International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

Sleep Measurement Tools: A Narrative Review of Recent Advances

Dr. Darshana Nariya

Department of Physiotherapy, SPB Physiotherapy College, Veer Narmad South Gujarat University, Surat, India.
Email: *darshananariya072[at]gmail.com*

Abstract: ***Background**: Accurate sleep measurement underpins clinical diagnosis, research, and consumer health tracking. Over the past decade the field has expanded beyond laboratory polysomnography (PSG) to include actigraphy, consumer wearables, contactless sensors, smartphone applications, and AI-driven signal analysis. **Objective**: This narrative review synthesizes recent advances (2020-2026) in sleep measurement tools, evaluates their comparative validity, and outlines future directions. **Methods**: A structured, non-systematic search of PubMed and Embase used MeSH terms and free-text keywords covering polysomnography, actigraphy, wearable and contactless sensors, artificial intelligence, and sleep questionnaires, restricted to English-language articles published between January 2019 and July 2026; reference lists of key reviews were hand-searched. **Results**: PSG remains the diagnostic reference standard but is limited by cost, access, and the first-night effect. Actigraphy reliably estimates total sleep time and efficiency but performs poorly at multi-stage classification. Consumer wrist wearables and smart rings combining accelerometry, photoplethysmography, and skin-temperature sensing approach 90-95% agreement with PSG for sleep-wake detection, though wake specificity remains variable. Contactless radar and under-mattress sensors permit unobtrusive, longitudinal cardiorespiratory monitoring. Deep-learning and foundation models increasingly automate PSG scoring and extract prognostic information. Subjective instruments such as the Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale remain valuable for perceived sleep quality but show limited criterion validity against objective sleep disorders. **Conclusions**: No single tool is universally superior; each occupies a niche defined by trade-offs among accuracy, scalability, cost, and patient burden. Rigorous, standardised validation across diverse populations remains the principal need as the field moves toward AI-augmented, multimodal, home-based sleep assessment.*

Keywords: polysomnography; actigraphy; wearable electronic devices; sleep apnea syndromes; artificial intelligence; sleep quality

1. Introduction

Sleep is a biologically essential process, and its objective and subjective measurement is central to both clinical sleep medicine and population health research. Historically, the gold standard for characterizing sleep architecture has been laboratory-based polysomnography (PSG), a multi-channel recording of brain, cardiac, ocular, and muscular activity, alongside respiratory and oximetric signals, scored according to standardised criteria [1,2]. However, PSG is resource-intensive, poorly scalable, and susceptible to the well-documented "first-night effect", in which an unfamiliar recording environment alters habitual sleep patterns [3].

These limitations have driven rapid innovation in alternative sleep-measurement tools over the past five years. Actigraphy, wrist-worn and ring-form consumer wearables, contactless radar and under-mattress systems, smartphone-based acoustic sensing, and artificial-intelligence (AI)-enabled signal-processing pipelines have all matured considerably, in parallel with a growing evidence base validating (and, in some cases, questioning) their accuracy against PSG [4]. Multimodal deep-learning "foundation models" trained on hundreds of thousands of hours of PSG data have recently demonstrated that sleep signals encode prognostic information relevant to numerous non-sleep diseases, further expanding the clinical relevance of high-quality sleep measurement [4].

This narrative review aims to (1) summaries the principles, recent technical advances, and validation evidence for the major categories of sleep measurement tools; (2) present a structured, reproducible literature search strategy across PubMed and Embase; (3) compare the performance characteristics of these tools using tabulated and graphical data; and (4) discuss the limitations, unmet needs, and likely future trajectory of the field.

2. Materials and Methods

2.1 Literature Search Strategy

This is a narrative review; nonetheless, a structured and transparent search strategy was used to identify relevant literature, consistent with recommended practice for high-quality narrative reviews. Searches were conducted in PubMed (MEDLINE) and Embase (Ovid interface) for articles published from 1 January 2019 to 18 July 2026, restricted to English-language, peer-reviewed publications (original research, systematic reviews, meta-analyses, and clinical practice guidelines/position statements). Conference abstracts without full text and non-peer-reviewed preprints were excluded unless no peer-reviewed equivalent was available.

