



HHS Public Access

Author manuscript

J Meas Phys Behav. Author manuscript; available in PMC 2021 August 25.

Published in final edited form as:

J Meas Phys Behav. 2021 ; 4(2): 151–162. doi:10.1123/jmpb.2020-0045.

Validity of Two Awake Wear-Time Classification Algorithms for activPAL in Youth, Adults, and Older Adults

Jordan A. Carlson,

Children's Mercy and University of Missouri Kansas City

Fatima Tuz-Zahra,

University of California, San Diego

John Bellettiere,

University of California, San Diego

Nicola D. Ridgers,

Deakin University

Chelsea Steel,

Children's Mercy

Carolina Bejarano,

Children's Mercy and University of Kansas

Andrea Z. LaCroix,

University of California, San Diego

Dori E. Rosenberg,

Kaiser Permanente Washington Health Research Institute

Mikael Anne Greenwood-Hickman,

Kaiser Permanente Washington Health Research Institute

Marta M. Jankowska,

University of California, San Diego

Loki Natarajan

University of California, San Diego

Abstract

corresponding author Carlson (jacarlson@cmh.edu).

Carlson and Tuz-Zahra are co-first authors. Carlson, Steel, and Bejarano are with the Center for Children's Healthy Lifestyles and Nutrition, Children's Mercy, Kansas City, MO, USA. Carlson is also with the Department of Pediatrics, Children's Mercy and University of Missouri Kansas City, Kansas City, MO, USA. Tuz-Zahra, Bellettiere, LaCroix, and Natarajan are with the Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. Ridgers is with the Institute for Physical Activity and Nutrition (IPAN), School of Exercise and Nutrition Sciences, Deakin University, Geelong, Victoria, Australia. Bejarano is also with the Clinical Child Psychology Program, University of Kansas, Lawrence, KS, USA. Rosenberg and Greenwood-Hickman are with the Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. Jankowska is with the Qualcomm Institute/Calit2, University of California, San Diego, La Jolla, CA, USA.

Background: The authors assessed agreement between participant diaries and two automated algorithms applied to activPAL (PAL Technologies Ltd, Glasgow, United Kingdom) data for classifying awake wear time in three age groups.

Methods: Study 1 involved 20 youth and 23 adults who, by protocol, removed the activPAL occasionally to create nonwear periods. Study 2 involved 744 older adults who wore the activPAL continuously. Both studies involved multiple assessment days. In-bed, out-of-bed, and nonwear times were recorded in the participant diaries. The CREA (in PAL processing suite) and ProcessingPAL (secondary application) algorithms estimated out-of-bed wear time. Second- and day-level agreement between the algorithms and diary was investigated, as were associations of sedentary variables with self-rated health.

Results: The overall accuracy for classifying out-of-bed wear time as compared with the diary was 89.7% (Study 1) to 95% (Study 2) for CREA and 89.4% (Study 1) to 93% (Study 2) for ProcessingPAL. Over 90% of the nonwear time occurring in nonwear periods >165 min was detected by both algorithms, while <11% occurring in periods ≤165 min was detected. For the daily variables, the mean absolute errors for each algorithm were generally within 0–15% of the diary mean. Most Spearman correlations were very large ($r > .81$). The mean absolute errors and correlations were less favorable for days on which any nonwear time had occurred. The associations between sedentary variables and self-rated health were similar across processing methods.

Conclusion: The automated awake wear-time classification algorithms performed similarly to the diary information on days without short (2.5–2.75 hr) nonwear periods. Because both diary and algorithm data can have inaccuracies, best practices likely involve integrating diary and algorithm output.

Keywords

accelerometer; non-wear; processing; sedentary

The activPAL (PAL Technologies Ltd, Glasgow, United Kingdom) is a thigh-worn accelerometer that distinguishes body posture and movement (i.e., lying, sitting, standing, and stepping) and is increasingly used in field-based sedentary research (Edwardson et al., 2017). Unlike the commonly used hip accelerometer (e.g., ActiGraph), the activPAL is most often employed using a 24-hr wear protocol (Troiano, McClain, Brychta, & Chen, 2014). This creates the problem of potentially misclassifying participants' true in-bed time as sedentary time, or true sedentary time as in-bed time, if precautions are not taken. In addition, as with any accelerometer, precautions must be taken to minimize the misclassification of true nonwear time as awake wear time, which can result in an overestimation of sedentary time. Such errors in the estimation of sedentary time could lead to inaccurate population prevalence estimates and introduce bias when testing associations between sedentary time and health or other factors. These awake wear-time considerations have traditionally been addressed in activPAL research by having participants track in-bed and nonwear times in a diary (Edwardson et al., 2017), but such diaries add participant, data collection, and processing burden, and the diary data are often incomplete due to challenges with compliance.

More recently, automated algorithms have been created to estimate the awake wear time from activPAL data. A third-party algorithm, “ProcessingPAL,” classifies periods of awake wear, in-bed, and nonwear time. (The latter two are not differentiated, but rather, are combined.) It was developed in one hundred and twenty-five 18- to 40-year-olds and has been shown to have good validity in a sample of 741 adults 35 years old (Winkler et al., 2016), but has not been tested in youth. The ProcessingPAL algorithm is available through a publicly available java-based software application that allows users to adjust certain parameter settings and review data visually to make corrections (Shashidar Ette, 2020). Another algorithm was developed by PAL Technologies Ltd., the maker of activPAL and the PAL processing software, with the aim of classifying periods of awake wear, in-bed, and nonwear time (the latter two are classified separately; PAL Technologies Ltd, 2020). This algorithm is available in the PAL software suite when using the “CREA (Enhanced analysis algorithm: non-wear, upright correction, lying, cycling, seated transport)” processing setting. The parameter settings are not adjustable by end users.

Despite the potential benefits of these algorithms for supporting physical behavior research that uses activPAL, little evidence is currently available regarding their validity, particularly for CREA and in youth, which limits their use in research. This paper presents two studies that aimed to investigate the convergent validity of the ProcessingPAL and CREA activPAL awake wear-time classification algorithms as compared with participant diaries in three age groups. Study 1 involved a small sample of youth and adults who wore the activPAL using a prescribed protocol that involved a mixture of 24-hr wear time, removal during in-bed time, and brief (\approx 30–120 min) nonwear periods across 3–5 days to capture various potential circumstances that can occur in field-based studies. The emphasis on internal validity—the wear protocol and the attention to the accuracy of the bed and nonwear logs—supported an investigation of between-method agreement with fine granularity (e.g., second-level epoch agreement), including during nonwear periods of different lengths. Study 2 involved a large sample of older adults who wore an activPAL for 4–7 days. This study employed a wear protocol that is typical in field-based epidemiology and intervention studies, supporting external validity, and generalizability of findings. In addition, the large sample and assessment of health measures supported an investigation of concurrent validity (i.e., associations with health markers).

