychophysiological insomnia: The behavioural model and a neurocognitive perspective. *Journal of Sleep Research*, *6*(3), 179–188. https://doi.org/10.1046/j.1365-2869.1997.00045.x

Riemann, D., Spiegelhalder, K., Feige, B., Voderholzer, U., Berger, M., Perlis, M., & Nissen, C. (2010). The hyperarousal model of insomnia: A review of the concept and its evidence. *Sleep Medicine Reviews*, *14*(1), 19–31. https://doi.org/10.1016/j.smrv.2009.04.002

Schmidt, J. H., & Klokker, M. (2014). Health effects related to wind turbine noise exposure: A systematic review. *PLoS One*, *9*(12), e114183. https://doi.org/10.1371/journal.pone.0114183

Sivertsen, B., Omvik, S., Havik, O. E., Pallesen, S., Bjorvatn, B., Nielsen, G. H., Straume, S., & Nordhus, I. H. (2006). A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep*, *29*, 1353–1358. https://doi.org/10.1093/sleep/29.10.1353

Smith, M. T., McCrae, C. S., Cheung, J., Martin, J. L., Harrod, C. G., Heald, J. L., & Carden, K. A. (2018). Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *Journal of Clinical Sleep Medicine*, *14*, 1209–1230. https://doi.org/10.5664/jcsm.7228

Smith, M. G., Ögren, M., Thorsson, P., Hussain-Alkhateeb, L., Pedersen, E., Forssén, J., Ageborg Morsing, J., & Persson Waye, K. (2020). A laboratory study on the effects of wind turbine noise on sleep: Results of the polysomnographic WiTNES study. *Sleep*, *43*, 1–14.

Sweetman, A., Lack, L., Lambert, S., Gradisar, M., & Harris, J. (2017). Does comorbid obstructive sleep apnea impair the effectiveness of cognitive and behavioral therapy for insomnia? *Sleep Medicine*, *39*, 38–46. https://doi.org/10.1016/j.sleep.2017.09.003

Thorne, B. (2011). The problems with “noise numbers” for wind farm noise assessment. *Bulletin of Science, Technology & Society*, *31*, 262–290. https://doi.org/10.1177/0270467611412557

Van De Water, A. T., Holmes, A., & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography–a systematic review. *Journal of Sleep Research*, *20*, 183–200. https://doi.org/10.1111/j.1365-2869.2009.00814.x

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, *147*, 573–577. https://doi.org/10.7326/0003-4819-147-8-200710160-00010

World Health Organization (2011). *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*. World Health Organization. Regional Office for Europe.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Liebich T, Lack L, Hansen K, et al. A systematic review and meta-analysis of wind turbine noise effects on sleep using validated objective and subjective sleep assessments. *J Sleep Res*. 2020;00:e13228. https://doi.org/10.1111/jsr.13228