A Comparison of Accelerometry Analysis Methods for Physical Activity in Older Adult Women and Associations with Health Outcomes Over Time

Katie J. Thralls[1,2], **Suneeta Godbole**[2], **Todd M. Manini**[3], **Eileen Johnson**[4], **Loki Natarajan**[2], **Jacqueline Kerr**[2]

[1]San Diego State University

[2]University of California, San Diego

[3]University of Florida, Gianseville, FL

[4]University of California, Berkeley

Abstract

This study compared five different methods for analyzing accelerometer-measured physical activity (PA) in older adults and assessed the relationship between changes in PA and changes in physical function and depressive symptoms for each method. Older adult females (N=144, M_{age}=83.3±6.4yrs) wore hip accelerometers for six days and completed measures of physical function and depressive symptoms at baseline and six months. Accelerometry data were processed by five methods to estimate PA: 1041 vertical axis cut-point, 15-second vector magnitude (VM) cut-point, 1-second VM algorithm (Activity Index (AI)), machine learned walking algorithm, and individualized cut-point derived from a 400-meter walk. Generalized estimating equations compared PA minutes across methods and showed significant differences between some methods but not others; methods estimated 6-month changes in PA ranging from 4 minutes to over 20 minutes. Linear mixed models for each method tested associations between changes in PA and health. All methods, except the individualized cut-point, had a significant relationship between change in PA and improved physical function and depressive symptoms. This study is among the first to compare accelerometry processing methods and their relationship to health. It is important to recognize the differences in PA estimates and relationship to health outcomes based on data processing method.

Keywords

physical function; SPPB; CESD; machine learning

Corresponding Author: Katie J. Thralls, 6672 Reservoir Lane, San Diego, CA 92115, (925)413-9791, kthralls@ucsd.edu.

Conflict of Interest

The authors do not have any conflicts of interest.

Introduction

The benefits of physical activity (PA) are widespread through all stages of life. Specific to aging, PA prevents the natural decline of all physiological systems, decreases fall risk, maintains physical function needed for activities of daily living (ADLs), and improves quality of life (Chodzko-Zajk, 2013; Fielding et al., 2007; Manini & Pahor, 2009; Neto & Fernandes de Castro, 2012; Taylor et al., 2004). Previously, estimates of time spent in PA relied on self-report. However, many studies now use accelerometers to directly measure the amount of activity in different intensities (i.e., sedentary, light, moderate, vigorous) (Troiano, McClain, Brychta, & Chen, 2014). Accelerometer devices, particularly with older adults and in intervention trials, help mitigate inherent biases of self-report. However, advancements in the data available from accelerometers and new methods to process these data make it challenging to compare studies that apply different processing methods.

With advancements in technology, accelerometry data processing has evolved over time. Early laboratory calibration studies, primarily conducted in young to middle-aged adults, developed specific cut-points based on accelerometer counts per minute (cpm) that serve to determine the amount of time spent in moderate-to-vigorous physical activity (MVPA) (Freedson, Melanson, & Sirard, 1998; Troiano, Berrigan, Dodd, Masse, & McDowell, 2008). However, the natural decline in physical capacity with age requires an older adult to move at a higher relative intensity to accomplish tasks that take less effort by a younger individual. Thus, cut-points that were developed in younger populations often misclassify the activity levels for older adults by underestimating the amount of physical activity. This underestimation is important when applied to physical activity interventions because the data will not capture meaningful behavior changes and may appear unresponsiveness to the changes in physical activity. Thus, recent studies have developed other cut-points and equations for intensities specific to older adults from laboratory datasets (Bai et al., 2016; Copeland & Esliger, 2009; Evenson et al., 2015). It is likely, however that activities undertaken in a laboratory setting or a clinical test environment (such as the 400-meter walk (MW)) do not represent activities undertaken by individuals when they are unobserved in their natural environment. Accelerometers worn in a free-living setting for multiple days can assess habitual physical activity. Rosenberg et al. (2017) also developed and validated a machine learned algorithm from free living data to classify physical activity walking behavior in older women, a preferable form of physical activity for older adults.

Other studies have also recognized the wide variability of physical capacity in older adults and have proposed a need for a relative, or "individualized" cut-point, based on an individual's fitness level or performance on a field measure (Miller, Strath, Swartz, & Cashin, 2010; Ozemek, Cochran, Strath, Byun, & Kaminsky, 2013; Pruitt, et al., 2008; Rejeski et al., 2016; Rejeski et al., 2017; Zisko et al., 2015). Pruitt et al., (2008) and Rejeski et al., (2016) applied an individualized cut-point based on the mean and median accelerometer cpm during a walking session, respectively. In the study by Rejeski et al., (2016), based on an individualized threshold using an equation with age, age^2, and gait speed from a 400-MW, they found different levels of MVPA, compared with traditional absolute cut-points (760, 1041, and 1952 cpm) (Rejeski et al., 2016; Troiano, Berrigan, Masse, & McDowell, 2008). The tailored equation also detected statistically significant

changes in physical activity level over their 6-month physical activity intervention that were not detected with traditional cut-points. There were additional differences between traditional cut-points and the individualized equation in levels of physical activity when divided by age group.. While this study did not have a physiological comparison with the accelerometer data (e.g., heart rate or metabolic equivalent threshold data), the wide variability of physical capacity in older adults supports individualized cut-points. However, previous studies have not reported an individualized cut-point and associations to health outcomes. It is important to understand the relationship between an increased activity level and improved health outcomes, particularly for maintaining independence in older adults. Further, application to epidemiological surveillance studies is challenging without accurate information about an individual's functional capacity.

