The Role of Patient-Reported Outcomes in Sleep Measurements

Dirk Pevernagie, MD, PhD[a,b,*], Fré A. Bauters, MD, PhD[a,b], Katrien Hertegonne, MD, PhD[a,b]

KEYWORDS

• Sleep • Sleep medicine • Questionnaire • Patient-reported outcome • Polysomnography

KEY POINTS

- Sleep disorders must be assessed subjectively and objectively.
- Subjective assessment includes the medical interview and administration of dedicated questionnaires or patient-reported outcome measures (PROMs).
- Generic and disease-specific PROMs are available for a variety of sleep disorders.
- PROMs have inherent limitations and future research should aim to improve them.
- PROMs have become increasingly important in clinical research and health outcomes assessments.

INTRODUCTION

Sleep, next to healthy nutrition and exercise, is the third fundamental pillar of good health. Disordered sleep is often associated with decreased health-related quality of life (HRQoL) and may predispose to socioeconomic adversity in many affected subjects.[1] Sleep disorders may constitute distinct medical conditions or may complicate other somatic or psychiatric diseases.[2] Adverse biomedical and psychosocial conditions[3–6] as well as unfavorable socioenvironmental factors[7] negatively affect sleep and may play a significant role in the clinical manifestation of sleep disorders. Owing to lack of education on the physiology and pathology of sleep in the curriculum of health care professionals, these disorders remain often underdiagnosed and, consequently, not well treated.[4,8] The use of questionnaires on sleep and sleep disorders may help the practitioner to compensate for this knowledge gap. Moreover, assessment of disordered sleep by applying structured enquiries may be instrumental for making suitable differential diagnosis and offering patient-centered care.

Sleep disorders are assessed the same way as any other medical problem. The history is key to formulating a working hypothesis that may be corroborated (or rejected) by physical examination and targeted technical investigations. To confirm a tentative diagnosis and to assess disease severity, sleep can be measured with different instruments. Polysomnography (PSG) is considered the gold standard for this purpose.[9] PSG is carried out by overnight recording of neurophysiological and cardiorespiratory signals, followed by detailed analysis of the content and finalized by interpretation of the results by a sleep specialist.[10] Thus, the biologic signals of PSG capture adverse events in sleep that compromise its quality. PSG is a reliable instrument for the objective assessment and quantification of sleep-related pathophysiological phenomena.

Surprisingly, in many patients, no robust correlation can be demonstrated between the "pathophysiological severity" of the disorder as

[a] Department of Respiratory Medicine, Ghent University Hospital, Gent, Corneel Heymanslaan 10, Gent 9000, Belgium; [b] Department of Internal Medicine and Paediatrics, Faculty of Medicine and Health Sciences, Ghent University, Corneel Heymanslaan 10, Gent 9000, Belgium
* Corresponding author.
E-mail address: dirk.pevernagie@ugent.be

Sleep Med Clin 16 (2021) 595–606
https://doi.org/10.1016/j.jsmc.2021.07.001

evidenced by markers on PSG and the "clinical severity" as indicated by the seriousness of symptoms and signs. Especially, the lack of association between the apnea-hypopnea index (AHI), a polysomnographic marker of severity in obstructive sleep apnea (OSA), and clinical manifestations of this condition has become evident in recent years.[11] Studies appraising associations between AHI and indices of HRQoL have also failed to demonstrate any significant relationships.[12–14] Such lack of correspondence may indicate that the pathophysiology-driven model of sleep-disordered breathing does not satisfactorily capture disease heterogeneity and does not identify the subtleties that constitute the individual's clinical picture. This lack of correspondence may hold true for nonrespiratory sleep disorders as well. The AHI and potentially other biomarkers emanating from pathophysiological paradigms may have insufficient power to predict clinical relevance and their use as surrogate markers for disease severity may be misleading.[11] In sleep medicine, as in other disciplines, it is mandatory to apply a broad range of examinations for establishing a correct diagnosis and for rating disease severity. In this respect, the medical interview still is the cornerstone of the clinical workup.

Treatment is primarily aimed to remedy the underlying cause of the diagnosed sleep disorder. In OSA, for example, the therapeutic goal is to lower the AHI by preventing passive collapse of the upper airway during sleep. Normalization of the AHI, however, is not always associated with sufficient improvement of daytime symptoms (**Box 1**).[15] In this case, alternative diagnoses or associated comorbidities must be further explored. Thus, restoration of the physiologic process is not an exclusive proxy for therapeutic success. Systematic reassessment of presenting symptoms is essential and questionnaires that gauge the patient's perceived alterations in symptoms and HRQoL may be used for that purpose.[16] Eventually, the patient's appreciation of her or his own health condition is what matters most.

The patient-reported outcome is instrumental in determining treatment success. A patient-reported outcome measure (PROM) is a questionnaire consisting of several patient-reported outcomes, designed to evaluate symptoms, functioning, and other attributes inherent to HRQoL. Such measures can be developed to assess the outcomes of a certain disease (disease-specific PROMs) or several diseases irrespective of their causes (generic PROMs). PROMs are used in combination with clinical outcome measures (COMs) to define overall therapeutic success.[17]

Box 1
Attribution bias

A 53-year-old obese male patient presents with complaints of loud snoring and breathing stoppage observed by the bed partner. He reports excessive daytime sleepiness (EDS), as he is unable to remain awake during staff meetings and driving. The AHI, assessed by PSG, is 35/h. The patient is compliant with prescribed CPAP therapy and reports that his snoring is well controlled, much to the satisfaction of his bed partner. His sleepiness, however, is not improved at all. A new PSG under CPAP therapy demonstrates a residual AHI of 2/h and a total sleep time of 674 minutes. An annex MSLT shows a mean sleep latency of 5 minutes, without any REM sleep in the 5 naps. Repeat history taking is remarkable for the persistence of EDS and the need to sleep more than 10 hours per night ever since his early teens. A diagnosis of "idiopathic hypersomnia with long sleep time" is established and treatment with methylphenidate 10 mg t.i.d. is commenced in addition to the already installed CPAP therapy.

