OPEN

Medicine & Science IN Sports & Exercise

The Official Journal of the American College of Sports Medicine
www.acsm-msse.org

. . . Published ahead of Print

The CNN Hip Accelerometer Posture (CHAP) Method for Classifying Sitting Patterns from Hip Accelerometers: A Validation Study

Mikael Anne Greenwood-Hickman[1*], Supun Nakandala[2*], Marta M. Jankowska[3], Dori Rosenberg[1], Fatima Tuz-Zahra[4], John Bellettiere[4], Jordan Carlson[5,6], Paul R. Hibbing[5], Jingjing Zou[4], Andrea Z. LaCroix[4], Arun Kumar[2#], Loki Natarajan[4#]

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA; [2]University of California San Diego, Department of Computer Science and Engineering, La Jolla, CA; [3]City of Hope, Beckman Research Institute, Population Sciences, Duarte, CA; [4]University of California San Diego, Herbert Wertheim School of Public Health and Human Longevity Science, La Jolla, CA; [5]Children's Mercy Kansas City, Center for Children's Healthy Lifestyles and Nutrition. Kansas City, MO; [6]University of Missouri Kansas City, Department of Pediatrics, Kansas City, MO

Accepted for Publication: 12 May 2021

Medicine & Science in Sports & Exercise, Publish Ahead of Print
DOI: 10.1249/MSS.0000000000002705

The CNN Hip Accelerometer Posture (CHAP) Method for Classifying Sitting Patterns from Hip Accelerometers: A Validation Study

Mikael Anne Greenwood-Hickman[1*], Supun Nakandala[2*], Marta M. Jankowska[3], Dori Rosenberg[1], Fatima Tuz-Zahra[4], John Bellettiere[4], Jordan Carlson[5,6], Paul R. Hibbing[5], Jingjing Zou[4], Andrea Z. LaCroix[4], Arun Kumar[2#], Loki Natarajan[4#]

[1]Kaiser Permanente Washington Health Research Institute, Seattle, WA; [2]University of California San Diego, Department of Computer Science and Engineering, La Jolla, CA; [3]City of Hope, Beckman Research Institute, Population Sciences, Duarte, CA; [4]University of California San Diego, Herbert Wertheim School of Public Health and Human Longevity Science, La Jolla, CA; [5]Children's Mercy Kansas City, Center for Children's Healthy Lifestyles and Nutrition. Kansas City, MO; [6]University of Missouri Kansas City, Department of Pediatrics, Kansas City, MO

*Co–first author

#Co–senior author

Corresponding Author

Mikael Anne Greenwood-Hickman

[1]Kaiser Permanente Washington Health Research Institute, 1730 Minor Avenue

Seattle, WA 98101

Mikael.Anne.Greenwood-Hickman@kp.org

This work was supported by grant number U01AG006781 from the National Institute on Aging and R01DK114945 from the National Institute of Diabetes and Digestive and Kidney Diseases. It was also supported in part by a Hellman Fellowship, an NSF CAREER Award under award number 1942724, and a gift from VMware. The content is solely the responsibility of the authors and does not necessarily represent the views of any of these organizations. **Conflict of Interest.** The authors have no conflicts of interest to declare. Results of the present study do not constitute endorsement by ACSM. Results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Abstract

Introduction: Sitting patterns predict several healthy aging outcomes. These patterns can potentially be measured using hip-worn accelerometers, but current methods are limited by an inability to detect postural transitions. To overcome these limitations, we developed the Convolutional Neural Network Hip Accelerometer Posture (CHAP) classification method. **Methods**: CHAP was developed on 709 older adults who wore an ActiGraph GT3X+ accelerometer on the hip, with ground truth sit/stand labels derived from concurrently worn thigh-worn activPAL inclinometers for up to 7 days. The CHAP method was compared to traditional cut-point methods of sitting pattern classification as well as a previous machine learned algorithm (Two Level Behavior Classification [TLBC]). **Results**: For minute level sitting vs. non-sitting classification, CHAP performed better (93% agreement with activPAL) than other methods (74%-83% agreement). CHAP also outperformed other methods in its sensitivity to detecting sit-to-stand transitions: cut-point (73%), TLBC (26%), and CHAP (83%). CHAP's positive predictive value of capturing sit-to-stand transitions was also superior to other methods: cut-point (30%), TLBC (71%), and CHAP (83%). Day-level sitting pattern metrics, such as mean sitting bout duration, derived from CHAP did not differ significantly from activPAL, whereas other methods did: activPAL (15.4 mins mean sitting bout duration), CHAP (15.7 mins), cut-point (9.4 mins), TLBC (49.4 mins). **Conclusion**: CHAP was the most accurate method for classifying sit-to-stand transitions and sitting patterns from free-living hip-worn accelerometer data in older adults. This promotes enhanced analysis of older adult movement data, resulting in more accurate measures of sitting patterns and opening the door for large scale cohort studies into the effects of sitting patterns on healthy aging outcomes. **Keywords:** Machine learning; healthy aging; sit-to-stand transitions; activPAL, ActiGraph; free-living

Introduction

Sedentary behavior is a severe and prevalent health risk for older adults comprising 10-14 hours of older adults' days (1–6). Recent evidence suggests that there may be additional risk associated with sitting for prolonged periods of time independent of the total time spent sitting (7–9). The latter findings have led to increased interest in the study of "sitting patterns", which refers to the number and duration of sitting bouts (i.e., continuous periods of sitting) versus non-sitting bouts (i.e., continuous periods of standing or stepping), as well as the postural transitions between them. Sitting patterns can be quantified using metrics such as number of daily sit-to-stand transitions, number of daily sitting bouts, number of daily prolonged sitting bouts (≥30 mins), mean sitting bout duration (total daily sitting time/total sit-to-stand transitions), and usual bout duration (the sitting bout duration at or above which 50% of an individual's sitting time is accumulated) (8,10).

Sitting patterns are generally measured using thigh or hip-worn accelerometers, however to date hip-worn accelerometry is the best approach to measure motion and movement (sedentary behavior) while thigh-worn devices are better at measuring posture and postural transitions (sitting patterns) (11–13). While systems using several sensors can measure both sedentary behavior and sitting patterns (14), it is desirable for participant ease and comfort to have one device that can measure both with high validity. Measures of sitting patterns derived from cut-point-based hip-worn accelerometer data do not adequately measure the postural transitions that form the basis of sitting pattern metrics, including overestimating the number of sit-to-stand transitions and underestimating prolonged sitting time (15–17). Progress in machine learning techniques may make it possible to address hip-worn accelerometry's major limitation and close

the gap in sitting pattern measurement between hip-worn and thigh-worn accelerometers, as evidenced by developments in related areas such as activity type and intensity classification (18–21). However, the ability of current algorithms to identify the postural transitions (sit-to-stand) needed to measure sitting patterns in free-living populations is low, and there is a lack of algorithms that are specifically trained to identify transitions (22–24).

Thigh-worn inclinometers such as activPAL have been shown to accurately capture sit-to-stand transitions and can be used as high-frequency ground truth in posture labeling because data are provided many times per second (25). In previous work we have demonstrated that activPAL data can be used to train machine learning models for capturing postural transitions in free-living hip-worn accelerometer data, though a small sample with low generalizability was used (26,27). Here we build on this previous work and describe the training and validation of a Convolutional Neural Network (CNN) + bi-directional long short-term memory network (BiLSTM) model designed to classify sitting patterns as well as sedentary behavior from hip-worn ActiGraph accelerometer data. We dub this algorithm the CNN Hip Accelereometer Posture (CHAP) method and detail its superior accuracy for identifying sit-to-stand transitions using data from 709 older men and women who concurrently wore hip-worn ActiGraph accelerometers and thigh-worn activPAL inclinometers for up to 7 days.

