Published in final edited form as:

Med Sci Sports Exerc. 2017 March ; 49(3): 617–624. doi:10.1249/MSS.0000000000001124.

Comparison of Accelerometry Methods for Estimating Physical Activity

Jacqueline Kerr¹, Catherine R. Marinac^{1,2}, Katherine Ellis³, Suneeta Godbole¹, Aaron Hipp⁴, Karen Glanz⁵, Jonathan Mitchell^{6,7}, Francine Laden^{8,9}, Peter James⁹, and David Berrigan¹⁰

¹UC San Diego, Department of Family Medicine & Public Health, La Jolla, CA

²Graduate School of Public Health, San Diego State University, San Diego, CA

³UC San Diego, Department of Electrical and Computer Engineering, La Jolla, CA

⁴Department of Parks, Recreation, and Tourism Management, North Carolina State University Center for Geospatial Analytics, Center for Human Health and the Environment, North Carolina State University, Raleigh, NC

⁵Perelman School of Medicine and School of Nursing, University of Pennsylvania, Philadelphia, PA

⁶Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA

⁷Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁸Departments of Environmental Health and Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

⁹Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

¹⁰National Cancer Institute, National Institutes of Health, Bethesda, MD

Abstract

Purpose—To compare physical activity estimates across different accelerometer wear locations, wear time protocols, and data processing techniques.

Methods—A convenience sample of middle aged to older women wore a GT3X+ accelerometer at the wrist and hip for 7 days. Physical activity estimates were calculated using three data processing techniques: single axis cut points, raw vector magnitude thresholds, and machine learning algorithms applied to the raw data from the three axes. Daily estimates were compared for the 321 women using generalized estimating equations.

Corresponding Author: Jacqueline Kerr; 9500 Gilman Dr., 0811; UC San Diego; La Jolla, CA 92093-0811; Phone: 858-534-9305; jkerr@ucsd.edu; Fax: 858-534-9304.

CONFLICTS OF INTEREST

None of the authors have conflicts of interest, and the results of this study do not constitute endorsement by ACSM.

Results—A total of 1,420 days were analyzed. Compliance rates for the hip vs wrist location only varied by 2.7%. All differences between techniques, wear locations and wear time protocols were statistically different ($p < .05$). Mean minutes per day in physical activity varied from 22–67 depending on location and method. On the hip, the 1952 count cut point found at least 150 minutes per week of physical activity in 22% of participants, raw vector magnitude found 32%, and the machine learned algorithm found 74% of participants with 150 minutes of walking/running per week. The wrist algorithms found 59% and 60% of participants with 150 minutes of physical activity per week using the raw vector magnitude and machine learned techniques respectively. When the wrist device was worn overnight, up to 4% more participants met guidelines.

Conclusion—Estimates varied by 52% across techniques, and by as much as 41% across wear locations. Findings suggest that researchers should be cautious when comparing physical activity estimates from different studies. Efforts to standardize accelerometry-based estimates of physical activity are needed. A first step might be to report on multiple procedures until a consensus is achieved.

Keywords

Actigraph; accelerometer; wrist; hip; machine learning; sedentary behavior

INTRODUCTION

Physical inactivity is the 4th leading contributor to death in the US (5). Most of the evidence to support a relationship between physical activity and health as well as the guidelines to improve physical activity are derived from self-reported data (3, 11, 17, 30). Self-reported data may be subject to bias and recall problems. Further, self-reported data does not allow researchers to explore precise intensities and timing of physical activity. A growing number of researchers have been employing objective measures of physical activity, including person-worn devices such as accelerometers.

Hip, rather than wrist-worn, accelerometers have been traditionally employed to assess physical activity because the trunk location was thought to provide a more accurate assessment of center of mass acceleration (4). Cut points to categorize the intensity of physical activity for hip worn accelerometers were developed based on laboratory protocols of physical activity and energy expenditure. These cut points used proprietary “count” data from the vertical x-axis. Early studies of wrist worn accelerometers using these “counts” found that the wrist location poorly predicted energy expenditure (25).

The technology and application of accelerometers in physical activity research is rapidly changing (19). Raw acceleration data are now available for three axes; novel computational techniques that move away from traditional count-based classification approaches are emerging (24); and a number of researchers have advocated placing accelerometers at the wrist to estimate physical activity because estimates are improving and more wrist worn devices are becoming available (28). While some of these developments seem likely to improve physical activity classification, they also present a number of logistical and analytic challenges.

To date, most of the new computational algorithms available for researchers have been developed for both hip and wrist worn accelerometers using data from individuals in laboratory settings (6, 23). One recent study, in the laboratory, compared different computation techniques (14). However, movement in the laboratory setting may not reflect free-living physical activity behavior because laboratory-based movements generally occur in sequences defined by the investigator and do not include outdoor activities such as driving and bicycling. Thus, it is not clear if these algorithms are applicable to epidemiological cohorts and intervention studies in free-living adults (2, 15). In one study, the performance of wrist- and hip-based algorithms from both laboratory and free-living training data were compared (7). The free-living algorithms to detect behaviors such as walking and sitting performed better than the laboratory based algorithms, and the hip algorithms performed better than the wrist algorithms when the behaviors were known from annotated observations of images collected by person-worn cameras (7). Other studies have shown little difference in hip- and wrist-based estimates of physical activity in laboratory settings (6, 23).