This case is remarkable for a spurious association between pathophysiologically relevant sleep-disordered breathing (with an AHI indicative of "severe OSA") and EDS. In this example, the hypersomnolence was primarily caused by an unrelated disorder of the central nervous system.

In this article, we will review the purposes of sleep questionnaires that are used as structured PROMs. Also, we will expand on the multiple purposes of PROMs, on their relevance for value-based health care, and on the necessity to establish standards for appraising the quality of these instruments. Conventional and special approaches to querying patients will be discussed, as well as inherent limitations and opportunities for future developments.

PATIENT-REPORTED OUTCOME MEASURES CAN BE USED FOR DIFFERENT PURPOSES

As a means of structured history taking, questionnaires have been introduced long ago in medical research. PROMs were initially developed for clinical trials, in which they were used to identify eligible participants, to monitor therapeutic efficacy, side effects, and safety of new medical products, and, eventually, to estimate their risk versus benefit ratios.[18] Currently, the use of PROMs has become mandatory in pharmaceutical research.[19]

In clinical practice, PROMs may have different purposes and may serve multiple goals. Screening questionnaires are typically administered in a

preclinical phase and are designed to establish the a priori likelihood of a certain diagnosis. Systematic reviews have been published on questionnaires that intend to screen for multiple sleep disorders,[20] and for single diseases such as OSA.[21,22] Further discussion of this matter is outside the scope of this review.

PROMS are especially useful for estimating the relative importance of different symptoms associated with a given clinical condition. Not all complaints are equally troublesome and gathering inclusive information on the different symptoms enables the practitioner to focus on details that matter most to the individual patient.[23] The characteristics of particular traits may provide actionable information suitable for personalized treatment.[24] Likewise, the PROMs that allow for this differentiation should be sensitive enough to monitor effects of treatment and to verify that therapeutic results correspond with the patient's expectations in terms of improvement of HRQoL.[25]

HRQoL can be concisely defined as "the personal health status of an individual."[26,27] Of note, symptom severity may compromise perceived health, but it is not per se synonymous with a decreased quality of life. Actually, HRQoL is a multiple domain concept not only referring to experiences of illness such as pain, fatigue, and disability but also considering broader aspects of the individual's physical, emotional, and social wellbeing.[16] Different constructs of HRQoL exist, but when used in a research domain, the chosen model should be consistently applied.[28]

PROMs are increasingly used to standardize medical practice and to assess the effectiveness of organized health care. Research on patient-centered outcomes makes use of aggregated PROMs data to compare the effectiveness of different providers with the aim to support quality improvement in health care.[29] Value-based health care is a prevailing health-economical model in which COMs and PROMs are combined into standard sets for appraising treatment outcomes of various diseases.[30] As outcomes are based on patients' priorities, the role of internationally validated, high-quality PROMs is paramount in the assessment strategy of value-based health care.[17] From an integrative perspective, there is a case for pooling the intentions and efforts of the various stakeholders (ie, clinicians, patients, researchers, and health care insurers) to endorse sustainable data collection systems in which PROMs are administered at intake and in the course of treatment. Such comprehensive approach is expected to stimulate meaningful use in research, clinical practice, and quality improvement programs.[31]

QUALITY AND REPORTING OF PATIENT-REPORTED OUTCOME MEASURES

Measurement properties of PROMs must comply with rigorous standards as shown in **Table 1**. The quality prerequisites of a PROM must be tested before release for large-scale use. Recommendations regarding the design and implementation of new questionnaires are available in the literature.[32–35] Also, there are guidelines on how to assess the methodological quality of existing PROMs.[36] The International Society for Quality of Life Research (ISOQOL) has published a set of standards itemizing different properties that a PROM should be tested for (see **Table 1**).[37] Together, these properties define the "validity," being the agreement between what a PROM actually measures in view of what it purports to measure. The role of patients is very important in determining the content aspect of validity—the target population should be involved already in the initial phase of designing a new PROM. Finally, PROMs should not be regarded as static instruments but should be updated in the course of time with the aim of improving their measurement properties in a "PROM cycle."[35]

As already mentioned, PROMs consist of separate questions that may be grouped into different domains reflecting various dimensions of a certain disease. To determine the dimensionality of a PROM in terms of different symptoms or HRQoL aspects, items can be grouped together based on the clinical relevance of symptoms. This is called a "clinimetric" method. In contrast, the application of principal component analysis (ie, statistical analysis of items in a covariant matrix) is known as a "psychometric" approach.[38,39] Both methods can be used to decide on the content of a PROM. The questions themselves test the severity of a phenomenon (a symptom, generally speaking) in terms of intensity and frequency over time. Furthermore, it can be appraised to what extent various symptoms pose a problem regarding different aspects of HRQoL. Usually, the results of the separate items are computed into scores per domain and/or a global score reflecting the overall subjective severity of the disease. Moreover, several options exist to visualize the results of PROMs. **Fig. 1** shows an overview of common graphical illustrations of symptoms and domains.