While there are now multiple new data processing methods for accelerometer data in older adults, studies have not applied these new methods to intervention data to compare the ability of each method to assess change in behavior and to relate that change to health outcomes (Bai et al., 2015; Evenson et al., 2015; Rejeski et al., 2016). The association of changes in physical activity levels with different methods and their comparison to health outcomes is crucial for activity recommendations for older adults. Thus, the purpose of the current study was two-fold:

1.) First, to compare five previously reported processing methods for analyzing physical activity in older adults, applied to free living accelerometer data during two time points (baseline and six months). We applied five different methods to analyze accelerometer data, including: a traditional cut-point for MVPA, a 15-second vector magnitude (VM) cut-point for MVPA, a 1-second VM algorithm, a machine learned algorithm for walking, and an individualized cut-point derived from a 400-MW, in 144 female older adults ages 67-98 years old.

2.) Second, to assess how the changes in physical activity with each of these different methods were associated to intervention-related changes in health outcomes of physical functioning and depressive symptoms over six months.

Methods

Design & Participants

This study consisted of a subsample of older adult females from a 6-month cluster-randomized controlled PA intervention in retirement communities (Kerr et al., 2012). Only females were included in the analyses because three of the methods were developed specifically for older women and have not been validated in older men. Inclusion criteria for the intervention included: a.) Age 65 years or older who could speak and read English; b) provided informed consent and completed a post-consent comprehension test; c) have no history of falls within the past year that resulted in a hospitalization; d) able to walk 20 meters without human assistance; e) complete the Timed Up & Go Test in less than 30 seconds. These criteria were established for the safety of the participants in this unsupervised walking program. All subjects gave written informed consent and approval for the study was obtained by the Institutional Review Board for the protection of human

subjects. Health outcomes and accelerometer measures were assessed at baseline and six months. Since this study was a methodological comparison, only complete cases were included so statistical comparisons could be made. Further, we selected two outcomes to study one physical, one emotional. These were chosen as examples to compare the different accelerometer methods and is not meant to represent all possible relationships.

Protocol.—The details of the intervention protocol can be found in clinical trial protocol # 150336 and have been described in Kerr et al., 2012.

Health Outcomes

Physical Function.—Physical function was assessed with the Short Physical Performance Battery (SPPB). The SPPB consists of three physical assessments testing balance, gait speed, and lower body strength. Each assessment is scored up to four points to calculate a total score out of 12 points; a higher score indicates better performance. Balance, gait, and lower body strength were examined by ability to stand with the feet together in the side-by-side, semi-tandem, and tandem positions; time to walk four meters at a normal pace; and time to rise from a chair and return to the seated position five times. The SPPB has shown to have evidence of validity and reliability in older adults (Gurlanik et al., 1994). Individuals with a score of ten or less have shown to have three times higher odds of walking disability within three years (Vasunilashorn et al., 2009).

Depressive Symptoms.—Depressive symptoms were measured with the Center for Epidemiologic Studies Depression Scare Short Form (CES-D). Participants responded to ten questions (e.g., I was bothered by things that usually don't bother me) on a 0–3-point scale based on how they have felt or behaved over the past week. Summed scores 0–30 were used for analysis; higher scores indicating greater symptoms (Andresen, Malmgren, Carter, & William, 1994).

Accelerometer Measures and Data Processing Methods

400-Meter walk test (400MWT).—Participants wore a triaxial accelerometer (GT3X+, ActiGraph) on their hip and were asked to walk 400 meters as quickly as they could, while remaining safe, on a standard course set up at their retirement community. The walking test was ended if the participant: completed the walk; 15 minutes elapsed; they no longer wished to continue; or if study staff felt the participant was unsafe to continue. For the current analyses, the accelerometer data from their walk were used to calculate an individualized cut-point for physical activity. To analyze accelerometer data to determine the individualized cut-points during the 400MWT, the time of day the walk occurred was recorded and used to identify accelerometer counts achieved during the test.

Free-living physical activity.—Physical activity during the six days after the walk test was also measured by the accelerometer. Participants were instructed to remove the device to sleep or if it would get wet (e.g., shower, swimming). Non-wear time was determined with the Choi et al. (2011) algorithm using 90 consecutive zero counts and individuals with 4 days of at least 10 hours of valid wear time were included in the analyses. The device recorded raw acceleration at 30hz and was converted to 60-second epochs, 15-second

epochs, and 1-second raw acceleration for data processing for the various methods using ActiLife 6.0. Table 1 summarizes the method, the epoch length, and the axis used for the data processing. These five methods were chosen based on previous use in large older adult studies (1041cpm), recent development and validation in older adult women (Evenson, Activity Index, Machine Learning), and feasibility to apply without costly metabolic equipment and testing (individualized from 400MW) (Bai et al., 2016; Copeland & Esliger, 2009; Evenson et al., 2015; Rosenberg et al., 2017).

Individualized cut-point.—This method was based on the counts per minute (cpm) during the baseline 400MWT. The first and last minutes of the 400MWT were removed from the analyses and then median cpm was used as the individualized cut-point. The median cpm was chosen, as this was also used in Rejeski et al. (2016). This cut-point would correspond to their activity above their "fast walking pace" for 400-meters, as directed during their 400MW test.