Methods

Parent Study

Data were obtained from the Adult Changes in Thought (ACT) study, an ongoing longitudinal cohort study that maintains an active enrollment of approximately 2,000 older adults (≥ 65 y old) in Washington State. The ACT study began in 1994 to investigate risk factors for development of dementia and has since provided a unique opportunity to additionally study a wide range of non-cognitive factors of healthy aging. Starting in 2016, the ACT activity monitor sub-study (ACT-AM) was initiated, adding a device-based activity monitoring component to capture the spectrum of sedentary and physically active patterns (28). Participants were excluded from ACT-AM if they were wheelchair bound, receiving hospice or care for a critical illness, resided in a nursing home, or if memory problems became evident during testing. The remaining participants were asked to wear a hip-worn ActiGraph wGT3X+ (ActiGraph LLC, Pensacola, FL, USA), activated using ActiLife software to capture 30 Hz triaxial (i.e. data captured from three spatial axes) data and worn on an elastic belt situated so the device rests on the right side at the level of the suprailiac crest, and a thigh-worn activPAL micro3 (PAL Technologies, Glasgow, Scotland, UK), activated using a 10s minimum threshold for labeling postural transitions and secured to the front, center thigh with waterproofed materials. Participants were asked to wear both devices 24-hours/day for 1 week. While some participants elected only to wear one device, most wore both simultaneously. Participants also recorded self-reported sleep logs throughout their device wear. Ethics approval was obtained from the Kaiser Permanente Washington institutional review board (approval #821300). All participants provided written informed consent.

Data Cleaning and Pre-Processing

In-bed and accelerometer non-wear time was removed from the device data. The collected self-reported sleep logs were used to identify and remove in-bed time. Missing sleep log information was imputed using person-specific means, when available, or using the sample average. To identify and remove periods of non-wear, ActiGraph accelerometer data were processed using the Choi algorithm (29,30) applied to vector magnitude counts per minute using a 90-minute window, 30-minute streamframe, and 2-minute tolerance.

For inclusion in this study, data was required from both the ActiGraph and activPAL devices simultaneously. Participants were excluded if data from either of the monitors were missing or invalid. No minimum wear time criteria were required; all days with concurrent device wear for any length of time were considered valid days and were included in the sample. After restricting to waking wear time on both devices, visual inspection was used to define invalid data based on time drift between the monitors, a phenomenon in which data collected from one device appears to gradually lose or gain time when compared to another device resulting in the two data streams no longer aligning (see Figure, Supplemental Digital Content–Appendix, which depicts an example of drift between activPAL and ActiGraph) (31).

CHAP Design

The CHAP method was developed using a deep neural network (32) to classify sitting versus non-sitting behavioral postures and postural transitions from 10 Hz triaxial ActiGrah data (downsampled from 30 Hz via boxcar aggregation to reduce the size of the dataset). All computations were made on 10-second non-overlapping windows of continuous 10 Hz data, each

containing 100 triaxial acceleration values. The 10-second window size was chosen to align with activPAL's 10-second minimum threshold for labeling postural transitions. We used a model architecture family called CNN-BiLSTM architecture (33), which has three main components: 1) a CNN base (34), 2) a BiLSTM network (35), and 3) a softmax output layer akin to a logistic regression classifier (36). The first component automatically extracted features for identifying sitting versus non-sitting for each time point; the second component refined these features by considering neighboring time points and the most likely sequence of events; the third component converted the extracted features to a final classification label (sitting or non-sitting).. Below, detailed descriptions are given for each component of CHAP and the unique way these components work synergistically.

CNN. After partitioning both activPAL and triaxial ActiGraph data into non-overlapping 10 second increments, features were extracted for each window. Unlike traditional machine learning models that target certain pre-defined features (e.g. time- or frequency-domain summary values), the CNN automaticaly learned its own features by repeatedly convolving the raw triaxial data, with each convolution using a different kernel. During training, the model learned the parameters of each kernel, which enabled the convolution-based features to capture the relevant information for the posture classification task.

BiLSTM. The CNN classifications were made under the assumption that all 10-second windows contained independent and identically-distributed data (37). Human behavior does not meet these conditions, as a given action will generally be influenced by the preceding actions. Therefore, it was important to account for this temporal dependence (38), which necessitated

layering the BiLSTM on top of the CNN. The BiLSTM component automatically learned temporal features from the patterns of variations across time to differentiate activities. The BiLSTM component took in a sequence of features produced by the CNN component for a window of input data and output another sequence of BiLSTM extracted features corresponding to each 10-second window of the input. During training, the parameters of the BiLSTM component were adjusted to properly smooth the output so that there was minimal opportunity for the model to insert spurious interruptions during continuous sitting or non-sitting bouts.

CNN and BiLSTM Featurization Relationship. The CNN and BiLSTM components have a complementary relationship in how they featurized the data for classification. The CNN captured behaviors at a lower temporal granularity using the immediate temporal relationships within the classification window (10 seconds). This helped identify sudden changes in the base accelerometer features, e.g., those caused by transitions. In a sense, similar to how 2-D CNNs exploit spatial dependencies in image pixels to extract relevant features, our 1-D CNN effectively treated time series as "1-D images" across time. The BiLSTM's memory cells "remembered" patterns in the extracted CNN features over time to discern higher-level behaviors with longer temporal relationships. This helped identify both non-changes in the base features, e.g., those during sitting (or non-sitting) bouts, as well as reoccurring changes, e.g., back-to-back transitions. Together, these capabilities demonstrated the power of modern deep learning in automatically featurizing low-level sequence data: myriad manually tuned temporal thresholds are replaced with compact end-to-end learned neural architectures.

Softmax Output Layer. The output of the BiLSTM component was a sequence of intermediate features corresponding to a window of input data. To perform the final behavior classification on the extracted features we used a Softmax layer. The Softmax layer converts input features to final probabilities of each 10 second time interval belonging to sitting or non-sitting behavior. We then selected the most probable label as the final classification.

CHAP Development and Evaluation

The sample was divided into a training sample (n = 399 participants), a holdout validation sample (n = 97), and a test sample (n = 213). The training and validation samples were used to determine the optimal settings for CHAP, while the test sample was withheld until final models were selected and used for a performance comparison of CHAP and two other commonly-used sitting pattern classification methods (described below). Given the large number of steps and parameter tuning that occurs when building CNN models, a test dataset was critical for obtaining unbiased estimates of model performance.

Model Development. The CHAP method was trained end to end using the backpropagation technique (32), meaning that output from each layer was sequentially fed into the subsequent layer to generate a final output. During training, we fed each window of input ActiGraph data through CHAP, generating classifications for each 10-second time interval in each input window. We then compared classifications with the activPAL-derived ground truth labels corresponding to the same 10-second input window in question, which are assigned based on the majority activPAL designated posture in a given 10-second window (note: in the case of a tie, the sitting label was chosen). Based on this comparison, we then used the backpropagation

method to update the learnable parameters in the model in order to minimize the cross-entropy of classifications (i.e., maximize accuracy) between the predicted classifications and the ground truth labels. This process was completed for all input training data and repeated several times.

