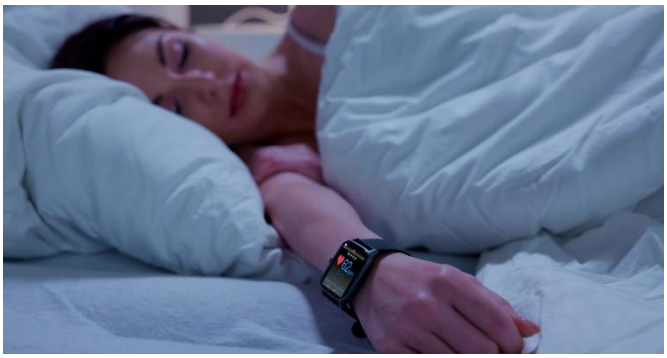
Gold-Standard Sleep Test Nearly Matched by Huawei Smartwatch

A recent study used data from participants from sleep clinics to assess the performance of a consumer-grade sleep-tracking device, relative to polysomnography (PSG).



Study

This study assessed the performance of the HUAWEI WATCH GT2 against PSG. The HUAWEI WATCH GT2 collects heart rate and movement variation signals for sleep detection. The smartwatch was tested across different sleep disorders. Adult participants who had completed demographic and sleep questionnaires were recruited between March 1st, 2021, and April 30th, 2023. Individuals who worked night shifts within the last 6 months, slept less than 4 hours, had cognitive conditions, or received treatments for sleep disorders were excluded.

All participants completed a single-night PSG monitoring, and their sleep stages (awake, N1, N2, N3, and rapid eye movement (REM) sleep) were scored. For epoch-by-epoch (EBE) analysis and sleep summary, epochs of 30s were used. Obstructive sleep apnea (OSA) was diagnosed using an apnea hypopnea index (AHI) of greater than or equal to 15/h. Other hypoxemia indices were collected, such as lowest pulse oxygen saturation (LSpO2) and oxygen desaturation index (ODI).

The smartwatch provided four stages of sleep recordings: Awake, light sleep (equalling N1 and N2), deep sleep (equalling N3), and REM sleep. The measurements of interest included total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE), which is defined as TST divided by minutes between lights off and lights on. Because raw data could not be exported from the device, researchers manually extracted sleep stage information from the app's summary graphs, ensuring synchronized timing with PSG data.

Results

A total of 98 participants met the inclusion criteria, with about 84% being male. The average age and body mass index (BMI) were 45 and 26.0kg/m2, respectively. More than half of the participants complained of daytime sleepiness and poor sleep quality. The median SE and TST were 85% and 405.8 minutes, respectively. The PSG results showed that 47 patients had moderate-to-severe OSA, 33 patients were normal, 12 suffered from comorbid insomnia and sleep apnea, and 30 had clinical insomnia.

The PSG and smart watches agreed more on wake and light sleep classifications. The smart watch seemed to classify PSG REM epochs as light sleep, and the error rates in this case were high. Furthermore, deep sleep and REM sleep were often classified as light sleep. Among other possible stage classifications, misclassification errors were relatively lower.

The smartwatch agreed with the PSG for EBE of sleep versus wake states with a specificity of 44.5% and a sensitivity of 95.3%, and the positive predictive value (PPV) was 72.20%. Overall accuracy reached 87.3%, and Cohen's κ value of 0.43 (prevalence- and bias-adjusted k = 0.75) indicated moderate-to-substantial agreement between the two devices.

Except for light sleep, the smart watch showed high accuracy for all sleep stages, i.e., greater than 70%. The smart watch significantly overestimated SE, TST, deep sleep, REM sleep, and sleep onset latency (SOL) while underestimating WASO. Specifically, it overestimated total sleep time by about +28.5 minutes and sleep efficiency by +5.9 percentage points, while it underestimated wake after sleep onset by about -37 minutes.

After adjusting for unstable sleep, the latency to persistent sleep (LPS) levels between PSG and the smartwatch were not significantly different. In patients with sleep disorders, t-tests revealed lower accuracy in insomnia patients and lower sensitivity in OSA patients relative to healthy controls. However, no significant differences among the disorder subgroups were found in sleep stage agreement.

When compared against published criteria for acceptable bias in wearable validation studies (≤30 min for TST and ≤5% for SE), the device's performance "almost reached" research-grade actigraphy standards, supporting its potential as a low-cost tool for sleep/wake detection in healthy individuals.

Conclusion

In summary, the HUAWEI WATCH GT2 device demonstrated high agreement in sleep/wake detection with PSG. While using smartwatches. Consumers and healthcare practitioners should be aware of sleep stage overestimations and underestimations.

A key limitation of the study revolves around the fact that it was conducted at a single center with a limited number of participants, thereby limiting generalizability. Other sleep and mental disorders, such as narcolepsy, depression, and periodic limb movement disorders, were not evaluated.

The device's algorithm was proprietary and may change with updates, requiring revalidation with each software iteration. Additionally, some participants experienced data loss due to device removal or movement during sleep. Rapidly updating algorithms requires new validations, and this issue limits the clinical application of wearable consumer devices.

Overall, the study suggests that while consumer wearables like the HUAWEI WATCH GT2 are not yet a suitable replacement for PSG in diagnosing sleep disorders, they can provide reliable two-stage (sleep/wake) monitoring for general sleep health tracking in real-world settings.

Source:

https://www.news-medical.net/news/20251007/Huawei-smartwatch-nearly-matches-gold-standard-sleep-test.aspx