PATIENT-REPORTED OUTCOME MEASURES IN SLEEP MEDICINE

Since the 1990s, several PROMs have been introduced for different purposes in the field of sleep

Table 1
Definition of PROM properties

Conceptual and measurement model	The conceptual model provides a description and framework for the targeted construct(s) to be included in a PRO measure. The measurement model maps the individual items in the PRO measure to the construct
Reliability	The degree to which a PRO measure is free from measurement error
Internal consistency	The degree of the interrelatedness among the items in a multi-item PRO measure
Test–retest reliability	A measure of the reproducibility of the scale, that is, the ability to provide consistent scores over time in a stable population
Validity	The degree to which a PRO instrument measures the PRO concept it purports to measure
Content validity	The extent to which the PRO measure includes the most relevant and important aspects of a concept in the context of a given measurement application
Construct validity	The degree to which scores on the PRO measure relate to other measures (eg, patient-reported or clinical indicators) in a manner that is consistent with theoretically derived a priori hypotheses concerning the concepts that are being measured
Criterion validity	The degree to which the scores of a PRO measure are an adequate reflection of a "gold standard."
Responsiveness	The extent to which a PRO measure can detect changes in the construct being measured over time
Interpretability of scores	The degree to which one can assign easily understood meaning to a PRO measure's scores
Minimal important difference	Minimal important difference—The smallest difference in score in the outcome of interest that informed patients or informed proxies perceive as important, either beneficial or harmful, and that would lead the patient or clinician to consider a change in the management
Burden	The time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (investigator or administrative burden)

Adapted from Reeve et al.[37] with permission from the publisher.

medicine. A distinction is being made between generic questionnaires that may be used in various medical disciplines versus disease-specific questionnaires, designed to gauge symptom severity and HRQoL effects for particular sleep disorders. With respect to disease-specific aspects of sleep medicine, we review frequently used questionnaires in the domains of OSA, insomnia, and restless legs syndrome (RLS).

Generic Questionnaires

Generic questionnaires such as the Medical Outcomes Study 36-item Short Form Health Survey (SF-36)[40] and the EuroQol 5 Dimensions Questionnaire (EQ-5D),[41] among others, have been used to assess HRQoL in patients with sleep disorders. As this target population was not specifically envisaged when these questionnaires were designed, there is little evidence regarding content and other features of measurement validity.[42] The relevance of generic HRQoL instruments for sleep medicine practice is limited and will not be discussed further in this article.

Generic Sleep- and Sleepiness-Related Questionnaires

Two sleep questionnaires, the Pittsburgh Sleep Quality Index (PSQI)[43] and the Epworth Sleepiness

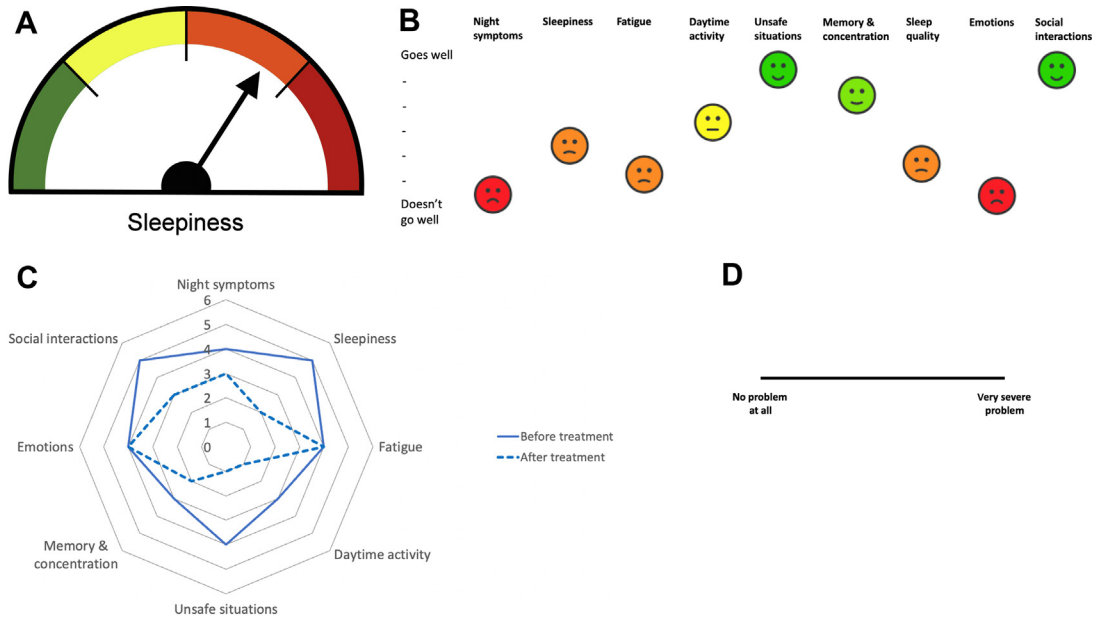

Fig. 1. Different graphical presentations of results from PROMs. (*A*) Graphical presentation of a trait, for example, sleepiness, as a value on a tachometer scale. (*B*) Presentation of different traits on a 7-point Likert scale in parallel columns, using smileys to enhance the visual effect; adapted from Abma and colleagues[39] with permission from the publisher. (*C*) Radar plot with positioning of different traits on a 7-point Likert scale, also showing treatment effects. (*D*) A 1-dimensional visual analog scale.

Scale (ESS),[44] are broadly used in different areas of sleep medicine. The PSQI is suitable for assessing sleep quality in sleep disorders as well as disturbed sleep in other conditions such as mood disorders or pain syndromes.[45] The PSQI consists of 19 questions on a 4-point Likert scale (0–3) and covers 7 domains: subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbance, use of hypnotics, and daytime dysfunction. The global score is the sum of all domain items and ranges from 0 to 21. The cut-off for abnormal sleep is >5, with worse sleep quality being associated with higher scores. A review and meta-analysis of the suitability of the PSQI for assessing sleep dysfunction in clinical and nonclinical populations has been published elsewhere.[46] According to this structured review, the PSQI shows strong reliability and validity, and moderate structural validity in a variety of samples, suggesting the tool fulfills its intended utility.

