International Psychogeriatrics (2024), 36:12, 1128–1156 © The Author(s), 2024. Published by Cambridge University Press on behalf of International Psychogeriatric Association
doi:10.1017/S104161022400070X

REVIEW

Sleep-related measurements to assess sleep disturbances among people living with dementia in nursing homes: a systematic review

Jonas Dörner,[1,2] Kathrin Wehner,[2] Margareta Halek,[2] and Martin N. Dichter[3]

[1]*German Center for Neurodegenerative Diseases (DZNE), Witten, Germany*
[2]*School of Nursing Science, Faculty of Health, Witten/Herdecke University, Witten, Germany*
[3]*Institute of Nursing Science, University of Cologne Medical Faculty and University Hospital of Cologne, University of Cologne, Cologne, Germany*

ABSTRACT

Background: There is a high prevalence of sleep disturbances among people living with dementia (PLWD) in nursing homes. Reliable and valid measurements are needed to assess these disturbances. The aim of this systematic review was to identify, analyze and synthesize studies of sleep-related measurements to assess sleep disturbances in PLWD.

Methods: The databases PubMed, CINAHL, and PsycINFO were systematically searched in 2019; the search was updated in March 2024. The inclusion criteria were as follows: participants with dementia or probable dementia in any care setting; and studies that reported at least one of the following aspects: (I) theoretical and conceptual frameworks, (II) user or patient involvement by type of users in measurement development, (III) feasibility and practicability of measurements, and (IV) results of psychometric analyses. The quality of the included studies was evaluated using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria and the quality appraisal tool for studies of diagnostic reliability (QAREL) tool.

Results: A total of 5169 studies were identified; ultimately, 15 studies describing three self-administered measurements, three proxy-administered measurements and two technological measurements were included. No sleep-related measurement showed acceptable psychometric properties in any of the COSMIN domains.

Conclusions: No measurement without adaptation can be recommended for PLWD in nursing homes. If existing measurements are used in clinical practice, the self-perspective of PLWD should be taken into account. If this is no longer fully possible, proxy-rating perspectives in combination could be an option. Future research on sleep-related measurements should be strictly based on international consensus-based psychometric quality criteria.

Key words: dementia, sleep disturbances, measurement, nursing home

Introduction

In 2020, approximately 50 million people worldwide were living with dementia (Livingston *et al.*, 2020). These numbers are expected to increase to 82 million in 2030 and to 152 million in 2050 (Patterson, 2018). Up to 36% of people living with dementia suffer from sleep disturbances (Garcia-Alberca *et al.*, 2013; Webster *et al.*, 2020a; Wilfling *et al.*, 2019). Many health conditions are associated with sleep disturbances (Fung *et al.*, 2016), such as depression, disinhibition and aberrant motor behavior (Garcia-Alberca *et al.*, 2013). Wakefulness at night and longer rapid-eye-movement sleep latencies are associated with poorer cognitive performance (Moe *et al.*, 1995), physical complaints, respiratory disabilities, poor self-reported health (Foley *et al.*, 1995) and mortality (Gehrman *et al.*, 2004). People living with dementia in nursing homes have reported that disturbed sleep is often associated with restlessness and pondering (Dörner *et al.*, 2023). This finding is consistent with the experience of nurses who characterize disturbed sleep of people living with dementia mainly by behavioral and psychological symptoms (Dörner *et al.*, 2023; Webster *et al.*, 2020b). The day after disturbed sleep, people living with dementia

Correspondence should be addressed to: Jonas Dörner, German Center for Neurodegenerative Diseases (DZNE), Witten, Germany. Tel: +49 2302926 238; Fax: +49 2302 88986 238. E-mail: Jonas.Doerner@dzne.de Received 11 Dec 2023; revision requested 20 Feb 2024; revised version received 27 May 2024; accepted 28 May 2024. First published online 19 September 2024.

experience poor wellbeing, confusion, reduced cognitive- and physical performance and exhaustion. In contrast, good sleep during the day is characterized by feeling well, improved cognition, full physical ability, increased interaction and being in balance (Dörner *et al.*, 2023). Approximately four out of five nurses have reported regularly observing sleep disturbances among people living with dementia in nursing homes (Wilfling *et al.*, 2020a).

Previous systematic reviews examined pharmacological interventions against sleep disturbances in people with Alzheimer's disease across all settings (McCleery *et al.*, 2014) and nonpharmacological interventions in nursing home residents (Wilfling *et al.*, 2020b). The primary studies included in those reviews used a variety of sleep-related outcome measurements to assess sleep variables or detect sleep disturbances, such as length of sleep. These measures included self-reported measures ($n = 4$) (Gattinger *et al.*, 2017; Kuck *et al.*, 2014; Serfaty *et al.*, 2002), proxy-reported measures ($n = 5$) (Alessi *et al.*, 2005; Alessi *et al.*, 1999; Gattinger *et al.*, 2017; Kuck *et al.*, 2014; Schnelle *et al.*, 1999; Serfaty *et al.*, 2002; Singer *et al.*, 2003) and technological devices ($n = 3$) (Alessi *et al.*, 2005; Alessi *et al.*, 1999; Camargos *et al.*, 2014; Dowling *et al.*, 2008; Gattinger *et al.*, 2017; Kuck *et al.*, 2014; Li *et al.*, 2017; NCT00325728, 2008; Richards *et al.*, 2011; Schnelle *et al.*, 1999; Serfaty *et al.*, 2002; Singer *et al.*, 2003). None of the measures used in those primary studies were developed specifically for people living with dementia and application in the nursing home setting. Therefore, it is unclear how appropriate these sleep-related measurements are for measuring sleep disturbances among people with dementia in the nursing home setting. To date, no systematic review has examined sleep-related measurements to assess sleep disturbances in relation to their psychometric properties for people living with dementia in nursing homes based on established consensus-based guidelines.

Therefore, the aims of this systematic review were as follows:

1. to identify sleep-related measurements to assess sleep disturbances that were developed for people living with dementia or that have been applied in this population,
2. to describe the theoretical basis, scope, domains, and extent of user involvement during the development process of sleep-related measurements to assess sleep disturbances;
3. to evaluate the reliability, validity and feasibility of the identified sleep-related measurements to assess sleep disturbances; and
4. to recommend sleep-related measurements to assess sleep disturbances among people living with dementia in nursing homes.

Methods

Design

This review is based on the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative for systematic reviews of patient-reported outcome measures (Mokkink *et al.*, 2018; Prinsen *et al.*, 2018). This systematic review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines for systematic reviews (Page *et al.*, 2021).

Literature search

A systematic search was performed in September 2019 and updated in March 2024 without any restrictions regarding publication date. The search strategy (Appendix A1) was developed iteratively based on the Population, Intervention, Comparison, Outcome (PICO) framework (Straus *et al.*, 2018). We initially used Google Scholar to perform an open search of the abovementioned databases, thus helping us to develop the syntax. Then, we systematically searched the PubMed, CINAHL and PsycINFO databases. Within the analyses, the first author conducted backward citation tracking of the included studies to obtain additional eligible studies. If a reference was not available, the authors and journals were contacted to ask for access. If sleep-related measurements were not sufficiently described in the identified studies, the authors were contacted for further information.

Study selection

Included studies had to be primarily focused on the development or psychometric evaluation of sleep-related measurements. The target group in our review was people diagnosed with dementia or possible dementia. Studies were also included if only a part of the target population had a possible dementia or was diagnosed with dementia. Studies without a dementia population were also included via backward citation tracking if they described the theoretical basis and the development of a sleep-related measurement. There were no restrictions regarding the care setting, thus enabling us to examine a wide range of sleep-related measurements among people living with dementia. Only studies published in English or German were included. We excluded studies that examined sleep-related

measurements that cannot be applied in nursing home care because they require extraordinarily high requirements, for instance, in terms of space-, personnel- or specific technical requirements that are not usually available (e.g., polysomnography). Appendix A2 provides an overview of the eligibility criteria. Two reviewers (JD, MND, KW) independently performed the study selection.

Data extraction

The data extraction was performed in two steps: (1) full-text analyses and data extraction were performed by one reviewer (JD), and (2) an independent cross-check of all extracted data and their accuracy was performed by a second reviewer (KW, MND). Any discrepancies were resolved by discussion between the reviewers or by consulting a third reviewer.

Synthesis and methodological quality of the extracted data

All data regarding the theoretical background and development of measurements, the characteristics of the measurements, the application of technological devices and the psychometric properties of the measurements were entered into standardized tables. Feasibility was analyzed based on recommendations from the literature in the following eight domains: *acceptability, demand, implementation, practicality, adaptation, integration, expansion and limited-efficacy testing* (Bowen *et al.*, 2009). Guidelines from the COSMIN initiative (Prinsen *et al.*, 2018) were used to assess internal consistency, test-retest reliability, construct validity and criterion validity. Additionally, the interrater reliability was assessed with the QAREL tool (Lucas *et al.*, 2010). One reviewer (JD) rated the quality of the studies (see Table 1), and the results were then cross-checked and discussed with a second reviewer (MND). Actigraphy results were descriptively analyzed and later discussed based on recommendations from the literature (Camargos *et al.*, 2013).

Results

Description of included studies

The systematic search conducted in 2019 identified $n = 3552$ records from the three databases. After removing duplicates, $n = 2642$ studies remained. Our updated search was performed at the end of March 2024 and yielded $n = 1617$ additional studies (thereof $n = 370$ duplicates). In total, $n = 3889$ studies were subjected to title and abstract screening. In the second step, $n = 53$ full-text articles were checked for eligibility. No measurement was excluded because it requires extraordinary high requirements that are not usually available in nursing homes. Ultimately, $n = 11$ records were included for data extraction. Three additional records were retrieved through backward citation tracking, and $n = 1$ paper was retrieved by contacting the author of an included paper who subsequently coauthored an additional psychometric manuscript which is published by now. Finally, $n = 15$ studies were identified that investigated $n = 8$ different measurements (Figure 1).

