OPEN

Classifiers for Accelerometer-Measured Behaviors in Older Women

DORI ROSENBERG[1], SUNEETA GODBOLE[2], KATHERINE ELLIS[2], CHONGZHI DI[3], ANDREA LACROIX[2], LOKI NATARAJAN[2], and JACQUELINE KERR[2]

[1]Group Health Research Institute, Seattle, WA; [2]Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA; and [3]Fred Hutchinson Cancer Research Center, Seattle, WA

ABSTRACT

ROSENBERG, D., S. GODBOLE, K. ELLIS, C. DI, A. LACROIX, L. NATARAJAN, and J. KERR. Classifiers for Accelerometer-Measured Behaviors in Older Women. *Med. Sci. Sports Exerc.*, Vol. 49, No. 3, pp. 610–616, 2017. **Purpose**: Machine learning methods could better improve the detection of specific types of physical activities and sedentary behaviors from accelerometer data. No studies in older populations have developed and tested algorithms for walking and sedentary time in free-living daily life. Our goal was to rectify this gap by leveraging access to data from two studies in older women. **Methods**: In study 1, algorithms were developed and tested in a sample of older women (N = 39, age range = 55–96 yr) in the field. Women wore accelerometers and SenseCam (ground truth annotation) devices for 7 d, yielding 3191 h and 320 d of data. Images were annotated and time matched to accelerometer data, and random forest classifiers labeled behaviors (sitting, riding in a vehicle, standing still, standing moving, and walking/running). In study 2, we examined the concurrent validity of the algorithms using accelerometer data from an observed 400-m walk test (2983 min of data available) and 6 d of wearing both accelerometers and global positioning systems devices in a sample of 222 women (age range = 67–100; 313,290 min of data available). Analyses included sensitivity, specificity balanced accuracy, and precision, as appropriate, averaged over each test participant at the minute level for each behavior. **Results**: In study 1, the algorithms had 82.2% balanced accuracy. In study 2, the classifier had 87.9% accuracy for predicting walking. Overall machine learning classifiers and global positioning systems had 88.6% agreement. **Conclusions**: Free-living algorithms for walking and sedentary time yielded high levels of accuracy and concurrent validity and can be applied to existing accelerometer data from older women. **Key Words:** MACHINE LEARNING, OLDER ADULTS, SITTING, WALKING, PHYSICAL ACTIVITY, SEDENTARY TIME

Physical activity promotes emotional, cognitive, functional, and physical health in older adults (21). Current estimates, however, suggest that few older adults meet physical activity guidelines, particularly when assessed by accelerometer cut points; estimates are fewer than 5% (30). Rates are higher when using self-reported metrics. For example, in the Women's Health Study, 67% of women reported meeting physical activity guidelines by questionnaires, whereas 13.4% were classified as meeting guidelines using the most commonly applied accelerometer cut point: 1952 counts per minute (27).

Accelerometer cut points allow acceleration data to be translated into activity intensity categories (8). This approach mapped well to current physical activity guidelines that state activities must be performed at a moderate or vigorous intensity. However, the most commonly used absolute cut points, developed on young adults, have not worked well in older adult samples who engage in activities at a relatively lower level of intensity. The use of absolute cut points results in common activities, such as walking, being misclassified as below the threshold for moderate intensity (8).

The exclusive focus on activity intensity can be problematic, however, as the public may not understand this concept and behaviorally specific goals may be easier to communicate than intensity-based ones (9). Understanding how specific patterns of behaviors such as walking relate to health outcomes could lead the field to more useful guidelines that older adults can realistically attain (2).

Computational techniques are now being applied to accelerometer data to develop classifiers that can distinguish time spent in actual behaviors, such as driving, walking, lifting weights, and sitting (10). If valid, these classifiers could be

Address for correspondence: Jacqueline Kerr, Ph.D., University of California, San Diego, 9500 Gilman Dr. #0811, La Jolla, CA 92093; E-mail: jkerr@ucsd.edu.