2.2 PubMed Search Strategy

Line	Search terms (PubMed / MEDLINE)
1	"Polysomnography"[MeSH] OR "polysomnograph*"[tiab] OR "PSG"[tiab]
2	"Actigraphy"[MeSH] OR "actigraph*"[tiab] OR "accelerometer*"[tiab] OR "wrist-worn"[tiab]
3	"Wearable Electronic Devices"[MeSH] OR "wearable*"[tiab] OR "smartwatch*"[tiab] OR "smart ring*"[tiab] OR "fitness tracker*"[tiab] OR "Fitbit"[tiab] OR "Oura"[tiab] OR "Apple Watch"[tiab] OR "Whoop"[tiab]
4	"contactless"[tiab] OR "non-contact"[tiab] OR "radar"[tiab] OR "ballistocardiograph*"[tiab] OR "under-mattress"[tiab] OR "bed sensor*"[tiab]
5	"mobile applications"[MeSH] OR "smartphone*"[tiab] OR "mobile app*"[tiab] OR "acoustic sensing"[tiab] OR "snoring"[tiab]
6	"Artificial Intelligence"[MeSH] OR "Deep Learning"[MeSH] OR "machine learning"[tiab] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR "neural network*"[tiab] OR "foundation model*"[tiab]
7	"Sleep Apnea Syndromes/diagnosis"[MeSH] OR "Home sleep apnea test*"[tiab] OR "HSAT"[tiab] OR "polygraphy"[tiab]
8	"Pittsburgh Sleep Quality Index"[tiab] OR "Epworth Sleepiness Scale"[tiab] OR "sleep questionnaire*"[tiab] OR "sleep diary*"[tiab]
9	2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8
10	"Sleep"[MeSH] OR "sleep"[tiab] OR "sleep stag*"[tiab] OR "sleep architecture"[tiab] OR "sleep quality"[tiab] OR "sleep-wake"[tiab]
11	"validation stud*"[tiab] OR "accuracy"[tiab] OR "sensitivity and specificity"[MeSH] OR "reliability"[tiab] OR "agreement"[tiab] OR "concordance"[tiab]
12	1 AND 9 AND 10 AND 11
13	12 AND (English[Language]) AND ("2019/01/01"[Date - Publication] : "2026/07/18"[Date - Publication])
14	13 NOT ("Comment"[Publication Type] OR "Letter"[Publication Type] OR "Editorial"[Publication Type])

2.3 Embase Search Strategy

Line	Search terms (Embase, Ovid interface)
1	exp polysomnography/ OR (polysomnograph* OR PSG).mp.
2	exp actigraphy/ OR (actigraph* OR accelerometer* OR "wrist-worn").mp.
3	exp wearable sensor/ OR exp wearable computer/ OR (wearable* OR smartwatch* OR "smart ring*" OR "fitness tracker*" OR Fitbit OR Oura OR "Apple Watch" OR Whoop).mp.
4	exp contactless sensor/ OR (contactless OR "non-contact" OR radar OR ballistocardiograph* OR "under-mattress" OR "bed sensor*").mp.
5	exp smartphone/ OR exp mobile application/ OR (smartphone* OR "mobile app*" OR "acoustic sensing" OR snoring).mp.
6	exp artificial intelligence/ OR exp deep learning/ OR exp machine learning/ OR ("artificial intelligence" OR "deep learning" OR "machine learning" OR "neural network*" OR "foundation model*").mp.
7	exp home sleep apnea test/ OR exp sleep disordered breathing/de OR ("home sleep apnea test*" OR HSAT OR polygraphy).mp.
8	exp Pittsburgh Sleep Quality Index/ OR exp Epworth Sleepiness Scale/ OR ("sleep questionnaire*" OR "sleep diary*").mp.
9	#2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10	exp sleep/ OR (sleep OR "sleep stag*" OR "sleep architecture" OR "sleep quality" OR "sleep-wake").mp.
11	exp validation study/ OR exp diagnostic accuracy/ OR (accuracy OR reliability OR agreement OR concordance OR "sensitivity and specificity").mp.
12	1 AND 9 AND 10 AND 11
13	12 AND (english language and ("2019" to "2026").yr.)
14	13 NOT (letter OR editorial OR conference abstract).pt.

2.4 Selection Process

Titles and abstracts identified through the combined PubMed and Embase searches, together with hand-searched reference lists, were screened for relevance by the review author. Priority was given to systematic reviews, meta-analyses, large validation studies ($n \geq 30$ participants with concurrent PSG or actigraphy comparison), and current professional-society guidelines. Studies were excluded if they addressed sleep constructs unrelated to measurement technology (e.g., pharmacological trials) or reported only device engineering without any human validation data. Duplicate records across databases were removed manually. This narrative approach does not follow a formal PRISMA protocol and was not registered; it is intended to provide a representative, expert-guided synthesis rather than an exhaustive systematic inventory.

3. Result

3.1 Polysomnography: The Reference Standard

Type-1, laboratory-attended PSG remains the diagnostic reference standard for most sleep disorders, combining electroencephalography (EEG), electro-oculography (EOG), submental electromyography (EMG), electrocardiography, airflow, respiratory effort, and pulse oximetry channels, scored in accordance with American Academy of Sleep Medicine (AASM) criteria[1,2]. This comprehensive montage permits classification of wake, N1, N2, N3, and rapid-eye-movement (REM) sleep in 30-second epochs and detection of respiratory, limb-movement, and arousal events[2]. A 2024 position statement from the Brazilian Sleep Association reaffirmed that only studies using this full electrode montage merit the designation "polysomnography," as distinct from simplified respiratory polygraphy[3].