Sleep-related measurements assessing sleep disturbances

The $n = 15$ studies assessed three self-administered measurements: the Epworth Sleepiness Scale (ESS) (Johns, 1991), the Alternative Version of the Epworth Sleepiness Scale (ESS-ALT) (Gronewold *et al.*, 2021) and the Pittsburgh Sleep Quality Index (PSQI) (Buysse *et al.*, 1989). Moreover, three proxy-administered measurements were investigated in the included studies: the Observational Sleep Assessment Instrument (OSAI) (Cohen-Mansfield *et al.*, 1989), the Sleep Continuity Scale in Alzheimer's Disease (SCADS) (Manni *et al.*, 2013) and the Sleep Disorders Inventory (SDI) (Tractenberg *et al.*, 2003). In addition, three studies investigated actigraphy (Ancoli-Israel *et al.*, 1997; Gibson and Gander, 2019; Van Someren, 2007), and one study investigated a wrist monitoring system (Nijhof *et al.*, 2012). The theoretical basis and characteristics of each measurement are presented in Tables 1 and 2. A description of the actigraphy characteristics is presented in Table 3. Detailed results of the psychometric characteristics and the reasons for the evaluation of methodological quality are shown in Table 4.

Theoretical basis of sleep-related measurements

Three out of six studies that examined the development of measurements used clinical experts as a source during the development process (Buysse *et al.*, 1989; Cohen-Mansfield *et al.*, 1989; Gronewold *et al.*, 2021). Other studies used previous literature as a source (Buysse *et al.*, 1989; Gronewold *et al.*, 2021; Johns, 1991; Tractenberg *et al.*, 2003). Moreover, one study included nurses, patients and relatives in the development process (Gronewold *et al.*, 2021). Two studies did not report the sources they used during the measurement development process (Cohen-Mansfield *et al.*, 1989; Sinforiani *et al.*, 2007). Two measurements were developed to assess symptoms of sleep disturbances

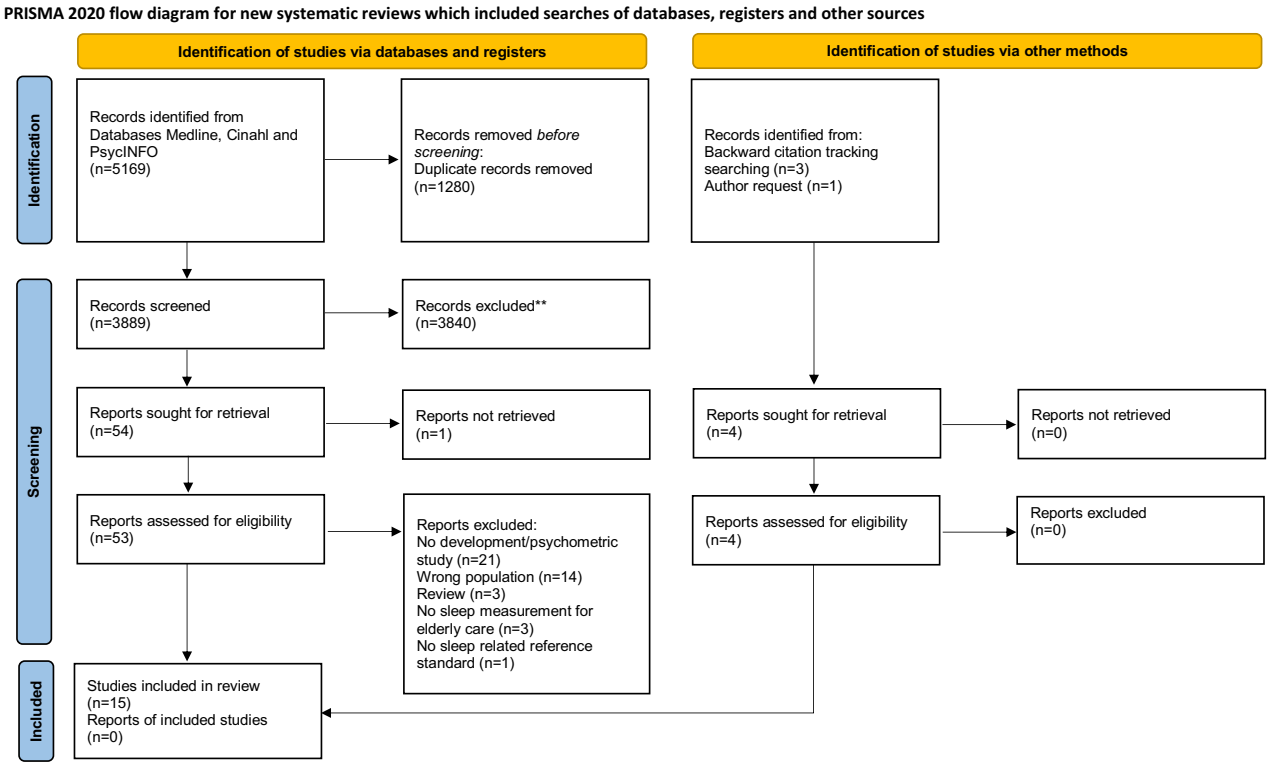

Figure 1. Flow diagram of identified studies (Page *et al.*, 2021).

during both nighttime and daytime (Buysse *et al.*, 1989; Tractenberg *et al.*, 2003); one measurement solely assessed sleep disturbances at night (Cohen-Mansfield *et al.*, 1989); two measurements solely assessed daytime sleepiness (Gronewold *et al.*, 2021; Johns, 1991); and one measurement remains unclear (Sinforiani *et al.*, 2007).

Risk of bias

Among the studies that included people living with dementia, the risk of bias was assessed for each self- and proxy-administered measurement by investigating the validity or reliability (Appendix A3).

The SCADS (Manni *et al.*, 2013) was found to have an adequate risk of bias for **structural validity** because only an exploratory factor analysis was performed.

The ESS-ALT (Gronewold *et al.*, 2021), SCADS (Manni *et al.*, 2013) and SDI (Hjetland *et al.*, 2020) were found to have very good risk of bias for **internal consistency,** as the Cronbach's alpha coefficients were calculated for each unidimensional scale. The PSQI (Curcio *et al.*, 2013) was found to have inadequate internal consistency, as the Cronbach's alpha coefficient was not calculated for each subdomain.

The PSQI (Curcio *et al.*, 2013) was found to have inadequate **cross-cultural validity** because the samples were not similar for relevant characteristics and because the subsamples were small.

The PSQI (Curcio *et al.*, 2013) was found to have a very good risk of bias for **criterion validity** because the area under the curve, sensitivity and specificity were calculated. However, the SDI (Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003) and OSAI (Cohen-Mansfield *et al.*, 1989) were found to have inadequate risk of bias for criterion validity because their reference standards were not consistent with international diagnostic criteria. Moreover, one study that assessed the criterion validity of the SDI (Tractenberg *et al.*, 2003) did not report the area under the curve, sensitivity or specificity, thus leading to the inadequate rating. The degree of hypothesis testing was rated as very good for the PSQI because the construct described by the comparator measurement is clear. The OSAI (Cohen-Mansfield *et al.*, 1989) and SDI (Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003) were found to have inadequate ratings because the construct of the comparator measurement was not clear; furthermore, unsatisfactory information about the measurement properties of the comparator measurement were presented in each study.

The SDI (Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003) was found to have an inadequate risk of bias for **responsiveness** in two studies because the quality of the comparator measurement was insufficient.

Table 1. Theoretical basis of identified sleep measurements

Source/ Country	Measurement	Aim	Theoretical framework and/or definition of sleep disturbances	Sources for construction	Secondary Paper people living with dementia
Johns, 1991, Australia	• Epworth Sleepiness Scale (ESS)	• Development of a measurement to assess daytime sleepiness among adults	• Derived from observations regarding the nature and occurrence of daytime sleep and sleepiness	• Clinical Experts • Literature	• Frohnhofen *et al.*, 2009 • Gronewold *et al.*, 2021
Gronewold *et al.*, 2021, Gemany	• Epworth Sleepiness Scale Alternative Version (ESS-ALT)	• Modification of the ESS to administer to geriatric patients exhibiting physical or mental disorders	• See Epworth Sleepiness Scale	• Literature review • Experts in sleep research and medicine • Input from nurses • Patients and their relatives	
Buysse *et al.*, 1989, UK	• Pittsburgh Sleep Quality Index (PSQI)	• Development of a psychometrically tested that can measure a variety of sleep disturbances and that can distinguish between good and poor sleepers among patients with mental disorders	• Derived from three sources: (1) clinical intuition and experience with patients with sleep disorders, (2) a literature review targeting previous sleep quality questionnaires and (3) clinical experience with the measurement during an 18 months field test • Sleep quality includes quantitative (sleep duration sleep latency and arousals in the night) and qualitative (depth of sleep, restfulness and perceived sleep quality) aspects of sleep	• Clinical Experts • Literature	• Curcio *et al.*, 2013 • Most *et al.*, 2012
Cohen-Mansfield *et al.*, 1990, USA	• Observational Sleep Assessment Instrument (OSAI)	• Development and validation of a measurement for observing sleep in nursing homes	• Sleep disruptions are defined as times that residents wakes up – even briefly – during a 3-minute observation period	• Not reported	
Sinforiani *et al.*, 2007, Italy	• Questionnaire for Hallucinations and Sleep-Wake Cycle in Alzheimer's disease (SCADS)	• Development of a measurement to investigate the relationship between hallucinations and the Sleep-Wake cycle among people living with Alzheimer´s disease	• The hypothesis (with respect to hallucinations in sleep disruptions) was that there is a relation between physiopathogenesis and sleep disorders presenting as visual hallucinations in neurodegenerative diseases	• Not reported	• Manni *et al.*, 2013 • Manni *et al.*, 2015
Tractenberg *et al.*, 2003, USA	• Sleep Disorders Inventory (SDI), developed based on the NPI [Cummings *et al.*, 1994]	• Development of a measurement to assess and quantify sleep disturbances and sleep disorders in people living with Alzheimer´s disease	• Definition: Sleep disturbances are defined as less than 6 hours of total sleep time in the night	• Literature (Cummings *et al.*, 1994) • No further information available	• Hjetland *et al.*, 2020

