Sleep Assessment Devices

Types, Market Analysis, and a Critical View on Accuracy and Validation

Abstract— <u>Introduction:</u> **Sleep assessment devices are essential for the detection, diagnosis, and monitoring of sleep disorders. This paper provides a state-of-the-art review and comparison of sleep assessment devices and a market analysis.**

<u>Areas covered:</u> **Hardware devices are classified into contact and contactless devices. For each group, the underlying technologies are presented, paying special attention to their limitations. A systematic literature review has been carried out by comparing the most important validation studies of sleep tracking devices in terms of sensitivity and specificity. A market analysis has also been carried out in order to list the most used, best-selling, and most highly-valued devices. Software apps have also been compared with regards to the market.**

<u>Expert opinion:</u> **Thanks to technological advances, the reliability and accuracy of sensors has been significantly increased in recent years. According to validation studies, some actigraphs present a sensibility higher than 90%. However, the market analysis reveals that many hardware devices have not been validated, and especially software devices should be studied before their clinical use.**

***Keywords*— Actigraphy; Sleep detection methods; Sleep devices; Sleep quality assessment**

***Article Highlights*—**

- Classification of sleep devices into contact and contactless devices. Review of underlying technologies.
- Comparative analysis of formal validation studies that measure sensitivity and specificity of sleep devices.
- Analysis of the evolution of sleep hardware devices and sleep apps in the market.

1. INTRODUCTION

Sleep is an essential part of a healthy lifestyle. Assessing the sleep pattern is fundamental to detect sleep disorders, but also to detect other diseases for which an abnormal sleep cycle is a symptom or indicator. The traditional methods to assess the sleep quality in hospitals and sleep centres are the polysomnography [56, 48, 2] (which is fully objective) and the sleep questionnaires and diaries (which are mostly subjective). A polysomnogram (PSG) is arguably the most advanced method for sleep assessment. While an individual sleeps, it combines sophisticated tests that monitor the brain (electroencephalogram), heart (electrocardiogram), and muscle (electromyogram) activity, snoring oxygen (pulse oximetry) and carbon dioxide (capnography), eye movement (electrooculogram), etc. For this reason, the PSG is the gold standard for sleep assessment (see, e.g., [61, 16, 18, 31, 1, 61]).

Unfortunately, the PSG is very expensive, it requires the assistance of specialized professionals and, often, studies are carried out only for short periods of time (e.g., one or two days). Moreover, patients do not sleep normally in a hospital bed with numerous sensors on their body, therefore, the assessment made by a PSG does not assess the usual sleep pattern the individual has at home. All of the above led to the need of a hardware device that could be used at home, and which produced objective and reliable reports of the sleep patterns. The first attempts to create such device date from the 1950s (as documented in [70], for example) where a mechanical device was already used to assess the sleep quality at home.

Nowadays, the most extensively used device of this kind is known as actigraph. The technique of sleep-wake cycle recording is known as *actigraphy*—word formed from the Latin term *actio* (action), which means "action" or "activity", and the Greek term γράφειν (gráphein), which means "writing" or "recording"—because it records the activity of (some part of) the body. Actigraphs have evolved significantly over time. Especially in recent years, new methods have emerged with the appearance of new technologies, such as mobile apps and advanced hardware sensors like galvanic skin response measurement.

In this article, we review current devices to detect sleep quality. All these devices can be classified into two groups: *contact devices* and *contactless devices*. Both groups have specific characteristics, and, thus, they are studied separately first, and later compared. In the study, we include hardware and software applications. In fact, the same hardware can be controlled by different software algorithms that produce different results when applied in the same context. This article complements previous reviews of sleep assessment methods introducing an up-to-date classification that includes a study of the evolution of this technology in the market and a critical view on the accuracy and validation of these methods. This complements those partial reviews that only focus on a specific subset of methods (e.g., sleep questionnaires [62, 16, 18, 50, 1, 64, 23], mobile apps [27, 47], contact sleep detection methods [25, 20, 35], etc.); and it also completes the information missing in general reviews [22, 60, 71, 8] and outdated reviews (see, e.g., [31, 24, 72]).

One objective of this article is to introduce this technology to the public and to professionals as a valid method for tracking sleep. For this reason, in this paper, there is a balance between the objective presentation of technical aspects and the reviews of previous studies. Each technique presented is summarized on a table with the most commonly used market products, and with previous studies that show their accuracy. Of course, not all the devices and software applications presented have the same functionality and accuracy. Therefore, critical discussion on validation is necessary. We also present the reliability and validity of the methods, analysing previous comparisons and validation studies.

As technology evolves, a new "commercial" device can be more accurate than an old "professional" one. For this reason, we have not classified the devices as professional or commercial. We have described them and their underlying technology; and we have compared them using two objective measurements:

1) Sensitivity (ability to detect sleep, true positive rate) and specificity (ability to detect wake, true negative rate) reported in formal validation studies.
2) Satisfaction level of real users (through a market study). This information is also very important for sleep professionals. It should be used in combination with objective number 1 and it is associated with usability or adherence, for example.

The market analysis is essential to know the evolution of the technology. For this reason, this article complements the study of validated devices, with a report about the most used devices according to the most important selling platforms (Amazon for hardware and Google Play for software). In both cases, the products evaluated have thousands of user reviews, and they have been monitored for a long period of time (almost two years). This allows us to extract statistically valid conclusions.

2. Systematic Methodology

We have followed the PRISMA guidelines for the transparent reporting of systematic reviews and meta-analyses [41]. Therefore, we started the literature review with a planning phase where we formulated research questions and defined inclusion and exclusion criteria. This phase was followed by search and screening of primary studies.

2.1. Research questions

We started with the formulation of two research questions:

- *What technologies for sleep assessment have been developed?*
  The aim of this question is to provide an overview of the technologies used in sleep assessment devices (especially those developed over the last 10 years).
- *What are the main characteristics of each sleep assessment device?*
  This question complements the previous one, giving a deeper understanding of sleep assessment devices.

2.2. Search process

The purpose of a literature review is to review the relevant studies in order to assess the body of knowledge that exists to support addressing the research questions. The whole process must be rigorous and unbiased so it must involve a wide coverage of relevant sources, such as online databases, journals, and conferences. We started the process of identifying relevant literature in the following electronic databases: PubMed, LILACS, TOXNET, SCOPUS, ScienceDirect, Web of Science, and Google Scholar.

After an analysis of the keywords from the relevant literature, which were found on several general searches in these resources, we created the following search string to retrieve information from the electronic resources and databases:

(device OR technology OR system) AND (sleep)

In various databases the search string produced thousands of results, thus, we refined it using TOXNET as:

(device) AND (sleep)

and on SCOPUS, ScienceDirect, Web of Science, and Google Scholar as:

"sleep device"

This search process identified 132 studies. Excluding unavailable and duplicated results, we obtained 84 studies.

The search process was also performed for the market analysis. From January 2017 to October 2019, we monitored the most popular devices in the main selling platforms of the market: Amazon and Google Play. Throughout that period, we registered the evolution (prices, number of user reviews, and average score of the user reviews) of the top ten most popular devices every month in both platforms.