Methods

Study 1—Participants and Procedures

Study 1 involved 20 youth and 23 adults who enrolled in an observational measurement study. The participants were recruited through an internal posting, targeting employees of a large research hospital system and their family members. The eligibility criteria were being within the age range of 9–17 years (youth) or 18–60 years (adults), able to read and understand English, comfortable wearing an activPAL for 5 days and two nights, and willing to complete a diary to track in-bed and nonwear times. On two nights, the participants were instructed to wear the activPAL overnight (i.e., while in bed). On the other two nights, they were asked to remove the activPAL before getting into bed and reapply it when getting out of bed for the day. The participants were also instructed to remove the activPAL for three

nonwear periods of different lengths, with the requested lengths being approximately 30, 60, and 120 min. The participants were told they did not need to adhere to these lengths exactly. The participants chose when to initiate each nonwear period based on their schedules, but were instructed to not initiate or end a nonwear period within an hour of another nonwear or in-bed period. All instructions were provided to the participants verbally and in a written document that accompanied the diary. The accuracy of the diary was emphasized to the participants as the most important part of the study. The participants received the activPAL and instructions in person during an office visit. The instructions covered the diary, nonwear periods, and how to remove and reattach the activPAL. The study was approved by the Children's Mercy institutional review board.

Study 2—Participants and Procedures

Study 2 leveraged the data from older adults who participated in the Adult Changes in Thought Activity Monitoring substudy. Adult Changes in Thought is a prospective cohort study that started in 1994 by enrolling approximately 2,000 adults aged 65 years or older without physician-adjudicated dementia. The participants were randomly sampled from the King County membership panels of Kaiser Permanente Washington (formerly Group Health), an integrated health care delivery system in the state of Washington. As the participants died or were diagnosed with dementia, new participants were randomly selected from the same King County sampling frame. From 2016 to 2018, 1,885 participants were eligible to wear an activPAL for 7 days and to concurrently complete a diary to indicate the time they went to bed and the time they got out of bed each day. A detailed STROBE diagram was published elsewhere (Rosenberg et al., 2020). The participants were asked to wear the device at all times, including overnight. Unlike in Study 1, the participants were asked to wear the activPAL device continuously for 7 days and not to remove it for any reason, if possible. The criteria for inclusion in the present analysis included the following: (a) availability of data from the concurrent wear of both the activPAL and ActiGraph, which enabled high-quality control through a rigorous visual inspection of overlay plots of the concurrently worn data, although ActiGraph data were not used in the present analyses; (b) complete diaries recording in-bed and out-of-bed times; and (c) device wear for 4 or more days, with 10 or more hours of awake wear time. Of the 954 participants who met these criteria, 204 were excluded due to having incomplete diaries, resulting in a final sample of 744 participants. The study was approved by the Kaiser Permanente institutional review board.

Measures

Diary.—All participants used a diary to track the date and time (to the minute) they got into bed for the night and out of bed for the day the following morning. In Study 1, the diary was also used to track the start and end date and time (in minutes) of all occurrences of the activPAL being removed, which determined valid wear time.

activPAL.—All participants wore an activPAL on their thigh with medical-grade adhesive. Study 1 used the activPAL4, and Study 2 used the activPAL3. Both studies required a minimum of 10 s in a new posture for that new posture to be registered. The data provided by each activPAL were processed to create a second-level epoch file that indicated for every

second whether the participant was sedentary (lying or sitting) or not sedentary (standing or walking).

Out-of-bed time, in-bed time, and nonwear-time processing.—Each second of data was classified as out-of-bed time, in-bed time, or nonwear time based on the (a) participant diaries, (b) ProcessingPAL algorithm (GitHub, 2020), and (c) CREA algorithm (PAL Technologies Ltd, 2020), with the exception that ProcessingPAL does not differentiate between in-bed and nonwear time, but rather, combines the two. Out-of-bed time represents awake wear time, but we used the terms “out-of-bed” and “in-bed” because the algorithms do not aim to capture when a participant was actually asleep or awake. A primary difference between the ProcessingPAL and CREA algorithms is that ProcessingPAL bases the aforementioned classifications on posture codes, and the algorithms are published (Winkler et al., 2016), whereas CREA uses proprietary algorithms applied to raw accelerometer data. Specifically, ProcessingPAL makes classifications based on periods without posture change/movement, using the event types sitting, standing, and stepping provided by the VANE (standard) classification algorithm in the PAL software suite. The default settings in ProcessingPAL (shortest bout threshold of 30 min, long bout threshold of 120 min, longest bout threshold of 300 min, check window length of 15 min, and maximum number of steps of 20) were used in the present study based on their established validity in adults (Winkler et al., 2016), with the exception that Study 1 used 25 min for the shortest bout threshold. This was to account for the possibility that the participants engaged in nonwear bouts that were slightly <30 min. ProcessingPAL also includes an option for users to visualize 24-hr data for each participant and, unlike CREA, to change and adapt settings for different populations and/or make manual corrections to the outputs of the algorithm. This semiautomated data processing approach was not used in the present study. The newer CREA algorithm available in the PAL software suite provides additional event types to those provided by the VANE algorithm, including primary lying, which is used to classify in-bed time, and nonwear. A secondary lying classification is also provided by CREA and is grouped with sitting when summing the daily sedentary time. Similar to the sitting, standing, and walking event types, these additional event types are derived from the raw accelerometer data from the activPAL using proprietary algorithms. However, no published research yet exists on the accuracy of these additional event types. The present study used the “events” files from the CREA algorithm, which have a separate row for each continuous period of a given event type (e.g., primary lying) and each individual step. In the events files, brief awakenings that involve movement (e.g., walking to the bathroom) can result in a gap in primary lying, thus being classified as out-of-bed time wear time. In-bed and nonwear times are also available in the day-level files from the CREA algorithm, with time in bed calculated as the duration of time between when the participant first got in bed (e.g., for the night) and last got out of bed (e.g., in the morning). Thus, in the day-level files, brief awakenings that include nonlying events (e.g., stepping) are regularly classified as part of in-bed time.

Day-level variables.—For both studies, the out-of-bed time (in minutes per day), total sedentary time (in minutes per day), percentage of awake wear time spent sedentary (total sedentary time ÷ out-of-bed time), mean sedentary bout duration (mean of all sedentary bouts; minutes), number of breaks per day from sedentary time, and in-bed time (in minutes

per day) were calculated at the day-level using events files. The PAL event files derived using the VANE (standard) PAL processing algorithm were used for creating the day-level variables that were based on awake wear time indicated by the diary (i.e., the two data sources were merged). ProcessingPAL adds awake wear-time information to the VANE-processed PAL events file, which was used to compute the ProcessingPAL-based day-level variables. CREA results in a PAL events file that includes information on sedentary events, nonwear events, and in-bed (primary lying) events, which was used to compute CREA-based day-level variables. A break from sedentary time (i.e., sit-to-stand postural transition) was defined as any time a standing/walking period lasting ≥ 10 s was preceded by a sedentary (sitting/lying) period lasting ≥ 10 s, and the sedentary periods preceding each standing/walking second were considered sedentary bouts (no tolerance). The day-level variables were based only on days with sufficient wear time to be considered a valid day (≥ 8 hr in Study 1 and ≥ 10 hr in Study 2).