Despite its diagnostic authority, PSG has well-recognised drawbacks: high per-study cost, limited laboratory capacity, the burden of technician-applied electrodes, and alteration of sleep due to an unfamiliar environment and monitoring apparatus (the first-night effect), which a recent study found

produced systematic discrepancies between home and in-laboratory subjective sleep-quality ratings[5]. These constraints have motivated the search for scalable alternatives to full PSG, as well as growing interest in extracting more value from PSG data itself: a 2026 multimodal "sleep foundation model" (SleepFM), trained via contrastive learning on more than 585,000 hours of PSG recordings from approximately 65,000 participants, demonstrated that latent representations of raw PSG signals can predict risk for 130 medical conditions - including all-cause mortality, dementia, and cardiovascular disease - from a single night of recording, while also performing competitively with specialised automated sleep-staging algorithms[4]. This illustrates a broader trend: PSG is evolving from a purely diagnostic tool toward a rich biomarker platform when paired with modern computational methods.

3.2 Actigraphy

Actigraphy estimates sleep-wake states from wrist- or ankle-worn accelerometer-derived movement counts, applying validated scoring algorithms (e.g., Cole-Kripke, Sadeh) to classify epochs as sleep or wake. Its principal advantages are low cost, minimal participant burden, and the capacity for multi-night, ecologically valid recording in the home environment, which has made it a mainstay of sleep epidemiology and circadian research for over three decades.
A comprehensive 2024 systematic review registered with PROSPERO and searching MEDLINE, Web of Science, Google Scholar, and Embase concluded that actigraphy provides reasonable estimates of two-stage (sleep/wake) classification, achieving moderate agreement with PSG for total sleep time and sleep efficiency, but that its ability to discriminate light, deep, and REM sleep stages was limited and inconsistently reported across the available literature[6]. A complementary scoping review that assessed 35 articles and 62 device-algorithm combinations similarly concluded that pure accelerometer-based devices are adequate for binary sleep/wake detection but fall short for multi-stage classification, whereas devices that combine accelerometry with photoplethysmography (PPG) perform substantially better at stage discrimination[7].

In paediatric populations, actigraphy has particular value because it can be worn over multiple nights in the home without disrupting a child's typical sleep schedule; early narrative and systematic reviews established generally good sensitivity (0.82-0.96) but persistently poor specificity for detecting wakefulness after sleep onset, a limitation attributed to children's characteristic combination of high movement during light sleep and prolonged immobile wakeful periods[8,9]. The AASM's 2025 clinical update on actigraphy technology reiterates that device fit, algorithm selection, and population-specific validation (particularly for paediatric, elderly, and overweight/underweight individuals) remain critical considerations for appropriate clinical and research use[10].

3.3 Consumer Wearable Devices: Wrist-Worn Trackers and Smart Rings

The past five years have seen substantial improvement in the sleep-tracking performance of consumer multisensor wearables, which typically combine three-axis accelerometry with PPG-derived heart rate and heart-rate-variability signals, and, increasingly, skin-temperature and blood-oxygen sensing. A 2025 head-to-head validation study compared six popular wrist-worn devices (Fitbit Charge 5, Fitbit Sense, Withings ScanWatch, Garmin Vivosmart 4, Whoop 4.0, and Apple Watch Series 8) against PSG in 62 adults across a single overnight laboratory stay, providing an updated, comparative benchmark of contemporary consumer device accuracy[11]. A 2025 meta-analysis pooling data across multiple consumer wrist-worn devices similarly concluded that, while sleep/wake detection has become highly reliable, specificity for detecting wakefulness remains the most persistent weakness across device brands[12].

Smart rings have emerged as a particularly promising wearable form factor. In a large validation study of 96 participants and 421,045 epochs of multi-night ambulatory PSG, the Oura Ring Generation 3 with its second-generation sleep-staging algorithm (OSSA 2.0) achieved 91.7-91.8% overall accuracy, 94.4-94.5% sensitivity, and 73.0-74.6% specificity for sleep-wake discrimination, with stage-specific accuracy ranging from 75.5% (light sleep) to 90.6% (REM sleep)[13]. An independent validation from the University of Tokyo and a separate study from Brigham and Women's Hospital corroborated that Oura's staging algorithm performs favourably relative to PSG and outperforms several competing wrist-worn devices in four-stage classification tasks[14], building on earlier multi-sensor architecture work demonstrating that combining accelerometer, autonomic nervous system-derived, and circadian features substantially improves four-stage classification accuracy (from 57% using accelerometer data alone to 79% with the full multi-sensor model)[15]. A single-night inpatient comparison of the Oura Ring Gen3, Fitbit Sense 2, and Apple Watch Series 8 against PSG in 35 adults found that all three devices exceeded 95% sensitivity for sleep detection, with stage-discrimination sensitivity ranging from 50% to 86% depending on device and sleep stage, and that the Oura Ring's estimates for wake, light, deep, and REM sleep did not differ significantly from PSG[16].

3.4 Comparative Summary of Sleep Measurement Tools

Table 3 summarises the underlying principle, typical setting, sleep-staging capability, and representative validation evidence for each major category of sleep measurement tool discussed in this review.

Table 3: Comparative overview of major sleep measurement modalities.