The ESS is a concise PROM composed of 8 questions on a 4-point Likert scale (0–3), yielding scores between 0 and 24, 11 and higher indicating excessive daytime sleepiness. The ESS was originally designed to assess subjective sleepiness in both normal subjects and patients with various sleep disorders.[44] In a separate study on measurement properties, adequate validity was demonstrated for this scale, based on which it was proposed as a reliable method for measuring daytime sleepiness in adults.[47] However, subsequent studies have shown limited internal consistency, rendering the ESS probably suitable for group but not for individual-level comparisons.[48] The reliability of the ESS in clinical settings is still unproven[49] and its unconditional application has been criticized.[50] Finally, the ESS seems to embody sleepiness better in men than in women who less often have a total score of 11 or higher, although they report feelings of sleepiness as often as men.[51,52]

Patient-Reported Outcome Measures for Obstructive Sleep Apnea

A whole array of questionnaires is currently available for use in OSA. Below, we only report on PROMs that have been subject to appropriate quality assessment and for which measurement properties have been reported.[42] These PROMs are listed in **Table 2**.

The Functional Outcomes of Sleep Questionnaire (FOSQ) is a PROM designed to assess HRQoL in adults suffering from excessive daytime sleepiness. It has been used for studying the effects of treatment with positive airway pressure in OSA patients.[53] The instrument comprises 30-items on a 4-point Likert scale assessing effects

Table 2
Validated PROMs in OSA

Questionnaire	Authors	Content	# Items	# Domains	Likert Scale	Direction
Functional Outcomes of Sleep Questionnaire (FOSQ)	Weaver et al,[54] 1997	Assessing the degree of difficulty for doing activities due to fatigue or being sleepy	30	5	4	↑
OSA Patient-Oriented Severity Index (OSAPOSI)	Piccirillo et al,[55] 1998	Magnitude and importance of problems related to impaired activities, feelings, situations, and behaviors	32	5	6	↓
Calgary Sleep Apnea Quality of Life Index (SAQLI)	Flemons et al,[56] 1998	The disease-related part of the questionnaire probes the amount of time, the amount of difficulty, or the severity associated with certain problems related to activities and functions	56	4	7	↑
Calgary Sleep Apnea Quality of Life Index (SAQLI)	Flemons et al,[56] 1998	The treatment-related part of the questionnaire probes side effects of CPAP therapy in terms of experienced problems	28	N/A	7	↑
Quebec Sleep Questionnaire (QSQ)	Lacasse et al,[57] 2004	Assessing the degree of problems related to impaired activities, feelings, situations, and behaviors	32	5	7	↑
Visual analogical well-being scale (VAWS)	Masa et al,[14] 2011	Rating the degree of the present well-being status between least favorable and most favorable by putting a marker on a horizontal line	1	1	N/A	→
Patient-Reported Apnea Questionnaire (PRAQ)	Abma et al,[39,59,60] 2017–2019	Rating the degree experiencing problems with activities, feelings, situations, and behaviors	40	10	7	↑

Abbreviations: ↑, Higher values indicate a better status; ↓, Higher values indicate a worse status; →, Value to the right indicates a better status; N/A, not available.

of being sleepy or tired on functional performance in 5 domains of health (activity level, general productivity, vigilance, intimate relationships, and social outcome).[54] A global score between 5 and 20 is obtained by computation of the subscales of the 5 domains, a lower score indicating worse HRQoL.

The OSA Patient-Oriented Severity Index (OSAPOSI) consists of 32 questions probing problems in 5 domains (nocturnal sleep, daytime functioning, emotions, productivity, and need of medical care).[55] Each item is assessed according to the severity of the problem on a 6-point Likert scale

and the impact on the HRQoL. Higher values correspond with higher impact.

The Calgary Sleep Apnea Quality of Life Index (SAQLI) comprises 56 disease-related and 28 treatment-related questions.[56] Each item is rated on a 7-point Likert scale. The following domains are covered: daily functioning, social interactions, emotional functioning, and symptoms. Also, unwanted treatment-induced side effects are registered. The questions encompass the amount of time a problem is present, the amount of difficulty a person experiences with a certain problem, or the severity of the problem itself. In contrast with the other OSA questionnaires that can be filled out by the patients themselves, this elaborate PROM was designed to be administered by an interviewer.

The Quebec Sleep Questionnaire (QSQ) lists 32 questions on a 7-point Likert scale, querying the degree of problems associated with daytime sleepiness, diurnal symptoms, nocturnal symptoms, emotions, and social interactions.[57] The mean scores of the 5 domains are computed to produce a total score, positively reflecting HRQoL. The instrument's responsiveness is adequate to show subtle changes induced by treatment. Although the SAQLI and QSQ bear similarities, a notable difference between the two is that the former is based on a "psychometric" factor analysis model, whereas the latter results from a "clinimetric" disease impact approach.[58]

Masa and colleagues[14] developed a PROM for OSA based on a simple visual analogical well-being scale (VAWS) and assessed its performance in respect of existing HRQoL questionnaires. VAWS correlated with all HRQoL tests but better with FOSQ and EQ-5D. Furthermore, VAWS and FOSQ correlated better with clinical variables (restlessness and snoring) than other HRQoL tests. VAWS captured effects of treatment similarly to FOSQ but better than other HRQoL tests. VAWS was promoted as a very simple tool for testing HRQoL in OSA before and after treatment.