Submitted for publication May 2016.
Accepted for publication September 2016.

0195-9131/17/4903-0610/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000001121

applied to existing longitudinal studies that include accelerometer assessments and well-documented health outcomes. For example, several large epidemiologic studies such as the Women's Health Initiative, Nurses' Health Study, Reasons for Geographic and Racial Differences in Stroke, and the Adult Changes in Thought studies are gathering substantial amounts of older adult accelerometer data (15,20,23).

Most studies using new computational techniques train and test the algorithms on different participants from the same sample and study behaviors in a laboratory setting or with participants following a fixed protocol in more naturalistic settings (28,29). Only two algorithms have been developed for older adults based on laboratory protocols (13,25). More recent studies, however, demonstrate that laboratory-based algorithms do not perform as well when applied to free-living data (1). Even algorithms from protocolized training data in naturalistic settings do not predict behaviors as accurately as totally free-living participants going about their normal behaviors across multiple days and hours (18). New algorithms trained in such totally free-living settings in adults are promising and can include important free-living behaviors that are difficult to conduct in laboratory settings, such as driving and bicycling (18). However, they have not yet been developed specifically for older adults and have not yet been validated in a completely independent sample of participants outside of the algorithm testing phase (7). Previous studies suggest up to an 8% difference in accuracy for training data sets that vary by age and gender. For researchers to be confident that they can apply such new algorithms to their free-living older adult cohort data, further validation efforts are required.

The purpose of our study was to develop and test a new computational algorithm to classify walking and sedentary time, including in a vehicle, in older adults. The algorithm was developed on data collected across multiple free-living days and validated in a completely independent cohort of older adults that were not involved in the algorithm development phase. We leveraged a unique opportunity in which older adults, age 65–100 yr, a quarter using walking aids, in a physical activity intervention trial completed an observed 400-m walk test while wearing accelerometers, providing a ground truth for comparison. Participants then wore an accelerometer and global positioning systems (GPS) devices for 6 d, providing further opportunity for investigating the algorithm's concurrent validity against free-living GPS-defined behaviors. The current work focused on older women to identify and validate an algorithm that could be applied to a large existing cohort of older women from the Women's Health Initiative (23).

METHODS

Both studies obtained ethics approval from the University of California, San Diego institutional review board. Participants completed written informed consent for both studies.

Study 1: Algorithm Development and Testing

Participants and procedures. A convenience sample of 39 older women were recruited to wear an ActiGraph GT3X+ accelerometer (ActiGraph, Pensacola, FL) on a belt over the right hip and a body-worn camera (the SenseCam, Vicon Revue, United Kingdom) on a lanyard around their neck during waking hours for 7 d. They were asked to continue their normal activities, but participants were trained in institutional review board–approved procedures to ensure privacy and confidentiality for themselves and others while the camera was being worn, such as turning the camera off or turning it over when needing privacy and only wearing the camera in public setting or with permission from others. The women were recruited to provide a diverse age range (56–94 yr), variability in self-reported functioning and physical activity levels, and a range of body mass index (19.74–45.62). All participants were ambulatory, able to provide informed consent, and complete surveys. Participants received and returned the devices in person at UCSD. They received wear time instructions to improve compliance and at the end were given the opportunity to delete any images they did not want included in the data set.

Ground truth annotation. The SenseCam camera, which captured first-person images approximately every 20 s, allowed researchers to capture ground truth information about participant behavior. SenseCam image data were downloaded and imported into the Clarity SenseCam browser, and researchers annotated the SenseCam images with ground truth behavior labels (6). A standardized annotation protocol was developed, and at least 80% agreement for each posture with a standardized day was established. More details on SenseCam image annotation can be found elsewhere (17), and the complete annotation protocol is available from the authors upon request. The SenseCam annotation protocol assigns mutually exclusive posture labels to each image: sitting, riding a vehicle, standing still with no movement), standing moving, i.e., walking within a confined space for example walking around in the kitchen, and walking/running, i.e., making progress to a distant point. Riding in a vehicle is separated out from other sitting because the accelerometer measurements differ in this context because of the vibration of the vehicle and the acceleration from driving. If a minute of data falls within a time window bound by images with identical activity codes, that activity label is applied to the minute. If a minute spans images with changing activity codes, no label is applied to the minute and it is not used for training the classifier.