Measurement tool	Principle	Typical setting	Sleep staging capability	Representative evidence
Polysomnography (PSG)	Multi-channel EEG/EOG/EMG/ECG + respiratory/oximetry	Sleep laboratory (Type 1) or attended home (Type 2)	Full 5-stage (W/N1/N2/N3/REM), reference standard	Berry et al. 2012 [2]; Palombini et al. 2024 [3]
Actigraphy	Wrist/ankle accelerometry with sleep/wake algorithms	Home, multi-night, ambulatory	Sleep/wake reliable; multi-stage limited	Yuan et al. 2024 [6]; Meltzer et al. 2012 [8]
Consumer wrist wearables	Accelerometer + PPG (± SpO_2, temperature)	Home, consumer use	4-stage possible; specificity for wake variable	Schyvens et al. 2025 [11]; Lee et al. 2025 [12]
Smart rings (e.g., Oura Gen3)	Finger PPG + accelerometer + temperature	Home, consumer/research use	4-stage; among best-performing consumer devices	Svensson et al. 2024 [13]; de Zambotti et al. 2024 [14]
Radar-based contactless sensors	Doppler radar detects chest-wall micro-movement	Home or laboratory, no device worn	Primarily respiratory events/AHI; limited staging	Gross-Isselmann et al. 2024 [17]
Under-mattress/ ballistographic sensors	Piezoelectric/pneumatic/fibre-optic bed sensors	Home, passive/unobtrusive	Cardiorespiratory + coarse sleep/wake; staging emerging	Boiko et al. 2023 [18]; Ravindran et al. 2024 [20]
Smartphone acoustic apps	Microphone-based acoustic event/snore detection	Home, consumer use	No formal staging; screening for snoring/SDB only	Brown et al. 2025 [22]
Home sleep apnea test (HSAT)	Simplified respiratory polygraphy (airflow, effort, SpO_2)	Home, physician-directed	Not designed for staging; AHI/respiratory focus	Kapur et al. 2017 [24]; Rosen et al. 2025 [23]
Subjective questionnaires (PSQI, ESS)	Self-report of perceived sleep quality/sleepiness	Any setting, minimal burden	Not applicable (subjective, not staged)	Buysse et al. 1989 [28]; Nishiyama et al. 2014 [30]

3.5 Sensitivity and Specificity Across Modalities

Figure 1 illustrates representative sensitivity for sleep detection and specificity for wake detection reported across five broad categories of sleep measurement technology, drawn from the validation studies reviewed above. Wake-detection specificity is consistently the weaker performance metric across nearly all non-PSG technologies, reflecting a shared tendency of movement- and photoplethysmography-based systems to misclassify quiet wakefulness as sleep.

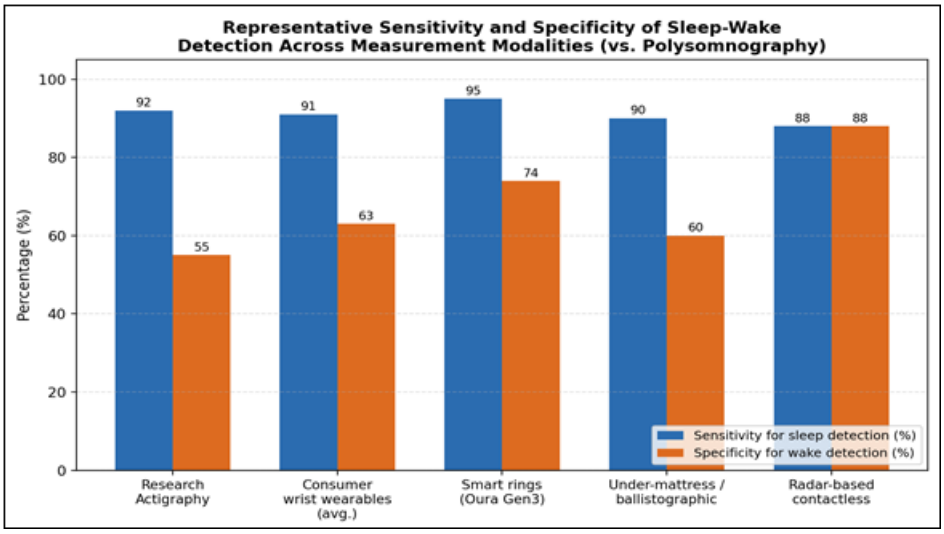

Figure 1: Representative sensitivity (sleep detection) and specificity (wake detection) of major sleep measurement modalities relative to polysomnography, synthesised from cited validation studies [6,11-13,17,20].

3.6 Four-Stage Sleep Classification Accuracy

Figure 2 compares reported accuracy for four-stage sleep classification (wake, light, deep, and REM sleep, relative to PSG) across representative device categories. Devices that combine multiple physiological signals (accelerometry, PPG, and, in some cases, skin temperature) consistently outperform single-sensor accelerometer-only actigraphy for this more demanding classification task.