Training neural networks is a complex process involving multiple parameters and tuning steps which could lead to models that overfit the data. Thus, it is unwise to use training data alone for model selection given that the goal is to apply the algorithm on future data that is independent of the training set (39). Therefore, we fitted several model configurations on the training data, and compared their performance when applied to the holdout validation data. Model configurations varied on four dimensions: BiLSTM window size (7 and 9 min), number of neurons in a CNN layer (3200 and 6400 neurons), learning rate (0.001 and 0.0001), and regularization coefficient (0.001 and 0.0001). All possible unique combinations of domain values were tested, for a total of 16 unique model configurations tested. These comparisons enabled us to identify the best model configuration, based on several performance metrics (Table 1). Metrics included overall and balanced classification accuracy, ability to accurately capture transitions (i.e., changes in posture), sitting and non-sitting bout deviations, and Kolmogorov-Smirnov statistics for comparing CHAP-predicted vs true (activPAL) probability distributions of sitting and non-sitting bouts. Models with low accuracy or high variance, relative to competing models, on any of these metrics were eliminated. Three models performed equally well on all metrics, and these models were used to create a hybrid ensemble model that made classifications based on the majority vote. This ensemble model represented the complete CHAP method. For each of the three models that performed best in the validation set, and the final ensemble model, we

calculated the means and standard deviations (SD) of the evaluation metrics described in Table 1.

Model Evaluation. Using data from the test set, we compared the performance of CHAP to the performance of two other classification approaches that are commonly used in the field: 1) the standard ActiGraph cut-point (AG cut-point) method, and 2) a previously developed two level behavior classification (TLBC) machine learned model designed to differentiate sitting from standing postures. The AG cut-point method is designed to capture sedentary, non-movement bouts, which are sometimes used as a proxy for sitting bouts (7). Sedentary bouts were defined using 1-minute epoch data, in which minutes were classified as sedentary if the vertical axis counts were less than 100 (40). Consecutive sedentary minutes were classified as bouts with no minimum duration required and no allowance for interruptions. TLBC sequentially applies a pre-trained random forest and hidden Markov model (hmm) to 30 Hz tri-axial accelerometer data and was trained using annotated images captured from person-worn SenseCams (41–43). TLBC first converts the 30 Hz tri-axial accelerometer data into a set of 41 engineered features that are used to classify minutes of sitting, riding in a vehicle (which collectively represent sitting), standing, and walking/running (which collectively represent non-sitting). We defined sitting bouts as any period labeled by TLBC as a sitting posture, specifically sitting and riding in a vehicle.

The methods were compared using the same classification metrics that were used during validation (see Table 1). Because TLBC and AG cut-point methods yielded results at minute-level, for model comparison purposes, CHAP's 10-second-level classifications were aggregated

to minute-level, using majority vote for sitting vs. non-sitting labels. We also included comparisons of common person-level sitting pattern metrics, including mean sitting bout duration (total sitting time/ number of sitting bouts), average daily sitting time (total sitting time/ number of days), and average daily number of sitting bouts (number of sitting bouts / number of days). A final performance indicator was how well each method was able to predict the timing of postural transitions at a 10-second granularity within a one-minute window. This analysis was done using the transition pairing method (44), which uses an extended Gale-Shapley algorithm to pair actual and predicted transitions together for analysis. The method allowed exclusion of non-sequential pairings and any pairings that exceeded a specified lag time (tolerance), which was 1 minute for this study. One minute was the minimum tolerance level after which the number of successful pairings levelled off (See Supplemental Table 1, Supplemental Digital Content– Appendix, which shows transition pair sensitivity and precision results at different tolerance levels, from no tolerance to 5 minutes, across methods). The pairings were analyzed to determine the true positive rate (recall) and positive predictive value (PPV; precision), of predicted transitions.

Performance metrics were calculated for each person and method. Summary statistics were then calculated across participants, and boxplots were used to visually examine variability across test subjects. In addition to model performance metrics, we also compared commonly used sitting pattern metrics (mean sitting bout duration, mean daily sitting time, and mean number of daily sitting bouts), derived using each method to the activPAL ground truth. General estimating equations (GEE), accounting for nesting of methods within participants, were used to evaluate differences of performance between methods as well as whether sitting pattern metrics

derived from different methods were significantly different from those derived from activPAL. GEE was implemented using an exchangeable correlation structure and robust standard errors. Finally, to allow inference about individual-level, in addition to sample-level, agreement, sitting pattern metrics derived from each modeling approach (AG cut-point, TLBC, and CHAP) were also compared to activPAL using mean absolute error (MAE).

Results

Sample Partitioning and Characteristics

Figure 1 summarizes data loss and partitioning, and Table 2 shows participant characteristics for the final sample. Participant characteristics for the included overall ACT-AM sample were similarly distributed in the training (N=399), validation (N=97), and test sets (N=213).

Model Accuracy

Ten-second-level summary statistics of the three best CNN model configurations (labeled A, B, C), as well as the CHAP model are displayed in Table 3. Here we focus on the accuracy and mean absolute percent error (MAPE) metrics defined in Table 1 between the three CNN model configurations, which estimate agreement and deviation between the actual and predicted values.

Across all performance metrics, CHAP was superior to the other methods (Figure 2) at the minute level. For balanced accuracy, which is the average of sensitivity and specificity, the AG cut-point method performed worst, with a value of 74%, followed by 83% for TLBC versus

93% for the CHAP model. All models had high sensitivity for classifying sitting, ranging from 88% (AG cut-point) to 97% (CHAP). Specificity varied markedly between models: 60% for AG cut-point, 74% for TLBC, and 89% for CHAP. The differences in performance in balanced accuracy, sensitivity and specificity between CHAP and the AG cut-point method, and between CHAP and TLBC were statistically significant at the 5% level. The MAPEs of sitting versus non-sitting classification were not similar. While all methods were able to accurately classify true sitting, the AG cut-point and TLBC methods classified between 25 – 40% of true (activPAL registered) non-sitting as sitting. Of note, the variation in these metrics was also higher for the AG cut-point and TLBC versus CHAP, indicating superior individual-level agreement for the latter method.

Participant-level Sitting Pattern Classification

Figure 3 shows results of the sitting pattern analyses. The average mean bout duration from CHAP, 15.7 minutes per day, did not significantly differ relative to activPAL, 15.4 minutes per day (MAE_{CHAP} = 2 minutes). Average mean bout duration using AG cut-point, 9.4 minutes per day, and TLBC method, 49.4 minutes per day, did significantly differ at the 5% level relative to activPAL ($MAE_{AG\ cut\text{-}point}$ = 6 minutes, and MAE_{TLBC} = 34 minutes). Average daily sitting time derived using AG cut-point, 643.2 minutes per day, and using TLBC method, 616.2 minutes per day, was significantly different relative to activPAL, 594.6 minutes per day ($MAE_{AG\ cut\text{-}point}$ = 75 minutes, and MAE_{TLBC} = 50 minutes), but average daily sitting time derived from CHAP, 595.4 minutes per day, was not significantly different relative to activPAL (MAE_{CHAP} = 31 minutes). Average daily number of sitting bouts using all three methods were significantly different from activPAL. Of the three methods, average daily number of sitting bouts derived

using CHAP, 41.8 per day, was the closest to activPAL, 43.9 per day (MAE_{CHAP} = 5) and the difference was not deemed to be relevant in practice. The average daily number of sitting bouts derived using AG cut-point, 79.2 per day, and TLBC, 14.1 per day, had much larger deviations relative to activPAL ($MAE_{AG\ cut\text{-}point}$ = 35, and MAE_{TLBC} = 30). The results suggest that the latter two methods are unable to accurately capture sitting patterns. AG cut-point over-predicted the number of transitions by two times explaining why its mean bout duration was lower than activPAL, whereas TLBC under-predicted relative to activPAL by two thirds, hence why its mean bout duration was higher. Despite its superior performance to the other two methods, the CHAP method had slightly lower person-to-person variability (i.e., lower SDs) compared to activPAL.

