24 Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices

Mary E. Rosenberger[1], **Matthew P. Buman**[2], **William L. Haskell**[3], **Michael V. McConnell**[4], and **Laura L. Carstensen**[1]

[1]Stanford Center on Longevity and Psychology Department, Stanford University, Stanford, CA

[2]School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ

[3]Stanford Prevention Research Center, Stanford University, Stanford, CA

[4]Cardiovascular Medicine, Stanford School of Medicine, Stanford University, Stanford, CA

Abstract

Getting enough sleep, exercising and limiting sedentary activities can greatly contribute to disease prevention and overall health and longevity. Measuring the full 24-hour activity cycle - sleep, sedentary behavior (SED), light intensity physical activity (LPA) and moderate-to-vigorous physical activity (MVPA) - may now be feasible using small wearable devices.

PURPOSE—This study compares nine devices for accuracy in 24-hour activity measurement.

METHODS—Adults (N=40, 47% male) wore nine devices for 24-hours: Actigraph GT3X+, activPAL, Fitbit One, GENEactiv, Jawbone Up, LUMOback, Nike Fuelband, Omron pedometer, and Z-Machine. Comparisons (to standards) were made for total sleep time (Z-machine), time spent in SED (activPAL), LPA (GT3x+), MVPA (GT3x+), and steps (Omron). Analysis included mean absolute percent error, equivalence testing, and Bland-Altman plots.

RESULTS—Error rates ranged from 8.1–16.9% for sleep; 9.5–65.8% for SED; 19.7–28.0% for LPA; 51.8–92% for MVPA; and 14.1–29.9% for steps. Equivalence testing indicated only two comparisons were significantly equivalent to standards: the LUMOback for sedentary behavior and the GT3X+ for sleep. Bland-Altman plots indicated GT3X+ had the closest measurement for sleep, LUMOback for sedentary behavior, GENEactiv for LPA, Fitbit for MVPA and GT3X+ for steps.

CONCLUSIONS—Currently, no device accurately captures activity data across the entire 24-hour day, but the future of activity measurement should aim for accurate 24-hour measurement as a goal. Researchers should continue to select measurement devices based on their primary outcomes of interest.

Keywords

Actigraph; GENEactiv; activPAL; accelerometers; activity monitors

Address for Correspondence: Mary Rosenberger, Department of Psychology, Jordan Hall, Building 01-420, Stanford University, 450 Serra Mall, Stanford, CA 94305; Telephone: 650-804-0843; Fax: 650-725-5699; maryr@stanford.edu.

Introduction

Substantial evidence has led to recommendations for adequate exercise, healthy sleep habits, and limited sedentary behavior for increased longevity, improved health, and disease prevention (7, 14, 21). Health research has focused intensely on these different daily activities, but in order for researchers, clinicians, and consumers to understand better these activity-health relationships, it is important to study the complete 24-hour activity cycle. Combined measurement of sleep, sedentary behavior and physical activity may be an important step in guiding activity recommendations throughout a 24-hour cycle. Current activity and sleep guidelines are limited to 30 minutes per day of exercise and 7–8 hours of sleep, leaving about 16 hours of unaccounted time with a non-quantified recommendation to avoid too much sitting.

The components of the 24-hour model, organized into domains of activity intensity, are sleep, sedentary behaviors (SED), light-intensity physical activity (LPA), and moderate-to-vigorous physical activities (MVPA, or "exercise"). For all non-sleep activities, sedentary behavior is defined as sitting or lying with energy expenditure less than 1.5 METs (32), LPA would include activities with energy expenditure between 1.5 and 3 METs (1), and MVPA includes moderate activity (3–6 METs) and vigorous activity (any activity greater than 6 METs) (1). A 24-hour model of activity was previously difficult to measure and incorporating the model into medical research was limited because of the error associated with the measurement. First, sleep, sedentary behavior and physical activity are traditionally studied in separate laboratories. Second, measurement technology had both limited memory and short battery life. Lastly, there has ben a lack of analytical methods to consider time spent in different activity levels and the relative relationships to health outcomes.

Sleep recommendations - to sleep for 7 to 8 hours per night – are based on observations that shorter or longer sleep durations are associated with risk factors for a range of diseases (7, 12,34, 35). Sedentary behavior recommendations are sparse (31), but objective monitoring of sedentary behavior has revealed relationships to a number of health outcomes (21), and several general recommendations have been published (13, 39). Exercise is also related to multiple health outcomes (14), and this has led to public health recommendation of 150 min of MVPA per week to contribute substantially to longevity and disease prevention (14). Increased LPA is associated with improved physical health and well-being measures in older adults (5). Decreased LPA contributes to several health risks including elevated plasma glucose (15) and higher blood pressure and lower HDL cholesterol (8), but not mortality rates (24). There are no recommendations for how much of the day should be spent in LPA compared to sedentary behavior.

Importantly, the relationships among these activity domains are not well understood. For example, physical activity can be used as a treatment for poor sleep (6), but research has not addressed the need for more sleep (or sedentary time) as recovery following several days of extended vigorous intensity exercise. The relationship among activity domains is also probably not stagnant, but changes across the life span, during specific physiological or disease states (i.e., pregnancy, diabetes), and with heavier physical training loads. Accurate

and reliable measurement of the 24-hour cycle could answer a number of these specific research questions that cannot be addressed with current measurement methods.

The collection of *objective* measures of sleep, SED, LPA, and MVPA has traditionally been costly, difficult, or non-existent. Technological advances now make these measurements possible, using small wearable devices. There has been a proliferation of wearable devices for the various components of daily activity, but little research into how these devices compare to one another and how valid and reliable they are compared to common research measurement methods. Figure 1 provides a representation of a 24-hour cycle of activity with current recommendations and a rough estimate of the proportion of SED to LPA. The 24-hour model is used in this study as a framework to help evaluate what these devices measure in each of the four activity domains. It should be noted that most devices are unable to produce this 24-hour chart with their current reported data. The purpose of this study is to compare the output from commercially available wearable devices using current standards for objective measurement of sleep, SED, LPA, and MVPA in the field. The ultimate goal of this research is to determine the best ways to measure the full 24 hours of activity behavior to guide future clinical studies and recommendations.