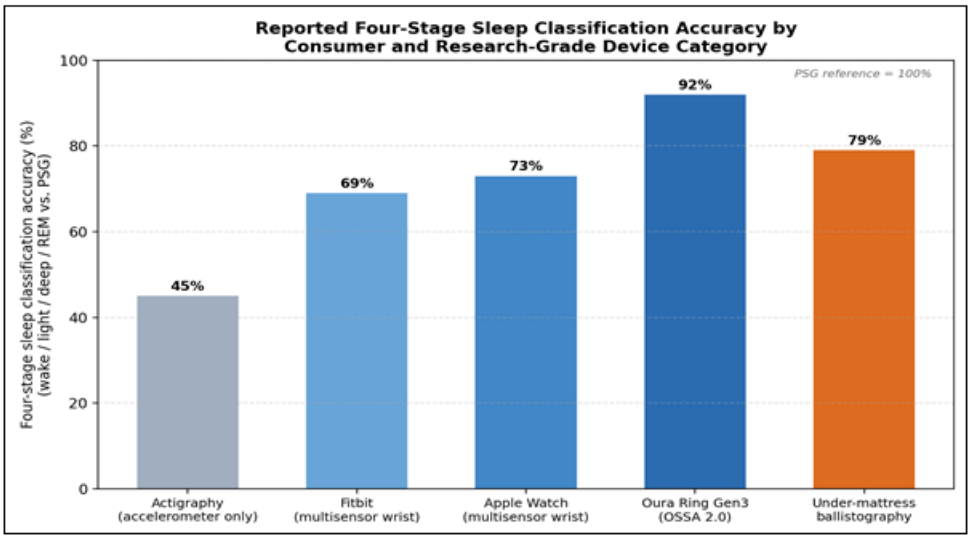

Figure 2: Reported four-stage sleep classification accuracy (vs. PSG) by device category, synthesised from cited validation studies [6,13,16,20].

3.7 Contactless and Non-Contact Sensing Technologies

Contactless technologies aim to eliminate the burden of wearing a device altogether. Radar-based systems emit low-power electromagnetic waves and analyse the Doppler-shifted reflections caused by chest-wall movement to derive respiratory and cardiac signals; a 2024 validation of the Sleepiz One+ radar device in 141 adults found a strong correlation with PSG-derived apnea-hypopnea index (r = 0.94 with supplementary pulse-oximetry data, r = 0.87 without), with sensitivity and specificity for sleep apnea detection reaching 88% and 98%, respectively, when combined with SpO_2 monitoring[17].

Under-mattress and bed-integrated systems use piezoelectric films, pneumatic sensors, or fibre-optic elements to capture ballistocardiographic signals - the micro-vibrations produced by cardiac ejection and respiration transmitted through the mattress. A comprehensive 2023 systematic review of contactless cardiorespiratory monitoring technologies catalogued radar, camera-based, and mattress-embedded approaches, concluding that while radar and fibre-optic/piezoelectric systems can achieve high accuracy for heart rate and respiratory rate, performance is sensitive to sensor placement, distance from the body, and susceptibility to motion artefact[18]. A related review of contact-free sensors as digital cardiovascular biomarkers similarly concluded that, although machine-learning approaches to under-mattress ballistocardiography show promising results for sleep-stage classification, EEG-based PSG remains the unchallenged reference standard for detailed sleep architecture assessment[19].

A digital health technology evaluation study comparing two under-mattress trackers (Withings Sleep Analyser and Emfit QS) and a bedside Doppler radar (Somnofy) against PSG and research-grade actigraphy in an ageing population found that contactless devices could reliably estimate heart rate and breathing rate in both laboratory and home settings, supporting their potential for continuous, low-burden monitoring of vulnerable populations such as older adults and people living with dementia[20]. Because these systems require no device to be worn or applied, they are particularly attractive for longitudinal home monitoring, paediatric and geriatric populations, and settings in which adherence to wearable devices is a concern.

3.8 Smartphone Applications and Acoustic Sensing

Smartphones equipped with microphones and accelerometers offer a near-zero-cost platform for passive sleep and snoring surveillance. Advances in flexible and wearable acoustic sensors, combined with machine-learning-based signal processing, have broadened the scope of acoustic health monitoring generally, including sleep-related applications such as snore and respiratory-event detection[21]. A 2025 validation study of the SleepWatch smartphone application's snore-detection algorithm, trained on more than 60,000 acoustic events, reported an average sensitivity of 86.3%, specificity of 99.5%, and overall accuracy of 95.2% for detecting simulated snoring events under controlled acoustic conditions, alongside a strong correlation (r = 0.97) between detected and true snore rates[22]. Other commercially available snoring applications have reported comparably high specificity (typically 93-97%) but more variable sensitivity, particularly at higher snoring frequencies, underscoring that acoustic sensing is best suited as a screening or adjunctive tool rather than a diagnostic replacement for formal sleep-disordered-breathing evaluation.

3.9 Artificial Intelligence and Deep Learning in Sleep Signal Analysis

Artificial intelligence, and deep learning in particular, has become integral to nearly every category of sleep measurement discussed above - from automating PSG scoring, to powering proprietary wearable sleep-staging algorithms, to enabling contactless signal extraction from noisy radar or ballistocardiographic data. A scoping review of 35 studies and 62 wearable-algorithm combinations found a clear evolution toward combining accelerometer and PPG data using increasingly sophisticated machine-learning classifiers, and proposed five recommendations to improve reliability reporting, including standardised algorithm validation across diverse populations and open-source data/code availability[7].