Classifying the timing of Sit-to-Stand Transitions

We examined accuracy in predicting sit-to-stand transitions within a 1-minute window by the three methods compared to the activPAL (Figure 4). Transition sensitivity estimates the percent of true transitions (as registered by the activPAL), that were captured by the different methods. Sensitivity for transition detection was similar for the AG cut-point (72%) and CHAP (83%), whereas it was only 26% for TLBC, likely due to over-smoothing. Transition PPV or precision estimates the proportion of predicted transitions which are true activPAL transitions. In contrast to the sensitivity results, PPV was similar for CHAP (83%) and TLBC (71%), whereas it was only 30% for the AG cut-point. The differences in performance in transition sensitivity and transition PPV between CHAP and the AG cut-point method, and between CHAP and TLBC were statistically significant at the 5% level.

Discussion

The CHAP model had higher accuracy than existing methods for classifying sitting bouts and sit-to-stand transitions from free-living hip-worn accelerometer data in older adults. As such, it represents an important step forward in the field of sitting pattern measurement in this population. CHAP will allow for less cumbersome protocols for studies in older adults by necessitating only one hip-worn device to measure both posture and motion. CHAP can be used to re-process previously collected hip-worn accelerometer data among older adults, resulting in more accurate measures of true sitting time and patterns in existing cohort studies as well as future studies that choose to use hip-worn accelerometers.

The AG cut-point method over-estimated true sitting time and failed to capture sit-to-stand transitions that are key to the measurement of sitting patterns (15–17,45). This underscores the importance of using methods for their intended use. That is, cut-point methods are meant to capture movement intensity and non-movement but not changes in posture. The main shortcoming of the cut-point method was that it misclassified approximately 40% of activPAL registered non-sitting time as sitting, while simultaneously over-predicting sit-to-stand transitions such that approximately 70% of the transitions it predicted were not activPAL transitions, resulting in inaccurate measures of sitting patterns. These findings are in line with other studies that support the use of hip-worn accelerometry for measuring motion and movement but suggest thigh-worn devices for measuring posture and postural transitions (11–13,15–17). Thus, evidence on sitting patterns measured using ActiGraph cut-points should be interpreted with caution. It is not clear whether such misestimation has major impacts on the ability to detect associations between sitting patterns and health. Nonetheless, there is sufficient evidence to

suggest that sitting pattern estimates, derived from ActiGraph cut-points should not be compared to studies that employed posture-based measures such as activPAL or used to inform specific thresholds of sitting patterns when generating intervention or public health recommendations.

Transitions have been a large issue for the field even with application of machine learned algorithms. Machine learning approaches most often rely on single label classification within a given window or period (e.g., 5 minutes), and therefore an inherent assumption is that only one activity type occurs within each interval window (22). Lab-based training data reduces the amount of transitions, resulting in algorithms with high predictive accuracy, but algorithms trained on data obtained from free-living populations must account for the inherent messiness of human postural changes and movement. The TLBC method was designed to address some of these limitations by training it against free-living images collected by a body-worn camera. However, the body-worn camera captured images triggered by changes in light and movement, meaning TLBC was unable to reliably capture postural transitions or their exact timing, leading to an underestimation of postural transitions (44). Solutions have been proposed in the literature to allow for better identification of transitions by machine learning models including activity-based windowing and adaptive sliding window segmentation, where for both solutions windows are adjusted to ensure one activity is represented per window and windows can vary in size throughout the dataset (46,47). Alternatively, CHAP uses a BiLSTM component with a fixed window that automatically learns to capture the transitions during training. We found that even though the model accuracy did not significantly vary (at most 2% variation) with the chosen BiLSTM window size, it significantly affected the ability of the model to capture transitions correctly. As the window size was increased from 1 minute to 9 minutes, the transition capturing

recall reduced by 6% from 83% to 77% and the PPV increased by 23% from 56% to 79%. In practice, we found that a window size of 7-9 minutes works well for our data, which had a mean activPAL sitting bout time of 15.4 minutes and mean non-sitting bout time of 7.9 minutes. More experimental results on the model sensitivity for the chosen BiLSTM window size are provided in Supplemental Table 2 (see Table, Supplemental Digital Content–Appendix).

Deep learning methods to improve measures derived from accelerometer data are of growing interest in the field. For instance, Nawaratne et al leverage a CNN model architecture to derive measures of physical activity intensity from wrist-worn ActiGraph that are of equal caliber to those measured from the hip-worn ActiGraph. While the goals of Nawaratne et al's model differ from those of CHAP, making the results not directly comparable, their work demonstrates the utility of CNN model architecture in constructing machine learned approaches to processing accelerometer data (48). CHAP builds on this approach, adding a BiLSTM layer for improved measurement of activity transitions.

We were able to find only one other study that uses hip-worn ActiGraph data to classify sedentary behavior and sitting patterns in a free-living population with high accuracy. Kuster et al developed an algorithm utilizing hip-worn ActiGraph data in a sample of office workers (N=38) to detect prolonged sitting bouts (≥5 and ≥10 minutes). Their method used a random forest classifier on 563 engineered ActiGraph signal features, followed by a bagged classification tree ensemble method. The model achieved a low bias of ≤ 7 minutes/d, when classifying time spent in prolonged sitting bouts (≥ 5 minutes and ≥ 10 minutes) relative to activPAL (49). CHAP builds on the model of Kuster et al in several ways. Most importantly, it was developed,

validated, and tested on a larger and more representative cohort (N=709) of free-living older adults. Through the CNN + BiLSTM architecture, CHAP was also able to automate the featureextraction process rather than relying on human engineered features. As a result, CHAP requires less human input than the Kuster et al. model and is a versatile and flexible model that can be used to derive various person-level sitting pattern variables beyond prolonged sitting bouts. This application in the older adult population of the ACT cohort represents only the first test-case for CHAP. Future work will apply this method in other populations to assess performance and generalizability of CHAP in other age groups, and refine the model for broader generalizability across age, sex, and other key demographic factors.

Researchers interested in more deeply exploring the CHAP algorithm or applying CHAP to their existing hip-worn accelerometer data to derive postural transition and sitting pattern metrics are invited to explore the study's GitHub repository. CHAP and associated user documentation are available for download from https://github.com/ADALabUCSD/DeepPostures.

Our study has several limitations that should be considered. We used thigh-worn activPAL data as ground truth rather than direct observation, which could lead to compounding of the activPAL's inherent measurement error. However, we believe the benefit of obtaining large amounts of free-living data outweighs limitations of activPAL. Furthermore, activPAL has been shown to be a highly valid instrument for measuring postural transitions (25). Notably, CHAP had slightly lower person-to-person variability (i.e., lower SDs for derived sitting pattern metrics) compared to activPAL, which could potentially result in reduced statistical power in

studies of associations between sitting patterns and health outcomes, and should be addressed in future studies. However, since our CHAP model predictions have similar probability distributions to that of the ground-truth (activPAL), in practice, we do not expect substantial negative effects on study power when using CHAP predictions. Despite these limitations, our study had considerable strengths, including the large sample size and rigorous machine learning procedures employed. Although CHAP allows posture-based classification from a single device, the hip-worn ActiGraph, it is important to acknowledge that methods for integrating both types of sensors (e.g., activPAL and ActiGraph) to achieve systems for postural and motion measurement have been previously developed (14). Additionally, recent studies have developed accurate classification methods of wrist-worn accelerometer data for both sedentary behavior and sitting patterns (50, 51).

