

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Fino, E., Mazzetti, M. (2019). Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence. *SLEEP & BREATHING*, 23(1), 13-24 [10.1007/s11325-018-1661-3].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/670927> since: 2019-02-22

*Published:*

DOI: <http://doi.org/10.1007/s11325-018-1661-3>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

## **Monitoring healthy and disturbed sleep through smartphone applications: A review of experimental evidence**

Edita Fino<sup>1</sup> PhD; Michela Mazzetti<sup>1</sup> PhD.

<sup>1</sup>Department of Experimental, Diagnostic and Speciality Medicine (DIMES), Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 5, 40126, Bologna, Italy.

Corresponding author: Edita Fino, Department of Experimental, Diagnostic and Specialty Medicine (DIMES), Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 5, 40126, Bologna, Italy; [edita.fino@unibo.it](mailto:edita.fino@unibo.it)

This work was performed in Department of Experimental, Diagnostic and Specialty Medicine, (DIMES) Alma Mater Studiorum Università di Bologna. Both authors have seen and approved the manuscript.

## **Abstract**

Smartphone applications are considered as the prime candidate for the purposes of large scale, low cost and long term sleep monitoring. How reliable and scientifically grounded is smartphone based assessment of healthy and disturbed sleep remains a key issue in this direction. Here we offer a review of validation studies of sleep applications to the aim of providing some guidance in terms of their reliability to assess sleep in healthy and clinical populations, and stimulating further examination of their potential for clinical use and improved sleep hygiene. Electronic literature review was conducted on Pubmed. Eleven validation studies published since 2012 were identified, evaluating smartphone applications' performance compared to standard methods of sleep assessment in healthy and clinical samples. Studies with healthy populations show that most sleep applications meet or exceed accuracy levels of wrist-based actigraphy in sleep-wake cycle discrimination, whereas performance levels drop in individuals with low sleep efficiency (SE) and in clinical populations, mirroring actigraphy results. Poor correlation with Polysomnography (PSG) sleep sub-stages is reported by most accelerometer-based apps. However, multiple parameter-based applications (i.e., EarlySense, SleepAp) showed good capability in detection of sleep-wake stages and sleep related breathing disorders (SRBD) like obstructive sleep apnea (OSA) respectively with values similar to PSG. While the reviewed evidence suggests a potential role of smartphone sleep applications in pre-screening of SRBD, more experimental studies are warranted to assess their reliability in sleep-wake detection particularly. Apps' utility in post treatment follow-up at home or as an adjunct to the sleep diary in clinical setting is also stressed.

**Key Words:** smartphone sleep applications, actigraphy, polysomnography, sleep diary, OSA, SRBD.

## **Introduction**

Sleep is a critical aspect of our health and wellbeing. Good quality sleep is essential for optimal cognitive functioning, physiological processes, emotion regulation and quality of life [1-7]. Current modern lifestyles, longer working hours and commute are constantly eroding our capacity to obtain and maintain good sleep with serious implications for emerging sleep related problems [8-12]. Therefore, looking for feasible methods able to provide objective, long term and large scale sleep monitoring remains on the highlight of the healthcare community and general population [13].

Unfortunately, objective measures of sleep, like the gold-standard polysomnography, are high resource consuming and therefore impractical for this purposes. As pointed out by Ko and colleagues [14], technological advancements allowing for a wide range of electronic devices to be used for health tracking functions, including sleep monitoring have brought the promise of a system able to provide low cost, large scale sleep assessment closer than ever. Amongst the most popular bearers of such promise are current generation smartphones, which through a series of inbuilt sensors (i.e., accelerometers, gyroscopes, microphones, cameras) and enhanced computational capacity are able to record and score sleep data in real time, providing immediate information on one's sleep and wellbeing [15]. Given their accessibility, ubiquity and personal nature, smartphones, amongst other technological devices, are considered the prime candidate to be utilized for these purposes. However, the first step in this direction requires addressing the issue of how reliable and how well and scientifically grounded are sleep reports yielded by smartphones. Recent experimental works and reviews [16-20] have noted how hardware and software technology for smartphone sleep monitoring is abounding, whereas validation studies on the reliability of their performance are far from catching up. Sleep applications of all kinds are currently available in the market, offering diverse functionality features, from helping individuals to improve their sleeping habits, to objectively assessing sleep parameters [17], and even aiding healthcare professionals in screening patients for sleep disorders (see [15] for a list of most common applications). In a recent work, Fietze [18] highlighted the necessity of further experimental studies, noting that despite the massive use and heightened public interest around this issue, there is a significant gap in research on sleep applications' functions and limitations. Given the fast growing developments in this field, and the need for validation studies with various populations and in practical contexts, in this review we offer a state of the art update of the experimental evidence gathered so far on smartphone based sleep monitoring. Studies

conducted with both healthy and clinical samples, that assess sleep analysis reports of smartphones compared to standard methods of sleep assessment are considered. Our aim is providing some guidance in terms of the reliability of sleep applications in assessing healthy and disturbed sleep and stimulating further examination of their potential for improving sleep hygiene.

## **Methods**

We searched PubMed with key terms including “smartphone applications”, “sleep monitoring”, “sleep quality”, “sleep related breathing disorder”. We eliminated articles that were not relevant to smartphone based sleep monitoring (e.g., other consumer sleep technologies, health tracking apps). To be included, the studies had to be in English language and meet the following criteria: (1) the technology considered regarded only sleep monitoring applications developed for smartphone using built-in and/or external (wearable or contact free) sensors and integrating a wide range of sleep parameters, (2) studies tested the performance of sleep applications that can be used without the need of a clinician, (3) studies examined the performance of sleep apps against (one or more) standard methods of sleep assessment such as PSG, actigraphy, sleep scales and questionnaires or clinical-diagnostic criteria (4) studies examined the performance of sleep applications with either healthy users or clinical populations, or both. The search was performed at/or before January 2018. We identified and discussed eleven validation studies published between 2012 and 2018, five conducted with healthy samples, five with clinical populations and one study conducted with both clinical and healthy samples (see Table 1 and Table 2).

## **Overview of literature**

Prior to a detailed analysis and discussion of experimental studies, in the next sections we offer an overview of traditional methods of sleep assessment, which are currently used as standard criterion for evaluating the outcome of smartphone based sleep monitoring. In so doing, we refer to extant literature examining this issue from various perspectives and further extend existing work by providing an up to date review of main findings.