Behavior classification algorithm. We used a behavior classification system that uses machine learning (ML) algorithms to predict five behaviors—sitting, riding a vehicle, standing still, standing moving, and walking/running—from raw triaxial accelerometer data. We have developed and tested this system in three other data sets (7,18). The classifier was retrained on the current data set of older women. Our system predicts a behavior label for each minute of accelerometer data. A 1-min window was chosen because we believe it is a

APPLIED SCIENCES

sufficiently detailed interval by which to represent public health relevant behaviors on a daily level. The behavior classification process is composed of three steps: feature extraction, minute-level classification, and time smoothing. A detailed description of these three steps can be found in our previous publications (7,18). A short summary is provided here.

Feature extraction. The raw (unfiltered) triaxial accelerometer data were split into 1-min windows. For each 1-min window, 41 descriptive features were calculated. For each sample in a data window, the vector magnitude (VM) of the acceleration signal was calculated, i.e., $v = (x^2 + y^2 + z^2)^{1/2}$. The following basic statistical descriptors of the VM were calculated over the data window: mean, SD (sd), coefficient of variation (coefvariation), minimum (min), maximum (max), and 25th, 50th, and 75th percentile (25thp, median, 75thp, respectively). The 1-s lag autocorrelation (autocorr) of the VM and the correlation between each axis were computed (corrxy, corrxz, and corryz). For each sample in the window, the roll, pitch, and yaw angles of the direction of acceleration were computed, as roll = $\tan^{-1}(y, z)$, pitch = $\tan^{-1}(x, \mathrm{z})$, and yaw = $\tan^{-1}(y, x)$. The average (avgroll, avgpitch, and avgyaw) and the SD (sdroll, sdpitch, and sdyaw) of these angles were computed over the window. A low-pass filter with a cutoff frequency of 0.5 Hz (preliminary experiments tested a few cutoff frequencies and found 0.5 Hz to perform best) was applied to the data window to estimate the average direction of gravity, and the roll, pitch, and yaw angles of this direction were computed (rollg, pitchg, yawg) (14). The fast Fourier transform was applied to the VM to decompose the time domain signal to its frequency components. The resulting power spectrum describes the contribution of a given frequency to the measured acceleration signal. The dominant frequency of the signal (fmax), i.e., the frequency with the highest power, and corresponding maximal power (pmax) were computed from the power spectrum. A similar calculation was conducted between the frequency bands of 0.3 and 3 Hz (fmaxband and pmaxband). The entropy of the frequency domain signal was computed. Finally, the power in each frequency band between 1 and 15 Hz (fft1–fft15) was computed.

Minute-level classification. Next, each feature vector was input into a random forest classifier. A random forest classifier is a commonly used ML algorithm made up of an ensemble of randomized decision trees, each of which is learned from a random sample of training data and a random sample of features. The decision tree outputs a probability of each behavior label for each feature vector. Test minutes are classified by averaging the output probabilities from each decision tree in the forest.

Time smoothing. After applying the random forest, a minute-by-minute sequence of probabilities of each behavior label results. These probabilities were smoothed over time using a hidden Markov model (HMM). The HMM uses the training data to learn the probability of transitions between behaviors, i.e., it can learn that it is more common to transition from sitting to standing than sitting directly to walking. The HMM was used to choose the most likely sequence of behaviors from the sequence of probabilities output by the random forest classifier.