Evenson cut-point.—This cut-point was developed in the laboratory study with nine lifestyle activities with older adult women from the Objective Physical Activity and Cardiovascular Health Study (OPACH). From the calibration study, the cut-point (>518 VM counts/15 sec epoch) was calculated based on the recorded Metabolic Equivalent Threshold (MET) levels of various activities using a resting MET value of 1.0 MET= 3.0mL/kg/min with 15 second epochs, vector magnitude (VM), and normal filter (Table 4 in Evenson et al., (2015)). This cut-point with the modified 1 MET value was chosen based on the lower fitness capacity of older adults and the activities performed in daily life (Kozey, Lyden, Staudenmayer, & Freedson, 2010).

Activity Index (AI).—The AI was developed in the same laboratory study as the Evenson cut-point. The AI uses the variance of all three axes on the device, rather than an aggregated mean. The variance allows the device to capture the magnitude and frequency of the device's oscillation signals when a person changes from walking to running. The AI was calculated using 1-second raw files, vector magnitude, and the Activity Index R package for TM2.0 as described in Bai et al., (2016). This activity level corresponds to treadmill walking at 2.0 miles per hour, a level of MVPA for this sample of older adults (https://github.com/javybai/ActivityIndex).

Traditional cut-point.—The traditional cut-point of 1041cpm was applied using 60-second epochs and the vertical axis. This cut-point is the most commonly used absolute cut-point to assess MVPA, derived in an older adult sample (Copeland & Esliger, 2009).

Machine learning (ML).—The raw (unfiltered) triaxial accelerometer data was split into minute-level windows. For each window, 41 descriptive features were calculated as described in Rosenberg et al., 2017. After applying the random forest, a minute-level sequence of probabilities of each behavior label results. These probabilities were smoothed over time using a hidden Markov model (HMM). The HMM learned the probability of transitions between behaviors (i.e., it learns that it is more common to transition from sitting to standing than sitting to walking). The HMM chooses the most likely sequence of behaviors from the sequence of probabilities output by the random forest classifier. The

classifier outputs minutes in ambulatory PA (walking), light PA, standing, and sitting. (https://github.com/kkatellis/TLBC). The physical activity variable for comparison for this study used the ambulatory PA (walking) time. This method will capture all walking, regardless of intensity. The algorithm was developed from free living data in older women and validated in an independent sample (Rosenberg et al., 2017).

Statistics

All analyses were performed in R 3.3.0 and significance was set at $p<.05$.

Part 1.   Comparing methods of physical activity.—For each method, minutes of physical activity per day were merged by participant number for valid days and then aggregated to mean minutes per day (min/d) of activity spent above the respective physical activity value for each method at both baseline and six months. Descriptive statistics with means, standard deviations, and medians were generated for each method of physical activity analysis and for the SPPB and CES-D.

To compare the physical activity estimates for each method ($n_j = 5$) to the others, we used generalized estimating equations (GEEs) with min/d of PA for each participant ($N_i = 144$) at baseline and with wear time (wt) and age as covariates:

$$Y_{ij} = B_0 + B_1 Method_j + B_2 WT_i + B_3 Age_i + \varepsilon_{ij}$$

This GEE model compared average min/d of physical activity across the methods. This model uses an exchangeable working correction structure which assumes that the variance between methods is equal. Next, to assess agreement with finer granularity, we generated a confusion matrix with each minute of accelerometry to report the percent overlap of minutes of PA for each method. For methods that were processed with less than 60-second epochs (Activity Index and Evenson), minutes with   30 seconds of PA were classified as PA minutes.

Finally, to compare intervention-related changes, we fit separate mixed effects models for each method, with a random intercept (a_{0i}), to model change over time (X_T) and conditions (X_C i.e., intervention; reference is control) with a time*condition ($X_C X_T$) interaction:

$$Y_{ij} = B_0 + B_1 X_{iC} + B_2 X_{iT} + B_3 X_{iC} X_{iT} + \alpha_{0i} + \varepsilon_{ij}$$

Plots for changes of PA at baseline and 6 months, stratified by condition, based on the models were created to show changes of PA by condition for each method.

Part 2.   Association for changes in PA to changes in SPPB and CESD over time.—To test the associations between changes of PA and changes in outcomes for each method we specified separate linear mixed effects models for each method of PA ascertainment and each outcome variable, adjusting for multiple measurements per person. Covariates of wear time, age, and condition (intervention or control) were added to the model:

$$\Upsilon_{ij} = B_0 + B_1 X_{ij} + B_2 WT_{ij} + B_3 Age_i + B_4 Condition_i + \alpha_{0i} + \varepsilon_{ij}$$

Y_{ij} = Outcome (SPPB or CESD) for ith participant and jth (0 or 6months)observations

X_{ij} = min/day for the selected method of PA for ith participant and jth observation

N_i= 144 (participants)

n_j= 2 (observations at 0 and 6 months)

For this model we also used an exchangeable variance structure with a random intercept and fixed slope. The specified model assumes the residuals are normally distributed as well as the random intercepts. Additionally, a fixed slope assumes the effect overtime is constant for each condition (i.e., intervention and control).