Table 2. Characteristics of sleep-related measurements to detect sleep disturbances in people living with dementia

Measurement, Source/country	Measurement period	Measurement content (Items and Domains)	Rating evaluation (Scores and Ratings)
Self Assessments			
Epworth Sleepiness Scale (ESS), Johns, 1991, Australia,	• Recent time	**Items*** • 8 **Sub domains** • No sub domains	**Scores** • Global score 0–24 **Ratings** • 8 items rated on a 4-point scale (would never doze – high chance of dozing)
Epworth Sleepiness Scale alternative version (ESS-ALT), Gronewold *et al.*, 2021, Germany	• Recent time	**Items** • 8 **Sub domains** No sub domains	**Scores** • Global Score 0–24 **Ratings** 8 items rated on a 4-point scale (would never doze – high chance of dozing)
Pittsburgh Sleep Quality Index (PSQI), Buysse *et al.*, 1989, United Kingdom	• Last 28 days	**Items** • 19 **Sub domains** • A: Subjective sleep quality (1 item) • B: Sleep latency (2 items) • C: Sleep duration (1 item) • D: Habitual sleep efficiency (3 items) • E: Sleep disturbances (9 items) • F: Use of sleeping medication (1 item) • G: Daytime dysfunction (2 items)	**Scores** • Global score 0–21, Scores >5 means disturbed sleep • Seven subscores 0–3 (A-G) **Ratings** • 12 items assessing frequency on a 4-point scale (not during the last month – three or more times a week) • 4 items, open-ended answers • 1 item assessing severity on a 4-point scale (very good - very bad) • 1 item rated on a 4-point scale (no problem at all – a very big problem) • 1 item rated on a 4-point scale (no bedpartner or roommate – partner in same bed)
Proxy			
Observational Sleep Assessment Instrument OSAI, Cohen-Mansfield *et al.*, 1990, USA	• 3 min observation period	**Items** • 17 **Sub domains** • Sleep (5 items) • Sleep patterns (12 items)	**Scores** • Descriptive scoring for each item • **Ratings** • 6 items, dichotomous (yes/no) • 9 open-ended items for counting of an event (numeric variables) • 1 item rated on a 4-point scale (low – irregular) • 1 item rated on a 5-point scale (open – both)

Table 2. Continued

Measurement, Source/country	Measurement period	Measurement content (Items and Domains)	Rating evaluation (Scores and Ratings)
Selection of specific items of the Questionnaire for Hallucinations and Sleep-Wake Cycle in Alzheimer's Disease (SCADS), Manni *et al.*, 2013, Italy	• Not reported	**Items** • 9 **Sub domains** • No sub domains	**Scores** • Not reported **Ratings** • Not reported
Questionnaire for Hallucinations and Sleep-wake cycle in Alzheimer's Disease, Sinforiani *et al.*, 2007, Italy	• Not reported	**Items** • 46 **Sub domains (no information available about the item distribution regarding sub domains)** • Sleep habits • Sleep hygiene • Symptoms of sleep disorders including snoring and apneas during sleep • Occurrence of violent related sleep episodes with or without dream mentation • Occurrence of hallucinations	**Scores** • Not reported **Ratings** • Dichotomous items (no information about the number) • Multiple choice items (no information about the number)
Sleep Disorders Inventory (SDI), Tractenberg *et al.*, 2003, USA	• Last 2 weeks	**Items** • 8 **Sub domains** • No sub domains	**Scores** • Global Score 0–12 (summed frequency score) • Global Scale 0–84 (summed product score product[§]) **Ratings** • Different response ratings: • 5-point scale rating the frequency of sleep disturbances (not present in the last 2 weeks – once or more per day (every night) • 5-point scale rating the severity of sleep disturbances (not present – marked: nighttime behaviors occur; several types of night time behavior may be present; the patient is very distressed during the night and the caregiver's sleep is markedly disturbed • 6-point scale rating caregiver distress (not at all – very severely/extremely)

*For items in detail of the measurements see Appendix A4.

§The summed product score was not calculated within the measurement development paper but in several later publications, e.g.: Tewary, S., Cook, N., Pandya, N., & McCurry, S. M. (2018). Pilot test of a six-week group delivery caregiver training program to reduce sleep disturbances among older adults with dementia (Innovative practice). Dementia (London), 17(2), 234-243. doi:10.1177/1471301216643191; Wilfling, D., Dichter, M. N., Trutschel, D., & Köpke, S. (2019). Prevalence of Sleep Disturbances in German Nursing Home Residents with Dementia: A Multicenter Cross-Sectional Study. J Alzheimers Dis, 69(1), 227-236. doi:10.3233/jad-180784.

Table 3. Application of technological devices in people living with dementia

Source/ Country	Technology & scoring	Recording time in weeks	Recorded sleep parameters	Wearing protocol or comparison measure	Localization
Ancoli-Israel *et al.*, 1997, USA	Algorithm: • Cole *et al.*, 1992* Signal: • Not reported Epochs: • 1-minute epochs	• Not reported	• Total sleep time • Total wake time • Percent sleep	• Polysomnography	• Wrist actigraphy, wrist not reported
Gibson *et al.*, 2018, New Zealand	Algorithm: • From the manufacturer of the Actigraph Signal: • Threshold Epochs: • 1-minute epochs	• 1	• Sleep epochs (%) • Wake epochs (%)	• Not reported	• Wrist actigraphy on nondominant wrist
Nijhof *et al.*, 2012, Netherlands	Algorithm: • Auto learning algorithm (no further information reported Signal: • Not reported Epochs: • Not reported	• 24	• Sleep time • Sleep periods • Circadian rhythm	• Sleep diary	• Wrist monitoring, wrist not reported
Van Someren, 2007, Netherlands	Algorithm: • From the manufacturer of the Actiwatch Signal: • Not reported Epochs: • 1-minute epochs	• 2	• Total sleep time • Sleep efficiency (%) • Sleep-wake rhythm parameters (interdaily stability, intradaily variability, amplitude of rhythm)	• Sleep logs	• Wrist actigraphy, wrist not reported

*Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep. 1992 Oct;15(5):461-9. doi: 10.1093/sleep/15.5.461. PMID: 1455130.

Table 4. Method and results of psychometric studies of sleep-related measurements to detect sleep disturbances in people living with dementia

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
Self-rated only						
Epworth Sleepiness Scale						
Frohnhofen *et al.*, 2009, Germany	Epworth Sleepiness Scale (ESS) • 8 items	• Investigation of the feasibility of the ESS for daytime sleepiness	• Hospital patients, geriatric department • Total $n = 433$ • Thereof PLWD* $n = 192$	**Feasibility** Acceptability • Not reported Demand • Not reported Implementation • Not reported Practicality • 36% were able to complete the scale ($n = 166$) • 26% had problems with single items ($n = 118$) • 38% couldn't answer any item ($n = 174$) • Incomplete questionnaires were significantly correlated with higher age and higher impairments in activities of daily living Adaptation • Not reported Integration • Not reported Expansion • Not reported Limited-efficacy testing • Not reported	**n.a.****	• Participants who were unable to complete the ESS were significantly older and had more severe disability • The authors concluded that the scale was only feasible in 36% of the sample • No specific analysis for PLWD

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
Gronewold *et al.*, 2021, Germany	Epworth Sleepiness Scale (ESS) • 8 items	• Investigation of the reliability of the ESS compared to Epworth Sleepiness Scale alternative version for daytime sleepiness • Investigation of the practicability	• Hospital patients, geriatric' department • Patients of the department ≥ 65 years (total $n = 52$) PLWD ($n = 14$)	**Feasibility** Acceptability Patients/relatives/medical staff • Brief, easy, assess daytime sleepiness symptoms • No version for relatives • Patients and relatives reported for missing items that they were not in the described situations Expert opinion • Easy to use, good acceptance on patient side, assesses daytime sleepiness symptoms, adequate for self-reporting • Proportion of missing values is high, patients and relatives disagree often, no reliable measure for daytime sleepiness, often assistance needed while rating Demand • Not reported Implementation • Not reported Practicality • 3 s to process (self-report $n = 52$) • 2 s to process (relative report $n = 7$) • 73.1% missing values total score of self-reported measurements (out of $n = 52$) • 57.1% missing values total score of relative-reported measurements (out of $n = 7$)	**n.a.**	• No specific analysis for PLWD

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				• Patients needed only little help by the interviewer to answer the items • 12% needed support on item 1: repetition of response format, repetition of situation, more detailed explanation of the situation • 2% needed support on item 4 Adaptation • Not reported Integration • Not reported Expansion • Not reported Limited-efficacy testing • Not reported **Reliability** IC • Cronbach's alpha = 0.23 ($n = 52$)	-	• No confidence interval reported
Epworth Sleepiness Scale Alternative Version						
Gronewold *et al.*, 2021, Germany	Epworth Sleepiness Scale Alternative Version (ESS-ALT) • 8 items	• Investigation of the reliability and validity of the ESS-ALT compared to the ESS for daytime sleepiness • Investigation of the feasibility	• In-hospital patients department of geriatrics • Patients of the department (total $n = 52$) • PLWD ($n = 14$)	**Feasibility** Acceptability • Not reported Demand • Not reported Implementation • Not reported	**n.a.**	• No sample size calculation • No PCA performed • Sample consists of PLWD and people without dementia • Time requirement, ease of administration and patient/relative opinion was only reported for potential new developed items for

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
						the ESS-ALT but not for the final version
				Practicality • 23.1% missing values total score of self-reported measurements (out of $n = 52$) • 28.6% missing values total score of relative-reported measurements (out of $n = 7$) Adaptation • Not reported Integration • Not reported Expansion • Not reported Limited-efficacy testing • Not reported		
				Reliability IC • Cronbach's alpha = 0.64 ($n = 52$)	-	• No confidence interval reported
Pittsburgh Sleep Quality Index						
Curcio *et al.*, 2013, Italy	Pittsburgh Sleep Quality Index (PSQI) • 19 items	Validation of the Italian version of the PSQI	• All participants were recruited in a sleep laboratory • Group I: Young healthy controls ($n = 10$) • Group II: Healthy elderly ($n = 10$) • Group III: PLWD ($n = 10$)	**Reliability** IC • Cronbach's alpha = 0.84 ($n = 50$ overall, all groups)	?	• No PCA performed • Scoring evaluation based on the original procedure (Buysse *et al.*, 1989) and a three-factor model (Cole *et al.*, 1992) • No sample size calculation • Sample consists of PLWD and people without dementia • Small sample size • No confidence interval reported