Evaluation. We evaluated the performance of our behavior classification algorithms using leave-one-participant out cross validation. This means each participant was used as the test subject in turn, using the remaining participants to train the classification algorithm. Sensitivity, specificity, and balanced accuracy (the mean of sensitivity and specificity) were averaged over each test participant at the minute level for each behavior (sitting, riding a vehicle, standing still, standing moving, and walking/running).

Traditional accelerometer count processing. For comparison with the machine-learned outputs, we processed the accelerometer data in Actilife 6. Median counts for each machine-learned behavior were also shown to provide an estimate of intensity, although counts were not a feature of the algorithm.

Study 2: Validation in a New Cohort

Two types of validation were investigated to establish that the algorithms developed in one cohort could be applied to another without loss of performance, demonstrating generalizability and validity. First, the algorithm performance was tested against a gold standard observation. Participants completed a timed 400-m walk and the start and end times were recorded. During this time, it was known that the participants were walking, although they were allowed to stop and rest as needed during the task and before and after. Stops were noted in the protocol. Second, the behavioral predictions from the algorithm were compared with GPS predictions to provide concurrent validity. The GPS predictions included walking, stationary, and vehicle time.

Participants and procedures. Data were from a sample of 222 older women (age 67–100 yr) living in 11 retirement communities and participating in a randomized control trial comparing a physical activity to a healthy aging comparison group were used for the validation phase (19). None of the women were included in the algorithm development phase. All participants were ambulatory but not at high risk for falling and able to provide informed consent. Women wore an ActiGraph GT3X+ accelerometer (ActiGraph) on a belt over the right hip and a Qstarz BT1000X GPS data logger during waking hours over 6 d.

Participants completed a timed 400-m walk test (26) using standard procedures as part of a physical functioning test battery. They were instructed to wear comfortable walking shoes to do the task. The course was set up indoors at each facility. All courses were flat but had various surfaces (some carpeted and some wood flooring). Participants were instructed to walk the course as quickly as possible while remaining safe and were allowed to have standing breaks to rest if needed throughout. The test was ended if the participant needed to sit down or more than 15 min were needed to complete the test. Participants wore the accelerometer device during the walk and observers recorded the time the test started and ended. Data from the baseline, 6-month, and 12-month measurement tests were combined and included in the current analyses to increase the number of walking minutes to be predicted per participant.

Data Processing

Accelerometer data during a 400-m walk. Acceler ometer data were truncated to the time within the recorded start and stop of the 400-m walk test using the sqldf package in R (11). The initial and last minute of the 400-m walk was removed before analysis to eliminate partial minutes where the walk was initiated and terminated. The behavioral categories of walking to a distant point and standing moving within a confined space were combined.

Accelerometer data in comparison with GPS-defined vehicle travel and walking. GPS and accelerometer data were merged at the minute level in the validated Personal Activity Location Measurements System (3,4). The Personal Activity Location Measurements System uses the 90th percentile of speed during a trip, the percent of time indoors during a trip, and the percent of time in a single location during a trip to predict walking, riding in a vehicle, and stationary time. Previous studies have shown this system to have 85% accuracy (3,4). Stationary time represents any behavior without movement in space; that is, less than 25-m distance in a minute. Only outdoor minutes of GPS were used because the GPS detection of activities can be hindered by poor signal strength indoors. Approximately 10.7% of outdoor time while wearing the GPS was spent walking, 22.5% riding in a vehicle, and 66.8% stationary. The ML behavior classifier described earlier was used to categorize each minute of accelerometer data as sitting, riding in a vehicle, standing, standing moving, or walking.