### *Standard measures of sleep assessment*

Polysomnography (PSG) is the golden standard of sleep assessment. As the best and most complex assessment of sleep it involves multiple parameter recording (i.e., the EEG, EOG, EMG, ECG, auditory recordings of snoring, and

video recording of movements in sleep) allowing for in depth analysis and reporting of sleep architecture, including sleep stages and main sleep parameters. The complexity and accuracy of PSG sleep evaluation has earned it the status of the ‘gold’ method, meaning also the most expensive in terms of related costs of medical equipment and expertise, which make it impractical for large-scale and long-term sleep monitoring [13].

Alternative methods like actigraphy offer a simpler approach with just one parameter recording. Actigraphy is an accelerometer based device that makes sleep-wake assessments based solely on movement detection and scoring of body activity. While it does not assess sleep stages actigraphy can reliably detect wakefulness from sleep [21-23] and is widely used as a second best alternative to PSG when sleep staging is not required [13]. However, because it relies only on movement detection, actigraphy has the tendency to underestimate sleep onset latency (SOL), which may be effectively masked by lack of body movement while awake in the bed. It also tends to overestimate total sleeping time (TST) for the same reasons. Indeed, research shows that its accuracy varies greatly with the amount of quiet wakefulness during the night and with specific clinical populations (e.g. elderly people or individuals with poor SE) [24,25]. Because people with sleep disorders tend to have a highly fragmented sleep architecture, this further deteriorates actigraphy performance in accurately detecting sleep-wake cycles in clinical samples compared to healthy subjects. Although widely used as a second best and low cost alternative to PSG, actigraphy remains heavily dependent on specialized expertise for data scoring and interpretation, and is thus not as feasible for long term and large scale sleep assessment.

It is well established that a comprehensive sleep assessment should include a comparison of both subjective and objective sleep measures. Subjective methods for assessing sleep involve data describing a person’s sleep patterns, usually captured through self-reports, sleep diaries and surveys. Such measures provide useful information and contribute to a comprehensive assessment of sleep quality, especially when combined with physiological monitoring (i.e., PSG), and may serve as pre screening layer for sleep disorders. For instance, the sleep diary is regarded as the “gold standard” for subjective sleep assessment and is widely used despite the lack of agreement on a common standard format [26]. While inexpensive and easily used for long term and large scale sleep assessment their reliability rests entirely on accurate self reports by the subject [27]. Sleep diaries remain fundamentally a measure of subjective perception of sleep allowing for an estimate of the possible rift between subjective perception and objective measurement of sleep, otherwise known as sleep misperception, which is a common phenomenon of

numerous sleep disorders [28-30]. Other examples of self reports include standardized questionnaires [31-34] to assess not only sleep quality but also eventual sleep disturbances. The Pittsburgh Sleep Quality Index (PSQI) [31] for instance is a widely used scale to assess sleep quality and disturbances over a one-month period. PSQI integrates a wide variety of factors associated with sleep quality, including subjective quality ratings, sleep time, efficiency (time spent trying to fall asleep), frequency, and severity of sleep-related problems. Another commonly used scale is the Epworth Sleepiness Scale (ESS) [32], which measures daytime sleepiness but is also reliably used as for screening sleep disorders [33]. Finally, other questionnaires are aimed to detect specific sleep disorders as is the case of the STOP-BANG questionnaire, which is a standard measure for screening of obstructive sleep apnea (OSA) [34].

#### *Smartphone-based modalities for sleep assessment*

Most common smartphone-based sleep applications rely on common principles of standard sleep assessment including movement detection, audio and video recording and questionnaires. Through the presence of inbuilt accelerometers the smartphone can act as a modern actigraph to discern wake and sleep from the movement detected by the phone's embedded sensors. Some smartphone applications compute their sleep assessments based on analysis of sound and noise present in the room while sleeping. While the accelerometer-based modality of sleep assessment through the smartphone is the closest reproduction of a standard method of sleep assessment, differences between actigraphy recorded from the phone and actigraphy used in standard sleep monitoring should not be overlooked. Research shows that actigraphic analysis results may depend not only on the type of actimeter used, but also on the targeted location of the device on a human body (i.e., wrists, waist etc.) [35-38]. Furthermore, sleep applications may consist in simple digital implementation of questionnaires such as sleep scales and be used for the purposes of assessing sleep quality as well as to distinguish those who actually poorly and only briefly sleep from those who suffer from sleep disorders. An advantage of questionnaire-based sleep applications compared to paper or web based sleep scales is the constant availability of the phone which highly increases adherence to self monitoring and self report rates of subjects [39-40].

Other sleep applications rely on multiple modalities (sensors plus questionnaires) and signal processing from a combination of built-in and external sensors that provide a wide range of physiological signal recordings. As a

result, such applications may yield more complex sleep analysis, including sleep stages (see review of Ong and colleagues [17]). Data from multiple sources of information can be directly derived through the phone in an unobtrusive way where the user is putatively removed from the monitoring process and does not need to interact with the recording device beyond normal phone user behavior. In this sense, smartphones would (ideally) represent a radically innovative, largely accessible and low cost sleep monitoring device able to record and score the data online without the need for specialized medical or technical assistance and possible to use for long term and large scale sleep assessment [15].

However, the scientific validity of sleep analysis yielded by smartphone applications remains an elusive notion as most sleep applications do not offer information on the analysis algorithm used for scoring sleep parameters [15]. Most of the apps' summary reports usually consist in visual graphs that give users a qualitative impression of how well they may have slept and give aggregate sleep scores labeled in lay language which is difficult to translate in terms of standard sleep parameters. According to these conclusions another recent work [17] examined features of 51 sleep assessment apps targeted for consumer use (excluding apps targeting health professionals) based on the highest user ratings received in respective store websites. Most of sleep applications provided data on sleep parameters, including duration, time awake, and time in light, medium, deep sleep, while reporting of REM and extra features was fairly limited. As noted by Behar and colleagues [15], such parameters *per se* are meaningless and unsuitable for direct comparison with standard sleep parameters calculated by standard sleep assessment methods. To overcome this barrier would require breaking in the 'black-box' of sleep applications and gaining access to the raw data.

Given the interest and potential clinical significance, Behar and colleagues [15] examined whether smartphone sleep applications available in the market can be effectively used for screening and diagnosis of obstructive sleep apnea (OSA). From the analysis of the apps' features and outputs, carried out in 2013, authors concluded that only applications implementing questionnaires commonly used for OSA screening such as STOP and STOP BANG [34] resulted valid for screening purposes, whereas accelerometer or microphone-based apps did not prove reliable for OSA screening. Recently, other authors [41] have focused on developing specific algorithm for smartphone enhanced snore and noise discrimination achieving good performance, potentially overcoming limits found by Behar and colleagues [15]. In the next sections we examine empirical evidence gathered so far on sleep application



validation studies conducted on healthy and clinical populations to test the reliability of sleep applications compared to standard sleep assessment methods (or clinical criteria).