Wrist-based accelerometers are now being used in a variety of studies to improve participant compliance, where compliance is defined as the total number of hours collected and number of people meeting valid day criteria (27, 28). For example, the 2011–2014 NHANES survey with wrist placement had substantially higher compliance than the 2003–2006 NHANES with hip placement (27). Higher compliance rates for wrist worn devices might be partly due to the wrist location being preferred by users because it is more comfortable. Also, when participants are asked to wear the devices for 24 hours per day, wrist-based devices are less likely to be removed at night and not re-placed in the morning. In most hip-based accelerometer studies, participants have been instructed to wear the hip accelerometer during waking hours only. Some hip-based studies have improved compliance by employing 24 hour protocols and have shown that participants are able to wear hip worn devices overnight (8, 13). Currently it is unclear whether the wrist or hip location results in the most complete and accurate 24 hour measurement of physical activity and sedentary behavior. Compliance may be higher with wrist placement, but some wrist-based algorithms appear to be less accurate in classifying activity (7). Another reason to consider the wrist placement for 24 hour activity monitoring is because it has been shown to provide a better approximation of patterns of polysomnography-defined sleep than the hip location (21), although few studies have attempted to use the hip location for sleep assessment or improve sleep estimates based on hip wear (16).

The current study studied hip and wrist based accelerometers in a free-living cohort of middle aged to older women across the US, including working adults, nurses, and breast cancer survivors. Novel machine-learned algorithms were applied to the raw data and compared with traditional count cut points and laboratory based algorithms using raw vector magnitude. This paper aims to identify differences in physical activity and sedentary behavior estimates between the wear locations, wear time protocols, and data processing techniques. Analyses will inform comparisons of accelerometer data across studies where different methods are likely to be employed.

METHODS

Population

Participants were recruited from four university sites (Harvard University (Harvard), University of California, San Diego (UCSD), University of Pennsylvania (UPenn), and Washington University in St. Louis (WUSTL)) involved in the NCI-funded Transdisciplinary Research in Energetics and Cancer (TREC) initiative (20). The sample included working adults (UCSD and WUSTL), nurses (Harvard), and breast cancer survivors (UPenn). The institutional review boards at each university site approved the study protocol and consent forms, and all participants provided written informed consent. Study protocols specific to the collection of accelerometer data were identical across all sites and all data used in these analyses were centrally pooled and uniformly processed at UCSD.

Protocol

All participants wore the Actigraph GT3X+ accelerometer for waking hours simultaneously on the hip and for 24 hours on the wrist. Monitors were mailed to 87% of participants with clear instructions, a link to an instructional video, followed by 2 check-in phone calls. The remaining participants received and returned the devices in person. Because participants also wore a GPS device secured on the same belt as the hip accelerometer (to avoid the two becoming separated), they were not asked to wear the hip accelerometer for 24 hours (as GPS devices have to be charged overnight and would be uncomfortable to wear in bed). Thus we cannot compare hip and wrist data over a 24 hour period. This paper will therefore compare hip- and wrist-based estimates of physical activity and sedentary behavior for the minutes when both were worn during waking hours with at least 600 minutes a day for the hip (34). The paper will also compare the non-wear defined waking hours to the waking hours from the 24 hour protocol in the wrist only. Participants completed sleep logs for 7 days. Sleep was excluded from the analysis based upon these self-reported sleep logs. Re-wear for devices was requested if the screener determined that there were fewer than 5 days of data with 600 minutes of wear time or fewer than 4 days with 3000 minutes total wear time for either the hip or wrist accelerometer. Additionally, the hip data was manually screened for less than 4 wear periods and human wear patterns to assess days of wear versus mailing days. The wrist day was screened for typical patterns of day and night wear, including almost continuous count values for day wear and longer series of zero counts punctuated by a few minutes of counts for night wear. Screening procedures were employed during data collection to ensure compliance and quality control across sites. Once all data was collected standardized data processing was performed at one site (UCSD). Not all participants in the final data set provided 5 days of data.