At the most ambitious end of this trend, multimodal foundation models such as SleepFM demonstrate that self-supervised, contrastively trained deep-learning architectures can learn generalisable representations of raw physiological sleep signals that transfer across cohorts and tasks, achieving competitive performance against specialised sleep-staging models (mean F_1 scores of 0.70-0.78) while simultaneously enabling long-horizon disease-risk prediction from routine PSG data[4]. This dual capability - accurate automated scoring plus disease-risk stratification from a single recording - represents a meaningful shift in how PSG data may be used clinically, moving from a narrowly diagnostic report toward a broader physiological biomarker. However, the developers of such models themselves note that the interpretability of these representations (i.e., precisely which signal features drive a given prediction) remains an open research question, and that external validation across more diverse, real-world populations is still needed before clinical deployment.

3.10 Subjective Sleep Assessment Tools: Questionnaires and Scales

Subjective instruments remain indispensable complements to objective measurement, capturing the patient's perceived sleep experience, which does not always correspond to objectively recorded parameters. The Pittsburgh Sleep Quality Index (PSQI), a 19-item self-report instrument yielding seven component scores (subjective sleep quality, latency, duration, habitual efficiency, disturbances, use of sleep medication, and daytime dysfunction), and the Epworth Sleepiness Scale (ESS), an 8-item measure of situational daytime sleepiness, remain by far the most widely used subjective sleep-assessment tools in both clinical and research settings[28,29].

However, evidence on their criterion validity against objectively diagnosed sleep disorders is mixed. A large study of 367 consecutive sleep-clinic patients undergoing PSG and multiple sleep latency testing found that receiver-operating-characteristic area-under-the-curve values for both the ESS and PSQI were below 0.9 for predicting obstructive sleep apnea, periodic limb movement disorder, REM sleep behaviour disorder, and narcolepsy, and that an accompanying meta-analysis confirmed the ESS had essentially no diagnostic value for obstructive sleep apnea specifically; the authors concluded that scores on both instruments were more strongly influenced by co-existing anxiety and depression than by objective sleep pathology, and recommended against their use as stand-alone screening or diagnostic tools for these conditions[30]. A multi-national study of 8,481 college students across Chile, Ethiopia, Peru, and Thailand similarly found that exploratory and confirmatory factor analyses did not fully support the traditional single-factor structure of either instrument, suggesting that global PSQI and ESS scores should be interpreted cautiously across diverse populations[31]. These findings do not diminish the practical utility of subjective questionnaires for tracking perceived sleep quality over time or for capturing patient-centred outcomes that objective sensors cannot measure, but they do argue against relying on them as substitutes for objective diagnostic testing.

3.11 Home Sleep Apnea Testing and Clinical Practice Guidelines

Home sleep apnea testing (HSAT), typically employing simplified respiratory polygraphy channels (airflow, respiratory effort, and pulse oximetry, with or without additional signals), has become an established alternative to in-laboratory PSG for diagnosing obstructive sleep apnea (OSA) in appropriately selected adult patients. The AASM's original 2017 clinical practice guideline for diagnostic testing of adult OSA endorsed HSAT for patients with an uncomplicated presentation and increased pre-test probability of moderate-to-severe disease[24], a position reaffirmed and refined in the Academy's updated 2025 position statement, which continues to restrict HSAT to physician-directed diagnosis in appropriately selected adults and to designate more complex cases (e.g., significant cardiopulmonary disease or suspected non-respiratory sleep disorders) as requiring full PSG[23]. The American Association of Sleep Technologists' technical guideline further specifies device and channel requirements necessary for an HSAT recording to be considered technically adequate for clinical decision-making[25].

Notably, the AASM has explicitly recommended against the use of HSAT for diagnosing OSA in children, citing the substantially different physiology, comorbidity profile, and scoring requirements of paediatric sleep-disordered breathing compared with adults[26]. More broadly, in a 2021 update addressing the proliferation of consumer sleep technologies (CSTs), the AASM emphasised that such devices should meet the same regulatory and validation standards as other medical devices before being used for clinical decision-making, while acknowledging their potential value as adjuncts that can enhance patient engagement and longitudinal symptom tracking throughout OSA management[27].

3.12 Special Populations: Paediatric Sleep Measurement

Sleep measurement in children presents unique methodological challenges, including device fit, differing sleep architecture across developmental stages, and the practical difficulty of obtaining multi-night laboratory PSG in young or behaviourally complex patients. Actigraphy has therefore been particularly valuable in paediatric sleep research and in perioperative contexts, where it enables objective tracking of sleep and activity across the hospital stay and recovery period without the burden of a full PSG montage[9]. Systematic appraisal of actigraphy validation studies in children indicates generally good sensitivity for detecting sleep, but consistently weaker specificity for detecting wakefulness compared with adult populations, together with substantial heterogeneity in scoring algorithms and outcome definitions across studies - a methodological inconsistency that complicates cross-study comparison and underscores the need for standardised reporting[8]. As consumer wearables and contactless sensors mature, several are now being formally evaluated against PSG in paediatric and adolescent cohorts, although this evidence base remains considerably smaller than the corresponding adult literature.