Other participant variables.—In Study 1, participant age and sex were reported by the participants. In Study 2, participant age, sex, race/ethnicity, and education status were assessed via survey, and self-rated health was assessed with a single item asking, “In general, how would you rate your health at this time?” (Ware & Sherbourne, 1992). The response options were excellent, very good, good, fair, and poor and were dichotomized to good, poor, fair versus excellent, and very good for the present analyses.

Statistical Analysis

For Study 1, contingency tables were used to calculate second-level agreement between diary and ProcessingPAL, diary and CREA, and CREA and ProcessingPAL, and F1 scores and overall accuracy were calculated. Next, the algorithms’ validity for classifying nonwear time was investigated separately by length of nonwear period, with diary-determined nonwear periods grouped as 25–40, 45–94, 120–165, and ≥ 165 min. These groupings were based on the data and generally reflected the flexible nonwear durations that the participants were instructed to engage in (30, 60, and 120 min and overnight). The periods prior to the time the participant first put the device on (tracked by research staff) and after the participant last wore the device (tracked in the diary) were removed from the data. Days with insufficient wear time to be considered a valid day were retained in the second-level epoch analyses because of the smaller sample size in Study 1, and the second-level agreement results were unlikely to be affected if a representation of the participants’ full waking hours was not captured.

For Study 2, binary classification measures (accuracy, F1 scores, sensitivity, specificity, and kappa) assessed second-by-second agreement in measuring out-of-bed wear time (yes/no) between each algorithm and the sleep diary. We were also interested in quantifying the within- and between-person variation in how accurate each algorithm classified out-of-bed wear as compared with the diary. Day-level accuracy values (percentage of seconds for which the algorithm and diary agreed) were computed for all 5,135 days and entered in an intercept-only mixed-effects regression model run separately for ProcessingPAL and CREA. The between- and within-person (i.e., across days) components of variation in accuracy were then compared using the following two ratio measures: $\text{ProcessingPAL}_{\text{within-person}}/$

$CREA_{within-person}$ and $ProcessingPAL_{between-person}/CREA_{between-person}$. Confidence intervals (CIs) for each ratio were estimated using bootstrapping. Day-level accuracy was logit transformed due to nonnormality, but this did not affect the interpretation of the ratio measures.

In both studies, the day-level means and *SDs* in the daily variables (e.g., minutes of out-of-bed time) were calculated for each processing method (diary, ProcessingPAL, and CREA). The mean absolute error (MAE), Lin's concordance correlation coefficient (CCC) (Lin, Hedayat, Sinha, & Yang, 2002), and Spearman correlations compared metrics between the diary and ProcessingPAL, the diary and CREA, and CREA and ProcessingPAL. The mean differences between the methods were assessed using mixed-effects linear regression, accounting for nesting of days within participants. In Study 1, analyses were conducted separately for days with and without prescribed nonwear time, and interactions were used to test whether agreement between methods differed by age group (youth vs. adult).

In Study 2, participant-level associations of total sedentary time, percentage of time spent sedentary, mean bout duration, and number of breaks with self-rated health were investigated for each wear-time processing method using mixed-effects regression, adjusting for age, sex, race/ethnicity, education status, and out-of-bed time. All sedentary behavior metrics—total sedentary time, percentage of time spent sedentary, mean bout duration, and number of breaks—were standardized to a mean of 0 and *SD* of 1 before analysis. These models were used to determine whether associations between sedentary time and health differed by the wear-time processing method. Horton's method (Horton & Fitzmaurice, 2004) was used to test whether the associations were significantly different between methods.

Statistical analyses were performed in SPSS (version 24.0; IBM Corp., Armonk, NY; IBM Corp., 2016) and R (The R Foundation for Statistical Computing, 2019). CCCs, Spearman correlations, and kappas were interpreted as small (< .40), moderate (.41–.60), large (.61–.80), and very large (.81–1.0; Landis & Koch, 1977).

Results

Sample Characteristics

In Study 1, the youth participants were 9–17 years old (mean = 13.3, *SD* = 2.5), and 40.0% were female. The adult participants were 20–56 years old (mean = 41.3, *SD* = 10.2), and 73.9% were female. Race/ethnicity was not collected in Study 1. The Study 2 sample had a mean age of 76.8 (*SD* = 6.5) years; 57.5% were female, 90% were non-Hispanic White, and 75.5% had a college degree.

Study 1

Overall, the second-level accuracy for classifying out-of-bed, in-bed, and nonwear time was 89.7% for CREA and 89.4% for ProcessingPAL, as compared with the diary. CREA correctly identified 95.1% of the diary-determined out-of-bed time (F1 score = 92.7%), 91.7% of the diary-determined in-bed time (F1 score = 87.2%), and 72.4% of the diary-determined nonwear time (F1 score = 82.0%) (Table 1). ProcessingPAL correctly identified 91.1% of the diary-determined out-of-bed time (F1 score = 91.3%) and 86.9% of the

diary-determined in-bed/nonwear time (F1 score = 86.6%). Agreement differed only slightly between youth and adults (CREA overall accuracy = 90.0% and 89.3%; ProcessingPAL overall accuracy = 89.0% and 89.9%, respectively). ProcessingPAL had similar second-level agreement with CREA, as it did with the diary, with an overall accuracy of 90.9% between algorithms for classifying out-of-bed and in-bed/nonwear time (Supplementary Table 1 [available online]). ProcessingPAL identified 90.5% of the CREA-determined out-of-bed time (F1 score = 92.7%) and 91.7% of the CREA-determined in-bed/nonwear time (F1 score = 88.0%).

When the devices were removed for shorter durations (bouts of 25–165 min), CREA misclassified 90.6–94.6% of the nonwear time as wear time. Only 8.6% of nonwear time in bouts 165 min in duration was misclassified as out-of-bed wear time (Table 2). ProcessingPAL misclassified 89.0–94.7% of the diary-determined nonwear time during nonwear bouts lasting 25–165 min, but only 5.1% of the nonwear time during nonwear bouts lasting 165 min, as out-of-bed time.

In the day-level analyses of days without prescribed nonwear time (i.e., devices were not removed), the mean differences between the diary and each algorithm were nonsignificant, and MAEs for each algorithm relative to the diary mean were 4.0% (Table 3). The exception was that the MAE for the CREA-estimated in-bed time was 12.7% of the diary mean. The CCCs were moderate (CREA) or small (ProcessingPAL) for the out-of-bed time and total sedentary time, and large or very large for the percentage of wear time spent sedentary, mean bout duration, number of breaks, and in-bed time (CREA and ProcessingPAL). The Spearman correlations were very large for all variables for both CREA and ProcessingPAL. ProcessingPAL had similar agreement with CREA, as it did with the diary. Even though the mean difference between the algorithms for out-of-bed wear time was small (<3 min/day), the larger MAE (36.7 min/day) indicates that ProcessingPAL could result in either a lower or a higher estimate of out-of-bed wear time as compared with CREA, albeit the average difference was small ($\pm 4.6\%$ of the CREA mean).

