(Not so) Smart sleep tracking through the phone: Findings from a polysomnography study testing the reliability of four sleep applications

(Not so) Smart sleep tracking through the phone: Findings from a PSG study testing the reliability of four sleep applications

Fino E*[1], Plazzi G[2,3]*, Filardi M[2], Marzocchi M[4], Pizza F[2,3], Vandi S[2,3], Mazzetti M[1].

Running head: **(Not so) Smart sleep tracking through the phone**

Word count: 3676

[1] Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Bologna, Italy.

[2] Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy.

[3] IRCCS, Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy

[4] Department of Psychology, Alma Mater Studiorum - University of Bologna, Italy.

***Corresponding authors:**

Edita Fino, PhD

Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Bologna, Italy. Viale Berti Pichat 5, 40126, Bologna, Italy;

Phone: + 39 051 2091330; E-mail: edita.fino@unibo.it

Giuseppe Plazzi, MD, PhD.

Department of Biomedical and Neuromotor Sciences (DIBINEM), University of Bologna, Bologna, Italy. Via Altura n°3, 40139, Bologna, Italy.

Phone: +39 051 4966929; Fax: +39 051 4966176; E-mail: giuseppe.plazzi@unibo.it

Author Contribution: Study conception and design: M.M. (last author), G.P. Data collection: E.F., S.V. Data preparation, statistical expertise, data analysis and interpretation M.M. (last author), E. F., S.V., M.M., M.F., F.P., G.P. Wrote the paper E.F., M.M. (last author), F.P., S.V., G.P.

Funding Statement: The first author was partly supported by a research grant of Foundation Altroconsumo.

Summary

An increasing number of sleep applications are currently available and are being widely used for in-home sleep tracking. The present study assessed four smartphone applications (Sleep Cycle-Accelerometer, *SCa*; Sleep Cycle-Microphone, *SCm*; Sense, *Se*; Smart Alarm, SA) designed for sleep-wake detection through sound and movement sensors, by comparing their performance with polysomnography (PSG). Twenty-one healthy participants (6 males, 15 females) used the four sleep applications running on iPhone (provided by the experimenter) simultaneously with portable PSG recording at home, while sleeping alone for two consecutive nights. Whereas all apps showed a significant correlation with PSG-Time in Bed, only *SA* offered significant correlations for Sleep Efficacy. Furthermore, *SA* seemed to be quite effective in reliable detection of Total Sleep Time and also Light Sleep, however it underestimated Wake and partially overestimated Deep Sleep. None of the apps resulted capable of detecting and scoring REM sleep. To sum up, *SC* (functioning through both accelerometer and microphone) and *Se* did not result sufficiently reliable in sleep-wake detection compared to PSG. *SA*, the only application offering the possibility of an epoch-by-epoch analysis, showed higher accuracy than the other apps in comparison with PSG, but it still shows some limitations, particularly regarding wake and deep sleep detection. Developing scoring algorithms specific for smartphone sleep detection and adding external sensors to record other physiological parameters may overcome the present limits of sleep tracking through smart phone apps.

Key words: Smartphone, sleep tracking, sleep applications, PSG,

1. INTRODUCTION

Equipped with an array of high-quality precision sensors, current generation smartphones are increasingly boasting computational powers similar to those of miniaturized sleep labs. Given their ubiquity and personal nature, smartphones are indeed widely considered as the prime candidate for the purposes of low cost, large scale and long-term sleep monitoring outside the laboratory setting, both in general and clinical population (Behar, Roebuck, Domingo, Gederi & Clifford, 2013; Fino & Mazzetti, 2019; Penzel, Schobel & Fietze, 2018).

Functioning through the support of in-built sensors (i.e., accelerometers, microphone, gyroscopes) or adjunct external sensor devices (some allowing even physiological signal recording such as heart rate and breathing frequency), sleep applications are able to analyze data online and yield complex sleep scoring reports that provide subjects with a real time feedback on their sleep quality and standard sleep parameters (Van de Water, Holmes & Hurley, 2011). Paralleling such developments, specialized data analysis algorithms are being developed by sleep researchers, in order to analyze and score sleep data recorded through smartphone-based sensor devices (Natale, Drejak, Erbacci, Tonetti, Fabbri & Martoni, 2012).