4. Discussion

4.1 Strengths, Limitations, and Cross-Cutting Challenges

Table 4 summarizes the principal strengths and limitations of each modality reviewed above, synthesizing the evidence presented in Sections 3.1-3.12.

Table 4: Strengths and limitations of major sleep measurement modalities

Modality	Key strengths	Key limitations
Polysomnography	Diagnostic reference standard; comprehensive multi-signal architecture and event detection	High cost, limited access, technician-dependent, first-night effect, poor scalability [1-3]
Actigraphy	Low cost, low burden, validated over decades, ideal for multi-night/home use	Poor specificity for wake, limited multi-stage discrimination, algorithm heterogeneity [6,8]
Consumer wearables/rings	High acceptability, continuous longitudinal data, improving 4-stage accuracy	Proprietary undisclosed algorithms, variable validation quality, specificity for wake still limited [11-13,16]
Contactless sensors	No device worn, ideal for vulnerable/non-adherent populations, good for cardiorespiratory signals	Sensitive to distance/motion artefact, limited standalone sleep-staging validation [17-20]
Smartphone acoustic apps	Near-zero cost, wide accessibility, useful for snoring screening	No formal sleep staging; sensitivity affected by ambient noise and device placement [21,22]
AI/deep learning models	Automates scoring, extracts novel prognostic signals, enables scalable analysis	Interpretability challenges, need for external/diverse validation, data/compute demands [4,7]
Subjective questionnaires	Captures perceived sleep quality, minimal burden, useful for tracking patient-centred outcomes	Poor criterion validity against objective sleep-disorder diagnoses; confounded by mood symptoms [30,31]

4.2 Future Directions

Several trends are likely to shape the next generation of sleep measurement tools. First, multimodal sensor fusion - combining accelerometry, PPG, temperature, and increasingly bioimpedance or radar signals within a single wearable or contactless platform - is likely to continue narrowing the performance gap with PSG for multi-stage classification, as already demonstrated by the incremental gains achieved through multi-sensor architectures relative to accelerometer-only devices.

Second, foundation models trained on very large, multi-institutional PSG datasets suggest that sleep signals may serve as a general-purpose physiological biomarker, extending the value of sleep measurement well beyond sleep-disorder diagnosis into broader disease risk-prediction and precision-medicine applications[4]. Realising this potential will require external validation across diverse demographic and clinical populations, careful attention to algorithmic bias, and improved model interpretability.

Third, regulatory and professional-society frameworks will need continued refinement to keep pace with the rapid iteration cycle of consumer devices; the AASM's ongoing position statements on consumer sleep technologies and home sleep apnea testing illustrate an evolving but still-developing consensus on how, and under what circumstances, these tools may substitute for or complement laboratory-based assessment.

Finally, greater methodological standardisation - consistent epoch lengths, transparent algorithm reporting, pre-registered validation protocols, and standardised performance metrics - remains a recurring recommendation across the wearable-validation literature, and will be essential for enabling meaningful comparison across the rapidly proliferating landscape of sleep measurement tools.

5. Conclusion

Sleep measurement has evolved from a single laboratory-based reference standard into a diverse ecosystem of complementary tools spanning polysomnography, actigraphy, consumer wearables, contactless sensors, smartphone applications, and AI-driven analytics. Each modality occupies a distinct niche shaped by trade-offs among diagnostic accuracy, patient burden, cost, and scalability: PSG remains unmatched for detailed diagnostic characterisation, actigraphy and consumer wearables enable ecologically valid longitudinal monitoring, contactless sensors minimise patient burden for vulnerable populations, and AI-enabled analysis is beginning to extract prognostic value well beyond traditional sleep metrics. Rigorous, standardised, and demographically representative validation remains the central unmet need across nearly all newer modalities, and will determine how effectively these tools can be integrated into both clinical sleep medicine and population-level health monitoring in the years ahead.

Declaration by Authors

Ethical Approval: Not applicable
Acknowledgement: None
Source of Funding: None
Conflict of Interest: The authors declare no conflict of interest.

References

[1] Kwon S, Kim H, Yeo WH. Recent advances in wearable sensors and portable electronics for sleep monitoring. iScience. 2021;24(5):102461.
[2] Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Darien, IL: American Academy of Sleep Medicine; 2012.
[3] Palombini L, Assis M, Drager LF, Mello LM, Pires GN, Zancanella E, et al. 2024 position statement on the use of different diagnostic methods for sleep disorders in adults – Brazilian Sleep Association. Sleep Sci. 2024;17(4):e476-92.
[4] Thapa R, Kjaer MR, He B, Covert I, Moore H 4th, Hanif U, et al. A multimodal sleep foundation model for disease prediction. Nat Med. 2026;32(2):752-62.