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
			• Group IV: Patients with obstructive sleep apnea syndrome ($n = 10$) • Group V: Patients with depression ($n = 10$)			
				Validity Criterion • Sensitivity: 94% cutoff score 5 • Specificity: 47% cutoff score 5 • AUC*** 0.71	?	• Reference standard for criterion validity: Polysomnography • Sample size for criterion validity is unclear • Participants underwent polysomnography for two consecutive nights in sleep laboratory. Only the second night was analyzed to avoid first-night effects
Proxy Assessments						
Observational Sleep Assessment Instrument						
Cohen-Mansfield *et al.*, 1990, USA	Observational Sleep Assessment Instrument (OSAI) • 17 items	Investigation of interrater reliability and criterion validity	• Nursing home ($n = 1$) • Patients ($n = 20$)	**Reliability** IRR • 92.7% agreement on average ($n = 33.5$ observations)	+	• Observer: trained registered nurse specialist • Small sample size • No confidence interval reported QAREL tool (Lucas *et al.*, 2010): • Unclear whether raters were blinded regarding any aspect of the study • Unclear whether order of examination varied
				Validity Criterion (for sleep efficiency <70%) • Sensitivity: 77.8% ($n = 17$) • Specificity 100% ($n = 17$)	-	• Reference standard for criterion validity: Portable sleep assessment measurement (including actigraphy for sleep variables on the wrist) • Not reported how many nights the actigraph was worn

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
						• No confidence interval reported
Sleep continuity scale in Alzheimer's disease						
Manni *et al.*, 2015, Italy	Sleep continuity scale in Alzheimer's disease (SCADS) • 9 items	Feasibility of the questionnaire	• Neurological Institute ($n = 1$) • Outpatients ($n = 275$)	**Feasibility** Acceptability • Measurement is a suitable, rapid and easy in daily clinical practice Demand • Not reported Implementation • Not reported Practicality • 5–10 minutes to process • N = 216 total, $n = 59$ participants were excluded because no relative or caregiver living with the participant exists or the dementia degree was to severe • The SCADS is not applicable in people living with severe dementia • Persons who are sharing the bed as proxy rater are necessary for rating Adaptation • Not reported Integration • Not reported Expansion • Not reported Limited-efficacy testing	**n.a.**	• Participation of the study from 10/2012 to 03/2014

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				• Not reported		
Manni *et al.*, 2013, Italy	Sleep continuity scale in Alzheimer's disease SCADS • 9 items	Investigate internal consistency and the construct validity between sleep patterns and Alzheimer's Disease clinical variables	• Neurological Institute ($n = 1$) • Outpatients ($n = 140$)	**Reliability** IC • Cronbach's alpha = 0.94 ($n = 140$)	?	• No confidence interval reported • No sample size calculation
				Validity Construct • PCA: one factor with an explained variance of 70%	-	• Questionnaire was developed based on another one (see (Sinforiani *et al.*, 2007)) and has similarities with the Sleep Disorders Inventory
Sleep Disorders Inventory						
Hjetland *et al.*, 2020, Norway	Sleep Disorder Inventory (SDI) • 8 items	Adaptation of the SDI for the nursing home context and validation with wrist Actigraphy	• Dementia nursing home units ($n = 8$) • PLWD ($n = 69$)	**Reliability** IC • Cronbach's alpha = 0.82 for frequency ratings ($n = 59$) • Cronbach's alpha = 0.87 for severity ratings ($n = 59$)	?	• No PCA performed • Small sample size • No sample size calculation • 95% confidence interval
				Validity Criterion • Sensitivity summed product score 70% ≥ 5 ($n = 59$) • Specificity summed product score 78% ≥ 5 ($n = 59$) • Sensitivity summed frequency score ≥ 5 67% ($n = 59$) • Specificity summed frequency score ≥ 5 81% ($n = 59$) • AUC summed product score 0.77 • AUC summed frequency score 0.78 Convergent ($n = 59$) • Total sleep time by actigraphy (AG)	+	• Reference standard for criterion validity: Actigraphy • At least five nights of recording with actigraphy • 95% confidence interval

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				SDI average total score: – .431 SDI summed product score: – .432 SDI summed frequency score: – .436 • Wake after sleep onset by AG SDI average total score: .389 SDI summed product score: .402 SDI summed frequency score: .395 • Neuro Psychiatric Inventory Nursing Home version - sleep item SDI average total: .746 SDI summed product: .751 SDI summed frequency: .754		
Tractenberg *et al.*, 2003, USA	Sleep Disorder Inventory (SDI) • 8 items	Report of study results and characteristics of the SDI	• Research centers ($n = 36$) • PLWD ($n = 104$)	**Validity** Convergent • Sleep quality rating (sleep diary) SDI summed frequency: – 0.277 • Night total sleep time (NTST) by actigraphy (AG) SDI summed frequency: – 0.244 • Day total sleep time (DTST) by AG • Sleep efficiency by AG	-	• Reference standard for convergent validity: Actigraphy • No hypothesis regard to expected correlations between SDI and actigraphy • No PCA performed • No sample size calculation • No characteristics for the sample of proxy-raters reported • Actigraphs were worn for 2-3 weeks • No confidence interval reported

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				SDI summed frequency: – 0.283 • Wake after sleep onset by AG SDI summed frequency: 0.243 • DTST/NTST by AG SDI summed frequency: 0.215 • 24 h Total sleep time SDI summed frequency: – 0.084		
Technological devices **Actigraphy**						
Ancoli-Israel *et al.*, 1997, USA	Actigraphy	Testing the reliability of actigraphy against polysomnography and observations	• Nursing homes • PLWD ($n = 10$)	**Reliability** IRR • Spearman's correlation with total sleep 0.87	**n.a.**	• Interrater reliability for a subset of EEG recordings through two experienced sleep technologists • Number of participating nursing homes is not reported • All participants were wheelchair bound • No confidence interval reported QAREL tool (Lucas *et al.*, 2010): • Only severe demented participants • Only 24-h recording of polysomnography • Unclear whether raters were blinded regarding any aspect of the study • Unclear whether order of examination varied

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				Validity Criterion • Sensitivity (detect wake): 87% ($n = 10$) • Specificity (detect sleep: 90% ($n = 10$)	**n.a.**	• Reference standard for criterion validity asleep and awake: Observations • Observer: Trained research staff • No confidence interval reported
Gibson *et al.*, 2018, New Zealand	Actigraphy	Evaluation of the reliability of actigraphy compared to sleep diaries	• Community setting • PLWD ($n = 15$)	**Validity** Criterion • Sensitivity (detected sleep, detected wake): 87%, 77% ($n = 15$) • Specificity (detected sleep, detected wake): 80%, 61% ($n = 15$)	**n.a.**	• Reference standard for criterion validity: Sleep diaries • Actiraphs were worn for one week • 95% confidence interval
Nijhof *et al.*, 2012, Netherlands	Wrist monitoring technology	Feasibility (usage and usability) of actigraphy (watch)	• Nursing home ($n = 1$) • PLWD ($n = 7$)	**Feasibility** Acceptability • Caregiver did not think that the introduction of the watch would take so long during a regular team meeting • Caregiver were skeptical in the beginning and did not think that the watch is helpful for their tasks • Not user-friendly, need an improvement • Computer system for generating data is easy to understand/ interpret data • In one case caregivers did not believe the data of the watch that the PLWD was truly awake but the camera in the room proved it (cameras were used for one night) Demand • Not reported	**n.a.**	• Actigraphs were worn for 6 months • The authors used a mixed methods approach: Monitoring data, observations (through researcher), Interviews (caregiver), diaries (caregiver) • $N = 7$ PLWD (severe dementia), $n = 1$ drop out and $n = 1$ "refilled" • The wrist monitoring system is very similar to actigraphy and measures the same sleep parameters

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				Implementation • Not reported Practicality • 140 working hours per external project manager (100 were planned) • Two caregivers were selected as opinion leaders in the project • Relatives were introduced in the topic from the project manager • Change from a removable strap to an irremovable failed because the resident open this strap as well • Caregiver were able to see out of the monitoring data if a resident's sleeping behavior was good or not • Computer system for generating data is easy to understand/ interpret data • Absent of good environment for printing out the data (printer was on another level in the nursing home) • Too big for small fragile arms of the elderly • Hard strap, which irritates patients' skin when it comes in contact with water • A normal clock face would be good for the elderly to look familiar		

Table 4. Continued

Source	Measurement (scales and items)	Objective(s)	Setting & Sample	Psychometric properties	Methodological quality	Methodological comments
				Adaptation • Not reported Integration • Not reported Expansion • Technological installation 1000 € • 315€ per watch Limited-efficacy testing • Not reported		
Van Someren 2006, Netherlands	Actigraphy	Investigation of the reliability in relation to recording duration (in days)	• Group care facilities for elderly (n = not reported) • PLWD ($n = 12$)	**Reliability** ICC • Total sleep time 0.97 after 10 days ($n = 12$) • Sleep efficiency 0.96 after 10 days ($n = 12$) • Interdaily stability 0.95 after 10 days ($n = 12$) • Intradaily variability 0.93 after 10 days ($n = 12$) • Amplitude 0.97 after 10 days ($n = 12$)	**n.a.**	• ICC values were assessed for 10 days in total and compared with each other: two single days, two periods of two days, periods of three days etc. • No confidence interval reported

*PLWD = people living with dementia.
** = Not applicable.
*** = Area under curve.

