Sleepiness in adults: An umbrella review of a complex construct
Vincent P Martin, Régis Lopez, Yves Dauvilliers, Jean-Luc Rouas, Pierre Philip, Jean-Arthur Micoulaud-Franchi

To cite this version:
Vincent P Martin, Régis Lopez, Yves Dauvilliers, Jean-Luc Rouas, Pierre Philip, et al.. Sleepiness in adults: An umbrella review of a complex construct. Sleep Medicine Reviews, 2022, 10.1016/j.smrv.2022.101718. hal-03864826

HAL Id: hal-03864826
https://hal.science/hal-03864826
Submitted on 22 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Sleepiness in adults:  
An umbrella review of a complex construct

Vincent P. Martin1,*,†, Régis Lopez2,3,*, Yves Dauvilliers2,3, Jean-Luc Rouas1, Pierre Philip4,5, Jean-Arthur Micoulaud-Franchi4,5,†

1 Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France
2 National Reference Center for Orphan Diseases, Narcolepsy, Idiopathic hypersomnia and Kleine-Levin Syndrome, Sleep Disorders Center, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier
3 Institute for Neurosciences of Montpellier (INM), Montpellier University, INSERM, Montpellier, France
4 Service de médecine universitaire du sommeil, CHU de Bordeaux, Place Amélie Raba-Leon, Bordeaux, France
5 Université de Bordeaux, SANPSY UMR 6033, F-33000 Bordeaux, France

* Authors with equal contribution

† Corresponding authors:
Jean-Arthur Micoulaud-Franchi
UMR 6033 SANPSY
CHU Pellegrin, Tripode 13ème étage, aile 3, Place Amélie- Raba-Léon
33076 Bordeaux cédex, France
Email: jarthur.micoulaud@gmail.com
Phone: +33622364019

and

Vincent P. Martin
LaBRI, F-33400 Talence, France
Email: vincent.martin@labri.fr
Phone: +33627343611

Conflict of interest
The author reports no conflict of interest.

Acknowledgements
We thank Christophe Gauld for his inspiring insights into the theoretical aspects of this article.

English Proofreading
The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:
http://www.textcheck.com/certificate/QlMbTh
SUMMARY (200 words)

Sleepiness involves many dimensions that require investigation. Since sleepiness is often defined operationally, we exhaustively inventoried all the assessment tools designed to measure it in an umbrella review, without any preconceptions, i.e. a review of reviews. We included all reviews and systematic reviews related to sleepiness assessment tools published up to March 2021. Three investigators independently assessed the eligibility of studies for inclusion and identified 36 relevant reviews. In total, 99 tools were identified and classified into 8 categories. We classified them depending on their category, their publication year and the number of mentions in the 36 included reviews. The 6 most frequently cited were the Epworth sleepiness scale, the multiple sleep latency test, the maintenance of wakefulness test, the Stanford sleepiness scale, the Karolinska sleepiness scale, and the psychomotor vigilance task. Despite the limitation that we may have missed some recently developed tools, this historical perspective on sleepiness measurement is a first step toward a better delineation of the different dimensions underlying the constructs of sleepiness, and will serve as a basis for further discussion in the clinical and research sleep community.

Keywords
Sleepiness; Hypersomnolence; Disorders of Excessive Somnolence; Hypersomnia; Fatigue; Assessment; Questionnaire; Physiological Monitoring
Abbreviations
BSI: Barcelona sleepiness index
EEG: electroencephalography
ESS: Epworth sleepiness scale
FOSQ: functional outcomes of sleep questionnaire
HI: hypersomnia severity index
IHSS: idiopathic hypersomnia severity scale
KSS: Karolinska sleepiness scale
MSLT: multiple sleep latency test
MWT: maintenance of wakefulness
OSDI: observation and interview-based diurnal sleepiness inventory
OSLER: Oxford sleep resistance test
PRISMA: preferred reporting items for systematic reviews and meta-analyses
PSG: polysomnography
PSQI: Pittsburgh sleep quality index
PVT: psychomotor vigilance task
RSS: resistance to sleepiness scale
SART: sustained attention to response task
SIQ: sleep inertia questionnaire
SSS: Stanford sleepiness scale
SWAI: sleep-wake activity inventory
THAT: Toronto hospital alertness test
INTRODUCTION