CHAP performed much better than currently available methods, and it established a novel and powerful framework for models that use hip-worn data. This advance will allow researchers to better understand the epidemiology of sitting patterns, including norms among healthy and unhealthy people and how sitting patterns are causally associated with a myriad of healthy aging outcomes. Additionally, it will reduce participant burden by allowing for accurate measurement of posture and motion using one hip-worn device, rather than necessitating several devices. Ultimately, this data will be needed to help inform future guidelines for sedentary behavior among older adults.

Acknowledgements

This work was supported by grant number U01AG006781 from the National Institute on Aging and R01DK114945 from the National Institute of Diabetes and Digestive and Kidney Diseases. It was also supported in part by a Hellman Fellowship, an NSF CAREER Award under award number 1942724, and a gift from VMware. The content is solely the responsibility of the authors and does not necessarily represent the views of any of these organizations. We thank the members of UC San Diego's Database Lab and Center for Networked Systems for their feedback on this work.

Conflict of Interest

The authors have no conflicts of interest to declare. Results of the present study do not constitute endorsement by ACSM. Results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

References

1. Copeland JL, Ashe MC, Biddle SJ, et al. Sedentary time in older adults: A critical review of measurement, associations with health, and interventions. Br J Sports Med. 2017;51(21):1539.
2. Biswas A, Oh PI, Faulkner GE, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: A systematic review and meta-analysis. Ann Intern Med. 2015;162:123–32.
3. Knaeps S, Bourgois JG, Charlier R, Mertens E, Lefevre J, Wijndaele K. Ten-year change in sedentary behaviour, moderate-To-vigorous physical activity, cardiorespiratory fitness and cardiometabolic risk: Independent associations and mediation analysis. Br J Sports Med. 2018;52(16):1063-1068.
4. De Rezende LFM, Lopes MR, Rey-Lopez JP, Matsudo VKR, Luiz ODC. Sedentary behavior and health outcomes: An overview of systematic reviews. PLoS One. 2014;9(8): e105620.
5. Matthews CE, Chen KY, Freedson PS, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol. 2008;167:875–81.
6. Harvey JA, Chastin SFM, Skelton DA. How sedentary are older people? A systematic review of the amount of sedentary behavior. J Aging Phys Act. 2015;23:471–87.
7. Healy GN, Dunstan DW, Salmon J, et al. Breaks in Sedentary Time: Beneficial associations with metabolic risk. Diabetes Care. 2008;31(4):661–6.
8. Chastin SFM, Granat MH. Methods for objective measure, quantification and analysis of sedentary behaviour and inactivity. Gait Posture. 2010;31(1):82–6.
9. Bellettiere J, Winkler EAH, Chastin SFM, et al. Associations of sitting accumulation

patterns with cardio-metabolic risk biomarkers in Australian adults. PLoS One. 2017;12(6):1–17.

10. Boerema ST, van Velsen L, Vollenbroek MMR, Hermens HJ. Pattern measures of sedentary behaviour in adults: A literature review. Digit Heal. 2020;6.
11. Janssen X, Cliff DP. Issues Related to Measuring and Interpreting Objectively Measured Sedentary Behavior Data. Meas Phys Educ Exerc Sci. 2015;19(3):116–24.
12. Kim Y, Barry VW, Kang M. Validation of the ActiGraph GT3X and activPAL Accelerometers for the Assessment of Sedentary Behavior. Meas Phys Educ Exerc Sci. 2015;19:125–37.
13. Montoye AHK, Pivarnik JM, Mudd LM, Biswas S, Pfeiffer KA. Validation and Comparison of Accelerometers Worn on the Hip, Thigh, and Wrists for Measuring Physical Activity and Sedentary Behavior. AIMS Public Heal. 2016;3(2):298–312.
14. Myers A, Gibbons C, Butler E, et al. A novel integrative procedure for identifying and integrating three-dimensions of objectively measured free-living sedentary behaviour. BMC Public Health. 2017;17.
15. Barreira TV, Zderic TW, Schuna JM, Hamilton MT, Tudor-Locke C. Free-living activity counts-derived breaks in sedentary time: Are they real transitions from sitting to standing? Gait Posture. 2015;42(1):70–2.
16. Carlson JA, Bellettiere J, Kerr J, et al. Day-level sedentary pattern estimates derived from hip-worn accelerometer cut-points in 8–12-year-olds: Do they reflect postural transitions? J Sports Sci. 2019;37(16):1899–909.
17. Bellettiere J, Tuz-Zahra F, Carlson J, et al. Agreement of sedentary behaviour metrics derived from hip-worn and thigh-worn accelerometers among older adults: with

implications for studying physical and cognitive health. J Meas Phys Behav. 2021;Advance online publication:1-10.

18. 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.

19. Marcotte RT, Petrucci GJ, Cox MF, Freedson PS, Staudenmayer JW, Sirard JR. Estimating sedentary time from a hip- And wrist-worn accelerometer. Med Sci Sports Exerc. 2020;52(1):225–32.

20. Wullems JA, Verschueren SMP, Degens H, Morse CI, Onambélé GL. Performance of thigh-mounted triaxial accelerometer algorithms in objective quantification of sedentary behaviour and physical activity in older adults. PLoS One. 2017;12(11).

21. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying Active Travel Behaviors in Challenging Environments Using GPS, Accelerometers, and Machine Learning Algorithms. Front public Heal [Internet]. 2014 Jan [cited 2014 Jul 29];2:36. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4001067&tool=pmcentrez&rendertype=abstract

22. Farrahi V, Niemelä M, Kangas M, Korpelainen R, Jämsä T. Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. Gait and Posture. 2019;68:285-299.

23. Narayanan A, Desai F, Stewart T, Duncan S, MacKay L. Application of raw accelerometer data and machine-learning techniques to characterize human movement behavior: A systematic scoping review. J Phys Act Heal. 2020;17(3):360–83.

24. Rosenberg D, Godbole S, Ellis K, et al. Classifiers for Accelerometer-Measured Behaviors in Older Women. Med Sci Sports Exerc. 2017;49(3):610–6.

25. Giurgiu M, Bussmann JBJ, Hill H, et al. Validating Accelerometers for the Assessment of Body Position and Sedentary Behavior. J Meas Phys Behav. 2020;3(3):253–63.

26. Kerr J, Carlson J, Godbole S, Cadmus-Bertram L, Bellettiere J, Hartman S. Improving Hip-Worn Accelerometer Estimates of Sitting Using Machine Learning Methods. Med Sci Sports Exerc. 2018;50(7):1518–24.

27. Nakandala S, Jankowska MM, Tuz-Zahra F, et al. Application of Convolutional Neural Network Algorithms for Advancing Sedentary and Activity Bout Classification. J Meas Phys Behav. 2021; DOI: https://doi.org/10.1123/jmpb.2020-0016.

28. Rosenberg DE, Walker R, Greenwood-Hickman MA, et al. Device-assessed physical activity and sedentary behavior in a community-dwelling cohort of older adults. BMC Public Health. 2020;20:1256.

29. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc. 2011;43:357–64.

30. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Med Sci Sports Exerc. 2012;44(10):2009–16.

31. Steel C, Bejarano C, Carlson JA. Time Drift Considerations When Using GPS and Accelerometers. J Meas Phys Behav. 2019;2(3):203–7.

