

## A performance validation of six commercial wrist-worn wearable devices for sleep stage scoring compared to polysomnography

**Schyvens A-M (1,2); Van Oost N (3); Aerts J-M (3); Masci F (3); Peters B (4); Neven A (4); Dirix H (4); Wets G (4); Ross V (4,5); Verbraecken J (1,2)**

1) Laboratory of Experimental Medicine and Pediatrics, University of Antwerp, Wilrijk, Belgium.

**Study Objectives:** Wearable devices for sleep tracking have gained immense popularity over the past few years. The aim of this study is to assess the performance of six different consumer wearable devices, namely the Fitbit Charge 5, Fitbit Sense, Withings Scanwatch, Garmin Vivosmart 4, Whoop 4.0 and the Apple Watch Series 8, for detecting sleep parameters compared to the gold standard, polysomnography (PSG).

**Methods:** Sixty-two adults (52 male, 10 female, mean age  $\pm$  SD = 46,0  $\pm$  12.6 years) spent a single night in the sleep laboratory with PSG while simultaneously using two to four wearable devices.

**Results:** The results indicate that most wearables displayed significant differences with PSG for total sleep time (TST), sleep efficiency (SE), wake after sleep onset (WASO) and light sleep (LS). Nevertheless, all wearables demonstrated a higher percentage of correctly identified epochs for deep sleep (DS) and REM sleep compared to wake (W) and LS. All devices detected >90 % of sleep epochs (i.e., sensitivity), but showed lower specificity (29.39 % to 52.15 %). The Cohens Kappa coefficients of the wearable devices ranged from 0.21 to 0.53, indicating fair to moderate agreement with PSG.

**Conclusions:** Our results indicate that all devices can benefit from further improvement for multi-state categorization. However, the devices with higher Cohens Kappa coefficients, such as the Fitbit Sense ( $\kappa$  = 0.42), Fitbit Charge 5 ( $\kappa$  = 0.41) and Apple Watch Series 8 ( $\kappa$  = 0.53), could be effectively used to track prolonged and significant changes in sleep architecture.