Sleepiness is a normal physiological state experienced by most individuals over a 24-h period (1,2). However, sleepiness occurring at inappropriate times, with increased frequency or interfering with daily functioning, is usually considered excessive (3). Since excessive sleepiness poses a danger to both personal and public safety, it has attracted increasing scientific, social, and political attention, especially in the context of driving (4). Excessive sleepiness is associated with a wide range of diseases, including sleep, metabolic, cardiovascular, neurological, and psychiatric disorders, which increase the risk of disability and mortality (5,6). In addition, excessive sleepiness is commonly associated with adverse social and economic consequences such as motor vehicle accidents, near misses, decrease in work productivity and quality of life (7–9).

Excessive sleepiness may be due to acute and/or chronic sleep deprivation, use of sedatives, or various underlying disorders, particularly sleep disorders. It is one of the most common complaints of people concerned by their sleep (10), and has received significant attention from sleep experts since the pioneering studies on sleep deprivation (11,12). Based on the presence or absence of excessive sleepiness, the first classification system for sleep disorders distinguished Disorders of Excessive Somnolence and Disorders of Initiating and Maintaining Sleep (13). Several decades after that system was published, excessive sleepiness still plays a central role in sleep-wake nosological classifications (14,15).

No consensus exists regarding the precise definition of sleepiness and the threshold above which it becomes excessive. This accounts, in part, for the marked variability in the estimated prevalence of excessive sleepiness in the general population, which ranges from 2.5% to more than 40% (3,16,17). One of the most widely accepted definitions of sleepiness is sleep propensity, i.e. the tendency to fall asleep. Sleep propensity is also referred to as ‘objective sleepiness’, which can be quantified on the basis of sleep latency using various tests based on polysomnography (PSG) that are performed in different experimental situations, e.g. the multiple sleep latency test [MSLT] and maintenance of wakefulness test [MWT] (18). Beyond sleep propensity, other symptomatic dimensions may also relate to excessive sleepiness. They include drowsiness, a continuous state associated with impaired cognitive performances underpinned by a low level of arousal and automatic behaviors, which is considered as an abnormal transitional state between wakefulness and sleep (19).

In recent decades, the study of severe medical conditions associated with excessive sleepiness (i.e. narcolepsy and idiopathic hypersomnia) has led to the emergence of the concept of hypersomnolence (20). Hypersomnolence is usually defined as the association to various degrees of excessive daytime sleepiness, excessive need for sleep and sleep inertia (20). Excessive need for sleep has been recently defined as “the complaint of a need for an excessive quantity of sleep over the full 24 h period […] associated with impairment and distress related primarily to deteriorated quality of daytime wakefulness” (20). Sleep inertia is characterized by difficulty in waking up, reduced vigilance and impaired performance lasting for up to several hours, and is sometimes associated with confusion and the inability to react appropriately to external stimuli on awakening (often called sleep drunkenness) (21,22). Finally, the boundaries between the concept of sleepiness and two other dimensions, fatigue and hypovigilance, remain blurred. Recent reviews on vigilance (19) and fatigue (2) have made it possible to delineate these constructs respectively as “the capability to be sensitive to potential changes in one's
environment” (19), and “a gradual and cumulative process associated with a disinclination towards effort, eventually resulting in reduced performance efficiency” (23,24).

Excessive sleepiness is thus a very complex construct and investigations into its multiple aspects are needed in order to allow clinicians and researchers to better understand them and to optimize therapeutic approaches. Thus, we propose to investigate and refine the notion of sleepiness by inventorying the tools designed to measure it. Contrary to the top-down approach usually proposed in the literature [e.g. (2,19,25,26)] which consists in associating tools with pre-defined constructs (27), we investigated the concept of sleepiness by a bottom-up approach by trying to infer the constructs by studying the assessment tools. To reflect the diversity of the approaches to evaluate sleepiness and to preserve the subtle variations in its measurement, we conducted an exhaustive review of the tools used to measure sleepiness in adults without any preconceptions. We used a broad definition of sleepiness to capture most of the tools. An exhaustive systematic review of the tools for assessing sleepiness would have required a review of the entire literature on the subject, which would be a challenging task. Therefore, we performed an umbrella review, i.e. a review of reviews, following a rigorous and standardized methodology adapted from systematic reviews (28) involving the collection of evidence from all reviews on a given topic (29), reflecting how the sleep community measures sleepiness.