Description of measurements

Epworth Sleepiness Scale (ESS). The ESS is a self-administered measurement for measuring daytime sleepiness – in particular, it assesses the nature and occurrence of daytime sleep. It was developed for adult patients in the hospital setting. It was constructed based on the literature and clinical expertise. The scale includes eight items with no subdomains (Johns, 1991). Its feasibility was tested in $n = 433$ geriatric patients, of whom $n = 192$ were people living with dementia. In total, only 36% of the patients were able to complete the measurement (Frohnhofen *et al.*, 2009). The internal consistency of the ESS was found to be insufficient in a study of geriatric patients ($n = 52$, including $n = 14$ people living with dementia) in a hospital setting (Gronewold *et al.*, 2021).

Epworth Sleepiness Scale – Alternative Version (ESS-ALT). The ESS-ALT is a self-administered modified version of the ESS for people with physical or mental disorders. This measurement can be completed by relatives and was developed in a hospital with geriatric patients. Clinical sleep experts, nurses, researchers and patients were included in the development process, and literature was used as well. In total, the scale has eight items with no subdomains, and five of the eight items were adapted from the ESS (Appendix A4). The feasibility, reliability and validity of this measurement were tested in $n = 52$ participants (including $n = 14$ people living with dementia) in the geriatric department of a hospital (Gronewold *et al.*, 2021). Patients or relatives needed no support to answer the items. The internal consistency of the ESS-ALT was judged as insufficient.

Pittsburgh Sleep Quality Index (PSQI). The PSQI is a self-administered measurement that assesses a wide range of sleep disturbances and was developed for inpatients and outpatients of a psychiatric clinic. It was constructed based on clinical expertise, a literature review and 18 months of field testing. The questionnaire contains 19 items subdivided into seven subdomains (Buysse *et al.*, 1989). One study (Curcio *et al.*, 2013) was conducted to assess the criterion validity of the PSQI, which was found to be indeterminate based on the criteria for good measurement properties. The reliability was also rated as indeterminate.

Observational Sleep Assessment Instrument (OSAI). The OSAI is a proxy-administered measurement for sleep disturbances (Cohen-Mansfield *et al.*, 1989). It was developed specifically for residents in the nursing home setting. Sources for the development process were not reported. The exact proportion of people living with dementia in the sample was not reported. The OSAI included 17 items across two subdomains: sleep and sleep patterns. The interrater reliability of the OSAI was unclear based on the QAREL. The criterion validity of this measurement was rated as sufficient.

Sleep Continuity Scale in Alzheimer's Disease (SCADS). The SCADS (Manni *et al.*, 2013) is a proxy-administered measurement that is completed by relatives or caregivers who usually share the bedroom with the person whose sleep will be assessed. It consists of selected items from the "Questionnaire for Hallucinations and Sleep-Wake Cycle in Alzheimer's Disease" (Sinforiani *et al.*, 2007), which was developed for outpatients living with dementia in a hospital. This original measurement consists of 46 items. The measurement aims to investigate the relationship between hallucinations and the sleep-wake cycle of people living with Alzheimer's disease. The theoretical assumption is that a relation exists between the physiopathogenesis and sleep disturbances when visual hallucinations are present in neurodegenerative diseases. Descriptions of the involved professionals, patients or other sources were not reported (Sinforiani *et al.*, 2007). The exact proportion of people living with dementia in the sample was also not reported. The SCADS (Manni *et al.*, 2013) includes nine items. The validity was rated as insufficient and reliability as indeterminate. In another investigation, the authors reported that in terms of feasibility, the SCADS is a feasible measurement that is rapid, easy to complete and suitable for people living with dementia (Manni *et al.*, 2015). However, as a limitation within this publication, it was mentioned that the SCADS cannot be administered to people living with severe dementia or people living with dementia with any relative or caregiver.

Sleep Disorders Inventory (SDI). The Sleep Disorders Inventory is a proxy-rated measurement (Tractenberg *et al.*, 2003). It was developed based on the Neuropsychiatric Inventory (NPI) (Cummings *et al.*, 1994) in research centers for Alzheimer's disease with people living with Alzheimer's dementia. The SDI aims to assess and quantify sleep disturbances and sleep disorders in people living with Alzheimer's disease. Sleep disturbances were defined as less than 6 hours of total sleep time at night. The NPI (Cummings *et al.*, 1994) was used as a source to develop the SDI, but no further information was reported. The measurement consists of eight items that are not divided into subdomains (Tractenberg *et al.*, 2003). The psychometric properties of the SDI were judged as insufficient. In another study (Hjetland *et al.*, 2020), the validity of the SDI was rated sufficient, and its reliability was rated as indeterminate.

Technological devices. In total, four studies (Ancoli-Israel *et al.*, 1997; Gibson and Gander,

2019; Nijhof *et al.*, 2012; Van Someren, 2007) investigated the use of technological devices for the measurement of sleep disturbances. Actigraphs were all worn on the wrist. Ancoli-Israel and colleagues (Ancoli-Israel *et al.*, 1997) compared actigraphy and polysomnography, and the interrater reliability was 0.87. The criterion validity parameters with observations as a comparison measure were 87% (sensitivity) and 90% (specificity). Gibson *et al.*, (Gibson and Gander, 2019) compared the interrater reliability of actigraphy vs. diaries to detect sleep epochs and found an overall agreement of 82% for sleep epochs and an overall agreement of 67% for wake epochs. In the same study, a sensitivity of 87% for sleep and 77% for wake and a specificity of 80% for sleep and 61% for wake were obtained. Another study examined the feasibility of wrist-worn measurements (Nijhof *et al.*, 2012). The most advantages are that nurses in a nursing home were able to see the data in the monitoring system to determine whether the sleep behavior of participants was good or not; furthermore, the computer system that generated the data was easy to understand, and the data were easy to interpret. A high amount of time to implement the system and user unfriendliness were reported as disadvantages (Nijhof *et al.*, 2012). In the study of Van Someren *et al.*, (Van Someren, 2007), the number of nights needed for a reliable measure of sleep disturbances in the use of actigraphy was investigated. The authors recommended more than 7 days of recording for an acceptable reliability of interdaily stability.

Synthesis of results of sleep-related measurements

Feasibility was assessed for the ESS (Frohnhofen *et al.*, 2009; Gronewold *et al.*, 2021), ESS-ALT (Gronewold *et al.*, 2021), SCADS (Manni *et al.*, 2015) and wrist monitoring (Nijhof *et al.*, 2012). The completion rate was the most commonly reported criterion for feasibility among self- and proxy-administered measurements (Frohnhofen *et al.*, 2009; Gronewold *et al.*, 2021; Manni *et al.*, 2015). Reasons for unsuccessful ratings or missing values included inadequate items related to the life of the participants (Gronewold *et al.*, 2021), health status and age (Frohnhofen *et al.*, 2009) and cognitive status or missing relatives for ratings (Manni *et al.*, 2015). According to Bowen *et al.* (2009), only three of eight domains were evaluated across all included feasibility studies.

Three aspects of **reliability** were assessed: internal consistency, interrater reliability and test-retest reliability. In total, four studies (Curcio *et al.*, 2013; Gronewold *et al.*, 2021; Hjetland *et al.*, 2020; Manni *et al.*, 2013) investigated internal consistency. The Cronbach alpha coefficients for the ESS and ESS-ALT were insufficient (Gronewold *et al.*, 2021). In other studies that investigated the internal consistency of the PSQI, SDI and SCADS, the results were found to be indeterminate. In these studies, small sample sizes (Curcio *et al.*, 2013; Hjetland *et al.*, 2020), the lack of sample size calculations (Curcio *et al.*, 2013; Hjetland *et al.*, 2020; Manni *et al.*, 2013) and the lack of information about confidence intervals (Curcio *et al.*, 2013; Manni *et al.*, 2013) were reasons for concerns.

Interrater reliability and test-retest reliability were only investigated for the OSAI (Cohen-Mansfield *et al.*, 1989). The interrater reliability was found to be indeterminate due to a small sample size and the lack of information about confidence intervals.

Two criteria of **validity** were assessed. Structural validity was only assessed for the SCADS (Manni *et al.*, 2013). The risk of bias was adequate, and measurement properties were considered insufficient. Criterion validity was assessed for the PSQI, OSAI and SDI (Cohen-Mansfield *et al.*, 1989; Curcio *et al.*, 2013; Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003). The criterion validity was found to be indeterminate for the PSQI because the sample included for the analysis of criterion validity was unclear (Curcio *et al.*, 2013). The criterion validity was sufficient for the OSAI (Cohen-Mansfield *et al.*, 1989). In one study, the criterion validity of the SDI was insufficient (Tractenberg *et al.*, 2003) due to missing sensitivity, specificity and AUC values. In another study, the criterion validity of the SDI was found to be sufficient (Hjetland *et al.*, 2020).

Discussion

This systematic review included 15 studies. The measurements are heterogeneous in terms of operationalization (items, subscales, scoring), rater perspective and psychometric properties. The theoretical frameworks or definitions of sleep disturbances in the measurements are also heterogeneous.

Identified sleep-related measurements and theoretical backgrounds

Three self-administered measurements, three proxy-administered measurements and two technological devices were identified. Among the papers that described measurements specifically designed for people living with dementia (Sinforiani *et al.*, 2007; Tractenberg *et al.*, 2003), none of them provided definitions or theoretical backgrounds for sleep disturbances, which could be derived from

specific national guidelines such as the NICE guideline for people living with dementia (National Institute for Health and Care Excellence, 2018), the AWMF guideline for dementia (Deuschl and Maier, 2016) or other similar guidelines that existed when the measurements were developed. Additionally, some of the studies specifically examined nursing home residents (Cohen-Mansfield *et al.*, 1989), adults (Johns, 1991), geriatric patients (Gronewold *et al.*, 2021) and patients with mental disorders (Buysse *et al.*, 1989); these studies did not use any important diagnostic criteria for sleep disturbances, such as the ICSD-3-TR (American Academy of Sleep Medicine, 2023), DSM-V-TR (American Psychiatric Association, 2022) or ICD-11 (World Health Organization, 2022) (or earlier versions), as sources when developing measurements. Using literature as a source during measurement construction was mentioned in three out of six development studies (Buysse *et al.*, 1989; Gronewold *et al.*, 2021; Johns, 1991). In two cases, it remains unclear which type of literature was used (Buysse *et al.*, 1989; Gronewold *et al.*, 2021). In particular, the specific literature on perspectives on sleep for people living with dementia in nursing homes that presents psychosocial factors and individual aspects in detail is important to consider in future measurement development or adaptation studies (Dörner *et al.*, 2023). Moreover, knowledge of healthcare professionals (Dörner *et al.*, 2023; Nunez *et al.*, 2018; Webster *et al.*, 2022; Webster *et al.*, 2020b) and family carers (Nunez *et al.*, 2018) could be an important source. Simultaneously, only one development study (Gronewold *et al.*, 2021) involved people living with dementia as stakeholders in the process. For research on complex interventions, the UK Medical Research Council requires the involvement of stakeholder engagement depending on context and phase of research and underlines the crucial importance for the selection of outcome measurements or evidence of change (Skivington *et al.*, 2021a, 2021b).