2.3. Inclusion and exclusion criteria

From the bibliography found in the databases, we defined the following inclusion and exclusion criteria to address the research questions:

- *IC1: papers that presented a sleep assessment device or its validation were included.*
- *IC2: papers that described the characteristics of a sleep assessment device were included.*
- *EC1: papers that did not describe a sleep assessment device were excluded.*

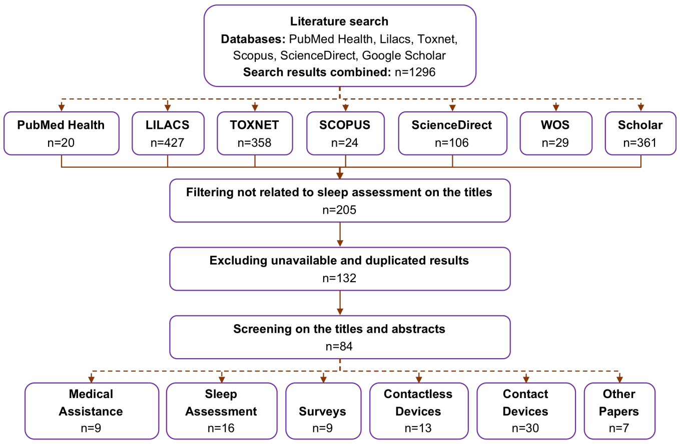
With respect to the market devices there was a unique inclusion criterion:

- *IC3: devices within the 10 best-selling in Amazon or Google Play.*

2.4. Studies Selection

Firstly, we screened through the titles and abstracts to decide whether to include or exclude each study. From the results found in the databases, we selected a total of 132 studies from which 48 were excluded. Then, we read in detail the full text of each primary study included in the preliminary selection to decide whether to include or exclude the study. The primary studies included in the final selection correspond to the relevant papers that met the research questions set out in this study. See the QUOROM flow chart of the reviewing process in Figure I.

FIGURE I
QUOROM FLOW CHART OF THE REVIEWING PROCESS



2.5. Data extraction

We extracted the information from the papers included in the primary studies set. We classified each paper as review, opinion, study, tool description, etc., and identified the kind of devices described. The data was grouped by types of sleep assessment devices (contact/contactless). The information extracted from Amazon and Google Play was the actual price and evaluation of each studied device published in the official Amazon and Google Play websites every month.

3. Hardware to Detect Sleep/Wake Patterns

There is a wide range of technologies used to detect sleep patterns. All of them can be classified as *contact devices* or *contactless devices*, depending on whether they need to be in contact with the individual's body during sleep or not. Those devices that are based on the echo produced by signals can be further classified into Sonar, Radar, and Lidar devices. All of them will be explained in a separate section.

3.1. Contactless sleep detection devices

Even though the accuracy of a contactless method may seem imprecise due to its obvious limitations, there are several contactless devices that are becoming relatively popular because the latest advances in technology have made them a valid alternative. Contactless sleep detection methods are based on one or (usually) several of the following technologies:

- *Microphone*. It monitors the volume and type of snoring [44]. It also monitors the sounds produced by the movements of the body and the bedding, and sleep talking. Advanced devices can record these sounds when they occur and reproduce them for the user or the healthcare provider. Recent research [73] shows that, with only a microphone, machine learning systems can detect sleep when they are properly trained to do so. Example: *S+*™.
- *Video camera*. It monitors the movements done while being asleep. These movements can also be recorded automatically. Example: *SAMi*™.
- *(Infrared) thermometer*. While standard thermometers can measure the room temperature, infrared thermometers can be directed to the user's body and monitor the temperature changes with a high sensitivity. Temperature charts are combined with other parameters to detect sleep phase changes. While standard thermometers are widely used, infrared thermometers are an emerging technology, still under development. Example: *Withings Aura*™.
- Different companies are researching on how to monitor the user's body temperature from the distance while sleeping. Infrared (or laser) thermometers are the most important technology currently under research. The three main challenges are:
  1. Improving accuracy (measurement errors are ±1.5% / 1-2ºC / 2-4ºF).
  2. Reducing the *distance-to-spot ratio* (D:S), ratio of the distance to the measurement surface D and the diameter of the temperature measurement area S. Current devices are 12:1.
  3. Detecting the user and their movements.
- *Pressure strap or belt.* It is placed underneath the sheet or mattress protector, and the user sleeps on top of it. In order to attach it to the bed correctly, it is clipped onto the sheets with magnets. It monitors the movements of the body. Example: *Sleepace Sleep Dot*™.
- *Accelerometer.* It is a small mechanism that measures proper acceleration and can be used to detect multi-axis motion (its underlying technology is discussed in Section 3.2.1). It is often distributed as a small device that clips onto the pillow or mattress and records the movements made throughout the night. Example: *HugOne Sleep Tracking System*™.
- *Echo based devices*. It detects the body movements by sending out periodic signals against the body and then measuring how long it takes for the signals to return or bounce back [28]. Currently, there are three different types of technologies that can be used to achieve this:

◊ *Sonar (short for sound navigation and ranging)*. It works by emitting ultrasound pulses (i.e., sounds at frequencies above the maximum range of human hearing) and measuring how long the echo takes to return. This exact same technology used by bats and dolphins, for example, and it can be implemented with just a speaker and a microphone. Example: *Sleep as Android*™.

◊ *Radar (short for radio detection and ranging)*. It uses the same principle used in sonar, but without the use of sound waves. Instead, it uses radio waves (electromagnetic radiation), invisible to the human eye, because they have longer wavelengths than visible light. Examples: *SleepScore Max*™ and *DoopleSleep*™ [54].

◊ *Lidar (short for light detection and ranging)*. Similar to the radar, but it uses light waves instead of radio waves. Sleep detection systems based on lidar technology increase resolution and accuracy compared to those based on radars. This emerging technology is still under development.

3.1.1. *Contactless sleep detection devices in the market*

We contrasted several experts' reviews (see, e.g., [26, 20, 3]) that provided comparison reports on the accuracy of different hardware devices in the market. The usual method to compare them is to use them all through a period of time (one or two weeks) to determine the sleep quality of several individuals being recorded with a video camera. Then, all the sleep reports are compared and checked against the actual recorded sleep. We combined the expert's reports with the market sales and customers' opinions to compare them and we realized that the expert's recommendations mostly coincide with the customers' opinions. It is of special relevance for sleep experts to know the market and its trends. In Table I, we list the main contactless sleep detection devices according to Amazon (they also appear among the top-rated devices in the experts' reviews [26, 20, 3]).

TABLE I
CONTACTLESS SLEEP DETECTION DEVICES
(PRICES AND REVIEWS ARE TAKEN FROM AMAZON.COM)

Device name	Developer	Price	Average review score	Main technology
Withings Sleep	*Withings*	99.95$	3.7 out of 5	Pressure belt
Ocho10006	Eight	199.99$	3.6 out of 5	Pressure Mattress
Withings Aura	Withings	69.99$	3 out of 5	Pressure strap
Beddit 3.0 Smart Sleep Tracker	Beddit	128.99$	3 out of 5	Pressure belt
S+	ResMed	55.95$	3 out of 5	Microphone, Camera, thermometer
HugOne Sleep Tracking System	SevenHugs	65.00$	2.9 out of 5	Mattress Accelerometer
Sleepace Reston	Sleepace	249.99$	1.9 out of 5	Pressure belt

A smartphone can also use its camera, microphone, and accelerometer to detect the sleep quality. For this reason, many apps have been developed and can be accessed from public repositories such as iOS' App Store (http://www.appstore .com) and Google Play (http://play.google.com). Table II lists the main contactless sleep detection apps according to the number of reviews in App Store and Google Play. Even though several experts' reviews [19, 25] also argue in favour of the apps in Table II, for some of them, there is not any research that formally validates the app against gold standard. In fact, various formal comparisons of sleep apps [25, 47, 51] indicate substantial skepticism about their effectiveness (see Section 4). As per the market, the most valued (customer's opinion) contactless hardware device to detect sleep is *Withings Sleep*. The best mobile app to detect sleep is *Sleep as Android unlock*. The app *Sleep Cycle Alarm Clock* was validated in [51] with 25 children that were undergoing overnight PSG simultaneously with that app running on a phone placed on their mattress. Correlation was not found between the app and the PSG, and the conclusion was that the app could be useful to raise the user's awareness about sleep issues, but was not yet accurate enough to be used as a clinical tool.