### **Reliability of smartphone apps in assessing healthy sleep: experimental evidence**

#### *Detection of sleep – wake cycle*

Two PSG studies have compared a smartphone assessment of healthy sleep with the gold standard PSG. Bhat and colleagues [16] evaluated the reliability of sleep analysis provided by Sleep Time app (Azumio Inc., Palo Alto, CA, USA) in detecting overall sleep-wake as well as individual sleep stages of twenty healthy adults undergoing an overnight in-laboratory PSG. For analysis purposes, authors divided both the PSG hypnogram and app graph into 15-min epochs which were then reassigned corresponding PSG and app stage. Absolute sleep parameters (SOL, TST, wake after sleep WASO, sleep stages and SE) were then scored and compared between the two methods. Results showed no correlations between the app and the PSG for SE, SOL, or sleep stage percentages for light sleep and deep sleep. The application underestimated light sleep, overestimated deep sleep and sleep latency and achieved very low accuracy in epoch-wise comparison (45.9%). However, sleep-wake accuracy (85.9%), sensitivity in detecting sleep (89.9%) and specificity in detecting wakefulness (50%) were similar to that observed with wrist actigraphy [42-45].

More recently, Tal and colleagues [46] tested the performance of EarlySense (by Ltd., Israel) an application for smartphone, which relies on an external sensor device (ES) validated for measuring movement, heart rate, and respiration in clinical settings [47-49] and adapted for personal home use. The study included a total of 63 subjects of which 43 were patients studied in the sleep laboratory and 20 were healthy subjects recorded at home for 1 to 3 nights with a portable PSG system in two conditions (7 participants were recorded while sleeping alone, whereas 13 while sleeping with partner). Heart rate (HR), respiratory rate (RR), body movement, and sleep-related parameters such as TST, sleep stages [Sleep Latency (SL), Wake After Sleep Onset (WASO), Rapid Eye Movement (REM) sleep and Slow Wave Sleep (SWS)] calculated from the app were compared to simultaneously generated PSG data. Combined results from the 20 healthy subjects (data from patients will be reviewed in the next section of present work) showed a 76.7% sensitivity to detect wakefulness, 95.2% sensitivity to detect sleep (REM + SL + SWS), and

a 92.5% overall accuracy of sleep–wake detection. Notably, separate analysis for both setups (single subjects in bed at home and subject recorded with partner in double bed) showed similar results with overall wake sensitivity of 72.1% and 79.0%, sleep sensitivity of 95.4% and 95.1%, and overall agreement 92.1% and 92.5% respectively.

In a study examining three validated algorithms [50-52] for actigraphy scoring, Natale and colleagues[53] directly compared raw data provided by an iPhone accelerometer with those provided by wrist actigraphy. Participants were 13 healthy subjects that completed four consecutive overnight recordings at home by wearing the actigraph on the non dominant wrist. Standard sleep statistics (TST, WASO and SE) were computed per each algorithm and compared across devices. Results showed satisfactory epoch by epoch agreement between the actigraph and smartphone accelerometer for all sleep parameters (with the exclusion of TST) and all algorithms, with the one improving that of Cole and colleagues[51] yielding a better performance. Another interesting finding of this study was the evidence that the ability of sleep application to detect TST, WASO and SE deteriorated with shorter TST (<6 hours) and lower SE (< 85%) and longer WASO (> 20 min), suggesting that the poorer the sleep, the less reliable results from sleep apps. This is in line with literature on wrist actigraphy showing relatively poor accuracy in detecting disturbed sleep or sleep-wake cycles in clinical populations [24, 25].

More recently, Scott et al [54] investigated the accuracy of Sleep On Cue (SOC, by MicroSleep, LLC), a novel iPhone application that uses behavioral responses to auditory stimuli to estimate sleep onset. SOC emits a low-intensity tone stimulus every 30 s via headphones to which the user responds by gently moving the phone. When an individual fails to respond to two consecutive tones, the app deems that the user has fallen asleep. Twelve young adults underwent polysomnography recording while simultaneously using the app, and completed as many sleep-onset trials as possible within a 2h period following their normal bedtime. Results showed a high correspondence between the app's and polysomnography-determined sleep onset ( $r = 0.79$ ,  $P < 0.001$ ). While the app generally overestimated SOL by 3.17 min (SD = 3.04), the discrepancy was reduced considerably when polysomnography SOL was defined as the beginning of N2 sleep. Despite the pilot nature of the study, authors highlight the potential relevance of using SOC for facilitating power naps in the home environment.

Overall, findings from PSG studies on healthy populations show that sleep-wake discrimination of sleep apps is similar and in fact quite better than that reported for wrist actigraphy [42- 44]: ~90% sensitive and ~50% specific for sleep. While Sleep Time by Azumio overestimated sleep comparably with actigraphy, it performed poorly with

respect to sleep stage analysis when compared to PSG. Early Sense on the other hand showed highly accurate sleep stage analysis compared to PSG. The app analyzed sleep using an algorithm based on HR, RR and motion detection, which probably gives it an advantage over actigraphy and enables analysis of sleep stages. Accurate sleep onset detection was offered by SOC suggesting that sleep apps utilizing behavioral input from the user may be a promising tool in this regard.

#### *Detection of snoring*

Self-monitoring of snoring is considered a useful tool for maintaining good health among the general population. Stippig and colleagues [55] tested the ability of three apps (SnoreMonitorSleepLab, Quit Snoring, and Snore Spectrum) to distinguish between snoring events and other noises present in the environment, such as cars driving past the window, conversations in the bedroom, or even just the rustling of sheets and blankets. They compared the three apps with the ApneaLink Plus (ResMed Germany Inc., Martinsried, Deutschland) screening device which was attached to a test subject spending one night with and one night without the Oral Appliance Narval (ResMed Germany Inc., Martinsried, Deutschland). Although these apps have features potentially advantageous for clinical purposes (like audio recording of snoring and counting of snore events), results did not correspond with the ApneaLink Plus screening device, which led authors to conclude that their reliability and accuracy is insufficient to replace common diagnostic standards.

#### *Electronic questionnaires to assess sleep quality*

Sleep applications that are based on implementation of electronic questionnaire to assess sleep quality represent another modality of sleep assessment via smart phone, which relies on user behavioral responses. To our knowledge, only one study [56] has compared the sleep application Toss N Turn with an electronic version of the Pittsburgh Sleep Quality Index (PSQI) [31] combined with a Sleep Diary. Sleep diary is a useful methodology for sleep assessment as it yields information about a number of relevant sleep parameters and has also been used to test sleep-detecting technologies including actigraphy [27]. In their study Min and colleagues [56] collected one month of phone sensor and sleep diary entries from 27 subjects in various sleep contexts and used this data to construct models for detecting sleep-wake cycles, daily sleep quality, and global sleep quality. More than 30 min differences

were found in bedtime, sleep duration, and wake time for all 3 parameters, which are larger than those of commercial actigraphs that have error rates lower than 10 minutes.