For days with prescribed nonwear time during waking hours, CREA significantly overestimated the out-of-bed time, total sedentary time, percentage of wear time spent sedentary, and number of breaks, and ProcessingPAL significantly overestimated the out-of-bed time and total sedentary time as compared with the diary. In addition, the differences for the percentage of wear time spent sedentary and number of breaks approached significance; P s = .102 and .077, respectively. The MAEs were larger and CCCs smaller for days with nonwear time than days without nonwear time, and similar between CREA and ProcessingPAL for most variables. The Spearman correlations were smaller than for days without nonwear time, but were large or very large for both CREA and ProcessingPAL. For these days with prescribed nonwear time, ProcessingPAL generally had slightly better agreement with CREA than it did with the diary, particularly with regard to several of the MAEs. The mean differences between the algorithms were nonsignificant.

None of the day-level mean differences between methods differed between youth and adults, as indicated by the nonsignificant interaction tests.

Study 2

Overall accuracy as compared with the diary for classifying each second as out-of-bed wear time was 95% for CREA and 93% for ProcessingPAL (Table 4). Sensitivity, which represents the proportion of diary-determined out-of-bed wear time correctly classified as out-of-bed wear time by the algorithms, was 97% for both CREA and ProcessingPAL. Specificity, which represents the proportion of diary-determined in-bed time correctly classified as in-bed wear time by the algorithms, was 91% for CREA and 86% for ProcessingPAL. The F1 scores were 96% for CREA and 95% for ProcessingPAL. Overall, kappa was 89% for CREA and 84% for ProcessingPAL. For CREA, 92% of the participants had a “very large” kappa ($>.80$), compared with 81% for ProcessingPAL. ProcessingPAL had similar second-by-second agreement with CREA, as it did with the diary, with an overall accuracy of 94% between algorithms for classifying out-of-bed and in-bed/nonwear time (Supplementary Table 2 [available online]). ProcessingPAL identified 97% of the CREA-determined out-of-bed time and 90% of the CREA-determined in-bed/nonwear time.

The ratio of the between-person variation in daily accuracy ($\text{ProcessingPAL}_{\text{between-person}}/\text{CREA}_{\text{between-person}}$) was 1.18 (95% confidence interval, CI [1.13, 1.23]), indicating that, on average, ProcessingPAL’s accuracy from one participant to the next varied more than CREA’s accuracy (i.e., CREA’s accuracy was more similar across participants than ProcessingPAL’s accuracy). The ratio of the within-person variation in daily accuracy ($\text{ProcessingPAL}_{\text{within-person}}/\text{CREA}_{\text{within-person}}$) was 1.29 (95% CI [1.26, 1.32]), indicating that, on average, ProcessingPAL’s accuracy varied more within participants (from day to day) than did CREA’s accuracy (i.e., CREA’s accuracy was more similar across days for the same participant than ProcessingPAL’s accuracy). Supplementary Figure 1 (available online) shows examples of when the algorithms performed perfectly and poorly.

At the day level, CREA overestimated the daily out-of-bed time and underestimated the total sedentary time, number of breaks, and in-bed time as compared with the diary ($p < .001$; Table 5). ProcessingPAL also overestimated the out-of-bed time and overestimated the total sedentary time and number of breaks ($p < .001$). The MAEs were 6.1–13.1% of the diary mean for CREA and 5.1–12.7% of the diary mean for ProcessingPAL. The CCC for out-of-bed time was moderate for CREA and small for ProcessingPAL. For both algorithms, the CCCs were large or very large for total sedentary time, mean bout duration, and number of breaks. The CCC for in-bed time was moderate (CREA only). The MAEs and CCCs were slightly more favorable (showing better agreement) for CREA than ProcessingPAL, except with regard to the number of breaks. The Spearman correlations were large or very large for all variables for both CREA and ProcessingPAL, with the exception of ProcessingPAL out-of-bed time, which had a moderate correlation with diary out-of-bed time. ProcessingPAL had a slightly worse agreement with CREA than it did with the diary; the MAEs were 7.2–17.9% of the CREA mean, except with regard to out-of-bed wear time. The mean difference between algorithms for out-of-bed wear time was moderate (19 min/day). The MAE was larger (68.2 min/day, 7.2% of CREA out-of-bed wear time), indicating that ProcessingPAL could result in either a lower or a higher estimate of out-of-bed wear time as compared with CREA.

The differences in associations between standardized sedentary behavior metrics—total sedentary time, percentage of time spent sedentary, mean bout duration, and number of breaks—and self-rated health by the wear-time processing method were not statistically significant (Figure 1).

Discussion

Overall, both the CREA and ProcessingPAL awake wear-time classification algorithms appeared to have good validity for classifying out-of-bed time, in-bed time, and long periods of nonwear time based on agreement with the participant diaries. These algorithms did not appear to be able to validly classify nonwear periods less than about 2.5–2.75 hr. The findings were similar between youth and adults and across samples. The findings also pointed to several factors that researchers should consider when deciding whether to use an awake wear-time classification algorithm. Because both the diary and algorithm data can have inaccuracies, best practices likely involve integrating diary and algorithm output.

A primary finding from Study 1 was that the validity of the awake wear-time algorithms was poorer for days with known nonwear time and for periods of nonwear time lasting less than about 2.5–2.75 hr. Since activPAL is most commonly employed using a 24-hr wear protocol and adhered directly to a participant's skin with waterproofing, it is often worn for multiple days without removal (Edwardson et al., 2017). Thus, nonwear time is typically of less concern when using activPAL than when using other devices, such as hip and wrist accelerometers (Troiano et al., 2014). However, given the possibility that the participants may remove the activPAL for relatively short periods of time (e.g., during certain sports or water activities) and the poor validity of the algorithms for classifying such nonwear periods, researchers should consider instructing participants to log their nonwear periods. While very short (e.g., 5 min) nonwear periods are not likely to have major impacts on the data quality for day- and participant-level measures, such as for estimates of minutes per day of sedentary time or the proportion of awake wear time spent sedentary, longer periods (e.g., 30 min) could lead to an over- or underestimation of such measures.

Study 1 was also valuable in showing that the algorithms' validity did not differ meaningfully between youth and adults. This is an important finding because youth are often excluded from the development and testing of such algorithms (Winkler et al., 2016), despite activPAL being commonly used in youth studies (Carson et al., 2016). The present findings suggest that these algorithms generally have validity for use in youth, taking into consideration their poor validity for classifying nonwear periods less than about 2.5–2.75 hr. This consideration may be particularly important for youth studies because some evidence has shown that device-wear compliance (i.e., removing the device), including for activPAL, is poorer in youth as compared with adults (Clemes et al., 2020; McCrorie, Duncan, Granat, & Stansfield, 2015; Sherry et al., 2019; Shi et al., 2019; Verloigne, Ridgers, De Bourdeaudhuij, & Cardon, 2018). Thus, the manual tracking of nonwear time, such as from participant logs, may be particularly important in youth. It is also important to explore objective ways to classify relatively short nonwear periods, such as through more sensitive classification algorithms (e.g., with lower minimum nonwear thresholds), given that logs can be challenging for participants and particularly youth.