Nonetheless, the growing number of sleep tracking applications and devices is hard to fathom as this is a constantly changing field (Van den Bulck, 2015). As pointed out by recent literature reviews (Baron, Duffecy, Berendsen, Mason, Lattie & Manalo, 2018; Choi, Demiris, Lin, Iribarren, Landis, Thompson, et al., 2018), the large number of sleep applications developed for assessing sleep is associated with relatively sparse validation studies and it seems the claims of most sleep tracking devices and applications currently outweigh the evidence to support them (Lee-Tobin, Ogeil, Savic & Lubman, 2017). Indeed, while a massive number of sleep apps are currently available at no cost in the market and are being widely used for sleep tracking functions

(Fietze, 2016; Ko, Kientz, Choe, Kay, Landis & Watson 2015; Ong & Gillespie, 2016), only a limited number have been successfully validated against polysomnography and/or wrist actigraphy (Bhat, Ferraris, Gupta, Mozafarian, DeBari, Gushway-Henry, et al., 2015; Natale et al., 2012; Toon, Davey, Hollis, Hons, Nixon, Home, et al., 2017). In particular, a satisfactory accuracy is found for a few applications that utilize multiple sensors and/or scoring algorithms that integrate such information, which probably gives them an advantage over single sensor-based applications (Gu, Yang, Shangguan & Liu, 2016; Paalasmaa, Toivonen & Partinen 2015). Regrettably, most of sleep apps do not grant access to the data collected or the scoring process, which remain locked together inside a black box, due to proprietary rights, thus posing formidable barriers to researchers in this field (Lorenz & Williams, 2017).

Considering the increasingly high number of sleep applications available in the market, their massive use by the general population and their potential clinical significance, further evaluating the reliability of smartphone-based sleep assessment remains a key issue and more validation studies with both healthy and patient populations are urgently needed, according with several recent researchers' conclusion (Baron et al., 2018; Choi et al., 2018; Fino & Mazzetti, 2019; Penzel et al., 2018).

The aim of the present study was assessing the reliability of four smartphone applications (namely: Sleep Cycle-Acceloremeter, *SC*a; Sleep Cycle-Microphone, *SCm*; Sense, *Se*; Smart Alarm, *SA*) designed for sleep-wake detection through sound and movement sensors either inbuilt or external to the phone, by comparing their performance with polysomnography (PSG). In addition, for one of the apps (*SA*) we examined the accuracy of overall sleep-wake detection as well as discrimination between individual stages within sleep.

2. METHODS

2.1 Participants

Participants were 21 (six males, fifteen females) healthy adults recruited from an initial sample of 25 volunteers (see Statistical analyses section). Inclusion criteria were absence of any neurological, psychiatric or sleep disorders and having generally seven to nine hours of sleep per night without any major disruption of the sleep–wake cycle during the two months preceding enrolment. The study was approved by the Institutional Review Board of University of Bologna and all participants provided signed informed consent.

2.2 Materials and Procedure

Four iPhone applications widely used and available in the market al low or no cost at the time of the data collection were selected based on sensor features: Sleep Cycle (by Northcube; v. 5.4) comes in two versions, functioning through accelerometer (*SCa*) and microphone (*SCm*), whereas Smart Alarm (*SA*, by Plus Sports; v. 8.4) and Sense (*Se*, by Hello; v. 2.1.0) function through an inbuilt and an external to the phone accelerometer, respectively.