[5] Wang JJ, Wu MF. Discrepancies in subjective sleep quality between home and hospital settings: insights of hypnotic agents use with post-polysomnography questionnaire. Diagnostics (Basel). 2025;15(24):3153.
[6] Yuan H, Hill EA, Kyle SD, Doherty A. A systematic review of the performance of actigraphy in measuring sleep stages. J Sleep Res. 2024;33(5):e14143.
[7] Birrer V, Elgendi M, Lambercy O, Menon C. Evaluating reliability in wearable devices for sleep staging. NPJ Digit Med. 2024;7(1):74.
[8] Meltzer LJ, Montgomery-Downs HE, Insana SP, Walsh CM. Use of actigraphy for assessment in pediatric sleep research. Sleep Med Rev. 2012;16(5):463-75.
[9] Conrad N, Karlik J, Holley AL, Wilson AC, Koh J. A narrative review: actigraphy as an objective assessment of perioperative sleep and activity in pediatric patients. Children (Basel). 2017;4(4):26.
[10] American Academy of Sleep Medicine. Staying current with actigraphy devices for sleep-wake monitoring [Internet]. Darien (IL): AASM; 2025 [cited 2026 Jul 18].
[11] Schyvens AM, Peters B, Van Oost NC, Aerts JM, Masci F, Neven A, et al. A performance validation of six commercial wrist-worn wearable sleep-tracking devices for sleep stage scoring compared to polysomnography. Sleep Adv. 2025;6(2):zpaf021.
[12] Lee Y, Lee J, Cho J, Kang Y, Choi J. Performance of consumer wrist-worn sleep tracking devices compared to polysomnography: a meta-analysis. J Clin Sleep Med. 2025;21(3):573-82.
[13] Svensson T, et al. Validity and reliability of the Oura Ring Generation 3 (Gen3) with Oura sleep staging algorithm 2.0 (OSSA 2.0) when compared to multi-night ambulatory polysomnography: a validation study of 96 participants and 421,045 epochs. Sleep Med. 2024;115:251-63.
[14] de Zambotti M, Zavecz Z, Goldstein RA. Validation of Oura Ring Gen3 sleep staging performance against polysomnography. J Med Internet Res. 2024;26:e53697.
[15] Altini M, Kinnunen H. The promise of sleep: a multi-sensor approach for accurate sleep stage detection using the Oura ring. Sensors (Basel). 2021;21(13):4302.
[16] Robbins R, et al. Accuracy of three commercial wearable devices for sleep tracking in healthy adults. Sensors (Basel). 2024;24(20):6532.
[17] Gross-Isselmann JA, Eggert T, Wildenauer A, Dietz-Terjung S, Grosse Sundrup M, Schoebel C. Validation of the Sleepiz One + as a radar-based sensor for contactless diagnosis of sleep apnea. Sleep Breath. 2024;28(4):1691-9.
[18] Boiko A, Martínez Madrid N, Seepold R. Contactless technologies, sensors, and systems for cardiac and respiratory measurement during sleep: a systematic review. Sensors (Basel). 2023;23(11):5038.
[19] Saner H, Knobel SEJ, Schuetz N, Nef T. Contact-free sensor signals as a new digital biomarker for cardiovascular disease: chances and challenges. Eur Heart J Digit Health. 2020;1(1):30-9.
[20] Ravindran KKG, della Monica C, Atzori G, Lambert D, Hassanin H, Revell V, et al. Reliable contactless monitoring of heart rate, breathing rate, and breathing disturbance during sleep in aging: digital health technology evaluation study. JMIR Mhealth Uhealth. 2024;12:e53643.
[21] Kong F, Zou Y, Li Z, Deng Y. Advances in portable and wearable acoustic sensing devices for human health monitoring. Sensors (Basel). 2024;24(16):5354.
[22] Brown J, Mitchell Z, Jiang YA, Archdeacon R. Accuracy of smartphone-mediated snore detection in a simulated real-world setting: algorithm development and validation. JMIR Form Res. 2025;9:e67861.
[23] Rosen IM, Kirsch DB, Carden KA, Malhotra RK, Ramar K, Aurora RN, et al.; American Academy of Sleep Medicine Board of Directors. Clinical use of a home sleep apnea test: an updated American Academy of Sleep Medicine position statement. J Clin Sleep Med. 2025.
[24] Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, 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.
[25] American Association of Sleep Technologists. AAST technical guideline: home sleep apnea testing (HSAT). Darien (IL): AAST; 2020.
[26] American Academy of Sleep Medicine. American Academy of Sleep Medicine position paper for the use of a home sleep apnea test for the diagnosis of OSA in children. J Clin Sleep Med. 2017;13(10):1199-203.
[27] Schutte-Rodin S, Deak MC, Khosla S, Goldstein CA, Yurcheshen M, Chiang A, et al. Evaluating consumer and clinical sleep technologies: an American Academy of Sleep Medicine update. J Clin Sleep Med. 2021;17(11):2275-82.
[28] Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213.
[29] Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. Sleep. 1991;14(6):540-5.
[30] Nishiyama T, Mizuno T, Kojima M, Suzuki S, Kitajima T, Ando KB, et al. Criterion validity of the Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale for the diagnosis of sleep disorders. Sleep Med. 2014;15(4):422-9.
[31] Gelaye B, Lohsoonthorn V, Lertmeharit S, Pensuksan WC, Sanchez SE, Lemma S, et al. Construct validity and factor structure of the Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale in a multi-national study of African, South East Asian and South American college students. PLoS One. 2014; 9(12): e116383.