Although the participant-level absolute agreement between the diary and algorithms (bias and MAE) was less than perfect, more so in Study 2 than Study 1, the participant-level Spearman correlations between the methods were high. This indicates that the participants were rank ordered similarly on the variables, such as sedentary time, when using the diaries versus algorithms to classify awake wear time. This finding, paired with the finding that associations of sedentary variables with self-rated health did not differ across the awake wear-time classification methods, suggests that the algorithms tested can be used validly in association-based research, such as epidemiology studies. However, some caution should be used when interpreting the absolute values of variables derived from data processed with these awake wear-time classification algorithms, for example, if seeking to identify a specific threshold of sedentary time (in minutes per day or percentage of time) on which to base a guideline or recommendation. Researchers should clearly report their processing methods, including how in-bed and nonwear times were accounted for, to support replication and comparison across studies.

Although the findings were generally similar between studies, Study 2 showed slightly poorer agreement between the diary and classification algorithms tested. It is possible that the poorer agreement observed in Study 2 was due to measurement error in the diary data. A primary purpose of Study 1 was to test these algorithms, and thus, the quality of the diary was emphasized throughout recruitment, enrollment, and data collection activities, whereas the scope of Study 2 was much larger and the requirements around the diary were not as strict as in Study 1. Thus, the higher estimates for out-of-bed time (in minutes per day) by the algorithms as compared with the diary in Study 2 may have been partly due to an underestimation of true out-of-bed time in the diaries. For example, some participants may have recorded actual sleep times rather than in- and out-of-bed times, despite the instructions to do the latter. In addition, although the bias values (differences between methods) were significant for most variables in Study 2, the magnitudes of the bias and MAE values were small (e.g., MAEs = 13.1%). The significant *p* values can be attributed to the large sample size for Study 2.

Although the CREA and ProcessingPAL algorithms had good agreement with one another, there appeared to be some minor differences between the output of the algorithms. This could be due in part to differences in the information used by each algorithm—with CREA basing in-bed classification on thresholds around whether the participant was in a lying position (i.e., primary lying events) and ProcessingPAL basing classifications on changes in posture/movement (i.e., across sitting, standing, and stepping events). There were minimal differences in the validity of CREA and ProcessingPAL as compared with the diary. The findings from the second-level analyses suggested slightly better agreement between CREA and the diary than between ProcessingPAL and the diary, although the differences were minimal and may not be meaningful. The ratios of variation in accuracy between the algorithms suggested that there was a small subset of individuals for whom ProcessingPAL did not perform as well as CREA. A visual investigation of the data (examples shown in Supplementary Figure 1 [available online]) corroborate the ProcessingPAL considerations provided by Winkler et al. (2016). These include that a less pronounced movement difference between sleep and wake may reduce algorithm accuracy and that accuracy may be limited in populations prone to extremely prolonged sitting/lying during their waking

hours, who step very little or who have interrupted sleep patterns, since the algorithm relies heavily on assuming that very long periods spent in a single posture predominantly occur during sleep or nonwear. Thus, ProcessingPAL appears to have the potential to perform poorly for people who are very sedentary, such as those with limited mobility, and people who are restless sleepers, move heavily during sleep, and/or have fragmented sleep patterns. Relatedly, ProcessingPAL had slightly more favorable day-level validity than CREA for several indicators in Study 1 (youth and adults), whereas CREA had slightly more favorable validity than ProcessingPAL for several indicators in Study 2 (older adults). These differences by age group may, at least in part, be explained by more participants in the older adult sample being very sedentary (Rosenberg et al., 2020) and are in line with previous evidence from adults that showed ProcessingPAL's accuracy was poorer in older participants (Winkler et al., 2016). One potential advantage of ProcessingPAL is that it includes a semiautomated data processing feature that allows users to visualize each participant's data. This process can inform refinements to the parameter settings and/or allow users to manually correct or reject the outputs of the algorithm. It is possible that the use of this feature could lead to improved validity, but it requires more effort and knowledge on behalf of the user. When deciding which algorithm to use, the output of the algorithm should also be considered. Since ProcessingPAL does not distinguish between in-bed and nonwear time, CREA would be preferred for studies that aim to measure these events separately, for example, if in-bed time is of specific interest. It is important to note that neither algorithm can detect when a participant is asleep as opposed to lying down or not moving. With interest in sleep and the 24-hr activity cycle (Rosenberger et al., 2019), future activPAL algorithm research should aim to capture sleep metrics around sleep duration and quality.

Implications for Practice

The awake wear-time classification algorithms tested allow researchers to omit burdensome participant sleep diaries from data collection protocols, which is a significant advancement in activPAL research. Both algorithms appear to have good validity for general use, but how and whether the use of the algorithms may impact the specific research question(s) of interest should be considered, as each study has unique needs. Given that both sleep diaries and the algorithms tested can exhibit measurement error, a conservative best practice approach to awake wear-time classification may involve combining data from the diaries and algorithms. For example, to minimize the occurrence of misclassifying in-bed time and sedentary time, a researcher could set a rule that requires both the diary and algorithm to predict that the participant is out of bed before classifying a given second as true out-of-bed time. Since both CREA and ProcessingPAL had a similar level of agreement with the diary, but less than perfect agreement with one another, it is possible that combining the information between the algorithms may improve the awake wear-time classification. The data streams could also be combined visually and used to facilitate a process by which human raters decide whether the most likely true value (for in-bed and out-of-bed time) are those provided by the diary, algorithm, or somewhere in between, which is a similar process to the ProcessingPAL semiautomation feature and what has been used in some sleep research (Ancoli-Israel et al., 2015; Blackwell et al., 2005). It is important to note that changing the parameter settings in ProcessingPAL is likely to impact the validity of

the output. With the exception of a slightly lower shortest bout threshold in Study 1, the present study used the default parameter settings, which were also used in the original ProcessingPAL validity paper (Winkler et al., 2016).