Subjects were recorded at home, while sleeping alone for two consecutive nights. The four smartphone applications (two per each night, in a cross-subjects counterbalanced order) were used simultaneously with portable PSG. All iPhones (iOS7) were provided by the experimenter and were placed near the pillow in a position that did neither restrict subject's movement, nor interfere with the PSG recording. Polysomnographic recording were performed with the System '98, Micromed® (Mogliano Veneto, Italy) including EEG (three channels, with a sample frequency of 128 Hz, Low- Frequency Filter set at 0.3 HZ and High Frequency Filter set at 35

Hz), right and left EOG, EMG mylohyoideus and tibial muscles and ECG. EEG electrodes were positioned according to the international 10-20 system.

2.3 Data Analysis

2.3.1 PSG Reports

PSG records were manually sleep-staged into 30s epochs according to the American Academy of Sleep Medicine (AASM, 2014), yielding the following sleep parameters: Time in bed (TIB: minutes between start and end of PSG recording), Sleep Period Time (SPT: minutes between sleep onset and morning awakening), Total Sleep Time (TST: minutes scored as sleep within SPT), Wake (minutes scored as wake within TIB), Wake After Sleep Onset (WASO: minutes scored as wake within SPT), Sleep Efficiency (SE: TST/TIB, expressed as a percentage), Sleep Latency (SL: minutes from start of TIB to the first epoch of sleep), duration of REM sleep and NREM sleep stages (N1, N2 and slow wave sleep – SWS).

2.3.2 Sleep Application Reports

The absolute sleep parameters provided by the four applications were: TIB and Sleep Quality for Sleep Cycle; TIB, TST, Sleeping Soundly, Deep Sleep, Sleep Latency and Sleep Score for Sense; TIB, Sleep Quality, Light Sleep, Deep Sleep, Wake for Smart Alarm. All apps provided a graphic representation of the sleep recording through the night (see Figure.1). Sleep Quality and Sleep Score can be assumed as a measure of SE, whereas we considered Sleeping Soundly as Light Sleep (LS), given that the other index provided by Sense was Deep Sleep (DS).

----- Insert Figure 1 about here -----

2.3.2.1 Epoch by epoch comparison data

None of the applications divided its recording into specific epochs and three out of four (*SCm*, *SCa* and *Se*) provided non-isomorphic hypnograms, hence an epoch-by epoch comparison with PSG was only possible for *SA*, whose sleep stage graph lend itself to being analyzed and was consequently transformed into 1-min epochs (see Figure 2) as per previous comparison studies (Bhat et al., 2016). Trimming of the data to match with the start and end times of the PSG was applied to SA as the start and end times of the sleep period (sleep period time, SPT).

----- Insert Figure 2 about here -----

For every minute of recording, PSG and SA data were manually recoded as either sleep or wake (0 = wake, 1 = sleep). PSG provides data in 30-s epochs, so each minute of PSG data was manually rescored as wake if either one or both of the 30-s epochs in each 1-min block were scored as wake, as per previous comparison studies (Tal et al., 2017; Toon et al., 2016). Further, in order to perform an epoch-by-epoch comparison of sleep stage distribution through the night between PSG and SA, epoch scored as sleep were further recoded as light sleep or deep sleep (utilizing three codes:1 = wake, 2 = light sleep, 3 = deep sleep). For PSG scoring, we considered N1 and N2 as Light sleep and SWS as Deep sleep. In addition, given that SA does not offer a scoring of REM, we considered REM sleep as part of PSG Light sleep, in order to performed epoch by epoch comparisons.

2.4 Statistical analysis

From a total sample of 25 participants, we excluded subjects with "SE < 85" (n=4). Seven nights of recording were lost due to either technical problems with PSG data storage (3

nights) or erroneous use of the app by participants (4 nights). Thus, our final sample consisted of 21 subjects and 42 nights, with each application being related to a different total number of PSG night recording, equally distributed between the first (n=16) and second (n=19) night session.

2.4.1 Correlational analysis between the four Apps and PSG parameters

Bivariate correlations (Spearman's coefficient) were performed between the absolute parameters provided by each app (*SCa, SCm, Se, SA*) and their correspondent indices obtained from the PSG recording. Spearman coefficient was applied as PSG and app's variables were not normally distributed, as indicated by Kolmogorov-Smirnov Test.