Analyses

The analyses assessed the concurrent validity of the ML classifier using the two sources of data available; observed 400-m walk test and free-living concurrent GPS data. First, we examined the minute-level sensitivity of the walking algorithm using accelerometer data from the observed timed 400-m walk test. Nonwalking behaviors were not noted so specificity metrics were not available. We then used generalized estimating equations (GEE) to examine predictors of achieving high (80% or higher) or low (<80%) sensitivity, using the "geepack" library in R (14). To explore potential reasons for high or low algorithm sensitivity, several predictors were examined based on prior work, which has shown that older adults sometimes have slow gait speed or other abnormalities in their mobility that could affect accelerometer signals (24). We explored the effect of age, which was self-reported at baseline. Furthermore, we examined several time-varying predictors measured at each time point: gait speed (calculated from the 400-m walk test), observer annotated use of a walking aid during the 400-m walk, short physical performance battery score (12), and fear of falling (falls efficacy scale) (16). We used an exchangeable working correlation structure to account for participant clustering and robust SE to provide valid statistical inference even if the working correlation might not hold. The predictors were age, gait speed, number of stops during the 400-m walk, use of a walking aid, short physical performance battery, and fear of falling.

Second, we examined the concurrent validity of the algorithm for detecting walking, vehicle time, standing moving, standing still, and sitting by time merging the machine-learned activity predictions to the GPS-based travel mode assignments. Because the machine-learned classification and GPS travel mode had different classes, we combine the minutes in the classes of standing moving, standing still, and sitting and compared it with the stationary GPS class. Two ratios were calculated to assess agreement. First, we examined the number of matching class minutes to the total minutes of the class by GPS, which is similar to the recall metric when defined the GPS classes as the standard for comparison, and second, we examined the number of matching class minutes to the total minutes of the class by ML, which is similar to precision. All analyses were conducted using the R statistical package (22).

RESULTS

Phase 1: Algorithm Development and Testing

Participants providing data included 39 older women (see Table 1). Table 2 shows the confusion matrix for the predicted minutes and known annotated behaviors. The most prevalent behavior was sitting, followed by riding in a vehicle and walking. Sitting behaviors were accurately predicted 89% of the time with misclassification as standing still occurring 7% of the time. Riding in a vehicle was accurately predicted 84% of the time with 6% of minutes being misclassified as sitting and 5% as standing moving. Walking had lower accuracy with 70% accurately being predicted and 24% being misclassified as standing moving. Standing still and standing moving had lower accuracy.

Table 3 demonstrates the sensitivity, specificity, and balanced accuracy of the algorithm for the five behaviors tested against the annotated SenseCam images, our ground truth. Overall, the algorithm performed with 82.2% average balanced accuracy, using the leave-one-participant-out cross validation. The median counts provided for comparison indicate that sitting, standing, and walking in this population occur

TABLE 1. Demographic and health characteristics of study samples.

	Study 1 Algorithm Development & Testing	Study 2 Observed the 400-m Walk Sample	Study 2 GPS Sample
N	39	195	219
Age, mean, range	69.4, 56–94	83.6, 67–100	83.8, 67–100
White, %	79.5	91.3	91.3
Use of walking aid, %	12.8	25.6	18.3

TABLE 2. Minute-level confusion matrix of predicted and annotated minutes.

	No. Minutes of SenseCam Annotated Activity (Percent Accuracy)				
ML Predicted Activity	Sitting	Riding in vehicle	Walking	Standing still	Standing moving
Sitting	83,111 (89)	891 (6)	126 (2)	2072 (20)	671 (6)
Riding in vehicle	1148 (1)	11,673 (84)	103 (1)	273 (3)	282 (2)
Walking	224 (0)	323 (2)	4994 (70)	458 (5)	1874 (16)
Standing still	6995 (7)	296 (2)	161 (2)	4104 (40)	1380 (12)
Standing moving	1889 (2)	755 (5)	1711 (24)	3238 (32)	7707 (66)

at lower intensities than would be detected by existing thresholds of <100 for sedentary behavior and >1951 for moderate to vigorous physical activity. Sitting in a vehicle recorded higher intensities than the sedentary behavior cutoff. The accuracy levels achieved by the algorithm were comparable with algorithms developed in laboratory studies (13,25). Given that this algorithm was developed on free-living data and laboratory studies applied to free-living data lose over 10% accuracy, we believed further validation in an independent cohort (Study 2) was warranted.