Strengths and Limitations

The study strengths included the use of two study samples to emphasize both internal (Study 1) and external (Study 2) validity and capture free-living activities in multiple age groups. The prescription and tracking of nonwear times in Study 1 supported a deeper understanding of how the algorithms performed with regard to classifying relatively short nonwear periods. A primary limitation of this study was that a criterion or gold standard measure was not available, so the analyses relied on testing convergent validity with sleep diaries, which are known to have inaccuracies and tracked times at the minute rather than the second level. However, Study 1 emphasized diary quality to improve accuracy. Due to potential measurement error in diary data, it was possible that the algorithms were more accurate than the diaries (i.e., better reflected the truth than the diaries). Future studies should consider exploring creative measures for objectively collecting ground truth information on in-bed and nonwear time, although such efforts can be challenging in field-based studies of free-living behavior. Future studies should also consider other approaches for classifying awake wear time from the activPAL, including those that leverage other sources of data, such as from novel hardware applications (e.g., wear sensors). While the present study used the activPAL event files, it is important to note that in-bed time can differ slightly between the CREA event files and day-level files, the latter of which groups brief awakenings with in-bed time, and thus, is likely to result in slightly higher estimates of in-bed time as compared with estimates from the events files. Future research should address the implications and considerations involved in using each approach. Finally, the present study did not investigate the algorithms' validity for classifying entire days of nonwear time, such as due to noncompliance. Since the algorithms had good validity for classifying longer periods (>about 2.5–2.75 hr) of nonwear, it is expected that they would have good validity for classifying entire days of nonwear. It is important to acknowledge that, when devices are in transit, such as in the mail, in a vehicle, or being carried by a person (e.g., in a bag or pocket), this time is likely to be misclassified by the algorithms as awake wear time.

Conclusions

The novel awake wear-time classification algorithms tested are major advancements in activPAL research. The present findings suggest that researchers can generally expect similar validity from these algorithms as can be expected from sleep diaries in children, adults, and older adults, though combining information from the diaries and algorithm output is likely to be the best practice when a conservative approach is desired. Nonwear and in-bed time are critical factors to account for in sedentary research, particularly because a large majority of an individual's waking day is spent sedentary. A better classification of awake wear time is likely to advance the field of sedentary research by reducing error in sedentary measurements, allowing for a better understanding of the impacts of sedentary time on health, influences of sedentary time, and interventions to reduce sedentary time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by NIH grants R01DK114513-01 and U01AG006781. N. Ridgers was supported by a Future Leader Fellowship from the National Heart Foundation of Australia (ID 101895). The authors are grateful to Ashleigh Galler, Kate Hoppe, and William Purdy for their work on Study 1.

References

- Ancoli-Israel S, Martin JL, Blackwell T, Buenaiver L, Liu L, Meltzer LJ, . . . Taylor DJ (2015). The SBSM guide to actigraphy monitoring: Clinical and research applications. *Behavioral Sleep Medicine*, 13(Suppl. 1), S4–S38. doi:10.1080/15402002.2015.1046356 [PubMed: 26273913]
- Blackwell T, Ancoli-Israel S, Gehrman PR, Schneider JL, Pedula KL, & Stone KL (2005). Actigraphy scoring reliability in the study of osteoporotic fractures. *Sleep*, 28(12), 1599–1605. doi:10.1093/sleep/28.12.1599 [PubMed: 16408420]
- Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput J-P, . . . Connor Gorber S (2016). Systematic review of sedentary behaviour and health indicators in school-aged children and youth: An update. *Applied Physiology, Nutrition and Metabolism*, 41(6), S240–S265. doi:10.1139/apnm-2015-0630
- Clemes SA, Bingham DD, Pearson N, Chen Y-L, Edwardson CL, McEachan RR, . . . Fray M (2020). Stand out in class: restructuring the classroom environment to reduce sitting time—Findings from a pilot cluster randomised controlled trial. *The International Journal of Behavioral Nutrition and Physical Activity*, 17, 1–16. [PubMed: 31898547]
- Edwardson CL, Winkler EA, Bodicoat DH, Yates T, Davies MJ, Dunstan DW, & Healy GN (2017). Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science*, 6(2), 162–178. doi:10.1016/j.jshs.2016.02.002 [PubMed: 30356601]
- GitHub I (2020). UOL-COLS/ProcessingPAL. Retrieved from <https://github.com/UOL-COLS/ProcessingPAL/releases>
- Horton NJ, & Fitzmaurice GM (2004). Regression analysis of multiple source and multiple informant data from complex survey samples. *Statistics in Medicine*, 23, 2911–2933. [PubMed: 15344194]
- IBM Corp. (2016). IBM SPSS statistics for windows, version 23.0. Armonk, NY: IBM Corp.
- Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. doi:10.2307/2529310 [PubMed: 843571]
- Lin L, Hedayat A, Sinha B, & Yang M (2002). Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, 97(457), 257–270. doi:10.1198/016214502753479392
- McCrorie PR, Duncan E, Granat MH, & Stansfield BW (2015). Seasonal variation in the distribution of daily stepping in 11–13 year old school children. *International Journal of Exercise Science*, 8(4).
- PAL Technologies Ltd. (2020). Why activPAL? Retrieved from <http://www.palt.com/>
- Rosenberg D, Walker R, Greenwood-Hickman MA, Belletiere J, Xiang Y, Richmire K, . . . LaCroix AZ (2020). Device-assessed physical activity and sedentary behavior in a community-based cohort of older adults. *BMC Public Health*, 20, 1256. doi: 10.1186/s12889-020-09330-z [PubMed: 32811454]
- Rosenberger ME, Fulton JE, Buman MP, Troiano RP, Grandner MA, Buchner DM, & Haskell WL (2019). The 24-hour activity cycle: A new paradigm for physical activity. *Medicine & Science in Sports & Exercise*, 51(3), 454. doi:10.1249/MSS.0000000000001811 [PubMed: 30339658]
- Ette Shashidar. (2020). UOL-COLS/ProcessingPAL. Retrieved from <https://github.com/UOL-COLS/ProcessingPAL/releases>

- Sherry AP, Pearson N, Ridgers ND, Barber SE, Bingham DD, Nagy LC, & Clemes SA (2019). activPAL-measured sitting levels and patterns in 9–10 years old children from a UK city. *Journal of Public Health*, 41(4), 757–764. doi:10.1093/pubmed/fdy181 [PubMed: 30346570]
- Shi Y, Huang WY, Yu JJ, Sheridan S, Sit CH-P, & Wong SH-S (2019). Compliance and practical utility of continuous wearing of activPAL™ in adolescents. *Pediatric Exercise Science*, 31(3), 363–369. doi:10.1123/pes.2018-0170 [PubMed: 30605620]
- The R Foundation for Statistical Computing. (2019). The R Project for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Troiano RP, McClain JJ, Brychta RJ, & Chen KY (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, 48(13), 1019–1023. doi:10.1136/bjsports-2014-093546 [PubMed: 24782483]
- Verloigne M, Ridgers ND, De Bourdeaudhuij I, & Cardon G (2018). Effect and process evaluation of implementing standing desks in primary and secondary schools in Belgium: A cluster-randomised controlled trial. *The International Journal of Behavioral Nutrition and Physical Activity*, 15(1), 94. doi:10.1186/s12966-018-0726-9 [PubMed: 30261883]
- Ware JE Jr., & Sherbourne CD (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483. doi:10.1097/00005650-199206000-00002 [PubMed: 1593914]
- Winkler EA, Bodicoat DH, Healy GN, Bakrania K, Yates T, Owen N, . . . Edwardson CL (2016). Identifying adults' valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. *Physiological Measurement*, 37(10), 1653. doi:10.1088/0967-3334/37/10/1653 [PubMed: 27652827]