Considering that two apps reported duration of 'light sleep' we obtained equivalent PSG values for 'light sleep' (LS) by adding the duration of N1 and N2 sleep. Thus, Deep Sleep was correlated to PSG-SWS, whereas Light Sleep with PSG-LS. Furthermore, as none of the apps provided a REM measure, we computed for Sleeping Soundly and Deep Sleep of Sense and for Light Sleep and Deep Sleep of Smart Alarm three additional correlations (with PSG-REM, PSG LS plus REM and PSG-SWS plus REM), to asses if the applications systematically considered REM within one sleep category or the other.

2.4.2 Comparison between PSG and apps' reports

We performed four one-way repeated measures MANOVA analysis, with device (1^{st} : PSG vs. *SCa*; 2^{nd}PSG vs. *SCm*; 3^{rd}: PSG vs. *Se*; 4^{th}: PSG vs. *SA*) as within-subjects factor and sleep indices as dependent variables (1^{st}: TIB and SE; 2^{nd}: TIB and SE; 3^{rd}: TIB, TST, SE, SL); 4^{th}: TIB, TST, SE, Wake), to compare absolute values of main parameters reported by PSG and applications.

2.4.3 Epoch-by-epoch comparison between PSG and SA

Inter-rater reliability (IRR) of overall epoch-by-epoch matching between devices was assessed. We determined the sensitivity (i.e., ability of *SA* to detect true sleep when the PSG also scores sleep), specificity (i.e., ability of *SA* to detect true wakefulness when the PSG also scores wakefulness) and accuracy (i.e., the ability of *SA* to correctly detect sleep and wake), applying the formulas reported in Table 1 and previously used by Toon and colleagues (Toon et al., 2016). In addition, to assess the difference between PSG and *SA* in identifying sleep stages, two supplementary analyses were conducted: first, a MANOVA was carried out on (recoded) TST, Wake, Light Sleep (LS) and Deep Sleep (DS), with Device (PSG vs. *SA*) as within-subjects -factor; and then, we used Bland-Altman plots (Bland & Altman, 1995) to examine the degree of agreement between (recoded) PSG and SA reports. A positive mean difference (or bias) indicates an underestimation of the sleep parameter, while a negative difference indicates an overestimation. All statistical analysis were conducted using SPSS version 22.0 (SPSS, Inc., Chicago, IL), results with $p < 0.05$ were considered statistically significant.

----- Insert Tables 1 & 2 about here -----

3. RESULTS

3.1 Demographic and polysomnographic data

Demographic characteristics and PSG data of participants are shown in Table2. Despite only participants with SE ≥ 85 were selected, the PSG data of the entire sample, in particular SL, WASO and SWS duration, seem to indicate the possibility that some subjects may have presented more sleep fragmentation or sleep pressure, compared to healthy population.

3.2 Correlational analysis between the four Apps and PSG parameters

Significant correlations (Spearman $r_{s)}$ between absolute parameters provided by the apps and the PSG are reported in Table 3. All apps showed a significant correlation with PSG-TIB, whereas only *SA* offered significant correlations for SE. *SA* seemed to be quite effective in reliable detection of sleep, wake and also Light Sleep. None of the apps was capable to detect and score REM values; but while Se seemed to spread REM sleep between Sleeping Soundly and Deep Sleep, *SA* systematically considered REM as Light Sleep, as shown by significant correlation between this index and PSG-N1, N2 plus REM.

3.3 Comparison between PSG and apps main sleep parameters

The four MANOVAs computed on sleep parameters and the subsequent ANOVAs are reported in Table 4. In general, TIB and values scored by the apps were comparable to ones showed by PSG, whereas SE resulted significantly lower for *SCa*, *SCm* and *Se* and higher for *SA* compared to PSG and durations of wake resulted significantly lower for *SA* compared to PSG. TST were comparable to PSG for *Se*, whereas *SA* showed higher values in comparison with PSG.