Results

A total of 180 females had both valid accelerometer data (>4d/10hrd) at baseline and 6 months. However, 36 participants did not have 400MW accelerometer data (either unable to complete walk or did not wear accelerometer during walk) and therefore did not have an individualized cut-point for that method. Sensitivity analyses showed significant differences between those with 400MW data and those without. Thus, to ensure a comparable sample for assessing differences in methods, only 144 females (Mean$_{age}$= 83.3yrs $\pm$ 6.4) with valid accelerometer data (>4d with 10hr/d) at baseline and six months and with complete data were used for the analyses. Thus, our results may only apply to women with functioning abilities to complete the 400MW. Future studies may need to replicate analyses in a larger sample of less able women. Descriptive statistics with means, standard deviations (SDs), and medians for each PA method and outcome variable at baseline and six months are presented in Table 2.

Part 1. Comparing Methods of Physical Activity

For comparison of minutes of PA between each method at baseline generalized estimating equations (GEE), when accounting for age and wear time, showed that there was no statistical difference between minutes of PA between the Activity Index (AI) and the Evenson or Machine Learning (ML) methods (ps>.05); the ML was also similar to the 1041cpm (p>.05) but not Evenson method (p< .05). There was a difference in minutes of physical activity between all methods and the individualized median cut-point (ps<.05).

Table 3 shows a confusion matrix of the percentage of overlap between each method at baseline and indicates which methods were statistically different from the GEE analyses (daily minutes). There was over 60% overlap in minutes of PA between 1041cpm, Evenson, and AI methods. Additionally, 1041 cpm had the most overlap in minutes of physical activity (42%) with the individualized median method, both were the two methods that used the vertical axis. While similar to the 1041cpm and AI in the GEE analyses, the Machine Learning (ML) method had only 43% overlap of minutes of physical activity with the

1041cpm and 60% overlap with the AI. This difference shows that while similar when analyzed at the daily level (min/d in the GEE), the ML and 1041cpm are classifying different minutes within the day as PA/non-PA between methods. The most overlap for the Machine Learning was the Activity Index (60%), both of which are methods that did not use a mean of the triaxial or vertical axis acceleration.

All methods, except the individualized median method had a significant time*condition interaction, supporting that changes in physical activity at baseline and six months differed between the intervention and control groups. Figure 1 shows the changes of PA for each method at baseline and six months, stratified by condition of intervention (n=71) and control groups (n=73), including the *p*-values for the interactions. The Individualized median method showed increases over time but no difference between the groups. The Evenson method detected more activity at baseline, as well as, larger differences between the groups, but less change over time. The 1041cpm showed the lowest levels of activity and an increase in the control group. The Activity Index and the Machine Learning showed smaller differences between the groups at baseline, and in particular the Machine Learning detected change in the intervention group over time.

Part 2. Association Between Changes in Physical Activity to Changes in SPPB and CESD Over Time

Physical Functioning (SPPB).—All methods, except for the individualized median method showed a significant relationship between minutes of physical activity and SPPB with a positive parameter estimate, supporting expected, positive associations— increases in physical activity also contributed to increases in functioning. The individualized median method was the only method to report a significant but negative coefficient.

Depressive Symptoms (CES-D).—All methods, except for the individualized median method showed a significant relationship between CES-D and physical activity with a negative parameter estimate, supporting that increases in physical activity contributed to decreases in depressive symptoms. The changes in time spent in physical activity using the individualized median method demonstrated a non-significant association with changes in depression.

The coefficient estimates, standard errors, and p-values for both outcomes are summarized in Table 4. Estimates represent changes in SPPB score for every one-minute in physical activity for each method. Estimates were higher for the methods that reflected higher intensities. However, this reflects the change that would be perceived if each minute were achieved. Given the difference in the change estimates for each method and the difference in minutes of physical activity achieved, we multiplied the actual minutes of change in physical activity of our intervention group by the coefficients to demonstrate the actual impact on outcomes by each method (Table 5). The Activity Index, Machine Learning, and 1041cpm methods had the highest changes in health outcomes showing 0.22 increases in SPPB score and 0.28 decreases in CES-D.

Discussion

The current study compares five different methods of analyzing accelerometer data for physical activity in older adults and reports each method's association with health outcomes of physical function and depressive symptoms in older adult women. We report differences between methods of analyzing PA at baseline and each method's changes over six months of an intervention and control group. Additionally, we found that improvements in health outcomes (i.e., physical function, depressive symptoms) were different, depending on the method of accelerometry analysis. This study reports different amounts of physical activity are needed to achieve significant improvements health of physical function and depressive symptoms. Our study demonstrates the importance of understanding different accelerometry analysis methods for PA when comparing studies, reporting objective physical activity estimates, and applying accelerometry analysis methods to detect changes in physical activity for interventions.

Individualized Median Cut-point

The Individualized median method was different from all other methods at baseline, was extremely variable across individuals, and did not detect differences over six months between the intervention and control groups that other methods detected. We hypothesize that older adults with lower function also have a lower baseline individualized cut-point, as shown by their 400MWT. This lower crossing threshold reflects a higher amount PA throughout the week because ADLs that are done throughout their day are above this level. In contrast, the higher functioning adults completed the test with a higher intensity walk than they normally achieve in daily life activities. Thus, the less functional older adults had higher minutes within the free-living individualized method and this then had an inverse relationship with health.