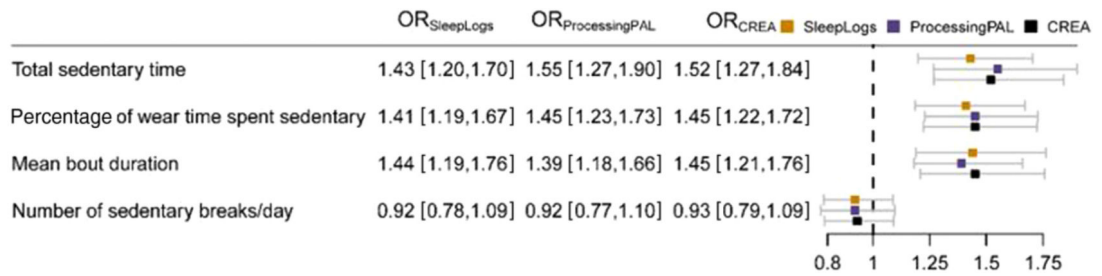


Figure 1 —. Associations between sedentary variables and self-rated health by wear-time processing method (Study 2). *Note.* Sedentary variables were standardized to a mean of 0 and *SD* of 1. Values shown are odds ratios and 95% confidence intervals and reflect having a self-rated health of good, poor, or fair as opposed to excellent or very good (the latter grouping is the reference).

Second-Level Agreement Between Diary and CREA and Diary and ProcessingPAL for Nonwear and In-Bed Classification (Study 1)

Table 1

Method and classification	Diary		
	Out of bed	In bed	Nonwear
All participants (N = 12,819,661 s)			
CREA			
Out of bed	7,397,443 (95.1%)	165,459 (7.1%)	611,583 (22.7%)
In bed	299,463 (3.8%)	2,143,320 (91.7%)	133,411 (4.9%)
Nonwear	85,053 (1.1%)	29,780 (1.3%)	1,954,149 (72.4%)
F1 score	92.7%	87.2%	82.0%
ProcessingPAL ^a			
Out of bed	7,087,092 (91.1%)		660,908 (13.1%)
In bed/nonwear	694,867 (8.9%)		4,376,794 (86.9%)
F1 score	91.3%		86.6%
Adults (n = 6,044,980 s)			
CREA			
Out of bed	3,608,272 (95.4%)	103,442 (9.4%)	280,614 (24.1%)
In bed	113,139 (3.0%)	963,576 (87.9%)	57,010 (4.9%)
Nonwear	60,572 (1.6%)	29,780 (2.7%)	828,575 (71.0%)
F1 score	92.8%	86.4%	79.5%
ProcessingPAL ^a			
Out of bed	3,475,310 (91.9%)		301,700 (13.3%)
In bed/nonwear	306,673 (8.1%)		1,961,297 (86.7%)
F1 score	92.0%		86.6%
Children (n = 6,774,681 s)			
CREA			
Out of bed	3,789,171 (94.7%)	62,017 (5.0%)	330,969 (21.6%)
In bed	186,324 (4.7%)	1,179,744 (95.0%)	76,401 (5.0%)
Nonwear	24,481 (0.6%)	0 (0%)	1,125,574 (73.4%)
F1 score	92.6%	87.9%	83.9%

Diary			
Method and classification	Out of bed	In bed	Nonwear
ProcessingPAL ^a			
Out of bed	3,611,782 (90.3%)		359,208 (12.9%)
In bed/nonwear	388,194 (9.7%)		2,415,497 (87.1%)
F1 score	90.6%		86.6%

^aProcessingPAL groups in-bed and nonwear time together.

Table 2
Second-Level Agreement of Nonwear Time by Nonwear Bout Duration for Diary Versus CREA and Diary Versus ProcessingPAL (Study 1)

Method and classification	Nonwear time indicated by diary, split by nonwear bout duration ^b			
	25–40 min	45–94 min	120–165 min	165 min
All participants (N = 2,699,142 s)				
CREA				
Out of bed	56,759 (90.6%)	128,680 (94.6%)	233,622 (92.4%)	192,522 (8.6%)
In bed	3,780 (6.0%)	0 (0%)	7,439 (2.9%)	122,192 (5.4%)
Nonwear	2,100 (3.4%)	7,279 (5.4%)	11,906 (4.7%)	1,932,864 (86.0%)
ProcessingPAL ^a				
Out of bed	56,879 (90.8%)	128,740 (94.7%)	225,123 (89.0%)	115,571 (5.1%)
In bed/nonwear	5,760 (9.2%)	7,219 (5.3%)	27,844 (11.0%)	2,132,007 (94.9%)
Adults (n = 1,166,199 s)				
CREA				
Out of bed	26,098 (92.9%)	54,976 (93.2%)	106,058 (89.9%)	93,482 (9.7%)
In bed	1,980 (7.1%)	0 (0%)	0 (0%)	55,030 (5.7%)
Nonwear	0 (0%)	4,003 (6.8%)	11,906 (10.1%)	812,666 (84.5%)
ProcessingPAL ^a				
Out of bed	26,278 (93.6%)	55,431 (94.0%)	96,983 (82.2%)	52,182 (5.4%)
In bed/nonwear	1,800 (6.4%)	3,548 (6.0%)	20,981 (17.8%)	908,996 (94.6%)
Children (n = 1,532,944 s)				
CREA				
Out of bed	30,661 (88.7%)	73,704 (95.7%)	127,564 (94.5%)	99,040 (7.7%)
In bed	1,800 (5.2%)	0 (0%)	7,439 (5.5%)	67,162 (5.2%)
Nonwear	2,100 (6.1%)	3,276 (4.3%)	0 (0%)	1,120,198 (87.1%)
ProcessingPAL ^a				
Out of bed	30,601 (88.5%)	73,309 (95.2%)	128,140 (94.9%)	63,389 (4.9%)
In bed/nonwear	3,960 (11.5%)	3,671 (4.8%)	6,863 (5.1%)	1,223,011 (95.1%)

^aProcessingPAL groups in-bed and nonwear time together.

