

Individual classification of elementary school children's physical activity: A time-efficient, group-based approach to reference measurements

Jan Kühnhausen^{1,2,3} · Judith Dirk^{1,2} · Florian Schmiedek^{1,2,4}

Published online: 6 July 2016
© Psychonomic Society, Inc. 2016

Abstract The objective measurement of physical activity using accelerometers is becoming increasingly popular. There is little consensus, however, about how to analyze acceleration data. One promising approach is the use of *reference measurements* in which the subjects conduct specific activities. This makes it possible to identify data patterns that indicate these activities for each subject. The drawback of this approach is its rather high cost, in terms of both time and money. We propose a new approach in which a group of children conduct the reference measurements at the same time. We trained *support vector machine* models on the accelerometer data of 70 children (ages 8–11 years) to predict their activities during those reference measurements. We correctly classified activities with an accuracy of 96.9 % when fitting the individual models for each subject, and 87.5 % when fitting general models for all subjects. The obtained accuracies were comparable to results reported in previous reference measurement studies, in which each subject was measured

individually. They were higher than the accuracies obtained by the traditional approach, which transfers accelerometer data to counts and classifies those on the basis of predefined cut points. We concluded that our approach can yield a valuable contribution, particularly to studies with larger samples.

Keywords Accelerometry · Raw data · Physical activity · Reference measurements · Support vector machines · Children · Group-based

The positive influence of physical activity (PA) on cognitive performance as well as psychological and physiological well-being has been established in numerous studies and meta-analyses (e.g., Ahn & Fedewa, 2011; Pasco et al., 2011; Ploughman, 2008; Reed & Buck, 2009; Sibley & Etnier, 2003; Strong et al., 2005; Tomporowski, Davis, Miller, & Naglieri, 2008; Trost, Blair, & Khan, 2014). For a better understanding of these relations, it is important to investigate them in real-life settings. To do so, ways to assess the habitual PA of people have to be found. This can be done either with subjective measures, like self-report questionnaires, or with objective measures. Objective measurements have the advantage that they can assess PA continuously and without the inaccuracies that come with self-report measurements (e.g., Baumeister, Vohs, & Funder, 2009; Ebner-Priemer et al., 2006). Objective measurements of PA can be collected with various devices, of which accelerometers are among the most popular (for an overview of objective activity measurement techniques, see Butte, Ekelund, & Westerterp, 2012).

Until now, researchers have not reached a consensus about the proper way to transfer acceleration data into meaningful measures of activity (cf. Lee & Shiroma, 2014). In the present article, we try to aid approaching such a consensus in three ways. In short, we are concerned with the questions of, first,

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0724-2) contains supplementary material, which is available to authorized users.

✉ Jan Kühnhausen
jan.kuehnhausen@uni-tuebingen.de

¹ Individual Development and Adaptive Education of Children at Risk (IDeA) Research Center, Frankfurt, Germany

² German Institute for International Educational Research (DIPF), Frankfurt, Germany

³ Eberhard Karls Universität Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

⁴ Department of Psychology and Sport Sciences, Goethe-University, Frankfurt, Germany

whether raw data or counts should be used; second, whether data should be classified into activities based on an individual model for each subject or the same model for all subjects; and third and most importantly, whether individual models can be obtained in an economical way that is reasonably applicable to real-life research. To answer the last question, we proposed to obtain individual models from reference measurements done in groups, rather than individually.

Traditionally, accelerometer data are interpreted with the help of *movement counts* (e.g., Freedson, Pober, & Janz, 2005). Those counts are an expression of the amplitude of movement in a given time interval. The activity in this interval, usually 1 min, can then be expressed as *counts per minute*, a unit that can in principle be translated to meaningful measures of PA with the use of established cut points. PA is often described in terms of the time in which people conduct moderate to vigorous physical activity (MVPA) and the time in which they are in sedentary behavior (e.g., Aznar et al., 2011; Basterfield et al., 2011; King et al., 2011; Puyau, Adolph, Vohra, & Butte, 2002; Reilly et al., 2008). It has become popular to express recommendations for how much PA is beneficial for mental or physical health, in terms of time per day spent in MVPA (e.g., Strong et al., 2005; Wong et al., 2012). The strength of the approach of measuring PA with movement counts lies in its apparent simplicity, yielding an easy-to-understand measure of physical activity. However, it has several shortcomings. First, the *counts* are calculated by the manufacturer's software. This means that researchers lose control over data processing. They cannot influence how the counts are calculated, nor can they readily calculate the counts themselves. This forces researchers to take the produced counts as granted. Second, data processing differs across devices (Butte et al., 2012; Chen & Bassett, 2005), which is especially problematic in combination with the first problem. Accordingly, different devices produce significantly different numbers of counts for the exact same movements (Rothney, Apker, Song, & Chen, 2008). One possible reason for that is the so-called *plateau phenomenon* (cf. John, Miller, Kozey-Keadle, Caldwell, & Freedson, 2012), which stems from the way in which counts are calculated in the widely used ActiGraphs (ActiGraph, LLC, Fort Walton Beach, FL). It describes the phenomenon that beyond a certain intensity of activity, the activity counts actually *decrease* when the intensity of activity increases. This means that activities of very different intensities can produce the same count data. Third, appropriate cut points for the interpretation of counts have to be used. Counts in themselves are not a very informative measure. To interpret them, the activities that correspond to certain values of, for example, counts per minute must be established in so-called *calibration studies* (e.g., Evenson, Catellier, Gill, Ondrak, & McMurray, 2008; Puyau et al., 2002), which are often rather expensive and time-consuming. Furthermore, their results are not extendable to different subpopulations,