32. Goodfellow I, Bengio Y, Courville A. Deep Learning [Internet]. Deep Learning. MIT Press; 2016. Available from: http://www.deeplearningbook.org

33. Yoon J, Kim H. Multi-channel lexicon integrated CNN-BILSTM models for sentiment

analysis. In: Proceedings of the 29th Conference on Computational Linguistics and Speech Processing, ROCLING 2017. 2017.

34. Kiranyaz S, Ince T, Gabbouj M. Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. IEEE Trans Biomed Eng. 2016;63(3):664-75.

35. Wang R, Liang X, Zhu X, Xie Y. A Feasibility of Respiration Prediction Based on Deep Bi-LSTM for Real-Time Tumor Tracking. In: IEEE Access. 2018. p. 51262–8.

36. Bridle JS. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In: Soulie FF, Jerault J, editors. Neurocomputing NATO ASI Series (Series F: Computer and Systems Sciences). Berlin: Springer; 1990.

37. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. Vol.1 No.1. Springer Series in Statistics. New York: Springer; 2009:191-218.

38. Ray EL, Sasaki JE, Freedson PS, Staudenmayer J. Physical activity classification with dynamic discriminative methods. Biometrics. 2018;74(4):1502–11.

39. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A. Hyperband: A novel bandit-based approach to hyperparameter optimization. J Mach Learn Res. 2017;18(1):6765–6816.

40. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. Sport Med. 2017;47:1821–45.

41. 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.

42. Ellis K. TLBC: Two-Level Behavior Classification. R package version 1.1 [Internet].

2016. Available from: https://github.com/sieberts/TLBC

43. 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 [Internet]. 2013 Mar [cited 2014 Jul 20];44(3):290–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23415127

44. Hibbing PR, LaMunion SR, Hilafu H, Crouter SE. Evaluating the Performance of Sensor-Based Bout Detection Algorithms: The Transition Pairing Method. J Meas Phys Behaviou2r. 2020;3(3):219–27.

45. Lyden K, Kozey Keadle SL, Staudenmayer JW, Freedson PS. Validity of two wearable monitors to estimate breaks from sedentary time. Med Sci Sports Exerc. 2012;44(11):2243–52.

46. Allahbakhshi H, Hinrichs T, Huang H, Weibel R. The key factors in physical activity type detection using real-life data: A systematic review. Front Physiol. 2019;10:75.

47. Noor MHM, Salcic Z, Wang KIK. Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer. Pervasive Mob Comput. 2017;38(1):41–59.

48. Nawaratne R, Alahakoon D, De Silva D, et al. Deep Learning to Predict Energy Expenditure and Activity Intensity in Free Living Conditions using Wrist-specific Accelerometry. J Sports Sci. 2021 Mar;39(6):683-690.

49. Kuster RP, Grooten WJA, Baumgartner D, Blom V, Hagströmer M, Ekblom Ö. Detecting prolonged sitting bouts with the ActiGraph GT3X. Scand J Med Sci Sport. 2020;30:572–82.

50. Straczkiewicz M, Glynn NW, Zipunnikov V, Harezlak J. Fast and Robust Algorithm for

Detecting Body Posture Using Wrist-Worn Accelerometers. J Meas Phys Behav. 2020;3(4):285–93.

51. Twaites J, Everson R, Langford J, Hillsdon M. Transition Detection for Automatic Segmentation of Wrist-Worn Acceleration Data: A Comparison of New and Existing Methods. J Meas Phys Behav. 2020;3(1):19–28.

Figure Captions

Figure 1. Flow diagram from ACT study for inclusion into this study and random division into training and testing data sets.

[1]Non-concurrent wear represents data in which the devices are not worn concurrently.

[2]Drift is a phenomenon in which data collected from one device appears to gradually loose or gain time when compared to another device, such that, over time, the two data streams no longer align. See Figure, Supplemental Digital Content for an example of drift in this sample.

Figure 2. Minute-level performance (balanced accuracy, sensitivity/recall, specificity) in classifying sitting versus not sitting comparing AG cut-point (peach), TLBC (blue), and CHAP (green).

Figure 3. Person-level sitting pattern metrics (mean sitting bout duration, average daily sitting time in minutes, average daily number of sitting bouts) comparing activPAL (orange), AG cut-point (peach), TLBC (blue), and CHAP (green).

Figure 4. Assessment of minute-level performance in timing of classification of sit-to-stand transitions within 1-minute window (tolerance) using paired actual and predicted transitions for AG cut-point (peach), TLBC (blue), and CHAP (green).