Data Processing

Raw accelerometer data at 30 Hertz were collected on 3 axes on the wrist of the non-dominant hand and hip. Wear time for the hip was processed using the Choi algorithm in Actilife 6.11, which assesses 90 consecutive minutes of zero *counts* as non-wear and includes a 30 minute small window to remove artefactual movement. For all analyses, only days with at least 600 minutes of hip wear time were included. All days with this criteria were included. We did not require a minimum number of days for analysis. We compared

compliance proportions with the wake time criteria across wrist and hip locations. We also provide wrist-based estimates with the 24 hour protocol for comparison, since researchers argue that the longer wear time will also provide more opportunity to capture physical activity (27). For estimates of activities, only wake time was employed. For the 24 hour data wake time was established by 7 days of self reported sleep logs. Table 1 summarizes the comparisons made across wear time protocols, wear locations, and data processing techniques. Matching the wrist and hip by wear time allows comparisons for location only, taking out variations in wear time. Additional comparisons of wrist data with wake time established by sleep logs, allows assessments of both wear time and location. This way we are able to establish independent effects of location and wear time, rather than confounding location differences by wear time. On the data presented to reflect a 24 hour wrist protocol, a non-wear time procedure of 60 continuous zeros on the minute level *vector magnitude* data was employed, after sleep time was removed. This approach is emerging for wrist data as the 90 minute count window used on the hip would not be appropriate on the wrist.

Counts per minute (CPM) from the X-axis of the hip accelerometer were aggregated and divided into three groups: Sedentary (0–99); Light (100–1951); moderate to vigorous physical activity (MVPA) (1952+) based on established cut points (9, 18). We also employed a new method called “GGIR” to classify MVPA in the hip and wrist from the raw vector magnitude of the three axes (32, 33). The thresholds (wrist 100 and hip 69) to classify MVPA in this algorithm were developed on a laboratory sample of 30 adults, both men and women (18–65 yr)(10). We selected the “GGIR” approach as it was the only available wrist and hip algorithm for the Actigraph GT3X+ at the time of data processing, and because we had validated its performance in the same set of women who provided training data for the machine learned algorithm employed here (7).

We also applied a machine-learned (ML) algorithm called two-level behavior classification (TLBC) to predict behaviors from tri-axial wrist and hip acceleration data. This classifier uses a combination of a random forest classifier and a hidden Markov model to predict four behaviors (sitting, standing, walking/running, and riding in a vehicle). In a previous study, the classifier was trained on a sample of 38 overweight or obese women (mean age = 55.2 ± 15.3 yrs; BMI = 32.0 ± 3.7) wearing two ActiGraph GT3X+ accelerometers (right hip, non-dominant wrist) for 7 free-living days. The women, from the UCSD site, were both breast cancer survivors and those at risk for breast cancer. Images captured by wearable cameras at 15 second intervals were used to classify activities of free living study participants and employed as the criterion (7).

Statistical Analyses

Separate generalized estimating equations (GEE) were used to make comparisons between methods while accounting for clustering within individuals (12). Comparisons were made at the day-level. All methods were compared across the hip for PA (MVPA or walking/running) and sedentary time (sitting or sedentary). We also provided single axis count data to allow readers to compare count intensities across the different behaviors and processing modes, although the count data was not employed in the GGIR or machine learned procedures. The GGIR hip and wrist algorithms were compared across the hip and the

matched wrist minutes. Then the hip-matched wrist minutes and the 24 hour protocol wrist minutes were compared. The same analyses were performed for each machine learned behavior (hip vs wrist, wake time wrist vs 24hr wrist protocol). We also compared the number of participants who would meet the 150 minutes weekly physical activity guideline by the wear locations, processing method, and wear time.

RESULTS

Of the 402 people recruited to the study, 368 returned both wrist and hip accelerometers with valid days of wear time, and an additional 3 people returned accelerometers with valid data after requesting a re-wear. 38 participants were not included in this analysis because they were used as the training sample used to define the machine-learned activity recognition classifier, resulting in a dataset of 333 women. Self-reported sleep logs were available for 321 participants over 1420 days and sleep time was removed based on these logs, leading to a final dataset of 321 women with 1420 days of data. The women who had sleep logs were more likely to have a lower BMI, more likely to be married, and had a higher education level ($p<.05$) compared to those without logs and those in the training sample. As shown in Table 2, participants were women with a mean age of 55 years ($SD=9.2$) and were predominantly white (~80%). About 70% of the sample were college graduates and only 12 participants were Latina. Half the sample was employed full time.

Table 3 shows the descriptive statistics for the number of hours the accelerometers were worn. It illustrates the wear time captured in the hip wake time protocol and the wrist 24 hour protocol including sleep time. As expected, substantially more hours per day were collected with the 24 hour protocol on the wrist. The percent of participants, however, who had wear time hours that met the 600 minute criteria for wake time was very similar for the hip and wrist locations.