Methods

Participants

Participants were recruited from the Stanford University community and surrounding areas through word-of-mouth with an effort to include equal numbers of men and women over a wide age range. Before participation, all participants signed a written informed consent approved by the Stanford University Institutional Review Board. Participants (N=40, 21 women) came to the laboratory for instructions, initialization, and device fitting; then wore the devices for 24 consecutive hours during normal activities, and returned to the laboratory on the second day to return the devices. The mean age of the participants was 36 years (the range was 21–76 years).

Standards for Free-Living Activity Measurements

Measuring activity domains over the 24-hour day cannot be limited to specific activities that can be measured in a laboratory, but is dependent upon measuring free-living activities. The standards selected for comparisons in this study were not laboratory-based gold-standard devices, but the closest standard that could be conveniently worn during a complete 24-hour cycle in a free-living environment. The Z-machine measures brain activity with an electroencephalogram (EEG) in a portable monitor and is thus a more comparable measurement to polysomnography, the laboratory-based measure of brain activity, than actigraphy, or an accelerometer on the wrist (20). For SED, posture measurement is a key component of the definition, which involves sitting or lying while awake with an energy expenditure of less than 1.5 METs, so the activPAL monitor was the standard for this domain (23). The Actigraph GT3X+ is a frequently used device for LPA and MVPA measurement (40). The Omron pedometer was selected as the standard because it has been validated as an accurate measure of steps (17), and is independent of our other standard

devices. For example, the GT3X+ is a standard for other domains in this comparison, but it is not regularly used as the step counter in epidemiological studies.

Measurements

In addition to the above devices used as standards, the following wearable devices were studied – the Fitbit One, Jawbone Up, Nike Fuelband, GENEactiv, and LUMOback. Table 1 shows a listing and description of the nine devices worn in this study. Devices were selected to represent both research devices and commercially available devices that were in widespread use at the outset of the study, and measured at least one domain of the 24-hour cycle with some specificity.

At the beginning of the study, participants came to the laboratory where height, weight, age, and gender were collected and recorded. Software described in Table 1 was used to submit participant-specific information to each device for initialization. Participants also received both written and oral instructions of when to put on the devices and how to wear them. The LUMOback also required initial calibration, where the participant walks and then sits in a slouching position while following directions on the mobile device. This was performed in the laboratory using an iPhone 4S, connected to the LUMOback via Bluetooth, and the participant was guided directly by the app on the phone. After initialization, a study kit was prepared for the participant. It included all nine of the devices plus both a hip and wrist strap for the GT3X+; one clip and one strap for the Fitbit; alcohol wipes, extra electrodes, electrode cables, and the user manual (supplied by General Sleep Corporation) for the Z-Machine; a clip and a leash for the Omron; and several stickies for the activPAL.

Participants were asked to wear all nine devices for a day consisting of one full day of activity and one full night of sleep. Devices were worn from approximately the time a participant woke up until the participant woke up the next morning. If the participant did not wake up at the same time on the two consecutive days, more or less than 24 hours are recorded. A daily log was used to record when the participant woke up, what time the devices were put on, if they were taken off for bathing or water activities, when the participant got into bed for the purpose of sleeping, and when the participant woke up and took off the devices. A verbal follow-up was also conducted when the participants returned the devices to confirm times were accurately recorded.

During daily and nightly wear, device feedback was not provided to the participant except in cases where the data was presented on the device itself. Omron has a steps display, the Fuelband displays steps and Nike Fuel, the Fitbit displays steps, floors climbed, calories burned, and activity level. All other devices did not provide feedback to the user. No interventions were introduced such as step goals, vibrations to interrupt sedentary behavior, or other guidelines for the participant.

Device data were downloaded after the participant returned the study kit. Participants could view their data after the conclusion of their participation if they were willing to stay through data download. No written reports were provided to the participant. Data were either downloaded to the computer (Fitbit, GT3X+, Fuelband, and activPAL) or through the phone application (LUMOback and Jawbone) for devices that lack desktop software. Additionally,

a separate research portal, provided by the company, was used to download data from the LUMOback to obtain five minute epoch summaries, which are not provided by the consumer phone application.

Sleep

Devices compared to the Z-machine for measuring sleep duration included the Fitbit,, Jawbone, GENEactiv, and GT3X+. All of these were worn for the entire 24-hour period with the exception of the Z-Machine (only during sleep periods). The Z-Machine uses 3 electrodes on the head/neck. Calibration of the Z-machine included inputs of height, weight, and age through a computer connected to the device. Once initialized, the user could apply the electrodes, check electrode connection, and start sleep measurement independently.

All other sleep measurement devices were worn on the wrist and rely on an accelerometer-based measurement algorithm to estimate total sleep time. Commercial devices have proprietary algorithms for sleep, so total sleep time was recorded directly from the summary. LUMOback and activPAL do not have specific sleep measurement because sedentary time and sleep are recorded based on posture, therefore these devices were not analyzed for total sleep time measurement. The Fitbit was moved from the trunk to the wrist and placed in a sweatband-style sleeve for sleep measurement. A button on the device was also pressed and held, putting the device into sleep mode, when the user got into bed for the purpose of sleeping. Similar buttons were used on the Jawbone and the GENEactiv to start sleep measurement. The GT3X+ was also moved from the waist band to the wrist in a specially designed sweat-style band with a pocket designed to hold the device. The GT3X+ does not "log" sleep with a button push, sleep time started when the participant started logging sleep on the Z-machine and stopped when the electrodes stopped recording. If the Z-machine malfunctioned due to user error, the sleep log as recorded by the participant was used to determine start and stop times of sleep.

Sleep can be measured using a variety of variables, but this comparison was limited to total sleep time because this is the variable universally measured by sleep devices and has also been shown to have a relationship to health outcomes (7). A sleep-specific algorithm, specifically, the Sadeh sleep algorithm (36), was used to analyze data for the research devices (GT3X+ and GENEactiv). The commercial device summaries (Fitbit, Jawbone) were downloaded using the device-associated software. The raw data extracted from the activPAL on the thigh cannot be analyzed with the same Sadeh algorithm because it was developed for actigraphy on the wrist and the activPAL is worn on the thigh.