Although the development of a consensual model of sleepiness and related constructs would have been useful, we hypothesize – given the fact that assessment tools are built to evaluate underlying constructs (30) – that this exhaustive inventory will serve as a basis for further discussion of the various aspects of sleepiness and will help to clarify them, with regard to both the clinical and research dimensions of sleep medicine.
METHODS

Search strategy and study selection criteria

The Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) (28) recommendations were followed.

Identification

We searched the PubMed, Web of Science, and Scopus databases to identify reviews (including systematic reviews) and meta-analyses related to sleepiness assessment or measurement published between 1950 and March 30, 2021. The search was performed using combinations of sleepiness-related keywords and measurement-related keywords. Seven keywords were related both to sleepiness (hypersomnolence OR sleepiness OR wakefulness OR drowsiness OR somnolence OR sleepy OR drowsy) and to measurements (measure* OR question* OR scale* OR evaluate* OR define* OR assess* OR diagnostic).

The search was limited to the titles of articles, and only reviews and systematic reviews were searched. In addition, manual searches of personal reference collections yielded two additional citations. The links to the results of database queries are given in the Supplementary Material S1.

Article screening

We manually searched the reference lists of the retrieved articles. Articles were initially screened by three investigators (VM, JAMF, and RL) based on the title, abstract, and journal in which they were published. Only references published in English, French, or Spanish were included. Then, the full-text versions of potentially eligible articles were scrutinized by the same investigators using the following criteria.

We selected reviews/systematic reviews of measures of sleepiness, irrespective of the study population. Editorials, correspondences, abstracts, and review summaries that were not automatically filtered by PubMed, Web of Science, or Scopus were excluded manually. We also excluded reviews that focused on specific diseases associated with sleepiness, i.e. those in which the main focus of the review was not sleepiness. Reviews that focused on the pediatric or adolescent population were also excluded. The same screening process was applied for each article. Disagreements during the inclusion of the reviews were resolved by consensus following cross analysis of the article concerned.

Data extraction and analysis

The sleepiness assessment tools mentioned in the included review articles were manually extracted by the three investigators (VM, JAMF, and RL). Based on the “operational definition” concept for a given term introduced by Hempel (31), we defined sleepiness assessment tools as “objective criteria by means of which any scientific investigator can decide, for any particular case, whether the term [sleepiness] does or does not apply,” i.e. a technical modality that provides a quantifiable measure related to sleepiness. Moreover, for each review article, we extracted its year of publication.

Data processing

Number of sleepiness assessment tools mentioned in the reviews

For each review article, we extracted the number of distinct sleepiness assessment tools included and computed the mean, standard deviation (SD), median, minimum, and maximum
number of tools in the reviews. Finally, we computed the Pearson correlation coefficient between the publication year and number of tools mentioned in the reviews.

**Categorization of sleepiness assessment tools**

There are many ways to classify sleepiness assessment tools. The most common in the literature are the distinction between introspective sleepiness, objective sleepiness and manifest sleepiness (performances) (25,26,32), or subjective sleepiness, performances, sleep propensity and arousal (33), and the more basic distinction between subjective and objective sleepiness (34–37). However, these classifications rely on the top-down definition of the construct. Consequently, in a bottom-up perspective, we categorized the sleepiness assessment tools according to the nature of the tool used to operationalize the measurement of sleepiness. Using the framework of grounded theory, we constructed categories by an iterative inductive aggregation process (38). Categories were created based on the common characteristics of the tools, and the criteria for membership of each category were updated. Two sleep medicine specialists (JAM and RL) classified the tools. Disagreements in category assignment were resolved by consensus. Otherwise, a third sleep specialist (PP) made the final decision.