For actigraphy, as a technological device, it is recommended to always describe the algorithm that is employed, the output procedure and scoring within the study (Camargos *et al.*, 2013). None of the included studies reported all of these data. Only one study transparently reported the algorithm (Ancoli-Israel *et al.*, 1997). Another study of wrist monitoring reported the developed algorithm not sufficient (Nijhof *et al.*, 2012), and two studies used manufacturer algorithms that were not described in detail (Gibson and Gander, 2019; Van Someren, 2007). Among the included studies, the algorithm for actigraphy described by Cole *et al.* (1992) can be recommended for future studies because, in contrast to the other identified algorithms, their algorithm is openly accessible and transparently reported. Furthermore, this algorithm has already been applied in several other actigraphy studies (Biegański *et al.*, 2021; Gao *et al.*, 2022; Hanowski *et al.*, 2007; Kikuchi *et al.*, 2011; Kim *et al.*, 2013; Quante *et al.*, 2018). The reported recording time within the included studies varied between one and 24 weeks. The recommendation for at least one week (Camargos *et al.*, 2013) recording time was reached in all studies that reported the recording time.

In relation to sleep parameters, the following variables should be recorded in actigraphy studies: night sleep time, number of nighttime awakenings, wake after sleep onset and sleep efficacy (Camargos *et al.*, 2013). None of the included studies assessed all of these parameters. The total sleep time was most often reported (Ancoli-Israel *et al.*, 1997; Nijhof *et al.*, 2012; Van Someren, 2007). Wake after sleep onset was reported in two studies (Ancoli-Israel *et al.*, 1997; Gibson and Gander, 2019), and sleep efficacy was reported in one study (Van Someren, 2007). The number of nighttime awakenings was not reported in any study. Furthermore, no study used validated questionnaires or scales to assess subjective sleep complaints, which is recommended as a comparison measure when actigraphy is used to measure sleep (Camargos *et al.*, 2013) or even described as a primary measure for diagnostic criteria other than the use of technological devices (American Psychiatric Association, 2022).

In summary, the heterogeneous and mostly insufficient reported theoretical basis indicates that concept clarification and strict reference to the evidence are vital for the further development of the measurements.

Psychometric properties of the included measurements

Related to **feasibility**, the completion rate according to the domain *practicality of Bowen et al. (2009)* was most often reported within the included studies (Frohnhofen *et al.*, 2009; Gronewold *et al.*, 2021; Manni *et al.*, 2015). Missing values or at least needed help for rating were commonly reported aspects in studies that measure health outcomes in people living with dementia (Jansen *et al.*, 2008; Khobragade *et al.*, 2022; Novella *et al.*, 2001a; Novella *et al.*, 2001b), and these phenomena should be considered when planning future studies with measurements. Compared to recommendations from the literature regarding feasibility studies (Bowen *et al.*, 2009), in summary, the self-reported ESS-ALT (Gronewold *et al.*, 2021) should be preferred because in terms of feasibility, it has a better completion rate (domain of *practicality*) and

seems to be better in the domain of *acceptability* than the ESS (Frohnhofen *et al.*, 2009; Gronewold *et al.*, 2021). Furthermore, after the adaptation process, the ESS-ALT (Gronewold *et al.*, 2021) seemed to have a better fit for people living with dementia and people in the nursing home setting, but it still needs to be further adapted to the nursing home setting. If the ESS-ALT is used in the nursing home setting, an implication for subsequent feasibility studies would be the adaptation to the setting (*domain of adaptation*). This approach goes along with the examination of whether the content of the items is completely appropriate. For example, item 4 *("as a passenger in the car")* (Gronewold *et al.*, 2021) may be inappropriate because this activity could no longer occur regularly for the majority of nursing home residents. Important in this context are among others, the user's satisfaction with the measurement, the successful application (depending on the success of the adaptation) and the costs (Bowen *et al.*, 2009).

The SCADS was the only proxy-administered measurement that was tested in terms of feasibility. The authors described two limitations in their study (Manni *et al.*, 2015) that are important to consider when assessing feasibility (Bowen *et al.*, 2009): the administration of this measurement is not possible among people with severe dementia, and bed-partners/caregivers were not always available for the rating. Therefore, this measurement should be adapted. In addition to the general adaptation to the setting *(domain of adaptation)* according to Bowen *et al.* (2009) one of the main implications is the adaptation to dementia severity. A key requirement regarding measurements is always to clarify for which concern and by whom they are used (Sheehan, 2012). This implies that a measurement that was developed or adapted for people living with dementia should cover all degrees of dementia severity or that it should be explicitly developed, for example, for the early stages of dementia. Moreover, the rating should not be determined only from the proxy-rating perspective. In general, it may be preferable to assess sleep disturbances by the person concerned. If this is not possible, e.g. due to dementia severity, a combined rating of the person concerned and relatives, caregivers or other healthcare providers could be a suitable solution. Therefore, an adaptation option would be to adjust the SCADS rating for self- and proxy ratings.

One study (Nijhof *et al.*, 2012) assessed the feasibility of wrist monitoring. First, the costs of the measurement were reported for the technological installation and for each watch. Second, the administration seemed to be difficult because it took more time than expected to implement the measurement for caregivers, and there was a skepticism in the beginning regarding whether actigraphy would help the caregiver with their tasks. Third, regarding the use and usability of wrist monitors for daily monitoring, caregivers printed out nightly results and put information in the healthcare records, but this topic was not discussed in team meetings. Regarding usability, it was reported that actigraphy is not user-friendly and needs to be improved (e.g., it is too large for small arms, the hard strap irritates the skin, a normal clockface would be more familiar for elderly individuals). Caregivers were able to monitor data regarding a resident's sleep quality, and the data were easy to interpret and understand (Nijhof *et al.*, 2012).The aspect of successful monitoring of sleep data with sleep monitoring devices such as actigraphy was mentioned in another study among elderly people (LeBlanc *et al.*, 2022). However, it was also stated in the same publication that actigraphy has been used less over time. Trust regarding the performance of the wrist monitor was reported in this study as supportive, ambivalent or negative. With this in mind, the wrist monitoring system has several limitations and is currently not a feasible opportunity to assess sleep disturbances.

In Summary, only three of the eight domains according to Bowen *et al.* (2009) were reported within the included feasibility studies. For future feasibility studies, this implies that all feasibility criteria should be evaluated.

Reliability was assessed for internal consistency, interrater reliability and test-retest reliability. None of the included measurements that were tested for reliability (Cohen-Mansfield *et al.*, 1989; Curcio *et al.*, 2013; Gronewold *et al.*, 2021; Hjetland *et al.*, 2020; Manni *et al.*, 2013) showed sufficient psychometric results for several domains of reliability.

A clear recommendation in terms of measurements with the best reliability for people living with dementia in the nursing home setting is therefore not possible for the proxy-administered measurements analyzed herein due to the limitations of each study. For self-administered measurements, the PSQI (Buysse *et al.*, 1989) seems to be potentially eligible. It is the most frequently used self-measurement in different settings (Fabbri *et al.*, 2021) but has various factor structures and needs to be adapted (Manzar *et al.*, 2018). For the population of people with dementia in particular, first, feasibility should be adapted with different guidelines (Beaton *et al.*, 2000; Bowen *et al.*, 2009) and psychometrically analyzed again.

Regarding validity, the structural validity for one measurement (Manni *et al.*, 2013) and criterion validity for three measurements (Cohen-Mansfield *et al.*, 1989; Curcio *et al.*, 2013; Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003) were assessed. Actigraphy

was used as a reference standard for the SDI in two studies (Hjetland *et al.*, 2020; Tractenberg *et al.*, 2003) and for sleep variables compared to the OSAI (Cohen-Mansfield *et al.*, 1989). The DSM-V-TR (American Psychiatric Association, 2022) states that for insomnia disorder diagnosis, the individual's subjective perception or a caregiver report is needed. Additionally, symptoms can be quantified by sleep diaries or actigraphy (American Psychiatric Association, 2022). Therefore, further research should be conducted to determine whether actigraphy is the best measurement approach to use as a reference standard in the included studies. Moreover, quantitative criteria are often used in research designs but cannot reliably distinguish between individuals with insomnia and normal sleepers. Therefore, it is recommended that quantitative guidelines for measuring the frequency and duration of sleep should only be used for illustrative purposes (American Psychiatric Association, 2022). Because of the insufficient results, the risk of bias analysis, the recommendations regarding actigraphy (Camargos *et al.*, 2013) and the statements from the ICSD-3-TR and DSM-V-TR, no recommendation can be provided due to the lack of validity of the measurements examined herein.

Recommendations for clinical practice

Sleep disturbances including disturbed sleep at night and daytime sleepiness (American Academy of Sleep Medicine, 2023). An important aspect when choosing a measurement for clinical practice is the consideration of both issues. Regarding the combination of different perspectives in terms of rating and day- and night of the included measurements, only the PSQI (Buysse *et al.*, 1989) as a self-measurement and the SDI (Tractenberg *et al.*, 2003) as a proxy-measurement are potentially eligible for these concerns. Both measurements could be used in combination with careful reflection of the results because of the lack of psychometric testing. The PSQI (Buysse *et al.*, 1989) could be challenging for people living with dementia because of its length and complexity. A negative aspect of the SDI (Tractenberg *et al.*, 2003) could be, that only one item for daytime sleepiness exists.