Of the other studies that analyzed individual cut-points, only two have reported changes in physical activity over time (Miller et al., 2010; Ozemek, et al., 2013; Pruitt et al., 2008; Rejeski et al., 2016; Rejeski et al., 2017; Zisko et al., 2015). Rejeski et al., 2017 is the only other study to also assess changes in PA with a health outcome (i.e., major mobility disability). Rejeski et al., 2016 compared their individualized approach to traditional methods (i.e., 760, 1041, 1952cpm) and found that their individualized method was most similar to 1041cpm. Similar to our sample, the individualized method and 1041cpm had the most percent overlap in our confusion matrix (Table 3). However, different than our study, their individualized method did detect changes in their intervention but not in their control group. This difference may be attributed to the inclusion of age and walk speed in their equation. Also, for higher functioning individuals they capped their individualized threshold at 1952cpm which would allow these participants to have more minutes of PA throughout their week than our sample who may engage in PA over 1952cpm but did not reach their individualized median cpm as often that they established during their 400MW.

Rejeski et al., 2017 examined changes in PA minutes overtime and risk for major mobility disability. They also did not see an intervention effect for the lowest functioning individuals. Similarly, while we did not split the current sample by functioning level, the individualized method was the only method that did not relate to expected increases in physical function or

depressive symptoms with increases in PA by this method. This seemingly contradictory result is likely not due to their unresponsiveness to the intervention. Rather, that low functioning individuals at baseline had a low PA threshold, yielding a high amount of minutes at baseline. Then, as they improve their functioning over the intervention, the low cut-point masks meaningful activity that is above that threshold that contribute to improved health outcomes. Thus, in follow-up measurements they have improved their functioning, yet application of their lower activity threshold does not reflect their increase in PA that contribute to their function.

Evenson Cut-point

In our sample, the Evenson cut-point show similar PA minutes per day to the Activity Index. However, there was more minute-level overlap with the 1041cpm method, as shown in the confusion matrix, reporting 69% of the same minutes from both methods were categorized as PA. Aside from the individualized median method, this method had the highest number of total minutes, resulting in the most overlap with the individualized cut-point. However, different than the individualized cut-point, it was still sensitive to change between the intervention and control groups and was associated with both SPPB and CES-D. The intervention group increased 6.5 min/day while the control group did not increase at all. The difference between intervention and control was large at baseline. Other methods did not indicate such a difference, which would have implications for determining the effectiveness of the randomization. Although this method had the lowest changes in SPPB score, it was comparable for the CES-D outcome. Similar to the individualized cut-point it may be because of the high number of minutes accumulated that each minute is less "meaningful" towards improving functioning.

This method was also developed and validated in older adult women (Evenson et al., 2015). Lacroix et al., (2017) compared this method to a traditional cut-point (1952cpm) revealing the traditional cut-point would significantly underestimate PA in their older adult sample (Lacroix et al., 2017). While this method has not been reported by another PA intervention, Buchner et al., (2017) found that those with a low or moderate SPPB (<9) in the lowest quartile of PA (<25.1min/d) were associated with higher fall risk, than those with higher PA and that those below the median PA (<44min/d) were significantly more likely to report an injurious fall (Buchner et al., 2017). Therefore, this may be an appropriate method to detect a reduction in fall risk.

Activity Index (AI)

Similar to the Evenson, the Activity Index was developed in the same sample of older women but had not been applied to intervention data (Bai et al., 2016). It was the only method in our sample with similar baseline PA estimates to the ML method at baseline. This similarity may reflect that both the AI and ML used different metrics than a mean count of the vertical axis or vector magnitude in the device. The AI also detected changes between intervention and control groups, showing about an 8 minutes/day increase in the intervention and no changes in the control. The AI showed the highest improvements in SPPB for minutes of PA. This finding may be because this method uses the variance of the three axes in the accelerometer which are more sensitive to changes in acceleration (e.g., moving from

walking to jogging) that are not detected by a mean value. Thus, more variability in acceleration and movement in older adults may be beneficial to physical function and meaningful to capture through accelerometer behavioral measurement. Although the AI has not yet been used frequently by other studies, Bai et al., (2016) reports several advantages in the simplicity of the algorithm. When applied to our sample, this method reveals promising advantages for detecting changes in PA as well as associations with health outcomes.

Traditional Cut-point (1041cpm)

The 1041cpm method was similar to the AI and ML at baseline in minutes per day and had more minutes of overlap than any other method with the individualized method. This similarity may be due to both methods using the vertical axis only. It also was sensitive to changes between the intervention and control groups overtime. While recent studies that have reported vector magnitude to be preferable over vertical axis, our results support this method is valuable for studies that have only used uniaxial devices. Further, it is important to show that this simple method that is the most prevalent in previous literature of all of our reported methods still reflects similar outcomes to the more complex proposed methods in both minutes of PA and relationship to health outcomes overtime.

Machine Learning (ML)

Although the ML method estimated similar total physical activity minutes per day with the AI and 1041cpm, when assessing each minute classified as physical activity, the ML showed much smaller overlap with each of these methods than the other methods to each other. That is, this difference demonstrates that the ML method is detecting minutes of physical activity that are different than other methods. This difference may be because it assesses a behavior, walking, regardless of the intensity. While this method specifically captures walking behavior, and minutes of PA below a level of intensity of other methods, we did not see an inflation in meaningless minutes per day as we did with the individualized threshold. Rather, this method detected the greatest increase in the intervention group over time, about 15 minutes, nearly twice as much as all other methods, aside from the individualized median cut-point. The ML method was also the only method to report decreases in the control group, which are expected in this older age group. While ML showed smaller changes in SPPB than the AI and 1041cpm for each minute of physical activity (Table 4), it was comparable to both methods when applied to the actual changes in physical activity minutes of the intervention group (Table 5). Walking is the most preferred form of PA for older adults and our intervention demonstrated older adults increased levels of walking PA that were meaningful to also improve their health. This method may be important to include in future interventions in conjunction with another method as it measures a specific behavior that captures meaningful activity that is not picked up by other methods.