**Original article metrics**

We identified the original article for each sleepiness assessment tool, defined as the oldest scientific article mentioning its use to assess sleepiness. We extracted the publication year of the original articles and computed minimum, maximum, and distribution values over the decades between 1960s and 2010s. Finally, we calculated the median publication year of the original articles for each category.

**Data synthesis**

As a basis for further discussion of the various aspect of sleepiness, we represented the sleepiness assessment tools according to the publication year of the associated original article and the category into which the assessment tool was classified. The size of each point corresponded to the number of reviews that mentioned the sleepiness assessment tools. To clearly represent the subtleties of the data, we generated an interactive bubble graph using the Python library **Plotly** and **Chartstudio API**. Hovering over the sleepiness assessment tools with the cursor provided a dynamic display of relevant information and the reference article; the interactive axis could be zoomed in and out.
RESULTS

Article selection process

Figure 1 illustrates the PRISMA flowchart used to select the review articles. After removing duplicate reports, 92 review articles were identified, of which a further 57 were excluded (34 were not reviews of sleepiness assessment tools, 11 did not have an available full-text version, 9 focused on pediatric populations, and 4 were not published in English, French, or Spanish). Finally, 36 review articles were included. A complete description of the included and excluded review articles is available in the Supplementary Materials S1.

Description of reviews and sleepiness assessment tools

The 36 review articles were published between 1982 and 2020, including 2 (5.5%), 2 (5.5%), 10 (27.8%), and 22 (61.1%) published in the 1980s, 1990s, 2000s, and 2010s, respectively. The mean and median numbers of sleepiness measures mentioned per review article were 11.8 (SD = 8.0) and 10 (range: 1–39), respectively.

Number of sleepiness assessment tools per category

Ninety-nine tools extracted from the selected reviews (cf. Supplementary materials S2 and S3) were classified into eight categories:
1) Questionnaires (n = 54, 54.5%): instruments that consist of a set of questions or other types of prompts, used to collect information from a respondent (the subject or an observer).
2) Electroencephalography (EEG)-derived assessments (n = 7, 7.1%): quantitative measures based on EEG signals.
3) Polysomnography (PSG)-derived assessments (n = 10, 10.1%): quantitative measures based on a combination of EEG, electro-oculography, and electromyography, used to distinguish sleep from wakefulness.
4) Performance-based assessments (n = 12, 12.1%): quantitative measures related to cognitive or psychomotor performance on a specific task.
5) Activity assessments (n = 7, 7.1%): quantitative and qualitative assessment tools based on observed behaviors and movements.
6) Eye-related assessments (n = 3, 3.0%): quantitative measures based on eye movements and eyelid blinking.
7) Autonomic assessments (n = 4, 4.0%): quantitative measures derived from physiological signals related to the activation of the autonomic nervous system.
8) Other assessments (n = 2, 2.0%): measures not fitting into one of the aforementioned categories (i.e. biological and magneto-encephalography-derived measures).

Number of sleepiness assessment tools per review

On average, each tool was mentioned 4.0 (SD 6.1) times per review. Fifty-nine tools (59.6%) were mentioned less than three times, including 38 (38.4%) that were mentioned only once. Six tools were mentioned in more than 50% of the review articles, namely the MSLT (n = 31, 86.1%), the Epworth sleepiness scale (ESS, n = 29, 80.6%), MWT (n = 26, 72.1%), the Stanford sleepiness scale (SSS, n = 25, 69.4%), the Karolinska sleepiness scale (KSS, n = 21, 58.3%), and the psychomotor vigilance task (PVT, n = 18, 50.0%).
None of the review articles mentioned more than 50% of the sleepiness measurements. The number of tools mentioned in a review article was positively correlated with its publication year ($r = 0.37$, $p = 0.02$).

Thirty-five review articles (97.3%) mentioned at least one questionnaire, 32 (88.9%) mentioned at least one PSG-derived assessment, 24 (66.7%) mentioned at least one performance-based assessment, 17 (47.2%) mentioned at least one autonomic assessment, 16 (44.4%) mentioned at least one EEG-derived assessment, 15 (41.7%) mentioned at least one activity assessment, and 12 (33.3%) mentioned at least one eye-related assessment. The mean number of categories mentioned per review was 4.3 (SD 1.7).