Strengths and limitations

A strength of this publication is that this is the first review that investigated the psychometric properties of self- and proxy-administered sleep-related measurements to assess sleep disturbances among people living with dementia. Although the aim of this review was to identify and recommend measurements for the nursing home setting, measurements for all settings were included to detect a higher number of measurements that can be potentially adapted for people living with dementia in nursing homes. One limitation of this study is that no protocol for this review was externally registered. Second, the samples of the included studies did not exclusively comprise people living with dementia. However, it was important to analyze all studies of measurements to measure sleep among people living with dementia to obtain information about potentially usable measurements. Third, the actigraphy algorithms were not compared, because three out of four algorithms were provided by manufacturers of technological devices without transparent reporting. Thus, it was not possible to compare the different approaches in more detail.

Conclusion

This systematic review identified eight measurements that have undergone psychometric analysis. The theoretical definitions of sleep disturbances were often poorly described within included measurements. Therefore, it is difficult to determine whether the construct of sleep was comprehensively considered in the development process with respect to content validity and the specific aim of the measurement. Moreover, none of the measurements were evaluated across all psychometric properties. Furthermore, the large number of measurements with insufficient or unclear reliability and validity shows that further research is needed to accurately assess sleep disturbances among people living with dementia. Currently, none of the measurements identified here can be recommended for use without further development in intervention studies.

Criteria-based decision making (e.g., the COSMIN methodology) is necessary for the selection of the optimal measurement. The identified technological measurements can be used to obtain secondary outcomes but not for primary outcomes. Previous studies used technological measurements to obtain primary outcomes, but this practice contradicts the recommendations of international diagnostic criteria. Future actigraphy studies should use open access algorithms to increase transparency. Researchers should quantify sleep for illustrative purposes and not as a primary outcome for detecting sleep disturbances or as a reference standard in diagnostic accuracy studies. This review indicates that no currently available sleep-related measurement can be recommended without strong reservations for assessing sleep disturbances among people living with dementia in nursing homes. However, a combination of self- and proxy assessments seems to be the best option to achieve valid measurements of sleep disturbances among people living with dementia.

Abbreviations

COSMIN: COnsensus-based Standards for the Selection of Health Measurement INstruments; DSM-V-TR: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision; ESS: Epworth Sleepiness Scale; ESS-ALT: Epworth Sleepiness Scale Alternative Version; ICD-11: International Classification of Diseases 11th Revision; ICSD-3-TR: International Classification of Sleep Disorders – Third Edition, Text Revision; NPI: Neuropsychiatric Inventory; PSQI: Pittsburgh Sleep Quality Index; QAREL: Quality Appraisal Tool for Studies of Diagnostic Reliability; SCADS: Sleep Continuity Scale in Alzheimer's Disease; SDI: Sleep Disorders Inventory.

Conflict of interests

None.

Funding

This review was undertaken at the DZNE, which receives basic funding from the Federal Ministry of Education and Research and the state of North Rhine-Westfalia.

Description of the authors' roles

Study Design: JD, MND, MH
Literature Search: JD, MND
Data Analysis: JD, KW, MND
First Draft of the Manuscript: JD, KW, MH, MND
Manuscript Preparation: JD, KW, MH, MND

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S104161022400070X.

References

Alessi, C. A., Martin, J. L., Webber, A. P., Cynthia Kim, E., Harker, J. O., & Josephson, K. R. (2005). Randomized, controlled trial of a nonpharmacological intervention to improve abnormal sleep/wake patterns in nursing home residents. *Journal of The American Geriatrics Society*, 53(5), 803–810.

Alessi, C. A., Yoon, E. J., Schnelle, J. F., Al-Samarrai, N. R., & Cruise, P. A. (1999). A randomized trial of a combined physical activity and environmental intervention in nursing home residents: do sleep and agitation improve? *Journal of The American Geriatrics Society*, 47(7), 784–791.

American Academy of Sleep Medicine. (2023). *International classification of sleep disorders* (3rd ed., text rev.).

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.), American Psychiatric Publishing.

Ancoli-Israel, S., Clopton, P., Klauber, M. R., Fell, R., & Mason, W. (1997). Use of wrist activity for monitoring sleep/wake in demented nursing-home patients. *Sleep*, 20(1), 24–27.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*, 25(24), 3186–3191.

Biegański, P., Stróż, A., Dovgialo, M., Duszyk-Bogorodzka, A., & Durka, P. (2021). On the unification of common actigraphic data scoring algorithms. *Sensors (Basel)*, 21(18), 6313.

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How we design feasibility studies. *American Journal of Preventive Medicine*, 36(5), 452–457.

Buysse, D. J., Reynolds, C. F., 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(2), 193–213.

Camargos, E. F., Louzada, F. M., & Nóbrega, O. T. (2013). Wrist actigraphy for measuring sleep in intervention studies with Alzheimer's disease patients: application, usefulness, and challenges. *Sleep Medicine Reviews*, 17(6), 475–488.

Camargos, E. F., Louzada, L. L., Quintas, J. L., Naves, J. O., Louzada, F. M., & Nóbrega, O. T. (2014). Trazodone improves sleep parameters in Alzheimer disease patients: a randomized, double-blind, and placebo-controlled study. *The American Journal of Geriatric Psychiatry*, 22(12), 1565–1574.

Cohen-Mansfield, J., Werner, P., & Marx, M. S. (1989). An observational study of agitation in agitated nursing home residents. *International Psychogeriatrics*, 1(2), 153–165.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5), 461–469.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. *Neurology*, 44(12), 2308–2314.

Curcio, G., Tempesta, D., Scarlata, S., Marzano, C., Moroni, F., Rossini, P. M., Ferrara, M., & De Gennaro, L. (2013). Validity of the Italian version of the Pittsburgh Sleep Quality Index (PSQI). *Neurological Sciences*, 34(4), 511–519.

Deuschl, G., & Maier, W. (2016). *S3-Leitlinie Demenzen. Deutsche Gesellschaft für Psychiatrie und Psychotherapie.* Psychosomatik und Nervenheilkunde (DGPPN) Deutsche Gesellschaft für Neurologie (DGN).

Dörner, J., Hüsken, J. M., Schmüdderich, K., Dinand, C., Dichter, M. N., & Halek, M. (2023).

Perspectives on sleep of people living with dementia in nursing homes: a qualitative interview study. *BMC Geriatrics*, 23(1), 331.

Dowling, G. A., Burr, R. L., Van Someren, E. J. W., Hubbard, E. M., Luxenberg, J. S., Mastick, J., & Cooper, B. A. (2008). Melatonin and bright-light treatment for rest-activity disruption in institutionalized patients with Alzheimer's disease. *Journal of The American Geriatrics Society*, 56(2), 239–246.

Fabbri, M., Beracci, A., Martoni, M., Meneo, D., Tonetti, L., & Natale, V. (2021). Measuring subjective sleep quality: a review. *International Journal of Environmental Research and Public Health*, 18(3), 1082.

Foley, D. J., Monjan, A. A., Brown, S. L., Simonsick, E. M., Wallace, R. B., & Blazer, D. G. (1995). Sleep complaints among elderly persons: an epidemiologic study of three communities. *Sleep*, 18(6), 425–432.

Frohnhofen, H., Popp, R., Willmann, V., Heuer, H. C., & Firat, A. (2009). Feasibility of the Epworth sleepiness scale in a sample of geriatric in-hospital patients. *Journal of Physiology and Pharmacology*, 60(Suppl 5), 45–49.

Fung, C. H., Vitiello, M. V., Alessi, C. A., Kuchel, G. A., Committee, A. N. S. C. P., & Faculty (2016). Report and Research Agenda of the American Geriatrics Society and National Institute on Aging Bedside-to-Bench Conference on Sleep, Circadian Rhythms, and Aging: New Avenues for Improving Brain Health, Physical Health, and Functioning. *Journal of the American Geriatrics Society*, 64, e238–e247.

Gao, C., Li, P., Morris, C. J., Zheng, X., Ulsa, M. C., Gao, L., Scheer, F. A. J. L., & Hu, K. (2022). Actigraphy-based sleep detection: validation with polysomnography and comparison of performance for nighttime and daytime sleep during simulated shift work. *Nature and Science of Sleep*, 14, 1801–1816.

Garcia-Alberca, J. M., Lara, J. P., Cruz, B., Garrido, V., Gris, E., & Barbancho, M. A. (2013). Sleep disturbances in Alzheimer's disease are associated with neuropsychiatric symptoms and antidementia treatment. *Journal of Nervous and Mental Disease*, 201(3), 251–257.

Gattinger, H., Hantikainen, V., Ott, S., & Stark, M. (2017). Effectiveness of a mobility monitoring system included in the nursing care process in order to enhance the sleep quality of nursing home residents with cognitive impairment. *Health and Technology*, 7(2-3), 161–171.

Gehrman, P., Marler, M., Martin, J. L., Shochat, T., Corey-Bloom, J., & Ancoli-Israel, S. (2004). The timing of activity rhythms in patients with dementia is related to survival. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(10), 1050–1055.

Gibson, R. H., & Gander, P. H. (2019). Monitoring the sleep patterns of people with dementia and their family carers in the community. *Australasian Journal on Ageing*, 38(1), 47–51.

Gronewold, J., Lenuck, M., Gülderen, I., Scharf, A.-C., Penzel, T., Johns, M W., Frohnhofen, H., & Hermann, D M. (2021). Developing an alternative version of the Epworth sleepiness scale to assess daytime sleepiness in adults with physical or mental disabilities. *Gerontologia*, 67(1), 49–59.

Hanowski, R. J., Hickman, J., Fumero, M. C., Olson, R. L., & Dingus, T. A. (2007). The sleep of commercial vehicle drivers under the 2003 hours-of-service regulations. *Accident Analysis & Prevention*, 39(6), 1140–1145.

Hjetland, G. J., Nordhus, I. H., Pallesen, S., Cummings, J., Tractenberg, R. E., Thun, E., Kolberg, E., & Flo, E. (2020). An actigraphy-based validation study of the sleep disorder inventory in the nursing home. *Frontiers in Psychiatry*, 11, 173.