Table 4 shows the mean daily minutes between the hip data processing methods, between the hip and wrist locations, and between wear time criteria across the wrist. There was a 10 minute per day difference in MVPA minutes between the GGIR raw vector magnitude algorithm and the single axis cut point method based on the hip accelerometer. The machine learned walking/running estimates were double the MVPA estimates although median CPM were similar across the GGIR and machine learned methods. There was almost a 3 hour per day difference in the hip cut point-based estimate of sedentary time and the hip machine learned sitting time. The mean MVPA minutes estimated using the GGIR algorithm were 20 minutes greater for data collected using wrist-based devices compared to devices worn on the hip. In contrast, the machine learned algorithm identified 10 fewer minutes of walking/running time for devices worn on the wrist, compared with the data collected using the hip accelerometers. The 24 hour wrist protocol (excluding sleep time) provided only 2–3 additional minutes of walking or physical activity per day for the machine learning and raw vector magnitude algorithms, when compared to the wrist wake time estimates derived from the hip wear time protocol. For the machine learned algorithm, estimates of standing still and standing moving were higher at the wrist than at the hip location and estimates of sitting were lower.

Table 5 shows the percent of participants that would be classified as meeting physical activity guidelines. The hip cut point classified the fewest participants as meeting guidelines and the hip machine learned walking classified the most people. The “GGIR” wrist algorithm and machine learned wrist algorithm classified similar numbers of participants as meeting the guidelines both in the wrist wake time and 24 hour protocols.

DISCUSSION

The current analyses compared three data processing protocols, two wear locations, and two wear time protocols. Future studies may use different techniques and understanding differences between them will inform our interpretation of population level physical activity. In our analyses, prevalence of meeting physical activity guidelines (by MVPA or walking at any intensity) varied by up to about 50% depending on which data processing technique was employed. Differences in meeting physical activity guidelines across body wear locations were as much as 20%. The current findings will inform consensus development for accelerometer wear and data processing protocols in future studies.

Researchers have recently called for standard applications of data processing techniques, at the same time as suggesting raw data processing should be employed (35). A recent laboratory study compared available computational techniques, demonstrating that different methods had different strengths depending on the type of measure being predicted (14). The development and real world validation of new algorithms in free living settings is on-going. Until we have more free living algorithms, applied to population level cohorts, and comparisons across methods to health outcomes it is not clear what method and wear location should be recommended. At this stage, comparisons across studies with varying methods may not be informative (28). Many large cohort studies have recently completed data collection on the hip, including some studies with 24 hour hip protocols (16). These studies are looking for new techniques to improve physical activity and sedentary behavior classification. Some major population studies, (e.g. NHANES), are now employing 24 hour wrist protocols and many projects may decide to adopt the wrist location. There are fewer algorithms available for the wrist and only a few studies have compared validation across the wrist and hip location (6, 7, 22, 23). Further, a limited number of new classification techniques have been developed on free living training data (15). This is one of the first studies to apply new classification techniques for the hip and wrist from free living training data matched to the sample, and compare the classification with a laboratory based algorithm for the hip and wrist.

Not surprisingly, wear time was greater in the wrist location when the 24 hour protocol was employed compared with the wake time wear protocol in the hip location. The number of compliance days was also slightly greater in the wrist location. The percent of participants who met the hip criteria (98%) was higher than in other reported studies where participants were more compliant with the wrist than the hip protocol and preferred this wrist location (13). This level of compliance was particularly impressive given that most participants (89%) were mailed their devices. However, it is notable that the current study included multiple days of training for staff in how to motivate compliance, including through 2 reminder phone calls. We also requested that participants re-wear devices if criteria were not

met, although this did not always result in full compliance. Such efforts may not be feasible in larger cohort studies.

Researchers have argued that, in addition to improved compliance with the wrist location and 24 hour protocol, the longer wear time will also provide more opportunity to capture physical activity. Although statistically different, our results do not show much gain in minutes or percent meeting guidelines (up to 4% more) with the 24 hour wrist protocol, compared to the hip derived wake time protocol on this wrist. Longer wear time may however, improve estimates of sedentary behavior, especially if worn for 24 hours at the hip. There were greater estimates of sitting time in the 24 hour protocol on the wrist, but greater estimates of sitting time with a shorter wear period on the hip. While this paper focuses on comparing the physical activity and sedentary behavior estimates of the hip and wrist locations, decisions on wear location may also include consideration of these locations to assess sleep duration and sleep quality. Sleep researchers have typically employed the wrist location and algorithms for detecting sleep duration and sleep quality and algorithms for these locations have been validated against gold standard polysomnography (36). While there are studies showing that sleep duration can be assessed at the hip (29), few studies have assessed the hip location compared to polysomnography (36). For example, one study by Zinkhan and colleagues found sufficient validity for the sleep duration measure at the hip, but very poor performance of the sleep quality metrics. This is not surprising as they were applying wrist-validated algorithms for sleep quality to the hip. More research is required to develop hip-based algorithms that can detect sleep quality. Nonetheless, researchers should be aware of the current differences in performance of sleep variables at the wrist and hip if they are considering a 24 hour protocol.