Sedentary Behavior

Devices compared to the activPAL for measuring SED duration included the GT3X+, GENEactiv, LUMOback, and Fitbit. Total minutes spent in sedentary behavior were found using the GT3X+ with a 150 count/min cutpoint (23), the GENEactiv worn on the right wrist with a <217 g*min cutpoint (10), the LUMOback with time spent in a sitting or lying posture, and the Fitbit with sedentary time defined on the dashboard (this feature was included in the original reporting, but removed when "tiles" were added to the dashboard).

The activPAL was used as the standard and adheres to the definition of sedentary behavior which includes sitting or lying. Devices that are accelerometer-based (GT3x+, GENEactiv, and Fitbit) will be measuring a lack of motion, not posture. Early sedentary research relied on motion measurement, yet a posture-based definition has evolved. This comparison will provide insight into the differences between posture and motion-based sedentary measurement. Therefore, they could not be included in comparisons of time spent in LPA.

Light Intensity Physical Activity

Devices compared to the GT3X+ for measuring LPA duration included the Fitbit and GENEactiv. A GT3X+ cutpoint of >150 and <1580 counts/min was used as the standard (11), and was compared to a GENEactiv cutpoint of 217–644 g*min (10), and time spent in light activity from the Fitbit. None of the other devices measured LPA, nor could it be derived from time spent in other behaviors.

Moderate to Vigorous Physical Activity

Devices compared to the GT3X+ for measuring MVPA duration included the Jawbone, Fitbit, GENEactiv, and. Fuelband A GT3X+ cutpoint of ≥1580 counts/min was used as the standard (11). The comparisons include active minutes from the Fuelband, active time from Jawbone, moderate plus vigorous minutes from the Fitbit, and a cutpoint of >644 from the GENEactiv (10). Other devices were not included in this comparisons because they did not measure time spent in MVPA.

Steps

Devices compared to the Omron for measuring steps included the Jawbone, Fitbit, Fuelband, GT3X+, LUMOback, and activPAL. All devices reported total steps per day.

Statistical Analysis

Table 2 summarizes the measurements provided by each device, which variables were used in this analysis and what device was used as a criterion measure for each activity domain. Standard sample calculations were conducted to set goals for subject recruitment, and alpha was set at .05 with the confidence interval set to 95%. Separate sample calculations were conducted for each domain. Statistical Analyses were performed to determine statistically significant differences as well as agreement among devices. Mean absolute percent errors (MAPE) are reported to establish differences between the devices and the "field-based" measurements, and determines accuracy. In addition, equivalence testing is reported to establish similarities between the devices and measurement standards. Bland-Altman plots were used to test biases between the standards and the other measurement devices. These measurements of differences, similarities and biases are similar to a recent study comparing devices to laboratory-based measurement of energy expenditure (25).

Results

Sleep Duration

Figure 2 illustrates the mean error analysis for the devices measuring sleep, ranging from 8.1% for GT3X+ to 16.9% for GENEactiv. Equivalence analysis, Figure 3, indicates the GT3X+ was equivalent to the Z-machine for sleep measurement, but the other devices showed significant differences. Bland-Altman plots had mean differences in measured sleep duration ranging from 4 min for GT3X+ to 36 minutes for Fitbit and GENEactiv. Summary data are provided in Table 3 and the original plots are contained in Supplemental Digital Content 1 (see Document, Supplemental Digital Content 1, Bland–Altman plots, including regression lines and average differences between the standard and the comparison device). The GT3X+ also had the lowest standard deviation (SD) on Bland-Altman analysis.

Sedentary Behavior

Figure 2 illustrates the mean error for sedentary behavior (i.e., sitting time), which ranged from 9.5% for LUMOback to 65% for GENEactiv. Equivalence testing (Figure 3) highlighted the LUMOback accurately measured sedentary behavior. All other devices produced significantly different estimates. Bland-Altman plots had mean differences ranging from 18 minutes for LUMOback to 162 minutes for GENEactiv (Table 3, and Supplemental Digital Content 1, Bland–Altman plots, including regression lines and average differences between the standard and the comparison device), with LUMOback also having the smallest SD. Since these numbers highlight a difference between posture-based measurement and motion-based measurement, results not reported here show if the GT3x+ was used as the standard, the GENEactiv would have significantly underreported sedentary behavior, but the Fitbit produced sedentary measurements equivalent to the GT3x+.

Light-Intensity Physical Activity

For LPA, mean absolute percent error (MAPE) from the GENEactiv was 20% and from Fitbit was 28%, as shown in Figure 2. Figure 3 illuminates significant differences in minutes of LPA from both GENEactiv and Fitbit. Lastly, the Bland-Altman summary in Table 3 gives an overestimation in LPA of 43 minutes for GENEactiv and underestimation of 64 minutes for Fitbit, with Fitbit having the smaller SD. The plots are contained in the Supplemental Digital Content 1 (see Document, Supplemental Digital Content 1, Bland–Altman plots, including regression lines and average differences between the standard and the comparison device).

Moderate-to-Vigorous Physical Activity

For MVPA, MAPE is illustrated in Figure 2 as ranging from 52% for Jawbone to 92% for Fuelband. All measurements were significantly different from the standard measure of MVPA. Mean differences from the monitors as determined by the Bland-Altman plots ranged from 48 minutes for Jawbone to 598 minutes for Fuelband, with the Jawbone also having the lowest SD.

Steps

Error rates for steps (as total steps per day) ranged from 14% for GT3X+ to 29% for Fuelband (Figure 2). All devices were significantly different from the standard for measuring steps (see Figure 3), and total step differences as large as 2500 steps. Bland-Altman plots had the smallest mean difference for GT3X+ at 698 steps, with the largest difference for activPAL at 2258 steps (Table 3), and the lowest SD for the GT3X+.

Discussion

Objective measurement of sleep, sedentary behavior, and physical activity is an important component of both research and feedback from consumer wearables. All of the activity domains are related to disease outcomes. This study suggests that measurement of these domains is highly varied among wearable devices when tested outside of the laboratory. While this may sound discouraging, the ability to measure very specific behaviors has greatly increased with the introduction of a large number of wearable devices. For sleep, this study shows that many of the devices can measure total sleep time with the predictable error that comes from comparing actigraphy to polysomnography. For sedentary behavior, this study highlights the differences between posture measurement (LUMOback being similar to activPAL) and an accelerometer measurement indicating a lack of motion (GT3X+, Jawbone, Fitbit, GENEactiv). For LPA and MVPA, this study also suggests there are major differences between the devices and that these devices may be using different measures of the behavior of interest. For example, LPA is usually defined as 1.5–3.0 METs, but not all devices may be trying to identify that intensity as LPA. For steps, many of the devices were different from the standard, but gave similar results to each other, implying some predictable agreement among devices.