TABLE II
CONTACTLESS SLEEP DETECTION APPS
(PRICES AND REVIEWS ARE TAKEN FROM GOOGLE PLAY)

App name	Developer	Price	Average review score	Number of reviews
Sleep as Android Unlock	Urbandroid Team	5.99$	4.8 out of 5	26,834
Sleep as Android	Urbandroid Team	0$	4.6 out of 5	291,618
Sleep Time	Oleg Filimonov	0$	4.6 out of 5	2,121
SleepAway	Samuel Banas	0$	4.6 out of 5	2,322
Sleep Cycle Alarm Clock	Sleep Cycle AB	0$	4.5 out of 5	92,814
Prime Nap Sleep Tracker	Boston IAB	0$	4.4 out of 5	1454
Jukusui	株式会社C2	0$	4.0 out of 5	19,717
Sleep Cycle	Azumio Inc.	0$	3.8 out of 5	30,683
Sleep Analyzer	A1 Brains Infotech	0$	3.2 out of 5	2,448

3.2. *Contact sleep detection devices*

Contact-based sleep detection devices are all those devices that use sensors to be placed in contact with the body. The most precise device is the PSG [48, 56], which is used in hospitals and sleep centres. But there are also reduced versions of a PSG that are portable and can be used at home. In the commercial side, there exist EEG monitoring devices, which have shown a sensitivity of 98% [6].

The most popular contact sleep detection devices are small gadgets often attached to the wrist, ankle, chest or head (e.g., Fitbit Charge 2™). These devices are often called *actigraphs* because they use Cartesian representation to record the activity of the body. Current actigraphs provide a fairly reliable (although not exact) measurement of when, and for how long, an individual sleeps. They use an accelerometer to collect body movement data, and they often incorporate other sensors to measure body temperature, heart rate, respiratory rate, or even the galvanic skin response.

It is often believed that actigraphy is a new method, exclusively based on electronic technology. However, it has been documented [70] that actigraphy was already used in the 1950s to assess psychological disorders (those actigraphs were exclusively based on mechanical sensors).

In 1995, the American Academy of Sleep Medicine

(AASM) recognized actigraphy as a useful research tool for the sleep study. Later, in 2002, it was improved and used as a clinical tool to measure sleep. In 2007, actigraphy was included in category 3 (emerging technology) of the Current Procedural Terminology (CPT) codes of the American Medical Association (AMA). Finally, in 2009, actigraphy was included in category 1.

3.2.1. Actigraphy: The underlying technology

Actigraphs use different body-fixed (or wearable) motion sensors:

- *Gyroscopes.* These sensors are used to determine orientation and angular velocity. A gyroscope is a device that consists of a freely-rotating disk called rotor, mounted onto a spinning axis in the centre of a larger wheel. As the axis turns, the rotor remains stationary according to the conservation of angular momentum. Therefore, a gyroscope can determine which way is "down", for example. Gyroscopes can be used to assess the sleeping position.
- *Accelerometers.* These are sensors used to measure the proper acceleration of objects in relation to a multi-axis. Measuring acceleration is particularly interesting because it is a physical magnitude that is proportional to the force that causes the acceleration itself. Therefore, it can be used to reflect the intensity of a human movement. By combining accelerometry data with time, we can also determine velocity and displacement [11]. Modern accelerometers use gravity to determine inclination. These tilt sensing can be used to identify body postures (orientation, such as standing up and lying down).

  Internally, accelerometers use a mechanical suspension system with a proof mass attached. Any inertial force due to acceleration (or gravity) causes the proof mass to deflect according to Newton's Second Law. Then, the acceleration is electrically measured by observing the movements of the proof mass with respect to a reference frame. The most common types of accelerometers are piezoresistive, piezoelectric, and differential capacitive accelerometers [46,21,12].
- *Other sensors.* Many actigraphs are exclusively based on motion detection (i.e., accelerometers and gyroscopes); but some modern actigraphs are multi-sensor, and they integrate other sensors (some of them are described in Section 3.1) that can enhance the sleep assessment with complementary information. Most common sensors measure the skin temperature, galvanic skin response, or heart rate (body measures); and also, ambient temperature, sound levels, or light (ambient measures). All these measures are synchronized and combined by an algorithm to assess the sleep and detect events such as, e.g., awakenings.

3.2.2. Accuracy of actigraphs

The word "actigraph" has been used to refer to many different types of devices with a wide variety of purposes: from activity trackers, which can track 24-hours activity to measure sleep/wake time, energy consumption, etc. to professional devices specialized in measuring sleep. It is important to distinguish between professional actigraphs and commercial actigraphs. Several articles have evaluated the accuracy of actigraphy, and they concluded that actigraphy is a valid method to assess the sleep. Some representative studies for different populations (older women, adults, children, individuals with mental health problems, etc.) are:

[7] *Blackwell T. et al., 2008*: this study targeted 68-year-old women (mean age of 81.9 years). Three different configurations of an actigraph were compared against PSG. They concluded that the actigraphy report corresponded reasonably well to PSG. The best results were produced by the proportional integration mode.

[31] *Martin J.L., Hakim A.D., 2011*: this study compared the accuracy of actigraphy to PSG. The main conclusion was that wrist actigraphy was a precise method to estimate total sleep time and waking time. Conversely, precision decreased when measuring awakenings, because wrist actigraphy could hardly differentiate between sleep and wakefulness when motionless (e.g., lying in bed watching television).

[30] *Marino M. et al., 2013*: the study validated actigraphy for detecting sleep and wakefulness using PSG as gold standard. The sleep quality of 77 patients was simultaneously measured with actigraphy and PSG. The authors concluded that actigraphy was a valid means for estimating total sleep time and wakefulness after sleep onset, with some limitations in specificity.

[4] *Baandrup L., Jennum P.J., 2015*: this study validated actigraphy with respect to PSG with a sample of 37 chronic, medicated patients with schizophrenia or bipolar disorder. The authors concluded that actigraphy produced reliable measurements of total sleep time for this population. Other parameters were sensitive to extensive periods of wakefulness after sleep onset.

[35] *Meltzer L.J. et al., 2016*: the study validated actigraphy with respect to PSG with a sample of 148 children (ages 5-12). The study concluded that actigraphy was a valid tool for this age group (see Table III). The main drawbacks were that actigraphy underestimated total sleep time by 30 min, and sleep efficiency by 5%; and it overestimated sleep onset latency by at least 10 min for a third of the children.