Phase 2: Algorithm Validation in New Cohort

Validation of ML walking algorithm. Participants providing data during the 400-m walk included 195 women who completed the test (see Table 1). At total of 90% of participants had one stop (range 1–4). The minute-level sample available for validation of the walking algorithm included 2983 min of the 400-m walk test data. Accelerometer counts per minute during the 400-m walk varied from 0 to 5264 counts per minute with a median value of 1591 counts per minute. This suggests that the commonly used 1952 cut point for moderate to vigorous activity would not have captured a substantial portion of walking that was performed at the older women's fastest safe pace. Overall, during the 400-m walk, the combined walking and standing moving classifier performed with an overall mean sensitivity of 87.9%. During the 400-m walk, the algorithm misclassified 9.2% of the test minutes as sitting, 1.2% as vehicle, and 1.2% as standing still. None of the included variables in the GEE analyses significantly predicted the algorithm sensitivity (Table 4). This suggests that the algorithm is robust across age, functioning, fall risk, and walking speed.

Concurrent validity of ML algorithms with GPS. A total of 219 women wore the accelerometer and GPS devices for six free-living days (mean age = 83.8, age range = 67–100, 91.3% white, 18.3% self-reported using a walking aid, 313,290 min of data available). Concurrent validity for behaviors during the 6 d of accelerometer and GPS wear are shown in Table 5. The overall agreement for the two methods was 88.6%. Precision (PPV) and recall (sensitivity) for walking was 68.1% and 85.5%. Precision and recall were 90.6% and 93.7%, respectively, for all stationary time (sitting, standing moving, and standing still) and 83.4% and 85.1%, respectively, for vehicle time precision and recall.

TABLE 3. Percent accuracy of classifiers for sedentary behaviors and physical activity using observed annotations of person worn camera images.

	Sensitivity	Specificity	Balanced Accuracy	Median Counts (IQR)[a]
Machine learned				
Sitting	89	91	90	0 (0–17)
Sitting in vehicle	84	99	91	72 (21–177)
Walking	70	98	84	597 (231–1210)
Standing moving	66	94	79	268 (97–562)
Standing still	40	93	67	56 (3–252)

[a] Counts were not used as a feature in the algorithm but are provided here as count data are commonly reported in traditional accelerometer studies as a metric of intensity. This demonstrates that behaviors are occurring at lower intensities than would be identified by traditional cut points (<100 for sedentary behavior; 1952 for moderate vigorous activity).

DISCUSSION

We developed a new classifier to predict five important health-related behaviors in free-living older women and demonstrated high performance of the algorithm (82.2%). Although our classifier accuracy is comparable with other algorithms developed in the laboratory with older adults (13,25), it could have been affected by several factors. Having less available walking data decreases accuracy by reducing the classifier's ability to generalize to walking patterns it has not seen before. Standing moving can include portions of walking, which can confuse the classifier.

We found excellent levels of sensitivity for our classifier in regard to the identification of walking behaviors during a 400-m walk field test (87.9%). This is the first time that machine-learned algorithms for physical activity and sitting, developed in a completely separate training sample, have been applied to a large, independent, and truly free-living validation sample. The sensitivity of the algorithm was not dependent on age, walking aid, falls risk, or physical functioning. This means that the algorithm can be applied in populations of women that vary in age, physical function, and gait speed.

In addition, the classifier had excellent concurrent validity with GPS data (88.6%). Our ability to accurately detect time spent in a vehicle is an advancement over the use of accelerometer intensity cut points which misclassify time spent in a vehicle as light-intensity about one-third of the time (7,18). Little is known about the health effects of vehicle time in aging-related health outcomes. Driving or riding in a vehicle

TABLE 4. Age and functioning predictors of algorithm performance (<80%) during the 400-m walk using GEE.