Limitations & Future Research

Some key characteristics of our sample may limit the generalizability to other older adult samples. First, we were only able to test the methods on women because the algorithms were developed in older women. Future studies may want to apply these methods, when validated, to older adult men, as their functioning and activity levels have shown to be different than older adult women (Marques et al., 2014). Additionally, we used only participants who

completed a 400MWT at both time points, excluding those who were not able to complete this assessment or who did not return at six months, thus our sample represents more healthy and compliant participants. Our sample also had a baseline CES-D mean score of 5.6 points, reflecting a generally emotionally healthy group. Thus, this outcome may have experienced ceiling effects and future studies may want to look at changes in depression scores with a more "depressed" population at baseline. Other studies have reported differences in physical activity with accelerometry between gait speeds and functioning levels in older adults (Corbett, Valiani, Knaggs, & Manini, 2016; Kherikahan et al., 2016; Rejeski et al., 2017). Future research also may continue to find differences in the application of the individualized median method to sub groups of older adults.

Conclusions

Our study demonstrates a comparison of five different methods of analyzing physical activity for older adults and their relationship to health outcomes over time. Though it is not yet known what is the "best" method, researchers should be aware of these differences when designing and evaluating interventions, comparing studies, and applying to older adult populations. Future research may want to apply several of these methods together and continue to find the most appropriate method to apply for specific older adult subpopulations.

Acknowledgments

References

Andreyeva T, & Sturm R (2006). Physical activity and changes in health care costs in late middle age. Journal of Physical Activity and Health, 1, 1–14.

Bai J, Di C, Xiao L, Evenson KR, LaCroix AZ, Crainiceanu CM, & Buchner DM (2016). An activity index for raw accelerometry data and its comparison with other activity metrics. PLoS ONE.

Buchner DM, Rillamas-Sun E, Di C, LaMonte MJ, Marshall SW, Hunt J, … LaCroix AZ (2017). Accelerometer-Measured Moderate to Vigorous Physical Activity and Incidence Rates of Falls in Older Women. Journal of the American Geriatrics Society, 65(11), 2480–2487. [PubMed: 28755415]

Choi L, Ward SC, Schnelle JF, & Buchowski MS (2012). Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Medicine & Science in Sports & Exercise, 44, 2009–2016. [PubMed: 22525772]

Chodzko-Zajk WJ (2013). Exercise and physical activity for older adults. Kinesiology Review, 3, 101–106.

Copeland JL, & Esliger JL (2009). Accelerometer assessment of physical activity in active, healthy older adults. Journal of Aging and Physical Activity, 17(1).

Corbett DB, Valiani V, Knaggs JD, & Manini TM (2016). Evaluating Walking Intensity with Hip-Worn Accelerometers in Elders. Elders. Med. Sci. Sports Exerc, 48(11), 2216–2221. [PubMed: 27327031]

Evenson KR, Wen F, Herring AH, Di C, LaMonte MJ, Tinker LF, … Buchner DM (2015). Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91years: The Women's Health Initiative OPACH Calibration Study. Preventive Medicine Reports.

Fielding RA, Katula J, Miller ME, Abbott-Pillola K, Jordan A, Glynn NW, … Rejeski WJ (2007). Activity adherence and physical function in older adults with functional limitations. Medicine and Science in Sports and Exercise, 39(11), 1997–2004. [PubMed: 17986908]

Freedson PS, Melanson E, & Sirard J (1998). Calibration of the computer science and applications, inc. accelerometer. Medicine & Science in Sports & Exercise, 30, 77–81.

Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG…Wallace RB. (1994). A short physical performance battery assessing lower extremity function: association with reported disability and prediction of mortality and nursing home admission. Journal of Gerontology, 49, 85–94.

Kerr J, Rosenberg D, Nathan A,Millstein RA., Carlson JA., Crist K., & Wasilenko K. (2012). Applying the ecological model of behavior change to a physical activity trial in retirement communities: description of the study protocol. Contemporary Clinical Trials, 33, 1180–1188. [PubMed: 22921641]

Kherikahan M, Tudor-Locke C, Axtell R, Buman MP, Fielding RA, Glynn NW …Manini TM. (2016). Actigraphy features for predicting mobility disability in older adults. Physiological Measurement, 37, 1813–1833. [PubMed: 27653966]

Kozey S, Lyden K, Staudenmayer J, & Freedson P, 2010 Errors in MET estimates of physical activities using $3.5 \text{ ml} \times \text{kg}– 1 \times \text{min}– 1$ as the baseline oxygen consumption. Journal of Physical Activity and Health, 7, 508–516. [PubMed: 20683093]

LaCroix AZ, Rillamas-Sun E, Buchner D, Evenson KR, Di C, Lee IM, … Herring AH (2017). The Objective Physical Activity and Cardiovascular Disease Health in Older Women (OPACH) Study. BMC Public Health, 17(1), 1–12. [PubMed: 28049454]

Manini TM, & Pahor M (2009). Physical activity and maintaining physical function in older adults. British Jounal of Sports Medicine, 43, 28–31.

Marques EA, Baptista F, Santos DA, Silva AM, Mota J, & Sardinha LB (2014). Risk for losing physical independence in older adults: The role of sedentary time, light, and moderate to vigorous physical activity. Maturitas, 79(1).