The OSA-specific questionnaires described earlier have been criticized for incomplete validation and the lack of certainty about measurement error.[42] Recently, a new PROM for OSA has been developed, involving patients in all the consecutive steps of instrument validation. The Patient-Reported Apnea Questionnaire (PRAQ) consists of 40 questions on a 7-point Likert scale, probing the degree of difficulties or problems with OSA-related symptoms over 10 health-related domains.[59] The measurement properties are appropriate and responsiveness to treatment seems adequate.[39] Although patients were generally positive about the usefulness of the PRAQ, health care providers reported minor impact on their practices and did not consider the PROM of great help with regard to improving patient-centeredness.[60]

Patient-Reported Outcome Measures for Insomnia

Although insomnia—the inability to fall asleep or to maintain sleep overnight—is a frequent complaint in many common diseases and sleep disorders, it can be a diagnostic entity in its own right. In the latter case, the term "chronic insomnia disorder" is used.[2] Several PROMs have been developed for this condition. The 2 most frequently applied questionnaires are discussed in this section. The aforementioned PSQI is suitable for assessing insomnia severity and for evaluating effects of treatment.[61] The Insomnia Severity Scale (ISI) is a 7-item questionnaire on a 7-point Likert scale, surveying difficulties with initiating or maintaining sleep and associated adverse daytime consequences. Results range between 0 (no insomnia) and 28 (very severe insomnia). Scores between 8 and 14 are considered sub-threshold insomnia.[62] There is convincing evidence to show that the ISI is a reliable instrument for detecting cases of insomnia in the general population, as well as for assessing treatment responses in clinical patients.[63]

Patient-Reported Outcome Measures for Restless Legs Syndrome

RLS is characterized by unpleasant feelings in the lower limbs, and sometimes also the arms or the trunk. These sensations cause an urge to move and are relieved by movement. The symptoms exacerbate in the evening and may prevent patients from falling asleep and/or cause awakenings.[64] PROMs are available to assess symptom severity and response to treatment.

The Restless Legs Syndrome Quality of Life Questionnaire (RLSQLQ) consists of 18 items that gauge the effects of RLS on the patient's functioning related to work, social and sexual interactions.[65] Ten of the items yield a global quality-of-life score between 0 and 100, a higher value indicating a better outcome. The other 8 questions deal more in depth with work and sexual interest. The RLSQLQ was found to be a reliable instrument for measuring HRQoL in RLS patients.[65]

The International RLS study group (IRLSSG) has developed a 10-question PROM.[66] This scale, the IRLSSG rating scale (IRLS), grades the severity, frequency, and impact on sleep of RLS symptoms, higher values indicating more severe complaints. The IRLS is not conceived as a screening tool and requires a prior diagnosis of RLS to be used properly. With this scale, spontaneous fluctuations

in symptom severity and treatment responsiveness can be assessed. The measurement properties of the IRLS are deemed appropriate.[66]

CONTROVERSIES

Although the methodology of measuring PROs by the systematic application of validated questionnaires has greatly improved our management of various sleep disorders, there are also downsides to this approach. The inexpert use of PROMs may lead to pitfalls that must be acknowledged and addressed.

The use of disease-specific questionnaires may be a source of nosologic bias. On the one hand, symptoms of disturbed sleep such as snoring, inability to sleep, and restlessness may be a manifestation of an underlying condition, for example, alcohol abuse, rheumatic pain, and constitutional eczema, respectively. On the other hand, each of these symptoms may be a key feature of a nosologically defined sleep disorder—in this example, OSA, chronic insomnia disorder, and RLS, respectively. How to interpret symptoms, either as elements of a multisymptomatic condition or as main traits defining a particular phenotype, largely depends on the clinical context. In sleep medicine, generic symptoms and "specific" symptom-based disorders are frequently mixed up.

Many questionnaires constructed around specific sleep disorders are a compilation of nonspecific symptoms that may occur in other conditions as well. A set of symptoms attributed to a particular sleep disorder may overlap with other disorders characterized by a different pathophysiological background. Yet, the assignment of a selection of symptom-based questions to a disease-specific PROM, invariably suggests that all items are causally related to the postulated disease, which is obviously not the case. Therefore, inexpert application of disease-specific PROMs may result in spurious diagnoses and, consequently, inefficient treatment (**Box 2**). In extreme situations, PROMs may generate information that potentially could be (ab)used in ways that disadvantage patients or to limit access to medical services.[67]

In contemporary sleep medicine, the diagnosis of nosologically defined sleep disorders is founded on a combination of a clinical presentation and evidence for pathophysiological abnormalities demonstrated by clinical sleep testing. Although both components may coincide or even be discordant, establishing a diagnosis is frequently straightforward and therapy is mostly effective. Not rarely, however, the clinical presentation is complex. Different sleep disorders may co-occur or be complicated by other diseases. Insomnia, for instance, is a common complaint in somatic and/or mental diseases. Moreover, insomnia and OSA co-occur in approximately 30% to 40% of cases.[68] Application of PROMs for specific sleep disorders will only partially map these complex conditions and associated HRQoL impairments. Also, patient-centered treatment outcomes will be incompletely assessed.

Box 2
Nonspecificity of symptoms

A patient with severe RLS obtains a total score of 25 on the Insomnia Severity Index, composed of the following subscores: (1) difficulty falling asleep—4; (2) difficulty staying asleep—4; (3) problems with waking up too early—2; (4) dissatisfaction with sleep—4; (5) sleep problem noticeable to others—3; (6) worry and distress—4; and (7) interference with daily activities—4.