Other validation studies such as [15, 40] have similar conclusions. In Table III, we summarize the results obtained in various validation studies. All the studies in the table coincide in stating that actigraphy is a useful tool, especially to detect sleep: sensibility is higher than 86% in all studies, but its ability to detect wake is still weak: specificity is below 66% in most cases, and even below 20% in some others [42]. The information contained in this table should be considered with caution. While each validation study provides useful information about the sensibility and specificity of a given device for a specific population, the results reported by different studies are in most cases incomparable. This means that the fact that one study reports a sensibility higher than another study (e.g., [14] reports a sensibility of 96% for

Jawbone and [37] reports a sensibility of 86% for Fitbit Ultra) does not necessarily imply that Jawbone is more sensitive than Fitbit Ultra, because these validation studies were made with different populations (youth people vs. insomnia and healthy people) and in different sleeping contexts.

TABLE III
STUDIES THAT VALIDATE SLEEP DETECTION HARDWARE DEVICES

	Reference	**Hardware device**	**Gold standard**	**Sample (Age)**	**Sample's health status**	**Results**
Professional Devices	[68] de Souza et al., 2003	Mini Motionlogger Actigraph - Basic 32	PSG	21 [18-33]	Healthy adults	Sensitivity: 0.97 Specificity: 0.44
	[66] Sivertsen B. et al., 2006	Actiwatch Plus	PSG	34 (60.5±4.5)	Chronic primary insomnia	Sensitivity: 0.95 Specificity: 0.36
	[30] Lichstein et al., 2006	AW64 Actiwatch	PSG	57 [21-87]	>6 months insomnia	Sensitivity: 0.94 Specificity: 0.61
	[49] Paquet J. et al., 2007	Actiwatch L	PSG	15 (39.3±15.1)	Caffeine consumers (1-3 coffee cups per day	Sensitivity: 0.95 Specificity: 0.54
	[6] Berthomier C. et al., 2007	Portable single EEG channel ASEEGA	PSG	15 (29.2±8)	Healthy adults	Sensitivity: 0.98 Specificity: 0.83
	[65] Sitnick S.L. et al., 2008	Actiwatch AW64	Video somnography	58 (47.8±12.7)	Parents of children with autism or development delays	Sensitivity: 0.97 Specificity: 0.24
	[45] Natale et al., 2009	Basic Mini-Motionlogger	Event recording (by users) and sleep diary	408 (40.39±14.28)	Insomnia patients	Sensitivity: 0.66 Specificity: 0.61
	[38] Meltzer et al., 2012	Motionlogger	PSG	115 (8.8±4.4)	Healthy youth or with sickle cell disease	Sensitivity: 0.92 Specificity: 0.65
	[38] Meltzer et al., 2012	Actiwatch-2	PSG	115 (8.8±4.4)	Healthy youth or with sickle cell disease	Sensitivity: 0.93 Specificity: 0.69
	[33] Marino M. et al., 2013	AW-64 and Actiwatch Spectrum	PSG	77 (35±12.5)	Insomnia patients, healthy adults, night-workers	Sensitivity: 0.97 Specificity: 0.33
	[43] Nakazaki K. et al., 2014	FS-750 actigraph	PSG	34 (21.9±1.7)	Healthy adults	Sensitivity: 0.94 Specificity: 0.57
	[63] Shin M. et al., 2015	Actiwatch-2	PSG	9 (23.3±4.1)	Healthy adults	Sensitivity: 0.95 Specificity: 0.45
	[63] Shin M. et al., 2015	SenseWear	PSG	9 (23.3±4.1)	Healthy adults	Sensitivity: 0.93 Specificity: 0.57
	[39] Meltzer L.J. et al., 2016	Actiwatch-2	PSG	148 (9.3±2)	Children born preterm	Sensitivity: 0.88 Specificity: 0.46
	[36] Matsuo M. et al., 2016	Actiwatch	PSG	20 (20.7±0.39)	Healthy adults	Sensitivity: 0.93 Specificity: 0.16
	[36] Matsuo M. et al., 2016	MTN-210	PSG	20 (20.7±0.39)	Healthy adults	Sensitivity: 0.78 Specificity: 0.57
	[36] Matsuo M. et al., 2016	SleepScope	PSG	20 (20.7±0.39)	Healthy adults	Sensitivity: 0.92 Specificity: 0.70
	[52] Pigeon et al., 2018	MyCadian	PSG and actigraphy	20 [18-64]	Healthy adults	Sensitivity: 0.94 Specificity: 0.75
	[53] Quante et al., 2018	GT3X+	PSG	35	Healthy children and adults	Sensitivity: 0.96 Specificity: 0.64
Commercial Devices	[42] Montgomery-Downs et al., 2012	Fitbit	PSG and actigraphy (Actiwatch-64)	24	Healthy adults	Sensitivity: 0.98 Specificity: 0.20
	[37] Meltzer et al., 2015	Fitbit Ultra	PSG and actigraphy	63 (9.7±4.6)	Youth people	Sensitivity: 0.86 Specificity: 0.52
	[14] De Zambotti et al., 2015	Jawbone	PSG	28 (50.1±3.9)	Insomnia and healthy people	Sensitivity: 0.96 Specificity: 0.37
	[7] Bhat et al., 2015	Sleep Time app	PSG	28 [22-57]	Healthy adults	Sensitivity: 0.90 Specificity: 0.50
	[69] Toon et al., 2016	Jawbone UP and MotionX 24/7	PSG and actigraphy	78 (8.4±4)	Suspected sleep disordered breathing	Sensitivity: 0.92 Specificity: 0.66
	[29] Lee et al., 2019	Fitbit Alta HR	PSG and actigraphy	17 [15-19]	Healthy adolescents	Sensitivity: 0.90 Specificity: 0.88
	[57] Roomkham et al., 2019	Apple Watch	Actigraphy (Actiwatch Spectrum Pro)	14 >18	Healthy adults	Sensitivity: 0.99 Specificity: 0.79

TABLE IV
CONTACT SLEEP DETECTION DEVICES
(PRICES AND REVIEWS ARE TAKEN FROM AMAZON.COM)

Device name	Developer	Price	Number of reviews	Average review score	2-Year variation
Fitbit Charge 3	Fitbit	169.99$	61	4.4 out of 5	-9.30%*
Bellabeat Leaf Urban	Bellabeat	137.11$	1,439	3.7 out of 5	-15.91%
Spire Mindfulness	Spire	129.95$	1,740	3.7 out of 5	-7.5%
Garmin Vivofit 2	Garmin	49.99$	1,830	3.6 out of 5	0%
Misfit Ray	Misfit Wearables	59.95$	620	3.0 out of 5	-9.09%
Withings Pulse O2	Withings	89$	1,782	2.9 out of 5	-17.24%
Misfit Shine 2	Misfit Wearables	59.95$	154	2.6 out of 5	-6.25%