<< Please insert table 1 near here >>

### **Reliability of smartphone-based assessment of disturbed sleep: experimental evidence**

#### *Detection of sleep – wake cycle*

Three PSG studies have tested reliability of smartphone-based sleep monitoring with clinical subjects. Patel and colleagues [20] examined the accuracy of Sleep Cycle (an accelerometer based app developed by Maciek Drejak Labs, now Northcube AB) by comparing its sleep analysis with PSG in a clinical population of 25 children (age 2–14) undergoing overnight PSG for clinical suspicion of OSA. Sleep parameters (TST, SL, and time spent in sleep stages) were obtained by converting graph segments into minutes through comparison with the entire length of the graph. App graphs were then compared with the PSG. No significant correlation was found between TST or SL between the app and PSG although visual inspection of the app graphs and the PSG showed some correspondence. Only sleep latency from the PSG and latency to deep sleep from the app had a significant relationship ( $p = 0.03$ ). Authors concluded that Sleep Cycle App is not yet accurate enough to be used for clinical purposes.

Toon and colleagues [19] compared performance of a smartphone sleep application (MotionX 24/7), against combined actigraphy (Actiwatch2) and polysomnography (PSG) in a clinical pediatric sample of children and adolescents suspected for OSA, with and without comorbidities. Sleep outcome variables provided by the app were SOL, TST, WASO, and SE. Results of the paired comparisons between PSG and MotionX 24/7 revealed that SOL and WASO were significantly underestimated by MotionX 24/7 (12 and 63 min, respectively), resulting in significantly longer TST and greater SE (106 min and 17%, respectively). Based on these results, authors concluded that the MotionX 24/7 did not accurately reflect sleep duration or sleep quality, and should therefore be considered carefully before use in a clinical setting. More recently Tal and colleagues [46] tested the performance of EarlySense (Ltd., Israel) to calculate sleep stages (wake, REM, LS (N1+N2), and SWS) with 43 adult patients with various sleep disorders undergoing one overnight in-laboratory PSG. Results for this group showed a wakefulness sensitivity

of 83.4%, sleep sensitivity of 89.7%, and overall sleep accuracy of 88.5%. Detailed sensitivities for each sleep state were 40.0% for REM, 63.3% for light sleep (LS), and 53.6% for SWS.

In sum, both Sleep Cycle and MotionX sleep applications performed poorly in terms of sleep stage analysis when compared to PSG, which may be due to the fact that most movement-based algorithms used in actigraphy and accelerometer based sleep applications cannot distinguish sleep stages. On the other hand, EarlySense performance was quite good in discriminating sleep stages with satisfactory results compared to PSG. This may be due to the scoring algorithm that integrated data from multiple signals including HR, RR and motion detection.

#### *Detection of snoring and SRBD*

Given the importance of snoring in signaling potential sleep disorders (i.e., OSA) and considering limitations of apps reviewed in their previous work [15], Behar and colleagues [57] developed SleepAp to the purpose of screening and monitoring of OSA. SleepAp uses internal phone sensors and an external pulse oximeter to record audio, activity, body position, and oxygen saturation during sleep, and implements the clinically validated STOP-BANG questionnaire. The app ultimately classifies the user as belonging to one of two classes: nonOSA (healthy and snorers) and OSA (mild, moderate and severe). The algorithms implemented by the app is based on signal processing and machine learning algorithms validated on a clinical database of 856 patients and was tested on 121 patients. Compared to the clinicians' diagnoses, the app's classification on the sample tested had an accuracy of up to 92.2% when classifying subjects as having moderate or severe OSA versus being healthy or a snorer. Classifying mild OSA resulted the hardest and was associated with the lowest accuracy (88.4 %). Authors concluded that SleepAp is a first step towards a clinically validated automated sleep screening system, which could provide a new, easy-to-use, low-cost, and widely available modality for OSA screening.

Nakano and colleagues [58] used a smartphone to monitor and quantify snoring and OSA severity. They used data from 10 patients to develop the program and validated it with 40 patients with mild, moderate and severe OSA. The smartphone acquired ambient sounds from the built-in microphone and analyzed it on a real-time basis using signal processing procedure similar that developed for tracheal sound monitoring to detect OSA. Results showed a high correlation of snoring time (percentage of total time) measured by the smartphone with the snoring time determined by the PSG ( $r = 0.93$ ). The respiratory disturbance index estimated by the smartphone (smart-RDI) highly correlated with the apnea-hypopnea index (AHI) obtained by PSG ( $r = 0.94$ ). The diagnostic sensitivity and specificity of the

smart-RDI for diagnosing OSA ( $AHI \geq 15$ ) were 0.70 and 0.94, respectively. Results were not as good for subjects with a less than 30 in AHI score, which indicates that its diagnostic accuracy may be insufficient for screening milder forms of OSA. Finally, Camatcho and colleagues [59] conducted a pilot study testing the performance of Quit Snoring app with two patients undergoing polysomnography. The second-by-second evaluation of smartphone snoring results compared with the snores detected by PSG showed substantial agreement with snoring sensitivity ranges from 63.6 % to 95.5 % and positive predictive values from 93.3% to 96.0%.

Overall, apps specifically designed for snoring and OSA detection performed quite well compared to PSG and/or clinical criteria. In particular, two studies [57,58] showed good results in classifying subjects with OSA compared to healthy snorers with a 92.2% accuracy and  $r = 0.94$ , respectively. Both performed lowest when detecting mild OSA, which indicates that the app's diagnostic accuracy may be insufficient for screening milder forms of sleep apnea.