**Original articles of sleepiness assessment tools by year and category**

The original articles associated with the tools were initially published between 1961 and 2019. In total, 7 (7.1%), 6 (6.1%), 21 (21.2%), 37 (37.4%), 13 (13.1%), and 15 (15.1%) original articles were published in the 1960s, 1970s, 1980s, 1990s, 2000s, and 2010s, respectively. The mean interval between the publication year of the original article and the year of its first mention in a review article was 14.0 years (SD = 10.8). The original articles of the questionnaires, EEG-based measurements, PSG-based measures, autonomic measures, behavioral and motor activity measurements, eye-related metrics, and performance-based measures were published in 1973–2019 (median: 1995), 1964–2008 (median: 1995), 1979–2018 (median: 1986), 1961–1982 (median: 1964), 1986–2001 (median: 1995), 1984–1989 (median: 1987), and 1976–2010 (median: 1989), respectively. **Figure 2** presents the sleepiness assessment tools according to the eight categories, publication year of the original article, and number of mentions within the selected review articles. A brief description of each tool is available in the interactive version, allowing access to the original article associated with each tool.
DISCUSSION
We identified almost 100 sleepiness assessment tools mentioned in review articles of sleepiness in adults published over the previous six decades, reflecting the multiple dimensions of sleepiness. More than half of these measures were questionnaires and six tools (MSLT, ESS, MWT, SSS, KSS and PVT) were mentioned by more than half of the included reviews. We also found that the sleep medicine community reports the use of many other tools. This testifies to the ingenuity of the sleep research community to delimit the broad construct of sleepiness, as reflected in the continuous historical development of various tools to measure this central symptom of sleep disorders (Figure 2).

Main sleepiness assessment tools
The six most frequently cited tools (the ESS, MSLT, MWT, SSS, KSS, and PVT) well reflect the large diversity of the 99 tools that we identified. They cover the two main contexts in which sleepiness is measured: the ESS, MSLT, and MWT are traditionally used in a clinical setting, while the SSS, KSS, and PVT are mainly used in experimental settings. Moreover, they also reflect the various constructs underlying the phenomenon of sleepiness. The ESS and MSLT are generally considered as measures of sleep propensity (26,39). In contrast, the SSS, KSS, and PVT evaluate drowsiness associated with sleepiness (40–42). Finally, the functional consequences of sleepiness are usually evaluated with the ESS and MWT (39,43), particularly in automobile accidentology related to sleepiness (4).

Notably, other dimensions of the hypersomnolence spectrum, i.e. excessive need for sleep and sleep inertia (20) are not explored by the six most commonly used tools. Two factors may account for this. First, sleep experts differentiate excessive need for sleep and inertia from excessive sleepiness, so the tools that measure them were outside the scope of this umbrella review, which focused on excessive sleepiness. Second, tools measuring excessive need for sleep and inertia may be underrepresented because these constructs were distinguished only recently, so insufficient time has passed for the associated tools to be widely adopted in sleep clinics and research. Therefore, previous systematic reviews may not have included them. Accordingly, recent studies highlighted the lack of a standardized evaluation of excessive need for sleep (44,45). Other recent studies showed that the PVT may provide a quantitative measure of sleep inertia in patients with idiopathic hypersomnia or other sleep disorders (21,46).

The ESS, SSS, and KSS were designed to reflect the subjective experience of sleepiness, while the MSLT, MWT, and PVT capture objective aspects of sleepiness. While these six tools are often considered as the gold standard, major concerns remain regarding the weak correlations between them, in particular for the ESS (47) but also for other questionnaires (e.g., SSS) and for performances (12).

Questionnaires represented more than half (54.5%) of the tools mentioned in the included review articles, and their diversity highlights the complexity of the subjective experience of sleepiness. Some tools, such as the SSS and KSS, were developed to measure sleepiness at a given time. These simple questionnaires appear to be sensitive to sleep deprivation and time of day (48,49), and they are widely used in experimental settings. Other questionnaires were designed to assess the daily life consequences of sleepiness, including the ESS, the observation and interview-based diurnal sleepiness inventory (ODSI) (50), and the sleep-wake activity inventory (SWAI) (51), and are more likely to be used in clinical settings.