Miller NE, Strath SJ, Swartz AM, & Cashin SE (2010). Estimating absolute and relative physical activity intensity across age via accelerometry in adults. Journal of Aging and Physical Activity, 18:158–170. [PubMed: 20440028]

Neto MG, & de Castro MF (2012). Comparative study of functional independence and quality of life among active and sedentary elderly. Revista Brasileira de Medicina Do Esporte, 18(4), 234–237.

Ozemek C, Cochran HL, Strath SJ, Byun W, & Kaminsky LA (2013). Estimating relative intensity using individualized accelerometer cutpoints: The importance of fitness level. BMC Medical Research Methodology, 13(1).

Pruitt LA, Glynn NW, King AC, Guralnik JM, Aiken EK, Miller G, & Haskell WL (2008). Use of accelerometry to measure physical activity in older adults at risk for mobility disability. Journal of Aging and Physical Activity, 16, 416–434. [PubMed: 19033603]

Rejeski WJ, Anthony P, Brubaker PH, Buman M, Fielding RA, Hire D, … Miller ME (2016). Analysis and Interpretation of Accelerometry Data in Older Adults: The LIFE Study, 71(4), 521–528.

Rejeski WJ, Walkup MP, Fielding RA, King AC, Manini T, Marsh AP, … Miller ME (2017). Evaluating Accelerometry Thresholds for Detecting Changes in Levels of Moderate Physical Activity and Resulting Major Mobility Disability. The Journals of Gerontology: Series A, 00(00), 1–8.

Rosenberg D, Godbole S, Ellis K, Lacroix A, Natarajan L, & Kerr J (2017). Classifiers for Accelerometer-Measured Behaviors in Older Women. Med. Sci. Sports Exerc, 49(3).

Taylor AH, Cable NT, Faulkner G, Hillsdon M, Narici M, & Van der bij AK (2004). Physical activity and older adults: a review of health benefits and the effectiveness of interventions. Journal of Sport Sciences, 22, 703–725.

Troiano RP, Berrigan D, Dodd KW, Masse LC, & McDowell M (2008). Physical activity in the United States measured by accelerometer. Medicine & Science in Sports & Exercise, 40,181–188. [PubMed: 18091006]

Troiano RP McClain JJ., Brychta RJ., & Chen KY. (2014). Evolution of accelerometer methods for physical activity research. British Journal of Sports Medicine, 48, 1019–1023. [PubMed: 24782483]

Vasunilashorn S, Coppin AK, Patel KV, Lauretani F, Ferrucci L, Bandinelli S, & Guralnik JM (2009). Use of the short physical performance battery score to predict loss of ability to walk 400 meters:

Analysis from the InCHIANTI study. Journals of Gerontology - Series A Biological Sciences and Medical Sciences, 64(2), 223–229.
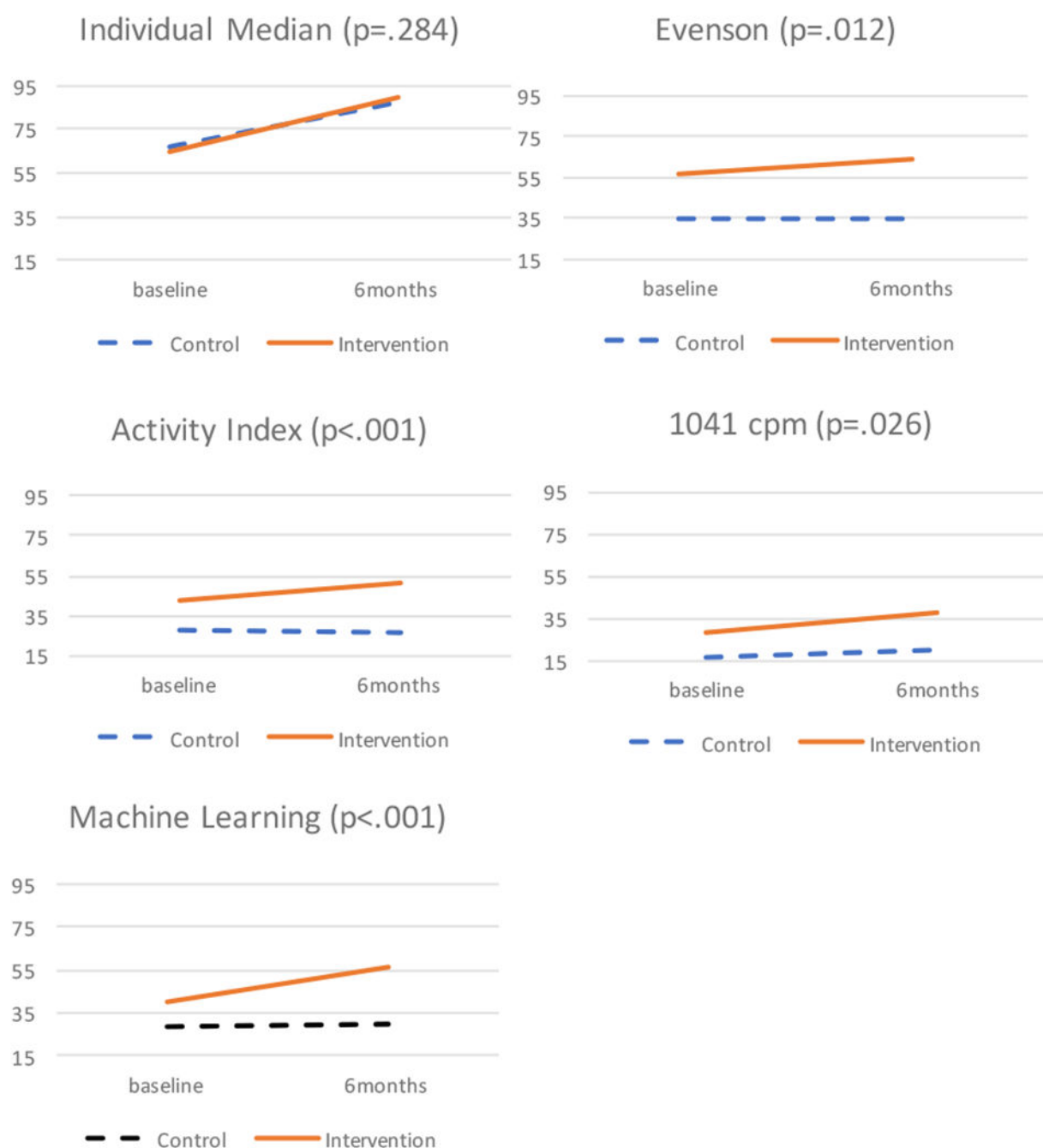
Zisko N, Carlsen T, Salvesen Ø, Aspvik NP, Ingebrigtsen JE, Wisløff U, & Stensvold D (2015). New relative intensity ambulatory accelerometer thresholds for elderly men and women: The Generation 100 study. BMC Geriatrics, 15(1), 1–10. [PubMed: 25559550]