Jansen, A. P. D., van Hout, H. P. J., Nijpels, G., van Marwijk, H. W. J., Gundy, C., de Vet, H. C. W., & Stalman, W. A. B. (2008). Self-reports on the IQCODE in older adults: a psychometric evaluation. *Journal of Geriatric Psychiatry and Neurology*, 21(2), 83–92.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*, 14(6), 540–545.

Khobragade, P., Nichols, E., Meijer, E., Varghese, M., Banerjee, J., Dey, A.B., Lee, J., Gross, A.L., & Ganguli, M. (2022). Performance of the Informant Questionnaire on Cognitive Decline for the Elderly (IQCODE) in a nationally representative study in India: the LASI-DAD study. *International Psychogeriatrics*, 36, 187–177.

Kikuchi, H., Yoshiuchi, K., Yamamoto, Y., Komaki, G., & Akabayashi, A. (2011). Does sleep aggravate tension-type headache?: An investigation using computerized ecological momentary assessment and actigraphy. *BioPsychoSocial Medicine*, 5(1), 10.

Kim, M. J., Lee, G.-H., Kim, C.-S., Kim, W. S., Chung, Y.-S., Chung, S., & Lee, S.-A. (2013). Comparison of three actigraphic algorithms used to evaluate sleep in patients with obstructive sleep apnea. *Sleep and Breathing*, 17(1), 297–304.

Kuck, J., Pantke, M., & Flick, U. (2014). Effects of social activation and physical mobilization on sleep in nursing home residents. *Geriatric Nursing*, 35(6), 455–461.

LeBlanc, R. G., Czarnecki, P., Howard, J., Jacelon, C. S., & Marquard, J. (2022). Usability experience of a personal sleep monitoring device to self-manage sleep among persons 65 years or older with self-reported sleep disturbances. *CIN: Computers, Informatics, Nursing*, 40(9), 598–605.

Li, J., Grandner, M. A., Chang, Y. P., Jungquist, C., & Porock, D. (2017). Person-centered dementia care and sleep in assisted living residents with dementia: a pilot study. *Behavioral Sleep Medicine*, 15(2), 97–113.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., Orgeta, V., Ritchie, K., Rockwood, K., Sampson, E. L. , Samus, Q., Schneider, L. S., Selbæk, G., Teri, L., & Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*, 396(10248), 413–446.

Lucas, N. P., Macaskill, P., Irwig, L., & Bogduk, N. (2010). The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of Clinical Epidemiology*, 63(8), 854–861.

Manni, R., Sinforiani, E., Terzaghi, M., Rezzani, C., & Zucchella, C. (2015). Sleep continuity scale in Alzheimer's

disease (SCADS): application in daily clinical practice in an Italian center for dementia. *Neurological Sciences*, 36(3), 469–471.

Manni, R., Sinforiani, E., Zucchella, C., Terzaghi, M., & Rezzani, C. (2013). A sleep continuity scale in Alzheimer's disease: validation and relationship with cognitive and functional deterioration. *Neurological Sciences*, 34(5), 701–705.

Manzar, M. D., BaHammam, A. S., Hameed, U. A., Spence, D. W., Pandi-Perumal, S. R., Moscovitch, A., & Streiner, D. L. (2018). Dimensionality of the Pittsburgh Sleep Quality Index: a systematic review. *Health and Quality of Life Outcomes*, 16(1), 89.

McCleery, J., Cohen, D. A., & Sharpley, A. L. (2014). Pharmacotherapies for sleep disturbances in Alzheimer's disease. *Cochrane Database of Systematic Reviews*, CD009178, 1–45.

Moe, K. E., Vitiello, M. V., Larsen, L. H., & Prinz, P. N. (1995). Symposium: cognitive processes and sleep disturbances: sleep/wake patterns in Alzheimer's disease: relationships with cognition and function. *Journal of Sleep Research*, 4(1), 15–20.

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, 27(5), 1171–1179.

Most, E. I., Aboudan, S., Scheltens, P., & Van Someren, E. J. (2012). Discrepancy between subjective and objective sleep disturbances in early- and moderate-stage Alzheimer disease. *Am J Geriatr Psychiatry*, 20, 460–467.

National Institute for Health and Care Excellence. 'Dementia: assessment, managment and support for people living with dementia and their carers (2018).

NCT00325728. A Double-Blind, randomized, placebo-controlled study of the efficacy, safety and tolerability of 8 week treatment of Rozerem 8 mg (QHS) in sleep disturbed, mild to moderately severe Alzheimer's disease subjects (2008)

Nijhof, N., van Gemert-Pijnen, J. E. W. C., de Jong, G. E. N., Ankoné, J. W., & Seydel, E. R. (2012). How assistive technology can support dementia care: a study about the effects of the IST Vivago watch on patients' sleeping behavior and the care delivery process in a nursing home. *Technology & Disability*, 24(2), 103–115.

Novella, J. L., Ankri, J., Morrone, I., Guillemin, F., Jolly, D., Jochum, C., Ploton, L., & Blanchard, F. (2001a). Evaluation of the quality of life in dementia with a generic quality of life questionnaire: the Duke Health Profile. *Dementia and Geriatric Cognitive Disorders*, 12(2), 158–166.

Novella, J. L., Jochum, C., Ankri, J., Morrone, I., Jolly, D., & Blanchard, F. (2001b). Measuring general health status in dementia: practical and methodological issues in using the SF-36. *Aging (Milano)*, 13(5), 362–369.

Nunez, K. M., Khan, Z., Testad, I., Lawrence, V., Creese, B., & Corbett, A. (2018). Current practice and challenges in night-time care for people with dementia living in care homes: a qualitative study. *International Journal of Geriatric Psychiatry*, 33(1), e140–e149.

Page, M. J., Boutron, I., Hoffmann, TC, Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M, Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA. 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71.

Patterson, C. (2018). *World Alzheimer Report 2018*. A. s. D. International.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B., (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157.

Quante, M., Kaplan, E.R., Cailler, M., Rueschman, M., Wang, R., Weng, J., Taveras, E.M., & Redline, S, (2018). Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms. *Nature and Science of Sleep*, 10, 13–20.

Richards, K. C., Lambert, C., Beck, C. K., Bliwise, D. L., Evans, W. J., Kalra, G. K., Kleban, M. H., Lorenz, R., Rose, K., Gooneratne, N. S., & Sullivan, D. H. (2011). Strength training, walking, and social activity improve sleep in nursing home and assisted living residents: randomized controlled trial. *Journal of The American Geriatrics Society*, 59(2), 214–223.

Schnelle, J. F., Alessi, C. A., Al-Samarrai, N. R., Fricker, R. D., Jr., & Ouslander, J. G. (1999). The nursing home at night: effects of an intervention on noise, light, and sleep. *Journal of The American Geriatrics Society*, 47(4), 430–438.

Serfaty, M., Kennell-Webb, S., Warner, J., Blizard, R., & Raven, P. (2002). Double blind randomised placebo controlled trial of low dose melatonin for sleep disorders in dementia. *International Journal of Geriatric Psychiatry*, 17(12), 1120–1127.

Sheehan, B. (2012). Assessment scales in dementia. *Therapeutic Advances in Neurological Disorders*, 5(6), 349–358.

Sinforiani, E., Terzaghi, M., Pasotti, C., Zucchella, C., Zambrelli, E., & Manni, R. (2007). Hallucinations and sleep-wake cycle in Alzheimer's disease: a questionnaire-based study in 218 patients. *Neurological Sciences*, 28(2), 96–99.

Singer, C., Tractenberg, R. E., Kaye, J., Schafer, K., Gamst, A., Grundman, M., Thomas, R., & Thal, L. J. (2003). A multicenter, placebo-controlled trial of melatonin for sleep disturbance in Alzheimer's disease. *Sleep*, 26(7), 893–901.

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021a). Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. *Health Technology Assessment*, 25(57), 1–132.

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N.,

French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021b). A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*, 374, n2061.

Straus, S. E., Glasziou, P., Scott Richardson, W., & Haynes, R. B. (2018). *Evidence-Based Medicine*, Elsevier.

Tractenberg, R. E., Singer, C. M., Cummings, J. L., & Thal, L. J. (2003). The Sleep Disorders Inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. *Journal of Sleep Research*, 12(4), 331–337.

Van Someren, E. J. W. (2007). Improving actigraphic sleep estimates in insomnia and dementia: how many nights? *Journal of Sleep Research*, 16(3), 269–275.

Webster, L., Costafreda Gonzalez, S., Stringer, A., Lineham, A., Budgett, J., Kyle, S., Barber, J., & Livingston, G. (2020a). Measuring the prevalence of sleep disturbances in people with dementia living in care homes: a systematic review and meta-analysis. *Sleep*, 43(4), 1–14.

Webster, L., Costafreda, S. G., Powell, K., & Livingston, G. (2022). How do care home staff use non-pharmacological strategies to manage sleep disturbances in residents with dementia: the SIESTA qualitative study. *PLoS One*, 17(8), e0272814.

Webster, L., Powell, K., Costafreda, S. G., & Livingston, G. (2020b). The impact of sleep disturbances on care home residents with dementia: the SIESTA qualitative study. *International Psychogeriatrics*, 32(7), 839–847.

Wilfling, D., Dichter, M. N., Trutschel, D., & Kopke, S. (2019). Prevalence of sleep disturbances in german nursing home residents with dementia: a multicenter cross-sectional study. *Journal of Alzheimers Disease*, 69(1), 227–236.

Wilfling, D., Dichter, M. N., Trutschel, D., & Köpke, S. (2020a). Nurses' burden caused by sleep disturbances of nursing home residents with dementia: multicenter cross-sectional study. *BMC Nursing*, 19(1), 83.

Wilfling, D., Hylla, J., Berg, A., Meyer, G., Köpke, S., Halek, M., Möhler, R., & Dichter, M.N. (2020b). Characteristics of multicomponent, nonpharmacological interventions to reduce or avoid sleep disturbances in nursing home residents: a systematic review. *International Psychogeriatrics*, 33, 245–273.

World Health Organization. (2022). *ICD-11: International Classification of Diseases (11th Revision)*.