**Fitbit Charge 3* is new and no data is available yet. This variation corresponds to Fitbit Charge 2.

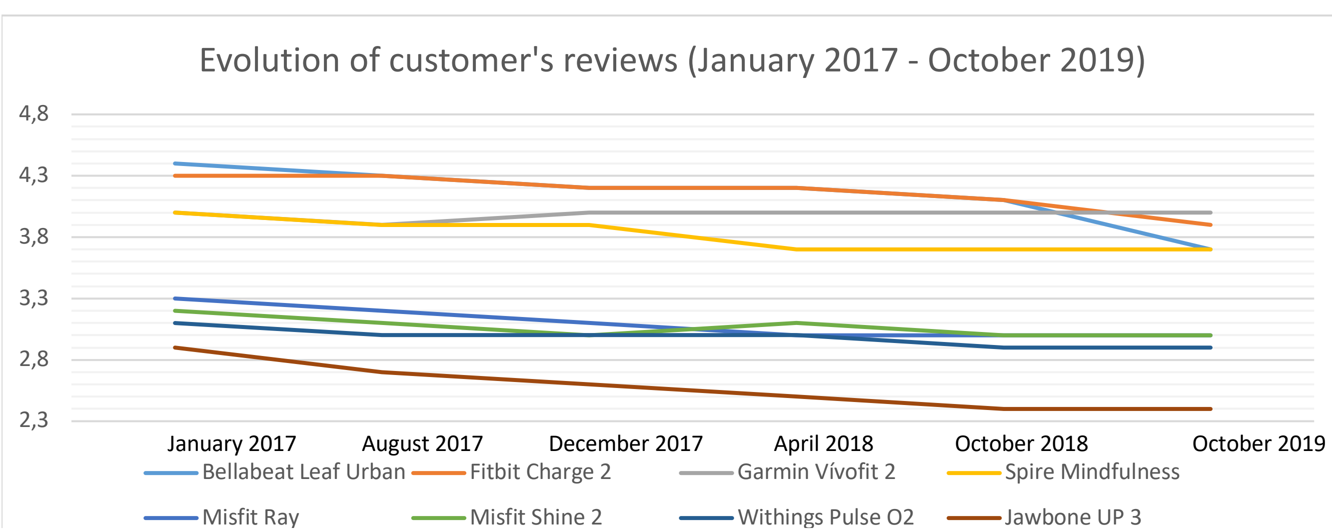
In order to fairly compare two sleep devices, we must design the experiment so that both devices are used at the same time and in the same arm. Moreover, a PSG should be used as gold standard because it is more reliable than actigraphy and video recording (e.g., PSG can detect the sleep phases).

3.2.3. Contact sleep detection devices in the market

Table IV lists the main contact sleep detection devices according to Amazon. These devices are also listed as the top most reliable devices in several experts' reviews such as [35], and have been compared in validation studies [17, 25]. Only the most advanced version of each device is shown. Note that this influences the number of reviews, because some devices are new in the market. For instance, while *Fitbit Charge 2* has 17,354 reviews in Amazon, *Fitbit Charge 3* only has 61. In the table, only *Fitbit Charge 3* has been included because it is the latest version (of course, most of the following comments will be written for the new version). Regarding the market, the most valued (customer's opinion) contact hardware device to detect sleep is *Fitbit Charge 3 (Fitbit Charge 2's average reviews score is 3.9/5)*.

We have been monitoring the market for over one year and a half to study the evolution of this technology in terms of price, lifetime, and reviews. This provided not a static picture of the state of the actual practice, but a record of its evolution and the speed of its changes. The main conclusion is that this is a dynamic and very competitive market. Throughout the last 34 months (from January 2017 to October 2019) many of the devices have aggressively changed their market price. For instance, *Misfit Shine 2* changed from 99.99$ to 59.95$ and *Withings Pulse O2* from 98.99$ to 89$. Moreover, prices fluctuate widely with a clear commercial purpose. For instance, in April 2018, *Misfit Ray* increased its price from 67.77$ to 99.99$. This allowed them to later decrease the price with a special offer that included a 45% discount. Not only prices, but also the products themselves are quickly replaced by new versions, or they just disappear (e.g., some devices that we were monitoring, such as Jawbone UP 3 disappeared for a few months from the market in October 2018). Another interesting observation is that the average score of all devices

FIGURE II
EVOLUTION OF THE CUSTOMERS' REVIEWS OF CONTACT SLEEP DETECTION DEVICES (REVIEWS FROM AMAZON.COM)



has decreased over time in almost all cases (the only exception is *Garmin Vívofit 2*), see the right column of Table IV, and Figure II. In particular, one device with a rate of 3.9/5 can be found to have a rate of 3.7 the next year just because it misses a novel functionality recently introduced by another device. Note that the review rate is accumulative, which implies a radical change in the new reviews.

For instance, consider a real case: a device had 1,100 reviews with an average rate of 3.9/5, and there have been 200 extra reviews in the last year. If the new average is 3.7/5, then the average score of the new 200 reviews is 2.6/5 ($3.7 * (1100 + 200) - 3.9 * 1100/200 = 2.6$).

4. A CRITICAL VIEW OF ACCURACY AND VALIDATION

Contact and contactless devices use different sensors to assess the sleep. Contact devices are often superior to contactless devices in accuracy [24,27] and, thus, produce more accurate results because most of the sensors used to detect sleep depend on their proximity to the user. Clear examples of this are accelerometers, because they are used in contact devices (such as wrist watches) and also in contactless devices (such as mobile phones). Even though the technology is the same in both cases, wrist watches are significantly more accurate than mobile phones [27]. The reason being is that a direct measurement of the body movements is more reliable than an approximation based on a measurement based on mattress or pillow movements. The situation is the same for sonars. The effective range of a phone that uses the microphone and the speaker to produce and receive ultrasounds as sonar is about 1 metre, and the reliable distance is about 0.5 metres. Obviously, the results become less precise the further away the user is. Moreover, the movements of the phone (sonar) itself do affect the results. Hence, it is preferable to place the device still, on a bedside table, instead of placing it on the mattress. This way the phone will be at least within half a metre distance. The same problem happens with similar technology that uses radio frequency to monitor the breathing and body movements.

4.1. Accuracy of hardware devices

Not only hardware, but also software has a big influence on accuracy. Currently, there are over 84,000 health app publishers and 325,000 health apps (health & fitness and medical apps) available in all major app stores [55]. Many of these apps implement different proprietary sleep detection algorithms. Therefore, one single device (e.g., a mobile phone with sensors such as an accelerometer) actually produces different results depending on the software used to detect sleep. The impact of these apps is immense. It has been estimated that 3.6 billion health apps were downloaded by users in 2017 [55].

What is even more important is that most of these sleep detection apps have not been validated—in fact, most of them were implemented by amateur programmers—and reliability of the software used is as important as the reliability of the sensors used. An interesting comparison and discussion on the accuracy of sensors can be found in [27]. Of course, there have been many studies performed in order to validate hardware devices and report on their accuracy and precision. Some important studies are presented in Table III. Validation of a device means "*confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use can be consistently fulfilled*" (Title 21 Code of Federal Regulations Part 820.3). In this area, the usual method for validation is the output comparison epoch by epoch between the device and a PSG, which is considered the gold standard. As a result of the validation, the study quantitatively reports how good the device is to assess sleep (i.e., sensitivity: time period when the participant is asleep and it is classified as asleep), and how good the device is to assess wakefulness (i.e., specificity: time period when the participant is awake and it is classified as awake). We refer the interested reader to [17] and [25] for systematic reviews of validation studies on sleep detection hardware devices.

The accuracy of actigraphy is dependent on the sampling frequency. In general, all electronic devices capture data using epochs. An epoch is a configurable period of time that usually ranges between 10 seconds and 1 minute. The use of epochs discretizes the timeline and it avoids continuous recording, hence saving a lot of space. For sleep tracking, 10 seconds is the recommended length of an epoch, because 15 seconds is the barrier used to determine sleep latency (sleep latency is the time for the first epoch with over 15 seconds of any stage of sleep [2]).