----- Insert Tables 3 & 4 about here -----

3.4 Epoch by epoch comparison between PSG and SA

Further epoch by epoch analysis performed on *SA* data compared to PSG (i.e., on epochs scored as sleep or wake) showed an inter-rater reliability (IRR) range of 87% -97%. *SA* showed a high sensitivity in detecting sleep (range 91 - 97.4%) but a low specificity in detecting wakefulness (range 0 - 48 %), and an overall accuracy of 92.8% (range 85% - 95%).

In addition, the MANOVA carried out on (recoded value of) TST, Wake, Light Sleep and Deep Sleep showed a main effect of Device (PSG vs. *SA*: Wilks' Lambda = 0.649; $F_{3,18}$=3.250;

p=0.046). As shown in Figure 3, subsequent ANOVAs revealed that the *SA* significantly underestimated Wake ($F_{1,20}$=8.297; p =0.009) and showed a tendency to overestimate Deep sleep, although the difference was not significant ($F_{1,20}$ = 3.938; p = 0.061). No significant differences were found for TST ($F_{1,20}$=2.942; p = 0.102), nor for LS detection ($F_{1,20}$ = 2.643; p= 0.120.

----- Insert Figure 3 about here -----

We further explored the per night distribution of sleep stages and sleep efficiency through Bland Altman plot profiles confirmed the direction of the bias (or difference) seen in Figure 3, with *SA* underestimating Wake and SE, while overestimating Deep sleep and TST compared with the PSG. The mean difference (or bias) and lower and upper limits of agreement (95% confidence interval = mean difference ± 2SD) are shown in Figure 4.

----- Insert Figure 4 about here -----

4. DISCUSSION

The aim of this study was assessing the overall performance of sleep tracking through the phone as well as the reliability of smartphone applications in sleep-wake and sleep stage detection compared to the reference PSG. Whereas all apps showed a significant correlation with PSG-TIB, only *SA* offered significant correlations for SE. Furthermore, *SA* seemed to be quite effective in reliable detection of Total Sleep Time and also Light Sleep, but it seems to underestimate Wake and partially overestimate Deep Sleep. In addition, none of the apps was capable of detecting and scoring REM sleep. Nonetheless, correlational analyses indicated that *Se* spread REM sleep between Sleeping Soundly and Deep Sleep, whereas *SA* systematically considered REM as Light Sleep.

Considering the overall data reported, *SC* and *Se* seem less valid (in terms of reliability and accuracy in identifying absolute values of wake and sleep periods) than *SA*. Indeed, *SA* yielded reliable estimations of SE as showed by significant correlation and allowed for an epoch-by epoch comparison with PSG, resulting in overall satisfactory agreement. The current sleep - wake analysis results are in line with previous studies on accelerometer based sleep applications (Bhat et al., 2016; Tal, Shinar, Shaki, Codish & Goldbart, 2017; Toon et al., 2016): *SA* showed greater sensitivity, accurately assessing sleep 97% of the time, than specificity, accurately assessing wake at best 48 % of the time, although the specificity results in the current study were lower than those reported in previous works which suggested that sleep tracking through smart phone fared well compared to traditional wrist actigraphy (Cole, Kripke, Gruen, Mullaney & Gillin 1992). Furthermore, the discrepancy observed between sleep stage detection by the SA compared to PSG revealed by the ANOVAs (Figure 3) was confirmed through the Bland-Altman plots (Figure 4). The app underestimated Wake by 15 minutes and overestimated Deep sleep by 42 min, which is consistent with previous validating studies on smartphone applications (Bhat et al., 2016; Natale et al., 2012). The low detection of wake as evidenced by the low specificity, and especially the difficulty in detecting wake during sleep could explain the app's overestimation of deep sleep. Not surprisingly, this affected the app's capacity to reliably asses sleep efficiency. Although correlations with PSG were significant the Bland Altman plot profile shows that *SA* underestimated SE by 5 %. In terms of sleep stage detection, our findings are in keeping with literature showing that sleep apps perform poorly when compared to PSG, putatively due to the fact that movement-based algorithms used in accelereometer-based sleep applications cannot reliably distinguish sleep stages (Choi et al., 2018; Lee-Tobin et al., 2017; Fino & Mazzetti, 2019).