	Beta Coefficient	SE	*P*
Age	−0.0232	0.0200	0.25
Gait speed during the 400-m walk	−0.00437	0.44902	0.99
No. stops during the 400-m walk	0.1235	0.2758	0.654
Use of a walking aid	0.288	0.354	0.42
Fear of falling	0.00381	0.02007	0.8495
Short physical performance battery overall score	−0.09403	0.05244	0.073

TABLE 5. Minutes in each category and percent agreement between GPS and machine-learned accelerometer classifier.

	GPS Personal Activity Location Measurements System Classification		
Accelerometer ML classification:	Pedestrian	Stationary	Vehicle
Sitting	725 (2.2)	126,050 (58.6)	3153 (4.9)
Standing still	249 (0.8)	17,446 (8.1)	1304 (2.0)
Standing moving	3049 (9.2)	51,595 (24.0)	4658 (7.2)
Riding in a vehicle	891 (2.7)	8616 (4.0)	54,143 (83.4)
Walking	28,194 (85.5)	11,553 (5.4)	1664 (2.6)

could promote increased life space and ability to engage in meaningful activities. Driving cessation is associated with depression and poor health outcomes (5). However, it could also substitute for time spent in more active pursuits and could negatively affect health.

The classifiers developed and validated here could now be applied to large samples of existing accelerometer data in older women in which there is rich data on health outcomes. We can then better understand whether intensity is more important or whether total walking could be as associated with health outcomes irrespective of intensity. With the recent surgeon general's Call to Action to Promote Walking, we can use accelerometers to determine whether there are improvements in walking behaviors because of public health interventions such as sidewalk installations. Our previous studies have indicated that the training sample and the type of training data are important predictors of algorithm performance (18). We, therefore, encourage use of algorithms that are appropriately matched to testing and validation samples at this stage. Future work may allow the development of an algorithm that is robust across genders, ages, and body types.

Limitations of our study include that we only had gold standard observational data available for walking and not for other important behaviors for which we have developed algorithms including driving, sitting, standing, running, and cycling. Because participants could stop during the walk, not all time may have been walking; however, we saw no effect of number of stops on the algorithm performance. In the future, we plan to compare machine-learned accelerometer algorithms for sitting and standing in older adults to the field gold standard for posture (activPAL) measures. Ongoing work with the activPAL as a ground truth will likely improve our estimates of standing still. The features of roll-pitch and yaw angles of the direction of acceleration were approximations because gyroscope or magnetometer data were not available.

Our study strengths include the first demonstration of how new algorithms can be developed outside of a laboratory or prescriptive free-living setting and applied and externally validated in a new sample. In addition, our focus on older adults is important because they engage in physically active behaviors at a range of intensities that are often far below the most commonly used cut points for moderate-intensity physical activity (8). Focusing on the identification of behaviors allows us to capture movements that are much more common and could still have effects on health outcomes. Furthermore, being able to recommend that older adults increase the time they spend walking is much more under individual control than recommending a certain level of activity intensity, something which most of the public is likely unable to clearly understand.

CONCLUSIONS

We found excellent sensitivity for identifying walking behaviors using accelerometer data in older adults. Furthermore, we found high levels of concurrent validity with GPS for sedentary, vehicle, and walking time. Our algorithms are available in R (https://cran.r-project.org/web/packages/TLBC/index.html) to researchers who have interests in applications to existing epidemiologic data sets in the validated age range.

Dr. Rosenberg was supported by K23HL119352. There are no conflicts of interest.