Our analysis of the constructs of the questionnaires revealed the need to investigate specific symptomatic dimensions of sleepiness. Like the ESS, most scales mainly address sleep propensity. However, some questionnaires, such as the Toronto hospital alertness test (THAT) (52) and the resistance to sleepiness scale (RSS) (53), were designed to evaluate the ability to stay awake under various conditions. Few questionnaires assess the experience of drowsiness
(SSS and KSS), sleep inertia (sleep inertia questionnaire [SIQ]) (54), or excessive sleep quantity (ODSI) (50). Finally, some questionnaires were designed to specifically assess the functional consequences of sleepiness, such as the ESS and the functional outcomes of sleep questionnaire (FOSQ) (55). However, most questionnaires only address a single dimension of sleepiness and there are very few multidimensional questionnaires. To overcome this issue, the idiopathic hypersomnia severity scale (IHSS) (56) and the hypersomnia severity index (HSI) (57) were developed almost simultaneously at the end of the 2010s to assess the whole spectrum of hypersomnia and its functional consequences. They have demonstrated excellent psychometric properties (58,59) so they could be widely used to assess hypersomnia in both clinical practice and randomized controlled trials (60).

**New sleepiness assessment tools and path dependence**

The widespread adoption of new, well-validated, efficient tools may be hindered by a broader *path dependence* effect on scientific knowledge production (61,62), involving that “what counts as empirical or theoretical knowledge at any time is a function not just of how the world is but of the specific material-conceptual-disciplinary-social-etc. space in which knowledge is situated” (63). This historical development has led to the over-selection of some tools to the detriment of others, as illustrated by our finding that 38 measures (38.4%) were mentioned only once in the included review articles. Several mechanisms may account for this effect. First, as in many other disciplines, the initial tools are used increasingly over time and are thus the most well-validated. Therefore, they are widely adopted by the research community. Moreover, path dependence may result from the historical choices of the scientists and learned societies that have shaped the recent field of sleep medicine. The over-representation of two PSG-based assessments (the MSLT and MWT) is probably related to the central role of PSG in sleep medicine (64). The MSLT is also an essential test in sleep medicine because of its inclusion among the diagnostic criteria of the first International Classification of Sleep Disorders (1990) and the fourth revision of the Diagnostic and Statistical Manual of Mental Disorders (1994). The widespread use of the MSLT, as well as the MWT, was also influenced by the establishment of practice parameters by the American Academy of Sleep Medicine, which have since been updated, i.e. in 2005 (18,65) and in 2021 (66). Most expert consensuses on the use of sleepiness assessment tools were concerned with objective tests rather than questionnaires. Our results suggest that this phenomenon also affects questionnaires, possibly through therapeutic trials, in which outcomes have to be assessed using reliable subjective assessment tools that are comparable between studies.

Our exhaustive list of sleepiness measurement tools puts the multiple aspects of sleepiness into an historical perspective. We propose that Figure 2 should be interpreted in the context of the previous six decades of sleep research. Before the 1970s, a psychophysiological approach appears to have driven the development of the first sleepiness assessment tools, with a focus on identifying physiological markers of sleepiness as a state of low level of arousal, such as EEG (67,68), respiratory rate (69), pupillometry (70), and skin conductance (71). The first performance-related measurements were developed to assess the functional consequences of sleepiness, such as real-life or simulated driving performance tests (72,73), even before the risk of sleepiness-related accidents emerged as a public health concern.

During the 1970s, few sleepiness assessment tools were developed. Nevertheless, that decade represented a paradigm shift and paved the way for psychometric testing, especially through the introduction of the SSS in 1973 (42), which became an essential tool for assessing sleepiness thereafter.