<< Please insert table 2 near here >>

## **Discussion**

Validation studies conducted so far with healthy populations show that sleep applications meet or exceed accuracy levels of wrist actigraphy in sleep-wake cycle discrimination, with most apps similarly tending to overestimate sleep. Accuracy of sleep-wake discrimination tends to drop the more SE levels go down, thus mirroring low actigraphy performance with clinical populations [24,25]. Most sleep applications reviewed here showed poor correlation with PSG sleep sub-stages, which is expected given that most accelerometer based sleep applications do not provide sleep stage analysis. A better performance was provided by Early Sense [46] which showed good sleep staging capability with similar values compared to PSG and a high correlation of estimated TST. It should be noted that this application uses a contact-free external sensor (ES) previously validated for clinical use and then adapted for personal home use through the support of a mobile phone. Specifically, ES has been validated for heart rate and respiratory rate measurement and analyzes sleep using an algorithm based on three parameter recordings (HR RR, and motion detection), which clearly gives this application an advantage over single parameter based sleep applications. As shown by Natale and colleagues [53], different algorithms can yield different results, hence, developing algorithms specifically for smartphone sleep assessment should be the focus of future efforts of both

sleep app developers and clinical research community. Notably, findings of Tal and colleagues [46] resulted from the analysis of combined data of 63 subjects including patients (N=43), with various sleep conditions tested in laboratory, and healthy subjects (N=20) recorded at home. However, separate group analyses showed similar results despite the different sleep conditions which further extend the validity of this application in accurately assessing healthy and disturbed sleep.

Among apps designed for snoring and OSA detection, SleepApp showed a good performance, reaching a 92.2% accuracy level in classifying subjects with OSA moderate and severe compared to healthy snorers [57]. Similarly, high correlation between smartphone and PSG was found by Nakamo and colleagues [58] in terms of total snore time ( $r = 0.93$ ) and AHI ( $r = 0.94$ ). In both cases a good diagnostic sensitivity and specificity was found for diagnosing severe and moderate OSA, whereas a lower performance in detecting patients with mild OSA was reported. Other applications designed for snore detection resulted generally not accurate enough in distinguishing snore from non snore events, especially when used in real-life settings. Although Quit Snoring [59] showed a good performance (accuracy range 63.6 - 95.5 % and positive predictive values range 93.3 - 96.0%), the pilot nature of the study makes it difficult to reach any conclusive results. As shown by Shin and Cho [41] developing snore detection algorithms for smartphone can increase apps performance in reliably distinguishing between snoring and non snoring noises. The algorithm they designed showed a 95.07% accuracy in detection of snoring and non-snoring sounds. Hence, more studies focusing on algorithms specifically developed for smartphone are needed in order to increase apps' reliability in monitoring and detecting snoring and SRBD.

A less taken validation path includes the use of sleep scales and self reports, considering that most of the sleep applications are designed to offer descriptive statistics of sleep quality and assist healthy users in improving sleep hygiene. More studies are needed in this direction. As put forth by Griesby–Toussaint and colleagues [60], sleep apps can serve as tools for behavior change through features specifically designed to encourage healthy sleeping habits. It is also possible that long term use of smartphone sleep monitoring can promote in the long run sustainable sleep hygiene among healthy users and also assist in the management of sleep related problems [59].

While representing an important step towards validation of smartphone sleep assessment, studies reviewed here present a number of limitations. For one thing, reliance on 'black box' phone actigraphy and lack of raw data (with the exception of Natale and colleagues [52], Behar and colleagues [57] and Namako and colleagues [58]) may have

limited studies' explanatory power. Raw data access is also crucial because new algorithms are continually being developed that can enhance information extraction from single parameter recording [38,53,41]. Lack of access to raw data and proprietary rights on algorithms used by sleep apps has lead authors to manually extract the app staging data in epochs of much larger duration than those used clinically [16]. In most studies reviewed here (except for Toon and colleagues [19] and Tal and colleagues [46],) data from the app were acquired by physically measuring the length of the graphs in an analog fashion. This process is not bias free since individual judgment may heavily influence the process of sleep stage reassignment. Furthermore, almost all studies have very small samples of variable age range, and may thus suffer from high internal variability in terms of sleep architecture, known to vary considerably with age [61]. In two studies the variability is further increased by presence of diverse sleep disorders in samples of 25 [20] and 43 patients [46]. In the end, most studies focused on single app testing by comparing it with one standard sleep assessment method (except for Toon and colleagues [19] that used PSG and actigraphy, Behar and colleagues [57] who used clinical criteria and standard questionnaire, and Stippig and colleagues [55] who tested 3 apps). Combining more methods including objective and subjective sleep assessment with healthy and clinical samples might be a useful approach in future validation studies of smartphone sleep monitoring.

## **Conclusion**

Altogether, results from validation studies support the conclusion that when it comes to reliable use of smartphones for monitoring healthy and disturbed sleep it may be useful to reframe the question as rightly pointed out by Bianchi [62] and ask which app, what for, and in what condition. For most of sleep applications reviewed here the space for reliable use may be that of traditional wrist actigraphy, which despite limitations has been widely accepted as appropriate for detecting sleep-wake cycles. In terms of sleep staging capacity, evidence shows that relying on external sensor devices (as in the case of EarlySense [46]) validated and adapted for personal home use may be advantageous and increase smartphone applications' accuracy in sleep-stage detection. Also, developing scoring algorithms specifically for smartphone sleep monitoring may enhance apps' capacity to yield accurate sleep-wake and SRBD detection from one or more parameter recordings (as in the case of SleepApp [57]). While the accuracy of most sleep applications in detecting sleep-wake cycles tends to drop in individuals with low SE and is generally scarce in clinical populations, studies reviewed here suggest a promising role of apps in detection of snoring and



sleep related breathing disorders (i.e., OSA). Using a smartphone to measure snoring may be useful not only for OSA screening but also for evaluating the status of snoring as a detrimental symptom for sleep and other health related problems in the general population. More validation studies are certainly needed for sleep apps to carve out a proper space large scale and low cost pre screening of poor sleep patterns and SRBD.

Nonetheless, smartphone sleep monitoring can be reliably used in adjunct to or as a substitute of sleep diaries in clinical setting or in home for post diagnoses long-term monitoring, which is especially relevant for sleep disordered individuals who wont or cannot adhere to self-reporting [40]. It can complement sleep diary when used as outcome for intervention studies, and can serve as a form of biofeedback, as reported previously for patients with misperception insomnia [28] or be used for administering specific sleep retraining therapies for persons suffering from chronic insomnia [59]. The potential of long term use of smartphone sleep monitoring to promote sustainable sleep hygiene among healthy users in real life contexts remain important avenues for future research.

**Compliance with Ethical Standards**

**Funding:** Fondazione Altroconsumo, Italy, provided partial financial support in the form of research funding (to EF). The sponsor had no role in the design or conduct of this research.

**Conflict of Interest:** The authors declare that they have no conflict of interest.

**Ethical approval:** This article does not contain any studies .with human participants or animals performed by any of the authors.