This test result could be inadvertently labeled as "very severe insomnia disorder." Yet, treatment with cognitive behavioral therapy would be ineffective in this case because RLS is the causative mechanism.

Nosologically defined sleep disorders may be heterogeneous in clinical presentation. For example, in OSA, at least 3 different phenotypes have been observed, namely patients with excessive daytime sleepiness, disturbed sleep, or minimal symptoms.[69] These subtypes cannot be discriminated by the AHI, as quite similar AHI values were shown across the 3 groups. Obviously, a case mix of different OSA phenotypes must be included in the validation process of PROMs for OSA. If not, the instrument may predispose to assessing the characteristics of only a certain subgroup. Particularly, subjects who participate in PROM research may belong to subclasses that are not representative of the entire target population.[59] As phenotypical heterogeneity of OSA has only recently been demonstrated, and postdates the publication of legacy OSA questionnaires, it is presumed that all these PROMs may suffer from selection bias to some extent. HRQoL assessment with the FOSQ, for example, only assesses effects of fatigue or being sleepy and does not include effects of disturbed nocturnal sleep.

FUTURE DIRECTIONS: THE NEED FOR NEW PATIENT-REPORTED OUTCOME MEASURES

Can we reliably and beneficially use the existing PROMs in clinical and investigative sleep

medicine? The answer is positive, if the user is sufficiently aware of the scope, strengths, and limitations of the different available instruments. Yet, the field lacks an easy-to-use tool—like a clinical thermometer—appealing to both patients and practitioners.[60]

Many sleep centers use a collection of different questionnaires, such as PSQI, ISI, ESS, FOSQ, and so forth, yielding excessive, redundant, and sometimes conflicting information, thus burdening doctors and patients. To overcome this exorbitance, several approaches may be envisaged. The first one is to pool items of existing PROMs that are already validated by patient input. Rather than rely on composite scores of the different domains in separate PROMs, the individual questions of the PROMs might be more suitable for alerting a health care professional to the most important problems of an individual patient.[59]

Another method may consist of extracting distinct traits from disease domains that have proven relevant and to disengage them from conventional—yet still putative—disease models. This way, the constellation of symptoms related to disturbed sleep and daytime dysfunction could be reduced to a minimal set of essential features, for example, insomnia, sleepiness, fatigue, bodily discomfort, and so forth. For each distinct feature, a degree of severity and impact on HRQoL can be assessed. Moreover, by making the PROM free of hypothesis as to a tentative medical diagnosis, preconceptions regarding causality—which is inherent to most disease-specific questionnaires—can be obviated. The expected elimination of bias together with opportunities for multipurpose utility would justify the development of a completely new sleep questionnaire.

When symptoms are nonspecific, a priori coupling with diagnostic outcomes may be speculative. In such conditions, a reference benchmark is required to assure certainty about causation. Although PSG may disclose certain pathophysiological markers, it is often uncertain whether pathophysiology and clinical symptoms are causally linked. Thus, PSG may fall short of providing the required benchmark. Therefore, attribution of causality remains elusive in many patients with sleep complaints. Favorable symptomatic response to treatment, for example, therapy with positive airway pressure for OSA, provides additional evidence regarding the relationship between the presenting symptoms and the purported sleep disorder. Diagnostic therapy is a means not only to assess the degree of symptomatic relief but also to suggest causality.[11] It has been emphasized that PROMs should be sufficiently sensitive to detect treatment-induced changes over time. Because the observed changes may support diagnostic evidence as well, responsiveness inherently reflects disease-specificity.

Finally, PROMs may become outdated as their content usually remains unchanged, whereas medical concepts and treatments will advance over time. To overcome static inertia, dynamic solutions for obtaining patient-reported outcomes have been developed. The Patient-Reported Outcome Measurement Information System (PROMIS) was established in 2004 with funding from the US National Institutes of Health (NIH).[70] In this configuration, patient-reported outcomes related to different diseases are collected and stored in item banks. These databases include large sets of single questions that comprehensively cover various symptom domains. The collection of items is accessible for computer-adaptive test systems that dynamically compose a (variable) set of patient-reported outcomes depending on the patient's characteristics and on the answers given to preceding questions. The aim is to introduce targeted approaches for capturing relevant patient-centric information, whilst reducing the respondent burden. PROMIS sleep disturbance and sleep-related impairments item banks have been created for assessing sleep disorders.[71] Excellent measurement properties were attributed to this PROM, which was considered useful for probing general aspects of sleep and sleep-related impairments in various groups of patients. This development holds promise for creating future patient-centered assessment instruments in the field of sleep medicine.

SUMMARY

As sleep disorders are highly prevalent, many patients seek appropriate medical help for their sleep problems. Although the medical interview is essential for establishing a diagnostic working hypothesis, questionnaires are valuable add-on tools with respect to clinical subtyping, differential diagnosis, identification of comorbidities, and assessing response to treatment. The term "patient-reported outcome measures" (PROMs) is standard for questionnaires that are validated along a spectrum of different measurement properties. PROMs must comply with rigorous psychometric standards and should be evaluated carefully before release. Sleep disorders can be assessed with generic or disease-specific questionnaires. The latter category comprises PROMs for specific sleep disorders such as OSA, insomnia, and RLS, among others. There are certain limitations to the use of PROMs. The composing traits are often nonspecific and overlap among different nosologic

entities. Moreover, PROMs may not capture the full spectrum of disease heterogeneity. Therefore, inappropriate use may yield spurious diagnoses and ineffective treatment. Besides the use of disease-specific instruments, the field of sleep medicine may envisage the introduction of domain-specific questionnaires—free from diagnostic preconceptions—targeting traits that are unique to the patient's condition. This observation may open up perspectives for innovative research on still better PROMs.