Shorter epochs lengths collect more activity data and they produce a higher resolution dataset. For instance, 10 minutes of 10 second epoch data will yield 60 data points whereas 10 minutes of 60 second epoch data will only produce 10 data points). Of course, this better resolution comes with a cost: shorter epochs create significantly larger files that can fill the memory of the device and that are slower to process. Shorter epochs also reduce the battery life. The size of the epoch depends on what we want to measure. To measure total sleep time, an epoch length of 1 minute is often used. However, for other measurements, a shorter epoch length is essential. For instance, if we want to measure the number of awakenings (wake period > 10 seconds), then we will need an epoch length of 10 seconds (because if an individual was awake for less than one minute that could not be detected with 60-second epochs). If we want to measure the number of arousals (wake period <= 10 seconds), then we would need an epoch length of 1 second. If we want to compare the actigraphy data with PSG data (e.g., to validate an actigraph), then 30 seconds is the recommended epoch length because it matches the 30 seconds epochs PSG scoring.

Another important factor to consider is the algorithm used to analyse the data. For instance, one largely used algorithm is Cole et al's algorithm [13], which can only process data in epochs of 60 seconds. If the epoch length is smaller than 60 seconds, then data is regrouped before the analysis. Therefore, the software that uses that algorithm (e.g., Actilife™) usually recommends a configuration with 60 seconds epoch length.

However, if, in the future, we want to analyse the data with another algorithm we will be limited.

Therefore, it is recommended to reduce the epoch length as much as possible whenever the device is not out of battery too fast. Most current devices can collect data with an epoch length of 10s for two weeks, and this is the reason why 10 seconds is the standard recommendation: good precision, long battery, and ability to detect all awakenings and many arousals.

While actigraphy has proven to be a useful clinical tool, with high sensitivity (see Table III), many types of populations have not been studied yet. Therefore, as a general advice, diagnosis of a disease should not be based only on actigraphy reports.

5. Conclusion

Sleep assessment and monitoring allows us to detect sleep disorders such as insomnia, parasomnia, hypersomnolence, or circadian rhythm sleep-wake disorders. Modern sensor-based devices such as actigraphs allow us to make this assessment at home, being able to produce objective reports on normal sleep for long periods of time. In this study, we have classified all current types of hardware devices to detect sleep into contact and contactless devices, and we have reviewed, compared, and discussed the state of the art (both, the literature and the current state of the practice and the market, providing up-to-date reviews on devices and apps).

The market analysis showed the most important hardware devices and apps in terms of popularity (higher score and higher number of reviews in Amazon and Google Play). The results of this study for contactless sleep detection devices are shown in Table I; for mobile apps, in Table II; and for contact devices, in Table IV. While most apps are free (all the studied apps are free except for the premium version of *Sleep as Android*, which costs 5.99$), the price of hardware devices vary widely. In general, contactless devices are more expensive. For instance, the price of the studied contact devices ranges between 45.95$ and 169.99$, while the price of the studied contactless devices ranges between 55.95$ and 249.99$.

There is not a perfect sleep assessment method. All methods have advantages and disadvantages, thus, they should be combined and adapted to the specific needs of the individual. However, as discussed, in terms of reliability and accuracy, contact devices are often superior to contactless devices. From the comparison of validation studies, we can conclude that current sleep detection hardware devices present an acceptable accuracy and reliability (see Table III). Traditionally, to avoid false negatives the industry gives preference to sensitivity with respect to specificity. In particular, the sensitivity of the validated devices ranges between 78% and 99%, while their specificity ranges between 16% and 88%.

Another important feature of this kind of methods compared with sleep questionnaires and sleep diaries, for example, is that there is a higher adherence to them, because they require less effort from the user (e.g., wrist actigraphy is mostly automatic).

Thanks to the technological advances, the reliability and accuracy of sensors has significantly increased in recent years. We have discussed the different existing, and also imminent, approaches to detecting sleep, such as lidars and infrared thermometers. A clear tendency in sleep hardware devices, no matter if they are contact or contactless devices, is that they can be directly managed and programmed with a smartphone. There already are commercial products often called *sleep trackers* that automatically provide reports directly visible in smartphones. Sleep trackers have been validated in different studies (see Section 3.2) targeting diverse populations. All studies coincide that current devices are reliable enough to be used as good sleep indicators, but they should be complemented with other methods to reach a diagnosis.

We have also reviewed the current most commonly used apps. Smartphone apps to assess sleep are widely used nowadays. In this respect, it is important to remark that, despite their use, the algorithms used in most apps are amateur implementations. Many proprietary algorithms do not pass enough quality controls, and some of them are even worse that the human inspection of the actigraphy data (see [10]). Hence, they must be validated at least before their clinical use. Fitness trackers and phone apps tend to underestimate sleep disruptions and overestimate total sleep times and sleep efficiency in normal participants [25].

The mean score of a sleep device in the market is not a good indicator of its quality because, in general, the opinion of customers regarding sleep devices is very contradictory. While some customers evaluate a device with a score of 5 out of 5, others evaluate the same device with a score 1 out of 5. This is for instance the case of the fitness tracker and sleep monitor *Misfit Shine.* It has 497 reviews in Amazon with an average score of 3.2 out of 5. However, the variance associated with this evaluation is huge: evaluations are 34% with 5 stars, 13% with 4 stars, 13% with 3 stars, 12% 2 stars, and 28% with 1 star. Table III shows that the same hardware device can show different accuracy depending on the target population (e.g., children vs. elderly people). Hence, even though the advances in technology are producing better sensors to assess sleep, most actigraphs are not yet prepared for clinical use. Their use should be preceded by validation studies that compare the specific hardware and software with a gold standard (e.g., a PSG) for a specific target population.

6. Expert Opinion

Technology is being improved continuously, which keeps producing better sleep devices and apps. Nevertheless, the rapid growth in the number of devices and apps outpaces the validation processes needed to assess their potential use as a clinical or research tool. In the case of software mobile apps, the market expansion is completely beyond any quality control. Currently, any user can register and distribute a new app in the online mobile marketplaces (e.g., Android's Google Play and Apple's App Store) without any technical or medical review. This situation is potentially dangerous because apps can diagnose, provide activity reports, and even medical advice without any real medical rationale behind it, and based on data produced by non-validated algorithms implemented in the app. As a result, it is of paramount importance to warn users about this potential danger and to define protocols that can validate, or at least provide, a quality score for these apps and devices before they are introduced into the market. This challenge is global, and goes beyond the legislation of one single country.

As per hardware devices, there are clearly two different leagues: those products designed as wearable devices for general consumers and those designed with the sole purpose of being used in clinical or research settings (we call these "formal actigraphs"). Formal actigraphs are easy to differentiate from the others because they are often less aesthetic and even heavy or uncomfortable (for example, Actigraph wgt3x-bt[1]). They only have one or two specific functions, and they are accompanied by a specific software for that device. Generally, they are also more expensive.

The question is: can a consumer device act as a formal actigraph? Formal actigraphs are, in general, more accurate and reliable because they have more precise sensors and are accompanied with more sophisticated sleep algorithms. But for certain measurements a commercial actigraph can behave as a formal actigraph, as shown in some validation studies (see Table III); for example, when measuring sleep start/end. Consumer devices can carry out gross measurements sleep time, sleep latency, sleep efficiency and number of awakenings. However, more precise or specific measurements are an issue. They are still unable to detect sleep phases. Even for formal actigraphs, there is a lack of research regarding sleep detection algorithms specific for different populations. For instance, two of the most commonly used algorithms (Cole-Kripke [13] and Sadeh [59]) are fairly old and specific for a kind of population.