We observed large differences in physical activity behavior estimates across hip and wrist locations, and the direction of these differences varied by the way the data were processed (e.g., using machine learning to derive estimates of physical activity behaviors ie. walking, or using the “GGIR” approach to derive physical activity intensities). For example, the machine learned algorithm classified more minutes of walking/running (at any intensity) for hip-worn accelerometers compared to accelerometers worn at the wrist. However, “GGIR”-derived estimates of MVPA for hip-worn accelerometers were lower than estimates from wrist-worn devices. Without a ‘gold standard’ it is not possible to confirm which position or processing approach is more accurate; however accuracies reported in the literature are higher for the machine learned algorithms than “GGIR” algorithms (machine learned accuracy is estimated at over 80%, compared to 59% for “GGIR”)(7, 10). Further, in the same women who provided training data for the machine learned methods applied here, we tested the GGIR algorithm against image annotations and reported similar estimates of MVPA across locations (17 and 18 minutes per day on hip and wrist), whereas the hip machine learned algorithm always performed better than the wrist location (7). This suggests that when applied to potentially more ‘noisy’ free living data, the GGIR (developed in the laboratory in adult and older men and women) may over estimate MVPA.

Both the vector magnitude and the data machine learned algorithms classified many more participants being active 150 minutes per week than the currently widely used cut points. These findings support other studies showing that the current hip based cut points may

underestimate common forms of physical activity (including walking) in older aged cohorts cohort (15). Current population level data with hip based accelerometer cut points suggest less than 5% of adults and less than 3% of older adults meet physical activity guidelines (26). In this sample of healthy middle aged and older adult women the cut points classified more than 20% of the sample as meeting guidelines. This suggests this sample may be particularly active and the results may not be generalizable. Comparatively, the more complex data processing techniques detected more minutes of physical activity and walking per day. New techniques may present a more positive picture for public health. Currently the physical activity guidelines are for moderate intensity physical activity, which the “GGIR” raw vector magnitude algorithm identifies. The GGIR MVPA estimates from the wrist identified more people meeting guidelines than the hip location. Depending on wear location, population estimates of MVPA could vary.

The behavior based machine learned algorithm identifies all locomotion, regardless of intensity. Although the count data were not employed in the ML algorithm, for comparison, median counts per minute for the walking classification were 1070, suggesting that most walking occurs below the 1952 MVPA threshold. The behavioral algorithms identified more than double the amount of walking than either the laboratory based algorithm or the cut points identified as MPVA minutes. Given the previously validated higher accuracy of the hip based free living algorithm to detect walking (7), these results suggest that the hip location will detect more walking behavior than the wrist location.

While current evidence demonstrates a dose response effect of exercise intensity, suggesting that higher intensity exercise provides the most benefits, these conclusions were drawn from studies that lacked objective information about total walking. With the Surgeon General’s new report on walking (31), it is important to understand the health effects of walking both at any intensity and at higher intensities. Longer durations than the current guidelines may be recommended if higher intensities cannot be achieved. Accurate behavior-based algorithms can further these investigations. If we wish to simplify public health guidelines with a ‘walking’ message, the behavior-based algorithms demonstrate that 74% of this sample walks for 150 minutes a week. If the relative difference in estimates were applied to national data sets we might see almost 20% of the population meeting walking guidelines. Clearly there is still great improvement needed in population levels of PA. More accurate estimates will hopefully better reflect real changes over time with continued public health efforts to improve PA levels. Many efforts will include built environment improvements that will focus on walking, hence accurate population estimates are needed.

Importantly, the behavior-based algorithms we applied to this cohort were able to distinguish sitting from standing. As studies demonstrate that excessive sitting time may be related to poor health, it is important to have more accurate estimates of sitting time. Cut point estimates of sitting are likely to include standing (which is not a sedentary behavior) shown by the median counts per minute of 96 for the machine learned algorithm defined standing category. While ActivPAL devices (placed on the thigh) are becoming the gold standard in sitting intervention studies, there are many large prospective epidemiological cohorts with high quality measures of health outcomes that have hip accelerometer data (16). With improved measures of sitting and standing, we can more accurately estimate the health

effects of prolonged sitting (i.e. sitting without breaks). Current accelerometer measures of breaks based on the 100 count per minute cut point have been shown to underestimate total sedentary time compared to criterion measures (15). In these analyses, the difference between the 100 count and machine learned sitting time was almost 3 hours per day. Such a magnitude of difference may have impacts on health outcomes. In addition to classifying standing time that falls within the 100 count threshold (15), the machine learned algorithm identifies vehicle time separately (as its pattern is different and the counts can exceed 100). After processing, we combine the vehicle minutes with the sitting (not in vehicle) minutes. While the cut point approach was not designed to specify vehicle time, this is clearly a sedentary behavior that contributes to inactivity levels. New methods that can distinguish categories of sitting can improve our understanding of such behaviors with health. Importantly, sitting can be detected in both hip and wrist locations. Some researchers believed hand movement during sitting would prevent detection of sitting, however, the patterns created by this type of movement are different from other activities. We saw almost an hour more sitting time from the hip algorithm than the wrist and this combined with the higher physical activity detection in the hip may weigh into researchers decisions to adopt the hip or wrist location.