## References

1. Walker, M. P (2009) The role of sleep in cognition and emotion. *Ann. N. Y. Acad. Sci.* 1156, 168–197.
2. Krause AJ, Simon EB, Mander BA, Greer SM, Saletin JM, Goldstein-Piekarski AN, Walker MP (2017) The sleep-deprived human brain. *Nat Rev Neurosci.* 18(7):404-418.
3. Raven F, Van der Zee EA, Meerlo P, Havekes R (2017) The role of sleep in regulating structural plasticity and synaptic strength: Implications for memory and cognitive function. *Sleep Med Rev.*; Doi:10.1016/j.smrv.2017.05.002.
4. Durmer, J. S., and Dinges, D. F (2005) Neurocognitive consequences of sleep deprivation. *Semin. Neurol.*, 25:117–129.
5. Killgore, W. D (2010) Effects of sleep deprivation on cognition. *Prog. Brain Res.* 185:105–129.
6. Tempesta, D., Couyoumdjian, A., Curcio, G., Moroni, F., Marzano, C., & De Gennaro, L., et al., (2010) Lack of sleep affects the evaluation of emotional stimuli. *Brain Res Bul.* 82, 104–108.
7. Panossian L A & Avidan A Y (2009) Review of sleep disorders. *Med. Clin. North Am.* 93: 407–25.
8. Morgan D, Tsai SC (2015) Sleep and the endocrine system. *Crit Care Clin.* 31(3):403-18.
9. Morgan D, Tsai SC (2016) Sleep and the Endocrine System. *Sleep Med Clin.* 11(1):115-26.
10. Cassoff J, Bhatti JA, Gruber R (2014) The effect of sleep restriction on neurobehavioural functioning in normally developing children and adolescents: insights from the Attention, Behaviour and Sleep Laboratory. *Pathol Biol.* 62(5):319-31.
11. Kecklund G., & Axelsson J (2016) Health consequences of shift work and insufficient sleep. *BMJ.* 1;355:i5210.
12. Wolkow A, Ferguson S, Aisbett B, Main L (2015) Effects of work-related sleep restriction on acute physiological and psychological stress responses and their interactions: A review among emergency service personnel. *Int J Occup Med Environ Health.* 28(2):183-208.
13. Van de Water, A. T. M., Holmes, A., & Hurley, D. A (2011) Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography—a systematic review. *J Sleep Res.* 20, 183–200.
14. Ko PR, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF (2015) Consumer sleep technologies: a review of the landscape. *J Clin Sleep Med.* 11(12):1455 –1461.
15. Behar J, Roebuck A, Domingo J S, Geder E, Clifford GD (2013) A review of current sleep screening applications for smartphones, *Physiol. Meas.* 34:29–46.
16. Bhat S, Ferraris A, Gupta D, et al.(2015) Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. *J Clin Sleep Med.*; 11:709–715.
17. Ong. A & Gillespie M. Boyd. (2016) Overview of smartphone applications for sleep analysis. *WJOHNS*; 2: 45-49.
18. Fietze, I (2016) Sleep Applications to Assess Sleep Quality. *Sleep Medicine Clinic.* 11 (4), 461–468. 19. Toon E, Davey MJ, Hollis SL, Hons BA; Nixon MG, Home R; Biggs NS (2016) Comparison of Commercial Wrist-Based and Smartphone Accelerometers, Actigraphy, and PSG in a Clinical Cohort of Children and Adolescents. *J Clin Sleep Med.* 12(3):343–50.

20. Patel P, Kim J. Y, Brooks L. J (2017) Accuracy of a smartphone application in estimating sleep in children. *Sleep Breath.* 21:505–511.
21. Marino, M., Li, Y., Rueschman, N.M., Winkelman, W., Ellenbogen, M., Solet, J. M., Dulin, H., Berkman, F. L., Buxton, M. O (2013) Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. *Sleep.* 1; 36(11): 1747–1755.
22. Meltzer LJ, Walsh CM, Traylor J, Westin AM (2012) Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep.* 35:159–66.
23. de Souza L, Benedito-Silva AA, Nogueira Pires ML, Poyares D, Tufik S, Calil HM (2003) Further validation of actigraphy for sleep studies. *Sleep.* 1;26(1):81-5.
24. Sadeh, A (2011) The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev.* 15(4):259-67.
25. Sadeh, A. & Acebo, C (2002) The role of actigraphy in sleep medicine. *Sleep Med Rev.* 6, (2)113–124.
26. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, Morin CM ( 2012) The consensus sleep diary: standardizing prospective sleep self monitoring. *Sleep.* 35:287–302.
27. Girschik J, Fritschi L, Heyworth, Flavie, Waters (2012) Validation of Self-Reported Sleep Against Actigraphy. *J Epidemiol.* 22(5): 462–468.
28. Tang NK, Harvey AG (2006) Altering misperception of sleep in insomnia: behavioral experiment versus verbal feedback. *J Consult Clin Psychol.* 74:767–76.
29. Herbert V, Pratt D, Emsley R, Kyle SD (2017) Predictors of Nightly Subjective-Objective Sleep Discrepancy in Poor Sleepers over a Seven-Day Period. *Brain Sci.* 9;7(3).
30. Kay DB, Buysse DJ, Germain A, Hall M, Monk TH (2015) Subjective-objective sleep discrepancy among older adults: associations with insomnia diagnosis and insomnia treatment. *J Sleep Res.* 24(1):32-9.
31. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ (1989) The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research,* 28:193–213.
32. Johns MW (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep.* 14:540-545.
33. Bonzelaar LB, Salapatas AM, Yang J, Friedman M (2017) Validity of the epworth sleepiness scale as a screening tool for obstructive sleep apnea. *Laryngoscope.* 127(2):525-531.
34. Chung F, Yegneswaran B, Liao P, Chung S, Vairavanathan S, Islam S, Khajehdehi A and Shapiro C (2008) Stop questionnaire: a tool to screen patients for obstructive sleep apnea *Anesthesiology.* 108, 812–21.
35. Berger AM, Wielgus KK, Young-McCaughan S, Fischer P, Farr L, Lee KA(2008) Methodological challenges when using actigraphy in research. *J Pain Symptom.Manage.* 36(2):191–199.
36. Sadeh A, Sharkey KM, Carskadon MA (1994) Activity-based sleep-wake identification: A empirical test of methodological issues. *Sleep,* 17:201-07.
37. Paavonen EJ, Fjallberg M, Steenari MR, Aronen ET (2002) Actigraph placement and sleep estimation in children. *Sleep,* 25:235–7.
38. Roebuck A, Monasterio V, Geder E, Osipov M, Behar J, Malhotra A, Penzel T , Clifford G D (2014) Signal processing of data recorded during sleep. *Physiol. Meas.* 35(1):1–57.