Supplemental Digital Content

APPENDIX: OlderAdult_ML_NIDDK_FINAL_2-10-2021_Supplemental Content.docx

Figure 1

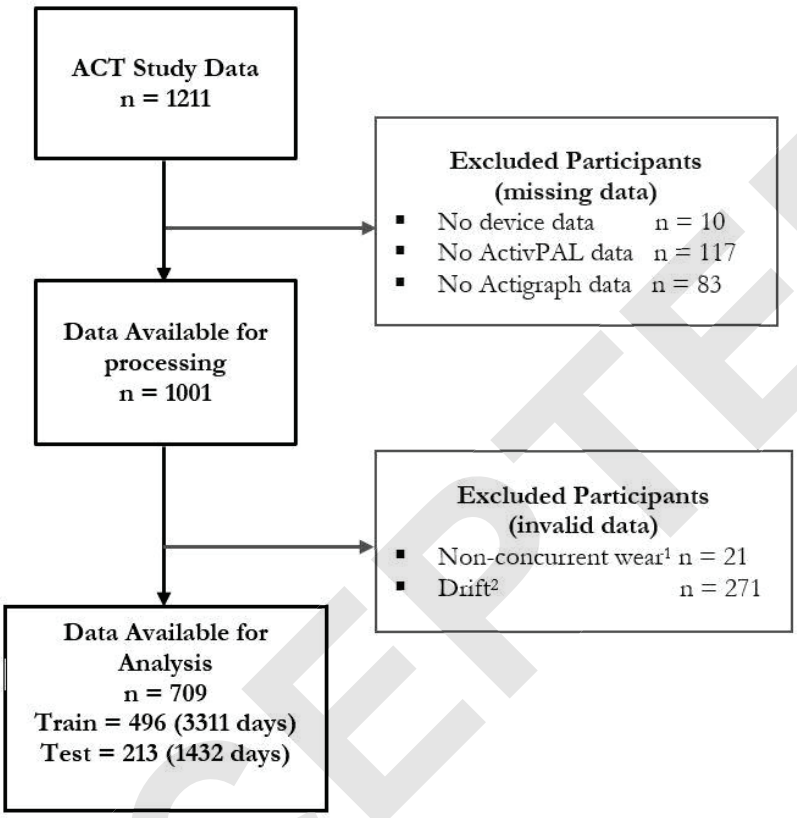

Figure 2

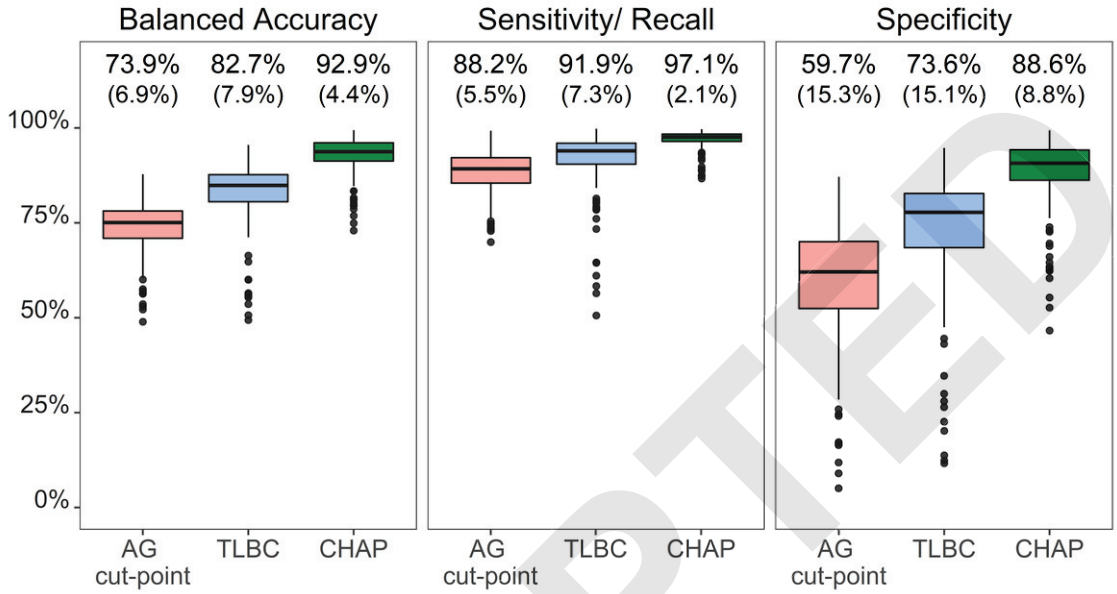

Figure 3

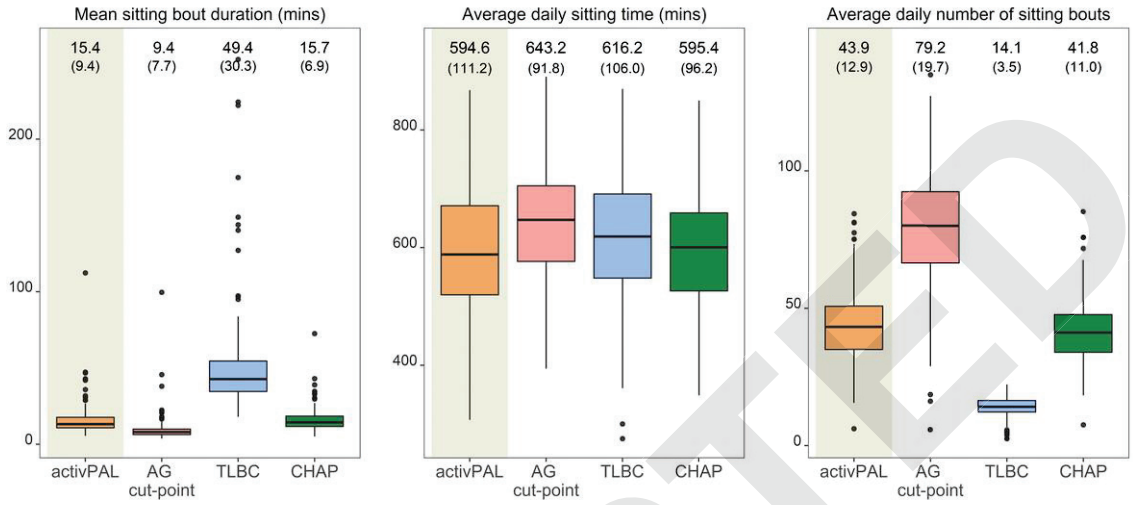

Figure 4

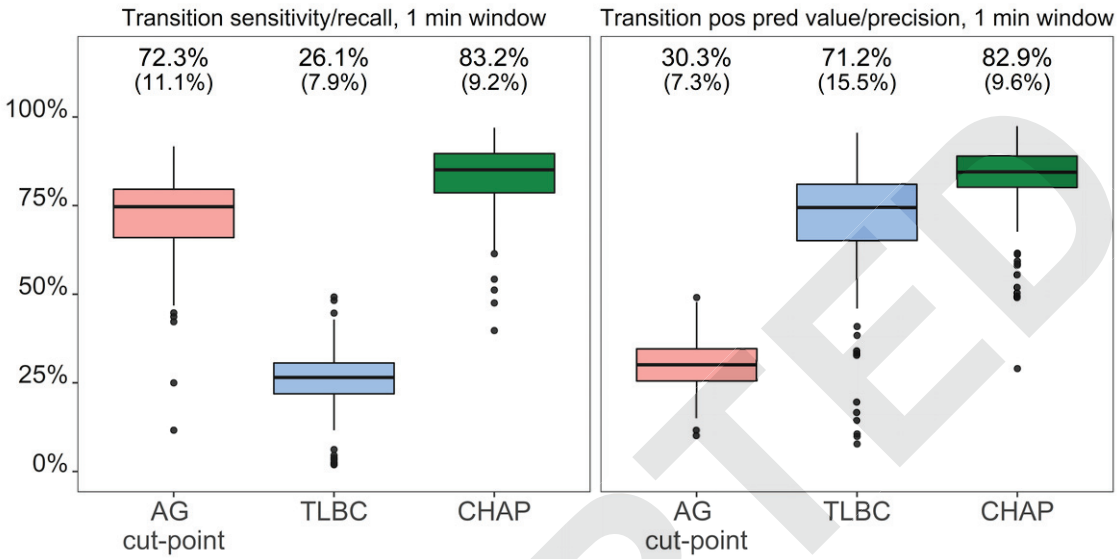

Table 1. Definitions and interpretations of accuracy and error metrics

Confusion matrix of actual and predicted 10s segments		
	Predicted Sitting	**Predicted Non-Sitting**
Actual Sitting	a	b
Actual Non-Sitting	c	d
Metric	**Definition**[1]	**Interpretation**
Accuracy	(a+d)/(a+b+c+d)	Proportion of segments correctly predicted.
Sensitivity	a/(a+b)	Proportion of activPAL sitting segments that were predicted sitting. Shows out of all the activPAL sitting segments how many were correctly precited as sitting.
Specificity	d/(c+d)	Proportion of activPAL non-sitting segments that were predicted non-sitting. Shows out of all the activPAL not sitting minutes, how many were correctly predicted as non-sitting.
Balanced Accuracy	0.5a/(a+b) + 0.5d/(c+d)	Average of sensitivity and specificity.
Sitting time Mean Absolute Percent Error (MAPE)	100*(|(a+b) – (a+c)|)/(a+b)	Absolute percent error in total predicted sitting time (versus total actual sitting time).
Not-Sitting time MAPE	100*(|(c+d) – (b+d)|)/(c+d)	Absolute percent error in total predicted non-sitting time (versus total actual non-sitting time).

[1]Refers to letters defined in the confusion matrix.

Table 2. Participant characteristics for the full, training, validation, and test sets.