CLINICS CARE POINTS

- History taking in patients with sleep disorders can be improved by using self-administered PROMs
- The selection of PROMs should comply with the tentative diagnosis obtained from the medical interview
- PROMs that have an optimal balance between amount of information versus respondent burden are to be preferred
- PROMs are a very important instrument to systematically assess effects of treatment
- Inadvertent use of PROMs is discouraged as such approach inevitably produces spurious diagnoses and inadequate treatment

ACKNOWLEDGMENTS

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no 965417 Sleep Revolution. The authors thank Prof. Dr. Erna Sif Arnardottir for reviewing the article and for providing pertinent suggestions for improvement.

DISCLOSURE

The authors have nothing to disclose.

REFERENCES

1. Fatima Y, Bucks RS, Mamun AA, et al. Sleep trajectories and mediators of poor sleep: findings from the longitudinal analysis of 41,094 participants of the UK Biobank cohort. Sleep Med 2020;76:120–7.
2. AASM. The international classification of sleep disorders. In: Sateia M, editor. Third edition. Darien (IL): American Academy of Sleep Medicine; 2014.
3. Colten HR, Altevogt BM. Sleep disorders and sleep deprivation: an unmet public health problem. Washinton (DC): National Institutes of Health; 2006.
4. Kapur V, Strohl KP, Redline S, et al. Underdiagnosis of sleep apnea syndrome in U.S. communities. Sleep Breath 2002;6(2):49–54.
5. Young T, Shahar E, Nieto FJ, et al. Predictors of sleep-disordered breathing in community-dwelling adults: the sleep heart health study. Arch Intern Med 2002;162(8):893–900.
6. Mai E, Buysse DJ. Insomnia: prevalence, impact, pathogenesis, differential diagnosis, and evaluation. Sleep Med Clin 2008;3(2):167–74.
7. Johnson DA, Billings ME, Hale L. Environmental determinants of insufficient sleep and sleep disorders: implications for population health. Curr Epidemiol Rep 2018;5(2):61–9.
8. Rosen RC, Zozula R, Jahn EG, et al. Low rates of recognition of sleep disorders in primary care: comparison of a community-based versus clinical academic setting. Sleep Med 2001;2(1):47–55.
9. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline. J Clin Sleep Med 2017;13(3):479–504.
10. Berry RB, Brooks R, Gamaldo CE, et al. The AASM manual for the scoring of sleep and associated events. Rules, terminology and technical specifications. Version 2.0. Darien (IL): American Academy of Sleep Medicine; 2012.
11. Pevernagie DA, Gnidovec-Strazisar B, Grote L, et al. On the rise and fall of the apnea-hypopnea index: a historical review and critical appraisal. J Sleep Res 2020;29(4):e13066.
12. Sanner BM, Klewer J, Trumm A, et al. Long-term treatment with continuous positive airway pressure improves quality of life in obstructive sleep apnoea syndrome. Eur Respir J 2000;16(1):118–22.
13. Weaver EM, Woodson BT, Steward DL. Polysomnography indexes are discordant with quality of life, symptoms, and reaction times in sleep apnea patients. Otolaryngol Head Neck Surg 2005;132(2):255–62.
14. Masa JF, Jimenez A, Duran J, et al. Visual analogical well-being scale for sleep apnea patients: validity and responsiveness : a test for clinical practice. Sleep Breath 2011;15(3):549–59.
15. Foster SN, Hansen SL, Scalzitti NJ, et al. Residual excessive daytime sleepiness in patients with obstructive sleep apnea treated with positive airway pressure therapy. Sleep Breath 2020;24(1):143–50.
16. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ 2002;324(7351):1417.
17. ICHOM. Standard sets - why measure outcome?. International consortium for health outcomes measurement 2020. Available at: https://www.ichom.org/standard-sets/#about-standard-sets.

18. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Control Clin Trials 2004;25(6):535–52.
19. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. Fundam Clin Pharmacol 2004;18(3):351–63.
20. Klingman KJ, Jungquist CR, Perlis ML. Questionnaires that screen for multiple sleep disorders. Sleep Med Rev 2017;32:37–44.
21. Nagappa M, Liao P, Wong J, et al. Validation of the STOP-bang questionnaire as a screening tool for obstructive sleep apnea among different populations: a systematic review and meta-analysis. PLoS One 2015;10(12):e0143697.
22. Prasad KT, Sehgal IS, Agarwal R, et al. Assessing the likelihood of obstructive sleep apnea: a comparison of nine screening questionnaires. Sleep Breath 2017;21(4):909–17.
23. Greenhalgh J, Dalkin S, Gooding K, et al. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Health Serv Deliv Res 2017;5(2).
24. Di Paolo A, Sarkozy F, Ryll B, et al. Personalized medicine in Europe: not yet personal enough? BMC Health Serv Res 2017;17(1):289.
25. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–9.
26. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. JAMA 1995;273(1):59–65.
27. Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 2003;6(5):522–31.
28. Bakas T, McLennon SM, Carpenter JS, et al. Systematic review of health-related quality of life models. Health Qual Life Outcomes 2012;10:134.
29. Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013;346:f167.
30. Porter ME, Lee TH. The big idea. The strategy that will fix health care. Harv Business Rev 2013; 91(10):50–70.
31. Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, et al. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. Milbank Q 2014;92(4):754–75.
32. Streiner DN. Health measurement scales: a practical guide to their development and use. 4th edition. Oxford, UK: Oxford University Press; 1995.
33. FDA. Patient-reported outcome measures: use in medical product development to support labeling Claims. Silver Spring, Maryland, USA: Food and Drug Administration; 2009.
34. de Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in medicine - a practical guide. New York, USA: Cambridge University Press; 2011.
35. Verkerk E, Verbiest M, van Dulmen S, et al. PROM-cycle (summary in English). National Health Care Institute of The Netherlands; 2018. https://www.zorginzicht.nl/ontwikkeltools/prom-toolbox/prom-cycle-summary-in-english.
36. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21(4):651–7.
37. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013;22(8):1889–905.
38. de Vet HCWT CB, Bouter LM. Clinimetrics and psychometrics: two sides of the same coin. J Clin Epidemiol 2003;56:1146–7.
39. Abma IL, Rovers M, IJff M, et al. Instrument completion and validation of the patient-reported apnea questionnaire (PRAQ). Health Qual Life Outcomes 2018;16(1):158.
40. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30(6):473–83.
41. EuroQol_Group. EuroQol–a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199–208.
42. Abma IL, van der Wees PJ, Veer V, et al. Measurement properties of patient-reported outcome measures (PROMs) in adults with obstructive sleep apnea (OSA): a systematic review. Sleep Med Rev 2016;28:18–31.
43. Buysse DJ, Reynolds CF 3rd, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193–213.
44. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14(6):540–5.
45. Osorio CD, Gallinaro AL, Lorenzi-Filho G, et al. Sleep quality in patients with fibromyalgia using the Pittsburgh sleep quality index. J Rheumatol 2006;33(9): 1863–5.
46. Mollayeva T, Thurairajah P, Burton K, et al. The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev 2016;25:52–73.
47. Johns MW. Reliability and factor analysis of the Epworth sleepiness scale. Sleep 1992;15(4):376–81.