In general, for clinical purposes, we can only be sure that the measurements of our actigraph are reliable if it is used with a population for which it has been validated. However, we should also consider that the word "validation" has been used polysemically in different studies (core validation, transitive validation, etc.). An interesting classification of validation types can be found in [58]), and not all of them achieve the required standards of clinical or research applications. In addition, a recent study [63] showed that the same device validated for a specific population can behave differently for that population at different ambient temperature conditions.

In summary, we envisage a future where sleep devices will accurately detect all variables related to sleep, including the sleep phases. The devices will include IA algorithms able to study the individual behaviour of the user and adapt themselves to produce even more accurate results. Even though we are still far from that future, this opens up new research paths: (i) definition of new sleep detection algorithms for each type of population, (ii) definition of software protocols and benchmarks that automatically provide a minimum validation phase to sleep devices and apps, (iii) regulatory requirements, agreements, or consensus for the application of the above protocols before the distribution of sleep devices and apps to customers, and (iv) validation of current and future actigraphs for their clinical and research use.

It is also important to point out that this future will come with new (already emergent) problems and challenges. One of them is the *Quantified Self* movement. The fact that people will wear advanced self-monitoring devices (e.g., EEG, ECG) and collect data for self-tracking and auto-analytics is potentially dangerous because it can lead to erroneous self-diagnosis and behaviours based on wrongly set objectives that only rely on the numerical data collected. The interpretation of the data collected by sensors require competencies to access, understand, appraise, and apply health information to make judgements and daily life decisions concerning healthcare, disease prevention, and health promotion [67]. Another potential problem is the increased focus on optimizing the sleep metrics reported by self-monitoring devices. This may lead to unexpected problems such as unsubstantiated concerns an anxiety, but also to worsened sleep produced by unhelpful changes, causing a condition called orthosomnia [5].

[1] https://www.actigraphcorp.com/actigraph-wgt3x-bt/

REFERENCES

[1] Adams R, Appleton S, Vakulin A et al. Comparison of the OSA50, STOP and STOP-BANG Questionnaires for Identification of Obstructive Sleep Apnea in a Male Population Cohort. American Journal of Respiratory and Critical Care Medicine. 2015;191:A5601.

[2] Armon C, Johnson KG, Roy A et al. Polysomnography. Medscape, 2016 February 23, http://www.medscape.com/.

[3] ASA. Sleep Devices. American Sleep Association (ASA). URL: https://www.sleepassociation.org/sleep-treatments/snoring-mouthpieces/sleep-devices/, 2018.

[4] Baandrup L, Jennum PJ. A validation of wrist actigraphy against polysomnography in patients with schizophrenia or bipolar disorder. Journal of Neuropsychiatric Disease and Treatment. 2015;11:2271–7.

[5] Baron KG, Abbott S, Jao N et al. Orthosomnia: are some patients taking the quantified self too far? J Clin Sleep Med. 2017;13(2):351–354.

[6] Berthomier C, Drouot X, Herman-Stoïca M et al. Automatic analysis of single-channel sleep EEG: validation in healthy individuals. Sleep. 2007 Nov;30(11):1587-95.

[7] Bhat S, Ferraris A, Gupta D et al. Is There a Clinical Role for Smartphone Sleep Apps? Comparison of Sleep Cycle Detection by a Smartphone Application to Polysomnography. Journal of Clinical Sleep Medicine. 2015;11:709-15, 2015.

[8] Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. Metabolism. 2018;84:99-108. doi: 10.1016/j.metabol.2017.10.008

[9] Blackwell T, Redline S, Ancoli-Israel S et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep. 2008;31:283-91.

[10] Boyne K, Sherry DD, Gallagher PR et al. Accuracy of computer algorithms and the human eye in scoring actigraphy. Sleep Breath. 2013;17:411-7. doi: 10.1007/s11325-012-0709-z.

[11] Chen KY, Bassett DR, The technology of accelerometry-based activity monitors: Current and future. Med. Sci. Sports. Exerc. 2005;37:490–500.

[12] Che-Chang Y, Yeh-Liang H, A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring. Sensors. 2010;10:7772–88.

[13] Cole RJ, Kripke DF, Gruen W et al. Technical note: Automatic sleep/wake identification from wrist actigraphy. Sleep. 1992;15:461–9.

[14] De Zambotti M, Baker FC, Colrain IM. Validation of Sleep-Tracking Technology Compared with Polysomnography in Adolescents. Sleep. 2015;38:1461-8.

[15] De Zambotti M, Claudatos S, Inkelis S et al. Evaluation of a consumer fitness-tracking device to assess sleep in adults. Chronobiology International. 2015;32:1024-8.

[16] El-Sayed IH. Comparison of four sleep questionnaires for screening obstructive sleep apnea. Egyptian Journal of Chest Diseases and Tuberculosis. 2012;61:433–41.

[17] Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. International Journal of Behavioral Nutrition and Physical Activity. 2015;12:159.

[18] Firat H, Yuceege M, Demir A et al. Comparison of four established questionnaires to identify highway bus drivers at risk for obstructive sleep apnea in Turkey. Sleep and Biological Rhythms. 2012;10: 231–6.

[19] Fossbytes.com. 8 Best Android Sleep Tracker App List For 2018. 2018 March 27. URL: https://fossbytes.com/android-sleep-tracker-app-list/

[20] Green E. The 10 Best Sleep Trackers in 2018. 2018 August 29, http://www.nosleeplessnights.com/best-sleep-tracker/

[21] Godfrey A, Conway R, Meagher D et al. Direct measurement of human movement by accelerometry. Med. Eng. Phys. 2008;30:1364–86.

[22] Ibáñez V, Silva J, Cauli O. A survey on sleep assessment methods. PeerJ. 2018;6:e4849 doi: https://doi.org/10.7717/peerj.4849.

[23] Ibáñez V, Silva J, Cauli O. A survey on sleep questionnaires and diaries. Sleep Medicine. 2018;42:90-6.

[24] Jessica MK, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. ISRN Neurology. 2012:768794.

[25] Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Review of Medical Devices. 2016:13:497-506.

[26] Langley H. Counting sheep: The best sleep trackers and monitors. 2018 October 8. URL: https://www.wareable.com/withings/best-sleep-trackers-and-monitors.

[27] Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. Studies in Health Technology and Informatics. 2015;210:458-60.

[28] Lee J, Hong M, Ryu S. Sleep Monitoring System Using Kinect Sensor. International Journal of Distributed Sensor Networks. October 25, 2015.

[29] Lee XK, Chee NIYN, Ong JL et al. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. J Clin Sleep Med. 2019;15(9):1337–1346.

[30] Lichstein KL, Stone KC, Donaldson J et al. Actigraphy validation with insomnia. Sleep. 2006:29:232-9.

[31] Lomeli HA, Pérez-Olmos I, Talero-Gutiérrez C et al. Sleep evaluation scales and questionnaires: a review. Actas Españolas de Psiquiatría. 2008;36:50-9.

[32] Luo J, Huang R, Zhong X et al. STOP-Bang questionnaire is superior to Epworth sleepiness scales, Berlin questionnaire, and STOP questionnaire in screening obstructive sleep apnea hypopnea syndrome patients. Chinese Medical Journal. 2014;127:3065-70.

[33] Marino M, Li Y, Rueschman MN et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013;36:1747-55.