Currently, 24-hour activity measurement is only possible with research devices, such as the GT3X+. None of the commercial devices provide all the measures of the 24-hour model. Tapping into richer data from application programming interfaces (or APIs) from commercial devices may allow complete 24-hour measurement, but it may be significantly different from previous measurement standards. For this reason, choosing a device specific to the primary outcome measure of interest will be of utmost importance. Calibration and evaluation of devices will be an ongoing research area because of the rapid changes in wearable technology. Evaluating devices for their ability to determine time spent at different intensities is highly relevant to optimal health, yet many devices are not created specifically with this focus in mind. This study highlights a lack of standards among commercial devices for important health-related objective activity measurement. The following discussion will highlight areas of interest in each activity domain and propose recommendations for manufacturers and device calibration experts.

Sleep

Actigraphy has previously been used to measure sleep/wake patterns with some reliability (37). Additionally, a single-channel electrode is an accurate method for sleep/wake detection relative to full polysomnography (20), and this was the method employed with the Z-machine. The portable electrode method of the Z-machine produced a similar difference in

total sleep time as the scoring of polysomnography (20, 30), and further exploration of the Z-machine may lead to better portable EEG sleep measurement in the field. While there are published algorithms for sleep scoring (36), none of the consumer accelerometer-based devices publish their algorithms for measuring sleep, creating an issue with comparisons of the devices. Previously, the Fitbit was found to overestimate total sleep time and lacked sleep/wake specificity similar to how other accelerometer-based devices compared to polysomnography (30). These results were replicated in this study, and in general, the sleep devices overestimated total sleep time. Since this study highlights some agreement between the sleep/wake measurement of consumer devices and research devices, use of these devices in research should be explored further. Algorithm development work is currently ongoing in this regard for the activPAL.

Sleep measurement from consumer devices covers aspects of sleep that were not examined in this study. Total sleep time was evaluated because stages of sleep, sleep efficiency or measurement of circadian rhythms are not recommended using actigraphy on the wrist (37). For example, the Jawbone Up has a number of sleep variables (light vs deep sleep) that contradict the recommendation for measurement with wrist actigraphy from sleep experts (37). Other variables that could be explored in future research include sleep latency, number of awakenings, time spent in different stages of sleep, and sleep efficiency. The evaluation of all sleep variables from these devices is dependent on either polysomnography in the laboratory or creation of a portable standard measure. Also, the sleep/wake measurement should be evaluated with different devices in broader populations.

Sedentary Behavior

Sedentary behavior measurement is complicated by varying definitions used to describe the lack of activity. Current definitions rely on a combination of posture (i.e. sitting) (32, 38), low levels of energy expenditure (32, 33), or specific activities (such as TV viewing, but not including sleep) (33). A promising outcome of this paper is the addition of LUMOback as an accurate measure of daily posture. Many health outcome studies that highlight the importance of limiting sedentary behavior found associations without the *postural* measurement defined in this paper (3, 15,16, 29), creating a debate on which measurement (postural or lack of motion) is important for health (32). Unfortunately, postural measurement devices are not necessarily the best devices for other components of the full 24-hour activity cycle, because they lack specificity in measurement of activity intensity. The design and goal of a study will determine whether a postural device should be used (e.g. sedentary interventions to reduce sitting) or whether 24-hour measurement should be prioritized (e.g. controlling for sedentary behavior in physical activity studies).

LPA

Relatively little is known about LPA because of the difficulty in obtaining accurate objective measurement (also true for assessing LPA by questionnaire) (4). In the past, LPA has been measured using a 7-day recall and subtracting sleep, sedentary time, and MVPA from 24 hours as opposed to having a direct estimate of LPA (4). Measuring LPA in the 24-hour cycle can be done with any device that can separate sedentary behavior and MVPA from LPA, but since there is no device that accurately captures LPA, a recommendation cannot be

made based on the results presented here. An important part of creating an accurate 24-hour measurement device will be the improved measurement of LPA during daily activities. 24-hour activity measurement could lead to a recommendation of how much time should be spent in LPA (which is also a major displacement of sedentary time) on a daily basis to optimize disease prevention.

MVPA

A surprising result of this study is that MVPA was not accurately measured by a number of devices. Given the small percentage of time spent in MVPA in many populations, even modest measurement error is clinically significant in a 24-hour period. One reason for the discrepancy in measurement could simply be the definition of MVPA. Many commercial device companies do not provide a definition of what they are measuring, so while the official definition of moderate activity includes any activity ≥3 METs and < 6 METs (1), there is no confirmation this is what the devices are attempting to measure. For example, the Jawbone UP defines their activity measurement only as "time spent moving" (19). In this study, MVPA had 51–91% error, most likely because the devices were measuring different activity than the official definition. One recommendation of standardizing activity measurement would be to adhere to commonly used definitions of intensity. Alternatively, the calibration of the Actigraph on the hip was one of the earliest calibration studies (11), and is still used as the standard in epidemiological research (2, 28). Research shows the relationship between these standards and health outcomes (16, 24, 28), making these an appropriate standard to use while calibrating devices.

The results of this study also call into question the ability of field methods to accurately measure MVPA. In recent evaluations of these devices for predicting energy expenditure, Jawbone and Fitbit were more accurate than the GT3X+ (25). The GT3X+ provides a measure of MVPA different from the measures of MVPA provided by the other devices, but it is not necessarily more accurate at measuring activities with energy expenditure above 3 METs. A recent study concluded that the cutpoint analysis of GT3X+ data underestimates time spent in MVPA compared to other methods (22). Cutpoint analysis is also not universally applicable and has known limitations, such as cutpoints for younger adults are not the same as those specifically created for older adults (28). This limitation is specific to the algorithm used, not to the device overall. In this case, a useful follow up will be to see if other device measures of MVPA have the same relationship to health outcomes as cutpoints on the GT3X+. Luckily, large databases of activity measurement are being created by the users of these devices. Defining the optimal amount of MVPA based on objective measurement may have to become device specific or, at the very least, current methods in physical activity epidemiology should consider additional standardization.