In the same lines, it needs to be noted that there is a (potentially) massive utilization of smartphone applications for at home sleep monitoring purposes of children and adolescent populations. Empirical evidence on smartphone sleep assessment on children and adolescents (Patel, Kim & Brooks., 2017; Toon et al., 2016) reveal a low reliability of apps compared to the PSG, which is in line with our results. It is therefore crucial to highlight that the apps' outputs should not be considered as diagnostic reports, especially in at home children sleep assessment, and that presence of sleep disorders especially at an early age should be evaluated exclusively in a clinical setting.

Our study is not exempt of limitations. The black box nature of apps' data collection and scoring analysis remains a key limitation in our study's explanatory power. Access to the raw data and the development of scoring algorithms specific for smartphone sleep monitoring may enhance the apps' capacity to yield accurate sleep-wake assessment (Natale et al.2012). It should also be noted that PSG reports seem to indicate the possibility that some of our subjects may present more sleep fragmentation or sleep pressure compared to healthy population**.** If fact, it may be that these very characteristics may have led the subjects to give their consent to monitoring their sleep quality. A random selection of a "not so normal" sample in our case may have influenced the results, given research showing increasing discrepancy to PSG the shorter and more fragmented the sleep (Tal et al., 2017; Fino and Mazzetti, 2019). Given that our sample cannot be considered as "very" representative of the normal population, future studies with healthy subjects are warranted to extend the validity of our findings together with the resulting implications. Another limitation of our study may refer to the lack of a group with sleep disturbances, which would have been a more stringent testing of the apps reliability to assess sleep. While this remains to be addressed by future research, findings from studies conducted on

both healthy and disordered sleep subjects (see Tal et al, 2017) have not yielded significantly different results in terms of reliability of apps compared to PSG and actigraphy in these two populations.

Despite these limitations the current study has a number of strengths, one of which being that of simultaneously testing four widely used sleep applications, one of which functioning through two different sensor modalities (accelerometer and microphone). Furthermore, subjects were recorded at home, in a naturalistic setting, which further extends the validity of present results.

5. CONCLUSION

To sum up, *SC* (functioning through both accelerometer and microphone) and *Se* did not result sufficiently reliable in sleep-wake detection compared to PSG. *SA*, the only application offering the possibility of an epoch-by-epoch analysis in comparison with PSG showed higher accuracy. Still, important validity limitations need to be considered. Compared to available data on sleep tracking through accelerometer-based smartphone applications, *SA* showed similar sensitivity but lower specificity, suggesting that it could not be 'smart' enough for a reliable use for these purposes or as an alternative to actigraphy. Although estimates of SE correlated significantly with those of the PSG, the significant differences observed in sleep stage detection shed light on the necessity to use SA with an understanding of its limitations, in particular the weakness in wakefulness and deep sleep identification. Developing scoring algorithms specific for smartphone sleep detection and adding external sensors to record other physiological parameters may certainly overcome the present limits of sleep tracking through the phone and correct inherent biases in sleep stage detection. More importantly, as the adoption of apps and devices

for sleep tracking purposes by the general public and potentially sleep disordered patients is on steady increase (as pointed out by Van den Bulck, 2015), it is of vital importance that future research keep monitoring developments in this field and examining sleep applications' effectiveness and reliability.

REFERENCES

AASM, America Academy of Sleep Medicine (2014). Manual for the Scoring of Sleep and Associated Events, Version 2.1. Westchester, IL: American Academy of Sleep Medicine.

Baron, K. G., Duffecy, J., Berendsen, M. A., Mason, I. C., Lattie, E. G., & Manalo, N.C. (2018). Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep. Sleep Medicine Reviews, 40, 151-159.