The nonwear bout duration ranges were based on the nonwear periods that the participants were asked to engage in (30, 60, and 120 min and overnight).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3
 Day-Level Agreement Across Methods for Classifying In-Bed Time, Out-of-Bed Time, and Sedentary Variables (Study 1; $N = 148$ days)

Method and statistic	Day-level variable					
	Out-of-bed time (min/day)	Total sedentary time (min/day)	Percentage of wear time spent sedentary	Mean bout duration (min)	Number of sedentary breaks/day	In-bed time (min/day) ^a
Days without nonwear ($n = 38$)						
Descriptive statistics						
Diary, observed mean (<i>SD</i>)	823.6 (164.1)	506.1 (130.6)	61.9 (12.7)	11.6 (6.2)	56.2 (30.5)	513.2 (77.0)
CREA, observed mean (<i>SD</i>)	802.5 (176.2)	510.1 (134.1)	64.2 (13.2)	11.5 (6.1)	56.7 (30.9)	503.2 (78.2)
ProcessingPAL, observed mean (<i>SD</i>)	805.1 (136.1)	499.2 (123.7)	62.3 (12.6)	11.4 (6.0)	55.4 (29.4)	—
CREA agreement with diary						
MAE	33.0	4.0	2.4	0.1	0.5	65.1
CCC [95% CI]	.46 [.32, .58]	.45 [.33, .58]	.78 [.71, .84]	.92 [.90, .95]	.91 [.94, .88]	.69 [.77, .60]
r_s	.91	.99	.88	1.0	.99	.95
ProcessingPAL agreement with diary						
MAE	29.3	11.8	1.0	0.2	1.7	—
CCC [95% CI]	.16 [-.01, .33]	.38 [.25, .51]	.81 [.75, .86]	.92 [.89, .94]	.89 [.86, .93]	—
r_s	.95	.97	.98	1.0	.99	—
ProcessingPAL agreement with CREA						
MAE	36.7	10.9	2.0	0.1	1.3	—
CCC [95% CI]	.20 [0, .49]	.37 [.10, .59]	.78 [.62, .88]	.92 [.85, .96]	.88 [.78, .93]	—
r_s	.90	.98	.89	.99	.99	—
Days with nonwear ($n = 110$)						
Descriptive statistics						
Diary, observed mean (<i>SD</i>)	809.2 (134.6)	483.3 (147.5)	59.8 (15.6)	12.6 (10.4)	51.1 (24.9)	537.4 (94.5)
CREA, observed mean (<i>SD</i>)	867.5 (167.0) [*]	569.4 (167.0) [*]	65.8 (14.5) [*]	13.0 (9.6)	55.8 (25.4) [*]	547.3 (143.4)
ProcessingPAL, observed mean (<i>SD</i>)	860.5 (195.2) [*]	531.9 (178.6) [*]	61.9 (14.6)	12.1 (6.3)	54.8 (25.1)	—
CREA agreement with diary						
MAE	104.0	86.1	6.2	1.4	4.7	127.2
CCC [95% CI]	.16 [.10, .23]	.38 [.30, .46]	.60 [.53, .67]	.68 [.62, .74]	.72 [.66, .77]	.37 [.24, .51]

Method and statistic	Day-level variable							
	Out-of-bed time (min/day)	Total sedentary time (min/day)	Percentage of wear time spent sedentary	Mean bout duration (min)	Number of sedentary breaks/day	In-bed time (min/day) ^a		
r_s	.68	.92	.83	1.0	.97	.76		
ProcessingPAL agreement with diary								
MAE	101.4	81.2	3.5	1.1	3.1	—		
CCC [95% CI]	.16 [.10, 0.23]	.28 [.20, .36]	.58 [.50, .66]	.61 [.53, .68]	.71 [.66, .77]	—		
r_s	.77	.83	.95	1.0	.98	—		
ProcessingPAL agreement with CREA								
MAE	69.6	37.6	3.5	0.7	1.8	—		
CCC [95% CI]	0.15 [0.02, 0.27]	0.29 [0.12, 0.44]	0.64 [0.48, 0.76]	0.64 [0.49, 0.76]	0.88 [0.78, 0.93]	—		
r_s	.78	.88	.87	.98	.99	—		

Note. CCC= concordance correlation coefficient; CI = confidence interval; MAE = mean absolute error; r_s = Spearman correlation.

^a75 days had in-bed time; 50% were on days with nonwear time, and 50% were in children.

* Significantly different from diary, $p < .05$.

Table 4
Participant-Level Summaries of Second-Level Agreement for Diary Versus CREA and Diary Versus ProcessingPAL for Classifying Out-of-Bed Wear Time (Study 2; N = 744 Participants)

Statistic	Mean (SD) across participants	
	CREA	ProcessingPAL
Binary classification measures		
Accuracy	0.95 (0.05)	0.93 (0.07)
Sensitivity	0.97 (0.06)	0.97 (0.08)
Specificity	0.91 (0.1)	0.86 (0.14)
F1 score	0.96 (0.05)	0.95 (0.07)
Kappa	.89 (.11)	.84 (.16)
Number of participants (%)		
Participant-level kappa category	CREA	ProcessingPAL
Very large (.81–1.0)	641 (86%)	536 (72%)
Large (.61-.80)	82 (11%)	159 (21%)
Moderate (.41-.60)	14 (2%)	34 (5%)
Small (.20-.40)	7 (1%)	15 (2%)

Note. Second-level agreement was calculated for each participant, and then summary statistics (mean, SD, and kappa) were computed across participants.

Table 5
Day-Level Agreement Across Methods for Classifying In-Bed Time, Out-of-Bed Time, and Sedentary Variables (Study 2; $N = 5,135$ days)

Method and statistic	Day-level variable						
	Out-of-bed time (min/day)	Total sedentary time (min/day)	Percentage of wear time spent sedentary	Mean bout duration (min)	Number of sedentary breaks/day	In-bed time (min/day)	
Descriptive statistics							
Diary, observed mean (<i>SD</i>)	928.2 (84.1)	591.4 (139.7)	63.8 (14.2)	15.2 (9.1)	45.0 (15.9)	511.8 (84.1)	
CREA, observed mean (<i>SD</i>)	947.8 (119.3)*	557.8 (162.0)*	58.8 (15.2)*	15.3 (9.2)	41.8 (15.4)*	491.2 (119.1)*	
ProcessingPAL, observed mean (<i>SD</i>)	966.2 (151.1)*	620.9 (173.9)*	64.1 (13.9)	15.1 (7.4)	46.0 (16.6)*	—	
CREA agreement with diary							
MAE	56.7	77.4	6.1	1.9	4.3	57.0	
CCC [95% CI]	.51 [.49, .53]	.71 [.70, .72]	.82 [.81, .83]	.9 [.90, .91]	.93 [.92, .93]	.50 [.48, .52]	
r_s	.64	.76	.87	.91	.94	.64	
ProcessingPAL agreement with diary							
MAE	82.8	75.1	2.4	1.7	2.3	—	
CCC [95% CI]	.34 [.32, .36]	.65 [.63, .66]	.95 [.95, .95]	.73 [.72, .74]	.97 [.97, .98]	—	
r_s	.52	.73	.96	.93	.97	—	
ProcessingPAL agreement with CREA							
MAE	68.2	99.7	6.5	2.1	4.6	—	
CCC [95% CI]	.55 [.53, .57]	.62 [.60, .63]	.79 [.78, .80]	.78 [.77, .79]	.92 [.92, .92]	—	
r_s	.64	.71	.86	.89	.95	—	

Note. Only days with 10 hr of wear time and participants with 4 days were included in the analysis. CCC = concordance correlation coefficient; CI = confidence interval; MAE = mean absolute error; r_s = Spearman correlation.

* Significantly different from diary, $p < .001$.