Assessment of consumer-grade wearable devices to track sleep in healthy individuals in free-living conditions.

Sanjay Rajput, Alexandra Jamieson, Nishi Chaturvedi, Alun Hughes, Michele Orini

MRC Unit for Lifelong Health and Aging, Institute of Cardiovascular Science, University College London, London, United Kingdom

Sleep is a vital physiological process critically implicated in maintaining physical and mental health. Novel consumer-grade devices provide the opportunity of tracking sleep, but their use for clinical and research purposes is limited due to lack of validation. This study aimed to validate sleep tracking in free-living conditions by comparing total sleep duration measured by popular wearable devices with sleep diaries. Twenty-seven heathy volunteers of median (interquartile range) age 21.0 (20.5-22.5), 71% male, and with wide range of skin tones (Fitzpatrick scale from 1 to 6) wore 5 devices for 2 consecutive nights and provided sleep diaries. Devices included the Garmin Vivoactive 4 (GV) and 4S (GVs), Fitbit Sense (FS), Withings Scanwatch (WS), and the Oura Ring (OR), which measured sleep duration using a standard and a β-software. Each night was considered as an independent observation. Agreement was assessed using the Spearman's correlation coefficients, average percentage error and Bland Altman plots. All devices expect WS reported some connectivity issues resulting in a maximum of 17 (GV) lost observations. 5% of GVs and 2% of FS estimates were considered outliers (error $>q_3+1.5 \times (q_3-q_1)$ or $<q_1-1.5 \times (q_3-q_1)$). Agreement is summarised in Figure 1. Correlation ranged from 0.41 for GVs to 0.76 for FS, while limits of agreement ranged from (-125, 71) minutes for OR-β to (-115, 256) minutes for GVs. Pair-wise comparisons (sinrank Wi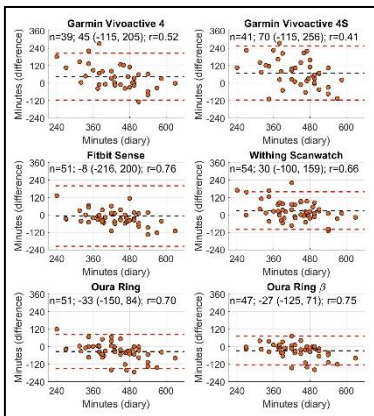lcoxon test with Bonferroni correction) showed that the absolute percentage error was not different in most cases, except for GVs (larger than FS, WS, OR and OR-β) and for OR-β (lower than OR). No association was found between the absolute percentage error and skin tone, body mass index or device position. These data show moderate to good agreement between wearable-enabled sleep tracking and sleep diaries in free-living conditions.

Figure 1: Total sleep duration. Bland-Altman plots reporting number of nights (n), bias and limits of agreement, and correlation coefficient (r).