Behar, J., Roebuck, A., Domingo, J. S., Gederi, E., & Clifford, G.D. (2013). A review of current sleep screening applications for smartphones. Physiological Measures, 34, 29–46.

Bhat, S., Ferraris, A., Gupta, D., Mozafarian, M., DeBari, V., Gushway-Henry, N., Gowda, S. P., Polos, P. G., Rubinstein, M., Seidu, H., & Chokroverty, S. (2015). 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, 11, 709–715.

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet, 346, 1085–7.

Choi, Y. K., Demiris, G., Lin, S. Y., Iribarren, S. J., Landis, C. A., Thompson, H. J., McCurry, S. M., Heitkemper, M. M., & Ward, T. M. (2018). Smartphone applications to support sleep self-management: review and evaluation. Journal of Clinical Sleep Medicine, 14(10), 1783–1790.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. Sleep, 15(5), 461–469.

Fietze, I. (2016). Sleep applications to assess sleep quality. Sleep Medicine Clinics, 11(4), 461–468.

Fino, E., & Mazzetti, M. (2019). Monitoring healthy and disturbed sleep through smartphone applications: A review of experimental evidence, Sleep and Breathing Journal, 23(1), 13-24.

Gu, W., Yang, Z. H., Shangguan, L., & Liu, Y. (2016). Sleep Hunter: Towards a fine grained sleep stage tracking with Smartphones. IEEE Transactions on Mobile Computing, 15, 1514-1527.

Ko, P. R., Kientz, J. A., Choe, E. K., Kay, M., Landis, C. A., & Watson, N. F. (2015). Consumer sleep technologies: a review of the landscape. Journal of Clinical Sleep Medicine, 11(12), 1455–1461.

Lee-Tobin, P.A., Ogeil, R.P., Savic, M., & Lubman, D.I. (2017). Rate my sleep: examining the information, function, and basis in empirical evidence within sleep applications for mobile devices. Journal of Clinical Sleep Medicine, 13(11),1349–1354.

Lorenz, C. P., & Williams, A. J. (2017). Sleep apps: what role do they play in clinical medicine? Current Opinion in Pulmonary Medicine, 23(6), 512-516.

Natale, V., Drejak, M., Erbacci, A., Tonetti, L., Fabbri, M., & Martoni, M. (2012). Monitoring sleep with a smartphone accelerator. Sleep Biological Rhythms, 10, 287–292.

Ong, A., & Gillespie, G.M. (2016). Overview of smartphone applications for sleep analysis. World Journal of Otorhinolaryngology-Head and Neck Surgery, 2, 45–49.

Paalasmaa, J., Toivonen, H., & Partinen, M. (2015). Adaptive Heartbeat Modeling for Beat-to-Beat Heart Rate Measurement in Ballistocardiograms IEEE Journal of Biomedical and Health Informatics, 19, 1945-1952.

Patel P., Kim J. Y., & Brooks L. J. (2017). Accuracy of a smartphone application in estimating sleep in children. Sleep and Breathing, 21, 505-511.

Penzel, T., Schobel, C., & Fietze, I. (2018). New technology to assess sleep apnea: wearables, smartphones, and accessories. F1000Researc, 7: 413.

Tal, A., Shinar, Z., Shaki, D., Codish, S., & Goldbart, A. (2017). Validation of contact-free sleep monitoring device with comparison to polysomnography. Journal of Clinical Sleep Medicine, 13:517–522.

Toon, E., Davey, M.J., Hollis, S.L., Hons, B.A., Nixon, M.G., Home, R., & Biggs, N.S. (2016). Comparison of commercial wrist-based and smartphone accelerometers, actigraphy, and PSG in a clinical cohort of children and adolescents. Journal of Clinical Sleep Medicine, 12(3),343–350.

Van den Bulck, J. (2015). Sleep apps and the quantified self: blessing or curse? Journal of Sleep Research, 24, 121–123.

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. Journal of Sleep Research, 20,183–200.