Limitations

The current study was a convenience sample of middle aged and older women, including cancer survivors and nurses from multiple locations across the US who may not represent the population as a whole. Physical activity was higher in this cohort than other population studies so our results may not be generalizable. With an average daily total count of 232,549, our sample would be in the 50% percentile for age (55 years) of the range recently reported by Bassett et al (1). Given that this study compared methods, we would not expect a differential bias across techniques or body placement. Unfortunately the current study did not have a gold standard for comparison (e.g. person worn camera data or in person observations) that would allow us to establish which technique provided a more accurate estimate of the behaviors and intensities examined. We, however, compared two algorithms that have been validated against known behaviors, in this study's female training sample, and this information helped us interpret the current results. Further, the women in this sample (mean age 55) were intentionally very similar to the women employed in the machine learning training and validation phase (mean age 55). BMI, however, differed between the training and application sample (32 vs 27) which may have affected the performance of the algorithm. The GGIR laboratory algorithm was developed on adults, both men and women (mean age 34), and the 1952 cut point on young adults (men and women). The differences in estimates may partly due to the fact that cut points and GGIR data processing techniques were not validated in middle aged and older women. Some women in the sample were breast cancer survivors, but cancer status was not known for all participants, limiting our ability to consider this in our analyses. Finally, we were not able to compare 24 hour protocols across the hip and wrist because our participants also wore GPS devices on the same belt as the hip accelerometer. They could have not worn this bulkier device overnight and it had to be charged overnight. Due to our need to have matched GPS and accelerometer data for other study purposes, we did not allow participants to remove the

GPS from the accelerometer belt. Future studies with combined GPS and accelerometer devices that could be worn overnight might prove fruitful.

Conclusion

Results of device-based measurement of physical activity in middle aged and older women vary greatly by data processing techniques and device placement. In our analyses prevalence of physical activity and walking/running varied by almost 50% depending on which data processing technique was employed. Differences across body wear locations were as much as 20%. Thus efforts to standardize processing methods, or include standard estimates regardless of the choices made in a particular study, are vital. Otherwise, comparisons across studies will remain extremely challenging. While concrete wear location and processing recommendations cannot be made at this stage, this study demonstrates differences in methods that need to be considered when comparing estimates of physical activity and behavior across studies using different device placement and processing techniques. Where possible, researchers should report estimates using multiple processing techniques so that more informed comparison can be made until a consensus is established. As more wrist based studies with 24 hour protocols are published, we need to understand differences in prevalence rates that may be due to the data collection and processing techniques, not necessarily population differences.

Acknowledgments

SOURCE OF FUNDING

This work was supported by the NCI Centers for Transdisciplinary Research on Energetics and Cancer (TREC) (U01 CA116850, U54 CA155496, U54 CA155626, U54 CA155435, U54 CA155850) and NIH Grants (UM1 CA176726) and (R01 ES017017). Dr. James was supported by Harvard Cardiovascular Epidemiology Training Grant T32HL098048 (NHLBI) and K99CA201542 (NCI). Dr. Mitchell was supported by Award Numbers F32CA162847 (NCI) and K01HL123612 (NHLBI). Dr. Marinac was supported by Award Number F31CA183125 (NCI). The opinions or assertions contained herein are the private ones of the authors and are not considered as official or reflecting the views of the National Institutes of Health. Results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. Bassett DR, Troiano RP, McClain JJ, Wolff DL. Accelerometer-based physical activity: total volume per day and standardized measures. *Med Sci Sports Exerc.* 2015; 47(4):833–838. [PubMed: 25102292]
2. Bastian T, Maire A, Dugas J, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. *J Appl Physiol* (1985). 2015; 118(6):716–722. [PubMed: 25593289]
3. Blair SN, Brodney S. Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. *Med Sci Sports Exerc.* 1999; 31(11 Suppl):S646–S662. [PubMed: 10593541]
4. Chen KY, Acra SA, Majchrzak K, et al. Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes Technol Ther.* 2003; 5(6):1023–1033. [PubMed: 14709206]
5. Danaei G, Ding EL, Mozaffarian D, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med.* 2009; 6(4):e1000058. [PubMed: 19399161]

6. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas*. 2014; 35(11):2191–2203. [PubMed: 25340969]
7. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and Wrist Accelerometer Algorithms for Free-Living Behavior Classification. *Med Sci Sports Exerc*. 2016; 48(5):933–940.
8. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Med Sci Sports Exerc*. 2016; 48(2):245–253.
9. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998; 30(5):777–781. [PubMed: 9588623]
10. Hildebrand M, van Hees V, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc*. 2014; 46(9):1816–1824. [PubMed: 24887173]
11. Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA. Physical activity and survival after breast cancer diagnosis. *Jama*. 2005; 293(20):2479–2486. [PubMed: 15914748]
12. Horton NJ, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Stat Med*. 2004; 23(18):2911–2933. [PubMed: 15344194]
13. Huberty J, Ehlers DK, Kurka J, Ainsworth B, Buman M. Feasibility of three wearable sensors for 24 hour monitoring in middle-aged women. *BMC Womens Health*. 2015; 15:55. [PubMed: 26223521]
14. Kate RJ, Swartz AM, Welch WA, Strath SJ. Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data. *Physiol Meas*. 2016; 37(3):360–379. [PubMed: 26862679]
15. Kerr J, Patterson RE, Ellis K, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. *Med Sci Sports Exerc*. 2016; 48(5):951–957. [PubMed: 27089222]
16. Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med*. 2014; 48(3):197–201. [PubMed: 24297837]
17. Manson JE, Rimm EB, Stampfer MJ, et al. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet*. 1991; 338(8770):774–778. [PubMed: 1681160]
18. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003–2004. *American journal of epidemiology*. 2008; 167(7):875–881. [PubMed: 18303006]
19. Matthews CE, Hagstromer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. *Med Sci Sports Exerc*. 2012; 44(1 Suppl 1):S68–S76. [PubMed: 22157777]
20. Patterson RE, Colditz GA, Hu FB, et al. The 2011–2016 Transdisciplinary Research on Energetics and Cancer (TREC) initiative: rationale and design. *Cancer Causes Control*. 2013; 24(4):695–704. [PubMed: 23378138]
21. Ray MA, Youngstedt SD, Zhang H, et al. Examination of wrist and hip actigraphy using a novel sleep estimation procedure. *Sleep Science*. 2014; 7(2):74–81. [PubMed: 25580202]
22. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. *Med Sci Sports Exerc*. 2013; 45(5):964–975. [PubMed: 23247702]
23. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. *J Appl Physiol (1985)*. 2015; 119(4):396–403. [PubMed: 26112238]
24. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc*. 2012; 44(1 Suppl 1):S61–S67. [PubMed: 22157776]
25. Swartz AM, Strath SJ, Bassett DR Jr, O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*. 2000; 32(9 Suppl):S450–S456. [PubMed: 10993414]

26. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008; 40(1):181–188. [PubMed: 18091006]
27. Troiano, RP., McClain, JJ. Proceedings of the 8th International Conference on Diet and Activity Methods Rome, Italy: NHANES. Rome, Italy: 2012. Objective measures of physical activity, sleep, and strength in U.S. National Health and Nutrition Examination Survey (NHANES) 2011–2014; p. 24
28. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med.* 2014; 48(13):1019–1023. [PubMed: 24782483]
29. Tudor-Locke C, Barreira TV, Schuna JM Jr, Mire EF, Katzmarzyk PT. Fully automated waist-worn accelerometer algorithm for detecting children's sleep-period time separate from 24-h physical activity or sedentary behaviors. *Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme.* 2014; 39(1):53–57.
30. U.S. Department of Health and Human Services. Washington, DC: Physical Activity Guidelines Advisory Committee Report, 2008; p. 1-683.
31. U.S. Department of Health and Human Services. Step it Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities. Washington, DC: Dept of Health and Human Services, Office of the Surgeon General; 2015. p. 1-56.
32. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol (1985).* 2014; 117(7):738–744. [PubMed: 25103964]
33. van Hees VT, Gorzelniak L, Dean Leon EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One.* 2013; 8(4):e61691. [PubMed: 23626718]
34. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc.* 2005; 37(11 Suppl):S582–S588. [PubMed: 16294121]
35. Wijndaele K, Westgate K, Stephens SK, et al. Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. *Med Sci Sports Exerc.* 2015; 47(10):2129–2139. [PubMed: 25785929]
36. Zinkhan M, Berger K, Hense S, et al. Agreement of different methods for assessing sleep characteristics: a comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. *Sleep medicine.* 2014; 15(9):1107–1114. [PubMed: 25018025]

Table 1

Summary of accelerometer-based comparisons made across wear time protocols and data processing techniques.

	Hip ^a	Wrist ^b	Wrist 24 hour ^c
	Wake time data defined by hip 600+ minutes a day	Wake time data defined by hip 600+ minutes a day	Wake time defined by sleep logs
Cut points, developed in laboratory based on X-axis only	Yes	No	No
Raw vector magnitude “GGIR” algorithm, developed in laboratory	Yes	Yes	Yes
Machine learned algorithm, developed in free-living based on multiple statistical features of all axes	Yes	Yes	Yes

^a hip device only worn during waking hours, non wear time excluded using Choi

^b wrist device matched to hip wear so that comparisons are based on location difference only, not wear time differences

^c wrist device during waking hours so comparisons between location and wear time can be made, 60 zeros on vector magnitude employed as non wear criteria

Table 2

Demographic characteristics of women in the study sample (n=321).