Steps

In this study, none of the wearables measured steps in the same way as the Omron, but a recent article found that the Fitbit One might be the most accurate device for measuring steps compared to researcher step counts (9). Many of the devices are dependent on the "bout" or number of steps you take in order for the device to count those steps and the "timeout" or the time between steps that will reset the "bout" (17). Given those two

variables, a recommendation should be developed as to what type of walking is most beneficial to health. For example, researchers may determine if a one-step bout requirement has a different relationship to health outcomes at 10,000 steps a day compared to an algorithm that requires a 4-step bout requirement.

Recommendations and Conclusions

Research has identified areas of our daily activity cycle that relate to health in many ways. Sleep research has focused on finding a healthy amount of sleep to prevent disease and optimize performance in our daily activities. Sedentary behavior research has cautioned about the detrimental health outcomes and metabolic disturbances that come from inactivity. LPA research has focused on the added benefit of burning extra calories through more movement in a 24-hour day. MVPA research, based primarily on survey data, has a very specific relationship to health in a dose-response manner with most benefits coming from getting 30 minutes or more of moderate-intensity physical activity in a day. At present, the most common activity intervention is to increase daily exercise, but for those who are sleeping less than six hours a night, increasing exercise may prove to be less important than increasing sleep to over 7 hours a night.

Given what we know about activity and the link to better health, these domains should be measured objectively, with accuracy, and in ways that can be compared to guidelines defined by the biomedical community. We should strive to make these activity definitions and measures match as closely as possible for both feedback to the user and for researchers to gain a better understanding of the rich datasets being generated by a barrage of new wearable users. The importance of 24-hour measurement in medical research, as well as for consumer application, raises a number of areas that should be considered for future device research. The explosion of new wearables, along with the addition of new devices, software upgrades and other changes, demand continuous updating of device evaluations. The expanding measurement capabilities of devices, with multiple physiological and contextual measures, will continue to expand how research can be conducted. Heart rate, for example is a common theme among the upcoming Apple Watch, Jawbone 3, Basis Peak, Microsoft Band, Fitbit Surge, and a number of other "smartwatch" devices. The addition of heart rate to the motion data presents a new avenue for defining sleep, sedentary behavior, and all levels of physical activity. Not only is there research needed in the validation of these devices, but there will be a number of proposed applications of these devices in medicine and public health. Wearables offer a great opportunity to obtain much more detailed data about how each person spends their life.

The results presented in this paper are a step toward accurate objective monitoring of full 24h spectrum of behaviors; yet this study does have significant limitations. First, the standards used in this study are based upon common field-based measures and do not represent gold standards used in the laboratory. Therefore, both the test device and criterion device introduce substantial error into the comparisons. Second, placement of activity monitors can affect how well these devices match up to standards, and location is an important consideration based on feasibility for long-term monitoring and wearability. Our focus was on accuracy of sensors based upon recommended placement, yet wearability must

also be considered. Last, the functions of these devices change with every software and hardware update, and therefore, not every possible update can be evaluated with the research at one particular point in time.

Importantly, with the volume and complexity of data generated by these 24-hour monitoring devices, researchers will need to expand the analytical techniques that are used to combine information when examining relationships among activities and health outcomes. Multiple data inputs from various devices can be quite complicated and the field lacks consensus about how to combine devices for an optimal daily activity cycle focused on promoting health while preventing negative health outcomes. An optimal activity cycle will be exceptionally important for quantifying activity as well as designing and evaluating interventions to promote health.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Ainsworth B, Haskell W. 2011 compendium of physical activities: a second update of codes and MET values. Med. Sci. Sport. Exerc. 2011; 43(8):1575–1581.

2. Atienza AA, Moser RP, Perna F, et al. Self-reported and objectively measured activity related to biomarkers using NHANES. Med. Sci. Sports Exerc. 2011; 43:815–821. [PubMed: 20962693]

3. Balkau B, Mhamdi L, Oppert J-M, et al. Physical activity and insulin sensitivity: the RISC study. Diabetes. 2008; 57(10):2613–2618. [PubMed: 18591396]

4. Blair SN, Haskell WL, Ho P, et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. Am. J. Epidemiol. 1985; 122:794–804. [PubMed: 3876763]

5. Buman MP, Hekler EB, Haskell WL, et al. Objective light-intensity physical activity associations with rated health in older adults. Am. J. Epidemiol. 2010; 172:1155–1165. [PubMed: 20843864]

6. Buman MP, King AC. Exercise as a Treatment to Enhance Sleep. Am. J. Lifestyle Med. 2010; 4(6): 500–514.

7. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur. Heart J. 2011; 32(12): 1484–1492. [PubMed: 21300732]

8. Carson V, Ridgers ND, Howard BJ, et al. Light-Intensity Physical Activity and Cardiometabolic Biomarkers in US Adolescents. PLoS One. 2013; 8:e71417. [PubMed: 23951157]

9. Case M, Burwick H, Volpp K, Patel M. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA. 2015; 313(6):625–626. [PubMed: 25668268]

10. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Med. Sci. Sports Exerc. 2011; 43(6):1085–1093. [PubMed: 21088628]

11. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med. Sci. Sports Exerc. 1998; 30:777–781. [PubMed: 9588623]

12. Gangwisch JE, Heymsfield SB, Boden-Albala B, et al. Sleep duration as a risk factor for diabetes incidence in a large U.S. sample. Sleep. 2007; 30(12):1667–1673. [PubMed: 18246976]

13. Hamilton MT, Healy GN, Dunstan DW, Zderic TW, Owen N. Too Little Exercise and Too Much Sitting: Inactivity Physiology and the Need for New Recommendations on Sedentary Behavior. Curr. Cardiovasc. Risk Rep. 2008; 2(4):292–298. [PubMed: 22905272]

14. Haskell WL, Lee I-M, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation. 2007; 116(9):1081–1093. [PubMed: 17671237]

15. Healy GN, Dunstan DW, Salmon J, et al. Objectively measured light-intensity physical activity is independently associated with 2-h plasma glucose. Diabetes Care. 2007; 30:1384–1389. [PubMed: 17473059]

16. Healy GN, Wijndaele K, Dunstan DW, et al. Objectively measured sedentary time, physical activity, and metabolic risk: the Australian Diabetes, Obesity and Lifestyle Study (AusDiab). Diabetes Care. 2008; 31(2):369–371. [PubMed: 18000181]

17. Holbrook EA, Barreira TV, Kang M. Validity and Reliability of Omron Pedometers for Prescribed and Self-Paced Walking. Med. Sci. Sports Exerc. 2009; 41(3):669–673.