39. Carter MC, Burley VJ, Nykjaer C, Cade JE (2013) Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. *J Med Internet Research*, 15(4):e32.
40. Min YH, Lee JW, Shin YW et al. (2014) Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. *J Med Internet Res*. 16:135.
41. Shin, H., & Cho, J. (2014). Unconstrained snoring detection using a smartphone during ordinary sleep, *BioMedical Engineering OnLine*, 13:116.
42. Martin JL, Hakim AD (2011) Wrist actigraphy. *Chest*, 139:1514–27.
43. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP (2003) The role of actigraphy in the study of sleep and circadian rhythms. *Sleep*, 26:342–92.
44. Marino M, Li Y, Rueschman MN, et al. (2013) Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*. 36(11):1747–1755.
45. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC(1992) Automatic sleep/wake identification from wrist activity. *Sleep*. 15(5):461–469.
46. Tal A, Shinar Z, Shaki D, Codish S, Goldbart A (2017) Validation of contact-free sleep monitoring device with comparison to polysomnography. *J Clin Sleep Med*. 13.
47. Ben-Ari J, Zimlichman E, Adi N, Sorkine P (2010) Contactless respiratory and heart rate monitoring: validation of an innovative tool. *J Med Eng Technol*. 34(7-8):393–398.
48. Zimlichman E, Szyper-Kravitz M, Shinar Z, et al. (2012) Early recognition of acutely deteriorating patients in non-intensive care units: assessment of an innovative monitoring technology. *J Hosp Med*. 7(8):628–633.
49. Zimlichman E, Shinar Z, Rozenblum R, et al. (2011) Using continuous motion monitoring technology to determine patient’s risk for development of pressure ulcers. *J Patient Saf*. 7(4):181–184.
50. Oakley NR. (1997) Validation with Polysomnography of the Sleepwatch Sleep/Wake Scoring Algorithm used by the Actiwatch Activity Monitoring System. Technical report, Bend, Ore., Mini-Mitter.
51. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC (1992) Automatic sleep/wake identification from wrist activity. *Sleep*, 15: 461–9.
52. Kripke DF, Hahn EK, Grizas AP et al. Wrist actigraphic scoring for sleep laboratory patients: algorithm development. *J. Sleep Res*. (2010) 19: 612–19.
53. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Martoni M Monitoring sleep with a smartphone accelerator. *Sleep Biol Rhythms*. (2012) 10: 287-292.
54. Scott, H. Methodology A pilot study of a novel smartphone application for the estimation of sleep onset, *J Sleep Res*. (2018) 27, 90–97.
55. Stippig, A., Hübers, U., & Emerich, M. (2015) Apps in sleep medicine, *Sleep Breath* 19:411–417.
56. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong IJ (2014) Toss 'n' turn: smartphone as sleep and sleep quality detector. In: Jones M, Palanque P, Schmidt A, et al, editors. Proceedings of the SIGCHI conference on human factors in computing systems. Toronto: Association for Computing Machinery. 477–86.

57. Behar, J., Roebuck, A., Shahid, M., Daly, J., Hallack, A., Palmius, N., Member, S. (2015). SleepAp : An Automated Obstructive Sleep Apnoea Screening Application for Smartphones, *IEEE J. Biomed. Health Inform.* 19(1), 325–331.
58. Nakano H; Hirayama K; Sadamitsu Y; Toshimitsu A; Fujita H; Shin S; Tanigawa T. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. *J Clin Sleep Med* (2014);10(1):73-78.
59. Camacho, M., Robertson, M., Abdullatif, J., Certal, V., & Kram, Y. A. (2015) Smartphone apps for snoring, , *J Laryngol Otol* 129, 974–979.
60. Grigsby-Toussaint, Shin, Reeves, Beattie, Auguste, Jean-Louis (2017) Sleep apps and behavioral constructs: A content analysis. *Prev Med Rep.* 6:126-129.
61. Ohayon MM; Carskadon MA; Guilleminault C; Vitiello MV. (2004) Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, 27(7):1255-73.
62. Bianchi MT (2015) Consumer sleep apps: when it comes to the big picture, it's all about the frame. *J Clin Sleep Med.* 11(7):695–696.

**Table 1****Characteristics and results of experimental studies on reliability of smartphone based monitoring of healthy sleep**

	Bhat et al. (2015)	Tal et al. (2017)	Natale et al.(2013)	Scott et al. (2018)	Stippig et al. (2015)	Min et al. (2014)
Sample	20 healthy subjects (age 22 to 57 years)	23 healthy subjects (age 24 to 65)	13 healthy subjects (Mean age=23 years)	12 healthy adults (mean age = 21.67 years old)	1 healthy subject	27 healthy subjects (age 20 to 59 years)
Sleep Application/ Device features	Sleep Time App (accelerometer based ) by Azumio Inc., Palo Alto, CA, USA.	Smartphone Application + EarlySense (Ltd., Israel) contact free sensor	Tri-axial accelerometer (LIS302DL) supported by iPhone (Apple)	Sleep On Cue (SOC, by MicroSleep, LLC) application emitting low intensity sounds through earbud head phones	SnoreMonitorSleepLab (Adactive AB, ©Leif Soderberg,Sweden);Quit Snoring (by Pointer Software Systems, Israel); Snore Spectrum (ZURLIN Technologies), all microphone based	Toss ‘N’ Turn (TNT), app for Android (7 sensor inputs: accelerometer, microphone, ambient light etc ).
Hardware features	Personal iPhones (models 4S and 5) running iOS5.0 or newer	One smartphone and one contact free, under-the-mattress piezoelectric sensor	iPhone (Apple). Placement near the pillow	iPhone 5s (iOS 8.0 operating system) with ear bud headphones (Skullcandy, model: S2DUHZ-335)	iPhone 4, iPhone 5, iPadmini, iPad 2ndGeneration	Android phone (version 4.0+)
Sleep Parameters yielded by Device/App	Graphic representation of Sleep Stages: Wake, Light Sleep, Deep Sleep.	HR; RR; TST; Sleep Stages: Wake, REM, nREM.	SOL; TST; WASO; SE	SOL	Snoring events (audio recording and count)	Bed time; Sleep duration; Wake time
Sleep Assessment	PSG	PSG	Wrist actigraphy: Actiwatch AW64	PSG	ApneaLink Plus (ResMed Germany Inc.,Martinsried, Deutschland) screening device	Subjective measures: Sleep Diary; PSQI