Age mean (SD), years	55.5 (9.2)
Body Mass Index mean (SD), (kg/m ²)	27.0 (5.94)
Race n (%)	
White	252 (78.5%)
Black/African-American	52 (16.2%)
Other	17 (5.3%)
Ethnicity n (%)	
Hispanic/Latina	12 (3.7%)
Not Hispanic/Latina	302 (94.1%)
Marital Status n (%)	
Never Married	27 (8.4%)
Divorced or Separated	46 (14.3%)
Widowed	9 (2.8%)
Presently Married or Living with Partner	239 (74.5%)
Education n (%)	
Some College or less	97 (30.2%)
College Graduate	224 (69.8%)
Employment n (%)	
Employed full-time, 35 hours or more per week	155 (48.3%)
Employed part-time, less than 35 hours per week	77 (24.0%)
Other	84 (26.2%)

Compliance estimates for participants with at least 600 minutes of accelerometer wear time per day (n = 321), stratified by hip and wrist locations.

Table 3

	Mean total time across all days (hours)	Mean # days (range)	Mean #hours/day (range)	% meeting hip wake time wear criteria (600 minutes/day)
All days with device data				
Hip	115	7.9 (2 – 16)	14.4 (11.3–19.4)	91.0%
Wrist	184	8.6 (6 – 16)	21.4 (17.5–24.0)	98.6%
All days with Sleep logs				
Hip	64	4.4 (1–8)	14.3 (10.4 – 17.3)	97.3%
Wrist	108	4.6 (1–8)	23.5 (18.8 – 24.0)	100%

Table 4

Day level comparison of minutes classified by algorithms, wear locations and wear time protocols for n=321 participants across n=1420 wear days waking hours only.

	Hip 600+ min wake time		Wrist 600+ min hip defined wake time		Wrist 24 hour wear/wake time only	
	Mean (SD) mins/day	Median CPM	Mean (SD) mins/day		Mean mins/day	
Wear time	865.2 (105.1)	--	865.2 (105.1)		946.7 (100.7)	
X-axis cut points						
Sedentary (0–99 CPM)	527.9 (116.6)	0	--		--	
Light (100–1951 CPM)	314.9 (98.7)	401	--		--	
MVPA (1952 + CPM)	22.4 (25.8)	2453	--		--	
“GGIR” raw VM algorithm						
MVPA	32.0 (53.8)	867	51.5 (47.1)	<i>c</i>	54.2 (48.2)	<i>d</i>
Machine-learned algorithm						
Sitting/lying	694.3 (166.4)	7	627.9 (149.1)	<i>e</i>	696.4 (157.9)	<i>f</i>
Standing Still	4.5 (11.1)	96	12.2 (21.1)	<i>e,g</i>	12.6 (21.3)	<i>f,g</i>
Standing Moving	90.9 (103.5)	500	163.6 (109.2)	<i>e,g</i>	173.6 (111.5)	<i>f,g</i>
Walking/Running	66.8 (53.2)	1070	58.7 (65.9)	<i>e,g</i>	60.7 (67.3)	<i>f,g</i>

* Abbreviations: VM = vector magnitude; CPM= counts per minute; SD= standard deviation

^a Comparison between hip estimates of MVPA 1952+ cutpoints, GGIR MVPA, and Machine-learned activity class walking/running, adjusting for wear time (p <0.05)

^b Comparison between hip estimates of Sedentary cutpoints, and Machine-learned sitting, adjusting for wear time (p <0.01)

^c Comparison between GGIR hip, & wrist (wake hours), after adjusting for wear time (p<0.01)

^d Comparison between GGIR hip, & wrist (24 hours), after adjusting for wear time (p<0.01)

^e Comparison between machine learned activity class hip, & wrist (wake hours), after adjusting for wear time (p<0.01)

^f Comparison between machine learned activity class hip, & wrist (24 hours), after adjusting for wear time (p<0.01)

^g Comparison between machine learned activity class wrist (wake hours), & wrist (24 hours), after adjusting for wear time (p<0.05)

Table 5

Percent of participants meeting guidelines of at least 150 minutes of physical activity per week (in 321 women) by data collection and processing techniques.

	Hip wake time protocol	Wrist wake time protocol	Wrist 24 hour protocol
	Pt, n (%)	Pt, n (%)	Pt, n (%)
Hip Cut points MVPA	70 (21.8) ^a	--	--
“GGIR” raw vector magnitude MVPA	101 (31.5) ^{a, b}	188 (58.6) ^b	201 (62.6) ^b
Machine learned walking (all intensities)	237 (73.8) ^{a, c}	191 (59.5) ^c	194 (60.4)

Pt: Participant

^aComparison between cut points, GGIR, and machine learned activity for meeting PA guidelines using hip estimates (p< 0.01)

^bComparison between hip, wrist wake time, and wrist 24 hour estimates of GGIR PA for meeting PA guidelines (p < 0.01)

^cComparison between hip, wrist wake time, and wrist 24 hour estimates of machine learned walking for meeting PA guidelines estimates (p < 0.01)