The results of the present study do not constitute endorsement by the American College of Sports Medicine. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. 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):716–22.
2. Centers for Disease Control and Prevention (CDC). Vital signs: walking among adults—United States, 2005 and 2010. *MMWR Morb Mortal Wkly Rep*. 2012;61:595–601.
3. Carey M, Markham C, Gaffney P, Boran C, Maher V. Validation of a point of care lipid analyser using a hospital based reference laboratory. *Ir J Med Sci*. 2006;175(4):30–5.
4. Carlson JA, Jankowska MM, Meseck K, et al. Validity of PALMS GPS scoring of active and passive travel compared with SenseCam. *Med Sci Sports Exerc*. 2015;47(3):662–7.
5. Chihuri S, Mielenz TJ, DiMaggio CJ, et al. Driving cessation and health outcomes in older adults. *J Am Geriatr Soc*. 2016;64(2):332–41.
6. Doherty AR, Kelly P, Kerr J, et al. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. *Int J Behav Nutr Phys Act*. 2013;10:22.
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):933–40.
8. Evenson KR, Wen F, Herring AH, et al. Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: the Women's Health Initiative OPACH Calibration Study. *Prev Med Rep*. 2015;2:750–6.
9. Floegel TA, Giacobbi PR Jr, Dzierzewski JM, et al. Intervention markers of physical activity maintenance in older adults. *Am J Health Behav*. 2015;39(4):487–99.
10. Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol* (*1985*). 2011;111(6):1804–12.
11. Grothendieck G. Perform SQL Selects on R Data Frames. 2014. Available from: https://cran.r-project.org/web/packages/sqldf/sqldf.pdf.

APPLIED SCIENCES

12. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55(4):M221–31.
13. He B, Bai J, Zipunnikov VV, et al. Predicting human movement with multiple accelerometers using movelets. *Med Sci Sports Exerc*. 2014;46(9):1859–66.
14. Hojsgaard S, Halekoh U, Yan J. Generalized Estimating Equation Package. 2016. Available from: https://cran.r-project.org/web/packages/geepack/geepack.pdf.
15. Howard VJ, Rhodes JD, Mosher A, et al. Obtaining accelerometer data in a national cohort of Black and White adults. *Med Sci Sports Exerc*. 2015;47(7):1531–7.
16. Kempen GI, Yardley L, van Haastregt JC, et al. The Short FES-I: a shortened version of the falls efficacy scale-international to assess fear of falling. *Age Ageing*. 2008;37(1):45–50.
17. Kerr J, Marshall SJ, Godbole S, et al. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. *Am J Prev Med*. 2013;44(3):290–6.
18. Kerr J, Patterson RE, Ellis K, et al. Objective assessment of physical activity: classifiers for public health. *Med Sci Sports Exerc*. 2016;48(5):951–7.
19. Kerr J, Rosenberg DE, Nathan A, et al. Applying the ecological model of behavior change to a physical activity trial in retirement communities: description of the study protocol. *Contemp Clin Trials*. 2012;33(6):1180–8.
20. 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.
21. Physical Activity Guidelines Committee. Physical Activity Guidelines Advisory Committee Report, 2008. Washington (DC): U.S. Department of Health and Human Services; 2008. pp. 1–683.
22. R Core Team. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing; 2013 [cited 2014]. Available from: www.R-project.org.
23. Rillamas-Sun E, Buchner DM, Di C, Evenson KR, LaCroix AZ. Development and application of an automated algorithm to identify a window of consecutive days of accelerometer wear for large-scale studies. *BMC Res Notes*. 2015;8:270.
24. Sandroff BM, Riskin BJ, Agiovlasitis S, Motl RW. Accelerometer cut-points derived during over-ground walking in persons with mild, moderate, and severe multiple sclerosis. *J Neurol Sci*. 2014;340(1–2):50–7.
25. Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, Freedson PS. Performance of activity classification algorithms in free-living older adults. *Med Sci Sports Exerc*. 2016;48(5):941–50.
26. Sayers SP, Guralnik JM, Newman AB, Brach JS, Fielding RA. Concordance and discordance between two measures of lower extremity function: 400 meter self-paced walk and SPPB. *Aging Clin Exp Res*. 2006;18(2):100–6.
27. Shiroma EJ, Cook NR, Manson JE, Buring JE, Rimm EB, Lee IM. Comparison of self-reported and accelerometer-assessed physical activity in older women. *PLoS One*. 2015;10(12):e0145950.
28. 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.
29. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *J Appl Physiol* (*1985*). 2009;107(4):1300–7.
30. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc*. 2008;40(1):181–8.