The two most widely used objective tests of sleepiness (the MSLT and MWT) were developed in the 1980s. With the introduction of PSG as a measure of sleepiness, the 1980s witnessed the
emergence of the concept of sleep latency as a continuous measure of sleep propensity. Developed in 1985, the PVT (41) simplified evaluation of the consequences of drowsiness for performance via the measurement of cognitive processes, such as attention, which is also the main measure of other important tests developed during the following decade, such as the Oxford sleep resistance test [OSLER, (74)] and the sustained attention to response task [SART, (75)]. The introduction of the Pittsburg Sleep Quality Index [PSQI, (76)] in 1989 as a tool for measuring sleep disturbance and disorders, and the ESS in 1991, paved the way for the era of psychometric testing in the sleepiness research field. Since the 1990s, sleepiness has been perceived not only as a physiological state, but also as a clinical symptom.

During the 2000s and especially the 2010s, particular attention was given to other dimensions of sleepiness, such as sleep inertia and excessive sleep, which are the main symptoms of central disorders of hypersomnolence. During this decade, new tools were mainly focused on pathological populations, since therapeutic innovations in the field of sleep disorders created a strong need for tools that are sensitive to changes and which can be used in randomized controlled trials for specific sleep disorders (60,77).

Some trends, especially those emerging over the last two decades, may have escaped our analysis due to methodological limitations. Indeed, the mean interval between the earliest mentions of a sleepiness measure in the scientific literature and its inclusion in a review article was 14 years. Accordingly, some recently developed tools may have been omitted from this umbrella review. For example, a new approach to the development of sleepiness assessment tools emerged in the late 2010s, involving refinements of existing tools based on the alternative measures of tests, without changing the procedure, as done recently for PSG (78). For instance, microsleep analysis (79–82) and video-based analysis of ocular movements (83) during the MWT may improve sleepiness-related driving risk assessments, as well as the conventional mean sleep latency measure. Other studies based on the MSLT have identified alternative cut-offs (84) and measures (e.g., sleep-wake transition dynamics (85,86) and the percentage of REM sleep (87)) as better measures of excessive sleepiness in narcolepsy.

Some studies using standardized PSG-based tests have focused on the correlation between subjects’ self-reported sleepiness and standard PSG-based parameters (47). Interestingly, the Barcelona Sleepiness Index (BSI) (88) was specifically designed to ensure correlations and external validity with objective sleepiness measurements such as the PSG, MWT, MSLT, and SART. However, its validity in clinical settings remains to be proven, as the original study lacked reproducibility. More recently, two studies investigated the self-perceptions of subjects performing the MWT. Although the first study showed the impact of Parkinson’s disease on the awareness of sleep onset (89), the second showed that misperception of sleep onset during the test was related to the risk of sleepiness-related road accidents (90). Finally, a recent study (91) investigating the relationship between PSG-measured sleep times and their self-reports found a significant influence of the type of sleep–wake disorder on the type of misperception (i.e., under- or over-perception) of total sleep time. Despite these advances, few studies have investigated predictors of awareness of sleepiness. Thus, further studies are needed to validate self-reported sleepiness with appropriate objective evaluation tools.

Limitations

First, as in all umbrella reviews, our analyses and conclusions are based on previously published reviews (29). Also, we did not set out to report all existing sleepiness assessment tools, and some do not appear in this review, including measures of sleepiness based on functional neuroimaging (92–94), genetic (95–97), and biological approaches (98–100). Furthermore, our selection criteria excluded sleepiness assessment tools developed for pediatric populations and specific diseases (101,102). Nevertheless, the umbrella review design allowed us to capture all
of the assessment tools widely adopted by the sleep community, with a view to establishing recommendations for clinical and research practice (29).

Second, due to the lack of a common organizational structure of the tools in the different reviews, we were unable to systematically categorize the sleepiness measures according to the symptoms or psychophysiological mechanisms that they explored, i.e. sleep propensity, drowsiness, vigilance, excessive need for sleep, sleep inertia, ability to maintain wakefulness, etc. Although recent articles have proposed clarifications of certain related constructs, such as fatigue (2), excessive sleep (20), vigilance, alertness, and arousal (19), sleepiness remains a poorly defined construct. Despite the difficulty of refining the very concept of sleepiness based on the tools we identified, our historical approach offers an original insight into their development and adoption (see Figure 2). This could represent a first step toward a consensus definition of the different aspects of sleepiness by proposing a clearer basis for further discussions.