48. Kendzerska TB, Smith PM, Brignardello-Petersen R, et al. Evaluation of the measurement properties of the Epworth sleepiness scale: a systematic review. Sleep Med Rev 2014;18(4):321–31.
49. Taylor E, Zeng I, O'Dochartaigh C. The reliability of the Epworth Sleepiness Score in a sleep clinic population. J Sleep Res 2019;28(2):e12687.
50. Omobomi O, Quan SF. A requiem for the clinical use of the Epworth sleepiness scale. J Clin Sleep Med 2018;14(5):711–2.
51. Baldwin CM, Kapur VK, Holberg CJ, et al. Associations between gender and measures of daytime somnolence in the sleep heart health study. Sleep 2004;27(2):305–11.
52. Arnardottir ES, Islind AS, Oskarsdottir M. The future of sleep measurements - a review and perspective. Sleep Med Clin 2021;16(3):447–64.
53. Billings ME, Rosen CL, Auckley D, et al. Psychometric performance and responsiveness of the functional outcomes of sleep questionnaire and sleep apnea quality of life instrument in a randomized trial: the HomePAP study. Sleep 2014;37(12):2017–24.
54. Weaver TE, Laizner AM, Evans LK, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep 1997;20(10):835–43.
55. Piccirillo JF, Gates GA, White DL, et al. Obstructive sleep apnea treatment outcomes pilot study. Otolaryngol Head Neck Surg 1998;118(6):833–44.
56. Flemons WW, Reimer MA. Development of a disease-specific health-related quality of life questionnaire for sleep apnea. Am J Respir Crit Care Med 1998;158(2):494–503.
57. Lacasse Y, Bureau MP, Series F. A new standardised and self-administered quality of life questionnaire specific to obstructive sleep apnoea. Thorax 2004;59(6):494–9.
58. Sheats RD. Health-related quality of life assessment tools and sleep-disordered breathing. J Dental Sleep Med 2016;3(2):49–55.
59. Abma IL, Rovers M, IJff M, et al. The development of a patient-reported outcome measure for patients with obstructive sleep apnea: the Patient-Reported Apnea Questionnaire (PRAQ). J Patient Rep Outcomes 2017;1(1):14.
60. Abma IL, Rovers MM, IJff M, et al. Does the Patient-Reported Apnea Questionnaire (PRAQ) increase patient-centredness in the daily practice of sleep centres? a mixed-methods study. BMJ Open 2019;9(6):e025963.
61. van Straten A, van der Zweerde T, Kleiboer A, et al. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. Sleep Med Rev 2018;38:3–16.
62. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297–307.
63. Morin CM, Belleville G, Belanger L, et al. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34(5):601–8.
64. Garcia-Borreguero D, Ferini-Strambi L, Kohnen R, et al. European guidelines on management of restless legs syndrome: report of a joint task force by the European Federation of Neurological Societies, the European Neurological Society and the European Sleep Research Society. Eur J Neurol 2012;19(11):1385–96.
65. Abetz L, Vallow SM, Kirsch J, et al. Validation of the restless legs syndrome quality of life questionnaire. Value Health 2005;8(2):157–67.
66. Walters AS, LeBrocq C, Dhar A, et al. Validation of the international restless legs syndrome study group rating scale for restless legs syndrome. Sleep Med 2003;4(2):121–32.
67. Wolpert M. Uses and abuses of patient reported outcome measures (PROMs): potential iatrogenic impact of PROMs implementation and how it can be mitigated. Adm Policy Ment Health 2014;41(2):141–5.
68. Janssen HCJP, Venekamp LN, Peeters GAM, et al. Management of insomnia in sleep disordered breathing. Eur Respir Rev 2019;28(153):190080.
69. Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. Sleep 2018;41(3):zsx214.
70. Smith AB, Hanbury A, Retzler J. Item banking and computer-adaptive testing in clinical trials: standing in sight of the PROMISed land. Contemp Clin Trials Commun 2019;13:005.
71. Buysse DJ, Yu L, Moul DE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010;33(6):781–92.