18. Jacobs DR, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med. Sci. Sports Exerc. 1993; 25:81–91. [PubMed: 8423759]

19. Jawbone. Jawbone UP [Internet]. 2014 Available from: https://jawbone.com/kb/articles/418.html.

20. Kaplan RF, Wang Y, Loparo KA, Kelly MR, Bootzin RR. Performance evaluation of an automated single-channel sleep-wake detection algorithm. Nat. Sci. Sleep. 2014; 6:113–122. [PubMed: 25342922]

21. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med. Sci. Sports Exerc. 2009; 41(5):998–1005. [PubMed: 19346988]

22. Keadle SK, Shiroma EJ, Freedson PS, Lee I-M. Impact of accelerometer data processing decisions on the sample size, wear time and physical activity level of a large cohort study. BMC Public Health. 2014; 14(1):1210. [PubMed: 25421941]

23. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med. Sci. Sports Exerc. 2011; 43(8):1561–1567. [PubMed: 21233777]

24. Lee I-M, Paffenbarger RS. Associations of Light, Moderate, and Vigorous Intensity Physical Activity with Longevity: The Harvard Alumni Health Study. Am. J. Epidemiol. 2000; 151(3):293–299. [PubMed: 10670554]

25. Lee J-M, Kim Y, Welk GJ. Validity of Consumer-Based Physical Activity Monitors. Med. Sci. Sports Exerc. 2014:1840–1848. [PubMed: 24777201]

26. Levine J, Eberhardt N, Jensen M. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. Science. 1999; 283(5399):212–214. [PubMed: 9880251]

27. Levine J, Schleusner SJ, Jensen MD. Energy expenditure of nonexercise activity. Am. J. Clin. Nutr. 2000; 72(6):1451–1454. [PubMed: 11101470]

28. Loprinzi PD, Lee H, Cardinal BJ, Crespo CJ, Andersen RE, Smit E. The Relationship of Actigraph Accelerometer Cut-Points for Estimating Physical Activity With Selected Health Outcomes: Results From NHANES 2003–06. Res. Q. Exerc. Sport. 2012; 83:422–430. [PubMed: 22978192]

29. Lynch BM, Dunstan DW, Healy GN, Winkler E, Eakin E, Owen N. Objectively measured physical activity and sedentary time of breast cancer survivors, and associations with adiposity: findings from NHANES (2003–2006). Cancer Causes Control. 2010; 21(2):283–288. [PubMed: 19882359]

30. Montgomery-Downs H, Insana S, Bond J. Movement toward a novel activity monitoring device. Sleep Breath. 2012; 16:913–917. [PubMed: 21971963]

31. Morris JN, Heady JA, Raffle PAB, Roberts CG, Parks JW. CORONARY HEART-DISEASE AND PHYSICAL ACTIVITY OF WORK. Lancet. 1953; 262(6795):1053–1057. [PubMed: 13110049]

32. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: the population health science of sedentary behavior. Exerc. Sport Sci. Rev. 2010; 38:105–113. [PubMed: 20577058]

33. Pate RR, O'Neill JR, Lobelo F. The evolving definition of "sedentary". Exerc. Sport Sci. Rev. 2008; 36(4):173–178. [PubMed: 18815485]

34. Patel SR. Reduced sleep as an obesity risk factor. Obes. Rev. 2009; 10(Suppl 2):61–68. [PubMed: 19849803]

35. Patel SR, Hu FB. Short sleep duration and weight gain: a systematic review. Obesity (Silver Spring). 2008; 16(3):643–653. [PubMed: 18239586]

36. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. Sleep Med. Rev. 2011; 15(4):259–267. [PubMed: 21237680]

37. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep. 1994; 17:201–207. [PubMed: 7939118]

38. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms "sedentary" and "sedentary behaviours". Appl. Physiol. Nutr. Metab. 2012; 37(3):540–542. [PubMed: 22540258]

39. Tremblay MS, Leblanc AG, Janssen I, et al. Canadian sedentary behaviour guidelines for children and youth. Appl. Physiol. Nutr. Metab. 2011; 36:59–64. 65–71. [PubMed: 21326378]