Study Design	One full night in-laboratory PSG study. Prior familiarization with app for 5 nights at home.	One full night in-laboratory PSG recording + PSG recording in home setting	Four consecutive overnight recordings at home	One over night in-laboratory PSG recording	One night with and one night without the OA Narval (ResMed Germany Inc., Martinsried, Deutschland)	One month study of sleep app sensor data collection in a variety of naturalistic sleep contexts + PSQI questionnaire
Sleep Parameters considered	Sleep Stages; TST; WASO; SE; Sensitivity and Specificity for sleep-wake detection.	TST; Sleep Stages: Wake, REM, nREM; Sensitivity; Specificity; Accuracy	SOL; TST; WASO; SE	SOL	Snoring events (audio recording and count)	Sleep quality features based on four factors of PSQI: Sleep duration, latency, efficiency, and disturbances + global sleep quality
Correlation between sleep parameters	Relatively high sensitivity (89.9%) for sleep-wake detection but poor specificity (50%). No significant correlation between PSG and sleep app absolute parameters.	High correlation between TST estimated with SE and PSG. Good sensitivity (92.5%), specificity (80.4%) and accuracy (90.5%) in sleep detection of SE compared to PSG. Good sleep staging capability.	Satisfactory epoch by epoch agreement for SOL, WASO, SE but not for TST between actigraph and smartphone accelerometer.	High degree of correspondence between PSG-determined sleep onset and Sleep On Cue sleep onset ( $r = 0.79, p < 0.001$ ).	The results of the apps did not correspond with the ApneaLink Plus screening device.	Differences of more than 30 min were found for all 3 parameters: in bedtime, sleep duration, and wake time.
Limitations	Small sample; Variability of setting; Variability of hardware; Raw data	High variability of sample; High variability of setting; Raw data	Small sample	Small sample; Raw data	Small sample; Raw data; Variability of hardware	Small sample; Variability of setting; Variability of hardware; Raw data

Abbreviations: PSG = Polysomnography; TST = total sleep time; SOL = sleep onset latency; WASO = wake after sleep onset; SE = sleep efficiency; HR = heart rate; RR = respiratory rate; REM = rapid eye movement; nREM = non rapid eye movement; LS= lights sleep; SWS= slow wave sleep; PSQI = Pittsburgh Sleep Quality Index; OA= Oral Appliance.



**Table 2****Characteristics and results of experimental studies on reliability of smartphone based monitoring of disturbed sleep**

	Patel et al. (2017)	Toon et al. (2016)	Tal et al. (2017)	Behar et al. (2015)	Nakano et al. (2014)	Camatcho et al. (2015)
Sample	25 children suspected for OSA (age 2-14 years)	78 children and adolescents (aged 3 to 18 years) suspected for OSA	43 patients with sleep disorders (age 17 to 72 )	121 patients	40 patients (mean age 47.9 years) with mild, moderate, and severe OSA	2 patients
Sleep Application/ Device features	Sleep Cycle App (accelerometer based) by Maciek Drejak Labs, now Northcube AB.	MotionX 24/7 App (accelerometer based) + Wrist worn actigraph UP by Jawbone	Smartphone Application + EarlySense (Ltd., Israel) contact free sensor	SleepApp Android phone application (accelerometer and microphone based) + an external pulse oximeter, WristOx2 3150 (Nonin Medical Inc., MN, USA) + STOP Bang questionnaire	Smartphone program (microphone based)	Quit Snoring' by Pointer Software Systems (microphone based) by Ramat Yishai, Israel
Hardware features	One iPhone 4S running SCA. Placement on the bed	iPhone 4 (Apple Inc., Cupertino, CA)	One smartphone and one contact free, under-the-mattress piezoelectric sensor	Android (versions 2.3.3 to 4.2.2)	SH-12C, Sharp Corp.Japan, Android system (version 2.33)	Adapted for for iOS 9.0 or later (compatible with iPhone, iPad, and iPod touch)
Sleep Parameters yielded by Device/App	Graphic representation of Sleep Stages: Wake, Sleep, Deep Sleep.	Time awake; Light Sleep; Deep sleep; Sleep Efficiency	HR; RR; TST; Sleep Stages: Wake, REM, nREM.	Classification in non OSA (healthy and snoring) and OSA (mild, moderate, severe)	Snore time (% total time); RDI	Snoring events (recording and playback); Recording Sensitivity
Sleep Assessment	PSG	PSG + Actigraphy: Actiwatch2	PSG	PG	PSG	PSG

Study Design	One full night in-laboratory PSG recording	One full night in-laboratory PSG recording + simultaneous actigraphy	One full night in-laboratory PSG recording + PSG recording in home setting	Clinician's diagnosis of home-based PG overnight recording	One over night in-laboratory PSG recording	One over night in-laboratory PSG recording
Sleep Parameters considered	Sleep Stages; TST, SOL; Time spent in each stage	SOL; TST; WASO; SE ; Sensitivity; Specificity; Accuracy	TST; Sleep Stages: Wake, REM, nREM; Sensitivity; Specificity; Accuracy	ODI; AC; AU; DE	Snoring time (% TST); AHI	Snoring events; Recording Sensitivity
Correlation between sleep parameters	Correlation of SL from the PSG and latency to deep sleep from the app was significant (p=.03). No significant correlation between TST or SL between the app and PSG.	SOL and WASO were significantly underestimated (12 and 63 min) resulting in significantly longer TST and greater SE (106 min and 17%) when assessed by the MotionX 24/7 compared to PSG.	High correlation between TST estimated with SE and PSG. Good sensitivity (92.5%), specificity (80.4%) and accuracy (90.5%) in sleep detection of SE compared to PSG. Good sleep staging capability.	Compared to the clinician's diagnosis the app showed an accuracy of 92.2% when classifying subjects as having moderate or severe OSA versus being healthy or snorer.	High correlation of snoring time measured by the smartphone with the PSG snoring time (% TST, r = 0.93, n = 40). High correlation of smart-RDI with the AHI obtained by PSG (r = 0.94). Diagnostic sensitivity of 0.70 and specificity 0.94 of smart-RDI for OSA (AHI ≥ 15).	Compared to PSG snoring sensitivities ranged from 64 to 96 %, and snoring positive predictive values ranged from 93 to 96 %.
Limitations	Small sample; Raw data	High variability of sample; Raw data	High variability of sample; High variability of setting; Raw data	Small sample size	Small sample size; Sound proof environment	Small sample; Raw data

Abbreviations: PSG = Polysomnography; PG = Polygraphy; TST = total sleep time; SOL = sleep onset latency; WASO = wake after sleep onset; SE = sleep efficiency; HR = heart rate; RR = respiratory rate; REM = rapid eye movement; nREM = non rapid eye movement; LS= lights sleep; SWS= slow wave sleep; PSQI = Pittsburgh Sleep Quality Index; OA= Oral Appliance; OSA= obstructive sleep apnea; RDI= respiratory disturbance index; AHI=apnea - hypo apnea Index; ODI= oxygen desaturation index; AC= accelerometer; AU= audio; DE= demographics.