[34] Martin JL, Hakim AD. Wrist Actigraphy. Chest 2011;139:1514–27.

[35] Maslakovic M. Ten gadgets for advanced sleep monitoring. 2017 January 9, http://gadgetsandwearables.com/2017/01/09/the-best-sleep-trackers/

[36] Matsuo M, Masuda F, Sumi Y et al. Comparisons of Portable Sleep Monitors of Different Modalities: Potential as Naturalistic Sleep Recorders. Front Neurol. 2016; 7: 110. doi: 10.3389/fneur.2016.00110.

[37] Meltzer LJ, Hiruma LS, Avis K et al. Comparison of a Commercial Accelerometer with Polysomnography and Actigraphy in Children and Adolescents. Sleep. 2015;38:1323–30.

[38] Meltzer LJ, Walsh CM, Traylor J et al. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep. 2012:35:159-66.

[39] Meltzer LJ, Wong P, Biggs SN et al. Validation of Actigraphy in Middle Childhood. Sleep. 2016;39:1219-24.

[40] Min JK, Doryab A, Wiese J et al. Toss ‘n’ turn: smartphone as sleep and sleep quality detector. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM. 2014:477–486.

[41] Moher D, Shamseer L, Clarke M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Sleep Systematic Reviews. 2015;4:1. doi: https://doi.org/10.1186/2046-4053-4-1.

[42] Montgomery-Downs H, Insana S, Bond J, Movement toward a novel activity monitoring device. Sleep Breath. 2012;16:913–17.
*** Validation study (against PSG) that reports the highest precision of a consumer sleep device: Fitbit**

[43] Nakazaki K, Kitamura S, Motomura Y et al. Validity of an algorithm for determining sleep/wake states using a new actigraph. J Physiol Anthropol. 2014 Oct 4;33:31. doi: 10.1186/1880-6805-33-31.

[44] Nakano H, Hirayama K, Sadamitsu Y et al. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. Journal of Clinical Sleep Medicine. 2014;10:73-8.
*** Study about how to detect the sleep from just the sound produced while sleeping**

[45] Natale V, Plazzi G, Martoni M. Actigraphy in the assessment of insomnia: A quantitative approach. Sleep. 2009;32:767–71.

[46] Öberg PA, Togawa T, Spelman FA, Sensors in Medicine and Health Care. Vol. 1. Wiley-VCH. Weinheim; Berlin, Germany. 2004:245–52.

[47] Ong AA, Gillespie MB. Overview of smartphone applications for sleep analysis. World Journal of Otorhinolaryngology-Head and Neck Surgery. 2016;2:45–49.
*** Review of apps for sleep detection and analysis**

[48] Pandi-Perumal SR, Spence DW, BaHammam AS. Polysomnography: An Overview. Primary Care Sleep Medicine. 2014:29-42.

[49] Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. Sleep. 2007;30:1362–9.

[50] Pataka A, Daskalopoulou E, Kalamaras G et al. Evaluation of five different questionnaires for assessing sleep apnea syndrome in a sleep clinic. Sleep Medicine. 2014;15:776-81.

[51] Patel P, Kim JY, Brooks LJ. Accuracy of a smartphone application in estimating sleep in children. Sleep Breath. 2017;21:505-11.

[52] Pigeon WR, Taylor M, Bui A et al. Validation of the Sleep-Wake Scoring of a New Wrist-Worn Sleep Monitoring Device. J Clin Sleep Med. 2018:14:1057-62. doi: 10.5664/jcsm.7180.

[53] Quante M, Kaplan ER, Cailler M et al. Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms. Nat Sci Sleep. 2018;10:13-20. doi: 10.2147/NSS.S151085.

[54] Rahman T, Adams AT, Ravichandran RV et al. DoppleSleep: a contactless unobtrusive sleep sensing system using short-range Doppler radar. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15). ACM New York. 2015:39-50.

[55] Research 2 Guidance. mHealth App Developer Economics 2017/2018. www.research2guidance.com, 2018 October.

[56] Robertson B, Marshall B, Carno MA. Polysomnography for the Sleep Technologist. Elsevier, ISBN: 9780323100199, 2014.

[57] Roomkham S, Hittle M, Cheung J et al. Sleep monitoring with the Apple Watch: comparison to a clinically validated actigraph. F1000 Research 2019, 8:754. DOI: https://doi.org/10.12688/f1000research.19020.1

[58] Russo K, Goparaju B, Bianchi MT. Consumer sleep monitors: Is there a baby in the bathwater? Nat Sci Sleep. 2015;5:147-57. doi: 10.2147/NSS.S94182.
**** Review of consumer sleep monitors**

[59] Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep–wake identification: an empirical test of methodological issues. Sleep. 1994;17:201–7.

[60] Saris WHM, Binkhorst RA. The use of pedometer and actometer in studying daily physical activity in man. Part I: Reliability of pedometer and actometer. Eur. J. Appl. Physiol. 1977;37:219–228.

[61] Silva GE, Goodwin JL, Vana KD et al. Obstructive sleep apnea and quality of life: comparison of the SAQLI, FOSQ, and SF-36 questionnaires. Southwest Journal on Pulmonary Critical Care. 2016;13:137-49.

[62] Silva GE, Vana KD, Goodwin JL et al. Identification of patients with sleep disordered breathing: comparing the four-variable screening tool, STOP, STOP-Bang, and Epworth Sleepiness Scales. Journal of Clinical Sleep Medicine. 2011;7:467-72.

[63] Shin M, Swan P, Chow CM. The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. Sleep Sci. 2015;8:9-15. doi: 10.1016/j.slsci.2015.02.003

** **Study that introduces the temperature conditions as a relevant variable to determine validity**

[64] Singh J, Mims N. Screening Tools for the Obstructive Sleep Apnea for the Cardiovascular Clinician. American College of Cardiology. 2015 Jul 14.

[65] Sitnick SL, Goodlin-Jones BL, Anders TF. The use of actigraphy to study sleep disorders in preschoolers: some concerns about detection of nighttime awakenings. Sleep. 2008;31:395–401.

[66] Sivertsen B, Omvik S, Havik OE et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. Sleep. 2006;29:1353–58.

[67] Sorensen K. "Health literacy and public health: a systematic review and integration of definitions and models". BMC Public Health. 2012;12:80: 80. doi:10.1186/1471-2458-12-80. PMC 3292515.

[68] de Souza L, Benedito-Silva AA, Pires ML et al. Further validation of actigraphy for sleep studies. Sleep. 2003:26:81-5.

[69] Toon E, Davey MJ, Hollis SL et al. Comparison of Commercial Wrist-Based and Smartphone Accelerometers, Actigraphy, and PSG in a Clinical Cohort of Children and Adolescents. Journal of Clinical Sleep Medicine. 2016;12:343-350.

[70] Tryon W, Bellak A, Hersen M. Activity Measurement in Psychology and Medicine. New York, NY: Plenum Press, 1991.

[71] Ulate-Campos A, Tsuboyama M, Loddenkemper T. Devices for ambulatory monitoring of sleep-associated disorders in children with neurological diseases. Children (Basel). 2017;5:E3. doi: 10.3390/children5010003

[72] Winter C. Personal Sleep Monitors: Do They Work? The huffington post, 2014 April 28.

[73] Wu H, Kato T, Numao M et al. Statistical sleep pattern modelling for sleep quality assessment based on sound events. Health Inf Sci Syst. 2017;5:11. doi: 10.1007/s13755-017-0031-z.