40. Troiano R, Berrigan D. Physical activity in the United States measured by accelerometer. Med. Sci. Sports Exerc. 2008; 40(1):181–188. [PubMed: 18091006]
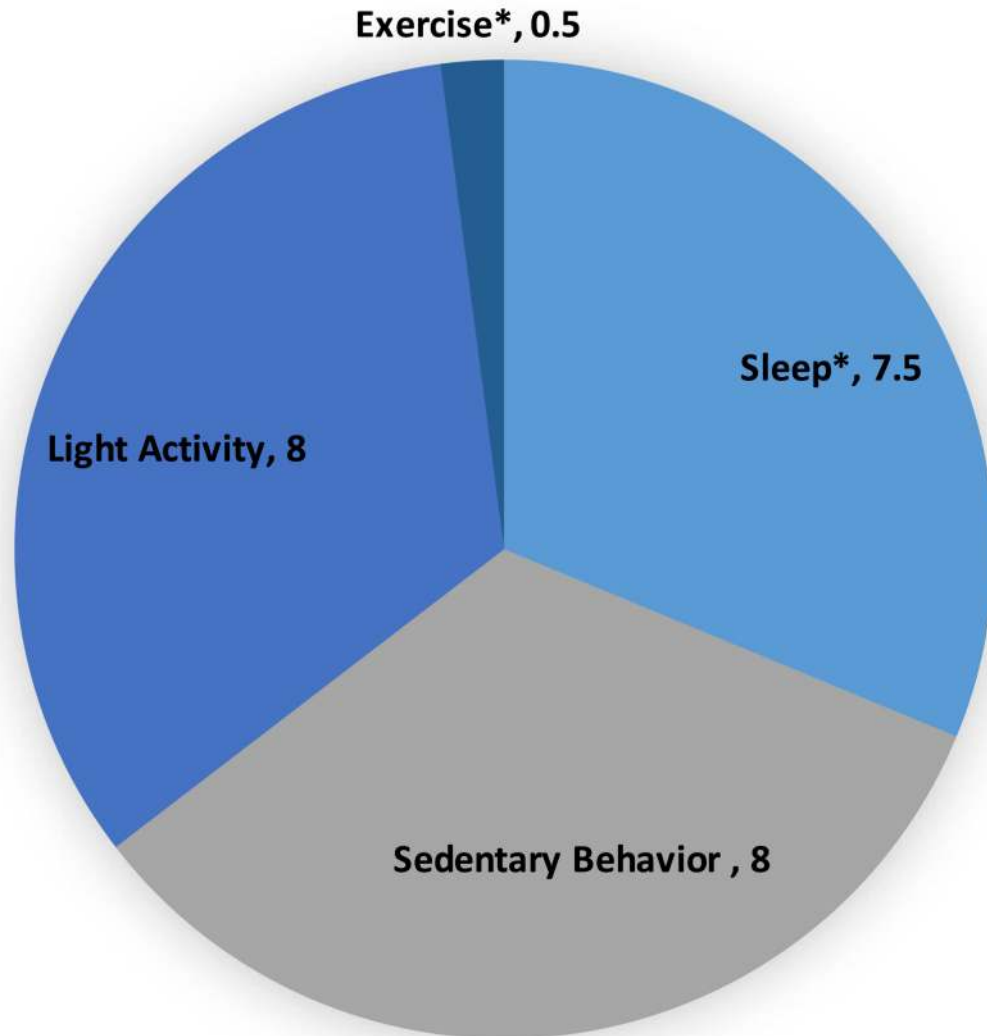
Figure 1.
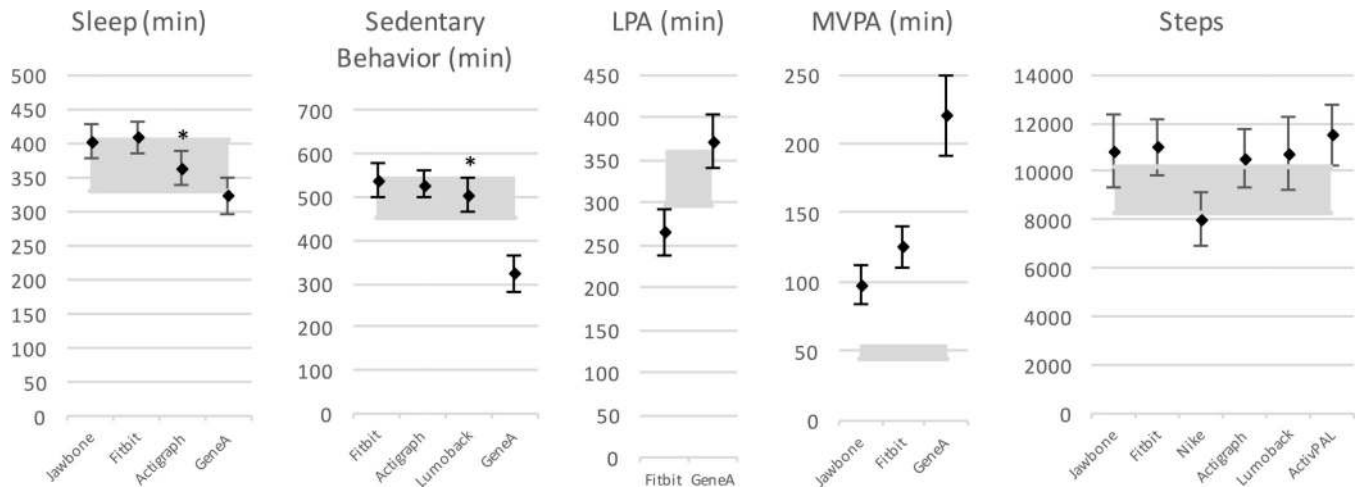Pie chart with current recommendations and estimates of the optimal 24-hour physical activity cycle.
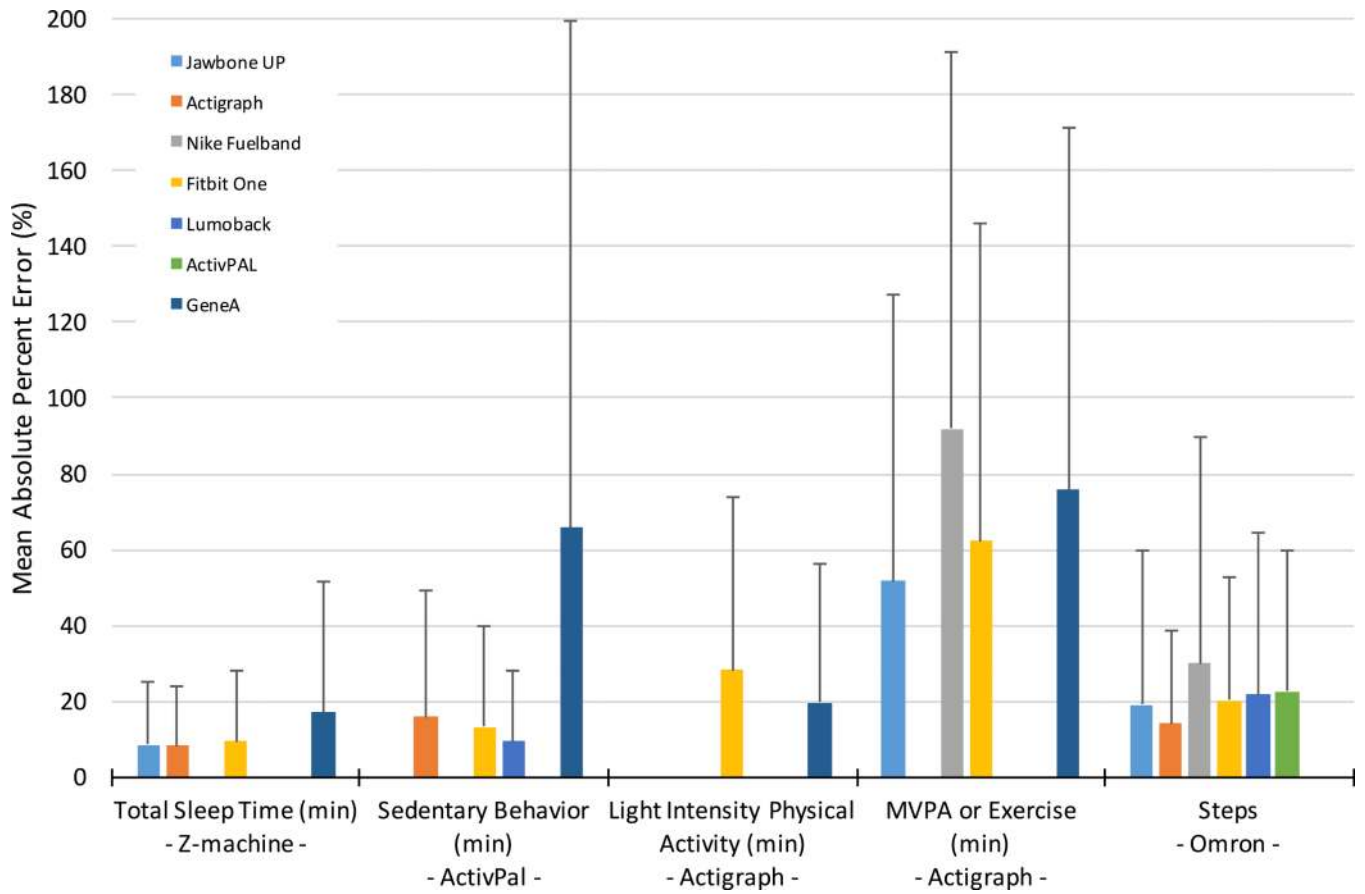
Figure 2.
Mean absolute percent error for the various devices and five activity domains.