Third, in line with the previous limitation, the data extraction methodology requires comment. The constructs explored in the reviews were not always specified by the authors, and the verbal descriptions included in the review articles varied widely (cf. Supplementary Materials S4). Large-scale, systematic approaches may be more appropriate to explore and classify sleepiness assessment tools to complement the bottom-up approach of sleepiness initiated in this paper. Among the techniques used for data-driven bioinformatics analysis, text mining enables the extraction of hidden knowledge from published articles, and may prove indispensable to understand and model the organization of the various constructs close to sleepiness (103).

Finally, we did not extract information regarding state vs. trait aspect, reliability, sensitivity to change, or thresholds for pathology, which are essential parameters for the clinical application of sleepiness measurements. Thus, further systematic analysis of these factors is required.

Conclusions

This umbrella review represents a first step toward a better delineation of the different dimensions underlying sleepiness. It reflects how the sleep medicine community measures sleepiness, in order to investigate and refine the notion of sleepiness inventoring such identified tools. In a forthcoming article (104), we have initiated an effort to reconcile this "bottom-up" approach with the traditional expert approach (“top-down” approach), by crossing the historical lineages identified in this article with the sub-dimensions of hypersomnolence defined in previous articles (20,35,105). From this cross comparison, we propose in Supplementary Materials S5 a synaptic table of the most used tools nowadays to measure sleepiness according to the expert-based approach, to which we have associated the category in which we have classified them, their use in clinical practice, and the sub-construct of hypersomnolence that they measure. This fruitful confrontation of approaches may encourage the sleep research community to create an integrated model of the different components of sleepiness based on a computational approach (106), similar to other brain-related symptoms (107). This may in turn prove beneficial, not only by identifying cut-off values for the diagnosis and treatment of hypersomnolence disorder, but also the neurophysiological mechanisms underlying sleep disorders and sleep-related impairment.
**Practice points**
- This umbrella review included 36 review articles published between 1982 and 2020 that investigated sleepiness assessment tools.
- In total, 99 tools were extracted and classified into eight categories.
- Six tools (the ESS, MSLT, MWT, SSS, KSS, and PVT) were overrepresented (mentioned by more than 50% of the review articles).
- The over-representation of these six tools may reflect a path dependence effect, in line with the recent trends in sleep medicine research.
- In total, 38 tools were mentioned only once in the review articles.
- More than 50% of the tools were questionnaires, which predominantly investigate the subjective experience of sleepiness.

**Research Agenda**
We encourage researchers in sleep medicine to:
- Conceive and validate alternative measures derived from the most frequently used tests without changing the assessment procedure.
- Promote the investigation of sleepiness by diversifying assessment tools, in particular those related to biological and neuroimaging techniques.
- Systematize the inter-validation of sleepiness assessment tools to better specify the different dimensions of the construct.
- Determine the variables influencing sleepiness awareness to investigate the link between perceived sleepiness and the outcomes of the classical tests.
- Determine the variables, i.e. causes of sleepiness, acute vs. chronic sleep deprivation, age, etc, which influence the reliability, sensitivity, and thresholds to better distinguish between physiological sleepiness and pathological sleepiness.
- Evaluate the interest of sleepiness assessment tools for diagnostic criteria in sleep disorder classifications.
- Encourage the use of innovative computational approaches to develop an integrated model of sleepiness in order to promote mechanistic understanding and for clinical use.
REFERENCES

Key references: (1), (2), (3), (19), (20)


22. Trotti LM. Waking up is the hardest thing I do all day: Sleep inertia and sleep drunkenness. Sleep Medicine Reviews. 2017;35:76–84.


71. Davies DR, Krkovic A. Skin-Conductance, Alpha-Activity, and Vigilance. AJP. 1965;78(2):304.


FIGURES

Figure 1.
PRISMA flowchart of article selection process.
Figure 2.

Sleepiness measures extracted from studies of adult population, by publication year of source article and category. Diameter of bubbles corresponds to number of times that measure was mentioned in included reviews. A brief description of each assessment tool is available in the interactive version of this figure which is available at:
https://chart-studio.plotly.com/~vincent.martin/7/##/