Figure 1.
Minutes/day of physical activity at baseline and 6 months by condition. P value reports the Time X Condition interaction.

Table 1

Method, Epoch Length, and Axis Used for Accelerometry Data Processing

Method	Epochs	Axis
Individualized 400MW (>Median cpm)	60 sec	VA
Evenson	15 sec	VM
Activity Index	1 sec	VM
1041 cpm	60 sec	VA
Machine Learning	30 Hz features aggregated to 60 sec	VM

Note. VM: vector magnitude; VA: vertical axis; cpm: counts per minute, Hz: hertz

Table 2

Means, Standard Deviations (SD), and Medians for Physical Activity (min/d) and Outcome Variables for complete sample (N=144)

Method	Baseline					6 Month				
	Mean	**SD**	**Median**	**LQR**	**HQR**	**Mean**	**SD**	**Median**	**LQR**	**HQR**
In.Median	65.9	70.4	44.6	21.0	76.3	87.7	77.3	63.3	35.5	112.5
Evenson	45.1	27.8	41.0	23.7	59.6	49.3	34.8	41.0	23.6	66.5
AI	35.1	23.3	32.4	17.9	45.8	38.9	29.3	32.6	17.2	52.4
1041cpm	22.4	20.3	16.7	8.5	30.3	28.8	23.8	20.9	12.1	41.3
Machine Learning	35.0	23.5	29.3	19.3	47.6	43.0	33.0	36.4	19.4	58.3
Health Outcomes										
SPPB	8.7	2.7	9.0	7.0	11.0	8.6	2.8	9.0	9.0	7.0
CES-D	5.6	4.0	5.0	2.0	8.0	5.6	3.9	5.0	5.0	3.0

Note. In.Median: Individualized 400MW median cut-point; AI: Activity Index; SPPB: 0-12 points (sum of 3 tests; 4 points each); CES-D: 0-30 points (sum of 10 questions; 3 points each); LQR: Lower Quartile Range; HQR: Higher Quartile Range

Table 3

Confusion Matrix at Baseline for Each Method with Percent (%) Overlap between Methods for Each Minute of PA and Similarities (*) between Methods for Minutes/day

	In. Median	Evenson	AI	1041cpm	ML
In. Median	100%	31% *	25% *	42% *	17% *
Evenson		100%	61%	69% *	42% *
AI			100%	63%	60%
1041cpm				100%	43%
ML					100%

Note.

* indicate significant differences based on total minutes/day by GEE (*p<.05);*

In. Median: Individualized 400MW median cut-point; AI: Activity Index; ML: Machine Learning

Table 4

Each Method Entered Separately in a Regression Model Adjusting for Wear time, Age, and Condition

	SPPB			CESD		
	Estimate	**SE**	*p value*	**Estimate**	**SE**	*p value*
In. Median	−0.009	0.002	0.000	0.003	0.004	0.413
Evenson	0.011	0.005	0.031	−0.031	0.009	0.001
AI	0.033	0.006	0.000	−0.035	0.011	0.002
1041 cpm	0.024	0.007	0.000	−0.037	0.012	0.003
ML	0.014	0.004	0.001	−0.017	0.008	0.039

Note. Models included covariates of age, weartime, and condition. Estimates represent changes in SPPB or CES-D score for every 1-minute change in PA.

Table 5

Changes in SPPB and CESD Score by the Mean Change in PA of the Intervention Group

Method	*PA (min/d)	SPPB	CESD
In. Median	24.03	−0.22	0.07
Evenson	6.48	0.09	−0.26
AI	7.49	0.28	−0.30
1041cpm	8.05	0.22	−0.33
Machine Learning	14.83	0.24	−0.28

Note. NS: Not Significant

*PA represents the mean change in minutes of PA (min/d) between baseline and 6 months for the intervention group