Figure 3.
Equivalence testing for all of the devices in all domains. Shaded areas are equivalence zones (±10% of the mean), and error bars indicate the 90% confidence interval for the mean measurement. * Equivalent measures.

Table 1

A list of devices included in this study by company, versions used, and location worn.

Company	Device/Version	Company Location	Location Worn	Software
General Sleep	Z-Machine	Euclid, OH	Head electrodes	Z-machine Data Viewer
PAL Technologies, Limited	activPALvt	Glasgow, UK	Right Thigh	ActivPAL3 v7.1.18
Actigraph, LLC.	GT3X+	Pensacola, FL	Day: Right Hip Night: Right Wrist	Actilife 6
Omron Healthcare, Inc.	HJ-112 Pocket Pedometer	Lakeforest, IL	Right Hip	On screen summary
Fitbit	One	San Francisco, CA	Day: Right Hip or Pocket Night: Left Wrist	Desktop sync, online feedback, also iPhone app
Activinsights, Ltd.	GENEactiv Original	Kimbolton, Cambs, UK	Right Wrist	GENEactiv PCSoftware, Version 2.2
Jawbone	Jawbone UP	San Francisco, CA	Right Wrist	iPhone App
LumoBodytech, Inc.	LUMOback	Palo Alto, CA	Lower back	iPhone App
Nike, Inc.	Fuelband	Beaverton, OR	Right Wrist	Desktop sync and online feedback, also iPhone App

Table 2

Device variables reported from the 9 devices included in the study. Variables used in analysis are highlighted in BOLD.

Device	Sleep	Sedentary	Light	Moderate/Vigorous	Steps
Z Machine (Sleep)	Sleep/Wake time[*]	N/A	N/A	N/A	N/A
activPAL	Included as Sedentary	Sitting Time[*]	Standing Time	Stepping Time	Steps
Actigraph GT3x+	Total Sleep Time	Sedentary < 150	Cutpoints[*] 150–1579	Cutpoints[*] 1580+	Steps
Omron	N/A	N/A	N/A	Moderate Steps/Time	Steps[*]
Fitbit One	Actual Sleep Time, Latency, Number of Awakenings, Efficiency	Sedentary Time	Light Time	Moderate Time + Vigorous Time	Steps
GENEactiv	Total Sleep Time	Sedentary cutpoint <217	Cutpoints 217–643	Cutpoints 644+	N/A
Jawbone Up	Sleep time, % of goal, Light sleep, Latency, Deep sleep, Awake time, number of awakenings	Longest idle	N/A	Longest Active, Active Time	Steps, Distance, percent of goal
LUMOback	Total Sleep Time, Right, Left, Side and Back	Sitting Time, Slouching Time, Straight Time, Standing Time	Walk	Walk	Steps
Fuelband	N/A	(Fuel by Hour)	(Fuel by Hour)	Active Time	Steps

[*] Variables used as a criterion measure. The units for GT3x+ cutpoints are counts/min and for GENEActiv are g*min.

Table 3

Bland-Altman Plot Summaries for all of the domains and all of the devices. The actual plots are contained in Supplemental Digital Content 1.

Domain	Device	Mean	SD	Slope	P-Value
Sleep (min)	UP	32	37.7	0.02	0.76
	One	36	37.8	−0.09	0.21
	GT3x+	4	35.8	−0.14	0.05
	GENEactiv	−36	51.8	−0.17	0.12
Sedentary Behavior (min)	LUMOback	18	52.1	−0.16	0.004
	GENEactiv	−162	110.0	−0.30	0.028
	One	34	81.2	−0.34	0.0006
	GT3x+	48	100.2	−0.47	<0.0001
LPA (min)	GENEactiv	43	91.8	−0.40	0.014
	One	−64	47.7	−0.18	0.03
MVPA (min)	One	76	39.2	−0.05	0.79
	UP	48	33.7	−0.06	0.69
	GENEactiv	170	89.3	−0.07	0.86
	Fuelband	598	134.2	−0.63	0.33
Steps	UP	1527	2708	0.19	0.047
	One	1878	1287	0.002	0.95
	Fuelband	−1267	1879	−0.06	0.02
	LUMOback	1281	1692	0.02	0.733
	GT3x+	679	1267	−0.07	0.160
	activPAL	2258	1452	−0.07	0.178