Characteristics	**Full Sample** N = 709	**Training** N = 399	**Validation** N = 97	**Test** N = 213
	mean (SD)			
Age (years)	76.70 (6.52)	76.87 (6.38)	76.60 (6.84)	76.44 (6.64)
Gender	**N (%)**			
Female	415 (58.5%)	234 (58.6%)	54 (55.7%)	127 (59.6%)
Race ethnicity				
Hispanic or non-white	70 (9.9%)	31 (7.8%)	16 (16.5%)	23 (10.9%)
Education				
Less than High School	10 (1.4%)	7 (1.8%)	1 (1.0%)	2 (0.9%)
Completed High School	52 (7.3%)	25 (6.3%)	8 (8.2%)	19 (8.9%)
Some College	113 (15.9%)	68 (17.0%)	13 (13.4%)	32 (15.0%)
Completed College	534 (75.3%)	299 (74.9%)	75 (77.3%)	160 (75.1%)
BMI				
BMI 29 or below	537 (77.4%)	293 (74.7%)	81 (88.0%)	163 (77.6%)
BMI greater than 29	157 (22.6%)	99 (25.3%)	11 (12.0%)	47 (22.4%)
Self-Rated Health				

Good, poor, or very poor	279 (39.4%)	164 (41.1%)	37 (38.1%)	78 (36.6%)
Difficulty in walking half a mile				
Some or more	168 (23.7%)	99 (24.8%)	21 (21.6%)	48 (22.5%)

[1]Differences between training + validation sets and the test set were not statistically significant at the 5% level using two-sample t-test for continuous variables and chi-square test for categorical variables.

Table 3. Test set performance of top three performing CNN models and ensemble CHAP at the **10-second level** (mean (SD) of metrics).

Models	**Accuracy (%)**	**Balanced Accuracy (%)**	**Sitting time MAPE (%)**	**Non-sitting time MAPE (%)**	**Transition sensitivity (recall) % at 1 minute tolerance[1]**	**Transition PPV (precision) % at 1 minute tolerance[1]**
A	93.5 (3.9)	91.8 (4.7)	5.3	7.7	76.7 (10.3)	74.5 (12.6)
B	93.7 (3.8)	91.9 (5.1)	5.2	8.7	76.2 (11.1)	76.7 (12.3)
C	93.7 (3.6)	92.4 (4.2)	5.5	9.8	75.8 (9.9)	77.0 (11.6)
CHAP (ensemble)	94.1 (3.6)	92.6 (4.5)	5.2	8.2	77.1 (10.8)	80.0 (12.5)

[1] Detection of transitions within ±6 10-s epochs of ActiGraph data.

Supplemental Digital Content

Example of drift.

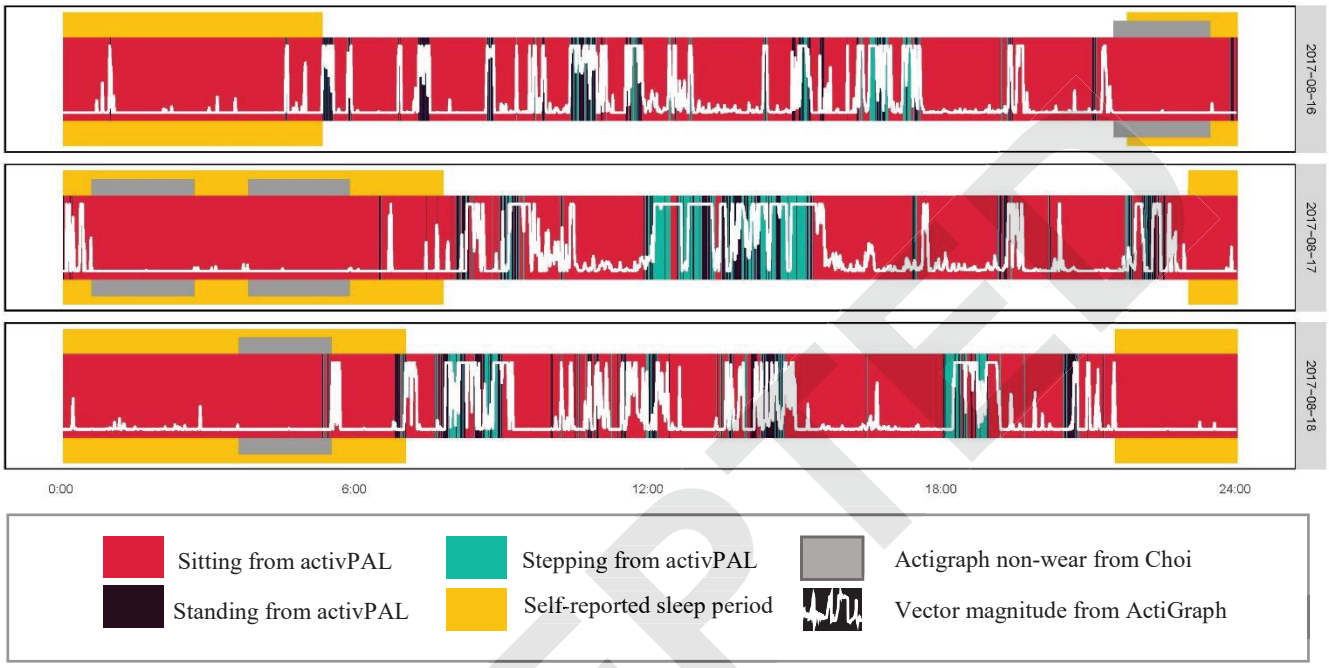

Figure. Example of Drift. Sample heat map of participant with device drift. Each rectangle is one 24-hour day. The horizontal outer bars represent sleep diaries (yellow bar), and Choi ActiGraph non-wear periods (grey bars). The horizontal inner bars represent activPAL postures; sitting (red), standing (dark blue), stepping (aquamarine). The white etching represents vector magnitude from the ActiGraph, plotted at one-minute epoch, truncated at 1500 counts per minute.

Transition recall and precision sensitivity to transition pairing tolerance window size

Supplemental Table 1. Minute-level test set transition analysis sensitivity to transition pairing tolerance window size (mean (SD) of metrics)

	Sensitivity / Recall (%)			**Positive Predictive Value / Precision (%)**		
Transition pairing tolerance	**AG cut-point**	**TLBC**	**CHAP**	**AG cut-point**	**TLBC**	**CHAP**
No tolerance	48.2 (12.9)	17.7 (7.2)	63.8 (16.2)	20.2 (6.6)	48.5 (17.2)	63.7 (16.5)
1 minute	72.3 (11.1)	26.1 (7.9)	83.2 (9.2)	30.3 (7.3)	71.2 (15.5)	82.9 (9.6)
2 minutes	79.3 (10.0)	27.8 (8.0)	84.7 (8.2)	33.2 (7.1)	75.9 (14.7)	84.4 (8.7)
3 minutes	82.7 (9.6)	28.8 (7.9)	85.3 (7.8)	34.7 (7.1)	79.0 (13.7)	85.0 (8.4)
4 minutes	84.9 (9.3)	29.6 (7.8)	85.9 (7.4)	35.6 (7.3)	81.4 (13.0)	85.5 (8.1)
5 minutes	86.4 (9.1)	30.2 (7.8)	86.2 (7.3)	36.3 (7.3)	83.3 (12.4)	85.8 (8.0)

Model sensitivity to the BiLSTM window size.

Supplemental Table 2. Test set performance of models at the **10-second level** with different BiLSTM window sizes when all other model configuration parameters kept fixed (mean (SD) of metrics).

BiLSTM Window Size (min)	**Accuracy (%)**	**Balanced Accuracy (%)**	**Transition sensitivity (recall) % at 1 minute tolerance**	**Transition PPV (precision) % at 1 minute tolerance**
1	92.1 (4.2)	90.2 (4.2)	83.5 (7.9)	56.0 (13.1)
3	93.1 (3.7)	91.5 (4.1)	79.4 (8.7)	69.3 (12.3)
5	93.1 (3.8)	91.9 (4.3)	77.4 (10.2)	74.3 (12.3)
7	93.2 (3.7)	92.1 (4.1)	75.9 (10.2)	73.8 (11.8)
9	93.9 (3.6)	92.4 (4.5)	76.9 (9.7)	78.5 (11.3)