meaning that in principle, different studies are needed for people of different ages, genders, and other relevant characteristics (Strath, Pfeiffer, & Whitt-Glover, 2012). However, even within a single subpopulation, the results from these studies vary greatly (e.g., Sherar et al., 2011). When comparing the results of different studies, Corder, Ekelund, Steele, Wareham, and Brage (2008) found that the cut points for MVPA varied between 615 and 3,200 counts per minute, even when they were specifically derived for use with young people. It has been demonstrated that the use of the different cut points can lead to vastly different descriptions of children's PA (Guinhouya et al., 2006). An additional problem with cut points is that there is no universal consensus about which exact activities correspond to the intensities PA is classified into. The intensity of an activity is usually described in terms of the *metabolic equivalent of task* (MET; e.g., Harrell et al., 2005). One MET is defined as the metabolic rate while sitting, called the *resting metabolic rate* (e.g., Harrell et al., 2005). Extensive research has been conducted to identify the energy expenditures of different activities in children (e.g., Harrell et al., 2005; Ridley, Ainsworth, & Olds, 2008). The results, however, are not always in agreement (cf. Janssen et al., 2013). For example, Harrell and colleagues (2005) found the MET for slow walking to be 3, whereas Ridley et al. (2008) found it to be between 2.9 and 3.6. At the same time, the cut points for important distinctions between different intensities of PA fall right in this area. For example, light PA is defined as an activity with less than three METs by Freedson and colleagues (2005), and less than four METs by Trost, Loprinzi, Moore, and Pfeiffer (2011). The definition of running as vigorous PA is clearer: Running is found to have an MET between 7.7 and 9.3, a much higher value than the usually used boundary of 6. It can, however, be argued that fast running should be considered very vigorous PA, defined as anything with an MET higher than 9 (Freedson et al., 2005). Fast walking is found to have METs between 3.8 (Harrell et al., 2005) and 4.6 (Ridley et al., 2008), and should thus be considered to be PA with *moderate* intensity.

Fourth, even if the most appropriate cut points for a given population were known, the same cut points would have to be used for each individual of this population. It has been stated that motion patterns differ considerably between individuals (Bussmann, Ebner-Priemer, & Fahrenberg, 2009), and even more so in children (Corder, Ekelund, Steele, Wareham, & Brage, 2008). A general model for all subjects does not account for these individual differences and can therefore be assumed to lack the necessary flexibility. The deduction of measures of PA from raw acceleration data might benefit highly from techniques that are individually tailored to each subject of a study (Bussmann et al., 2009; Foerster & Fahrenberg, 2000).

Despite the awareness of these problems, in the past, researchers had little choice but to use counts and general cut

points, due to a lack of alternatives. Only with recent developments in accelerometers has it become possible to get raw acceleration data as output. This enables researchers to work independently of derived counts and gives them the opportunity to take full control over the data and their analyses (cf. Peach, van Hoomissen, & Callender, 2014). It is thus not surprising that using raw acceleration data for investigating PA has been strongly recommended (e.g., Corder et al., 2008; Freedson, Bowles, Troiano, & Haskell, 2012; Freedson et al., 2005; John & Freedson, 2012; Wijndaele et al., 2015). The potential to take full control over all analysis steps, from the raw acceleration data to a measure of PA, also entails the obligation to find the optimal means to do so. An important part of this is the development of algorithms to transform raw data into measures of activity individually for each subject. This can be done with the use of *reference measurements*.

Reference-pattern-based classification (e.g., Foerster & Fahrenberg, 2000) does provide a way to deal with the diversity of motions of different individuals. For reference-pattern-based classification, accelerometer data are connected to different types of activities individually for every single subject. To do that, reference measurements have to be conducted, much like the aforementioned studies to determine cut points. The difference is that data are not collected with an independent sample, but with the same one that is to be investigated in the actual study. Subjects wear accelerometers while conducting a number of predefined activities. From these measurements, it is possible to find connections between the data and different (intensities of) activities. In principle, different methods are suited for classification (for an extensive overview of classification techniques, see Preece et al., 2009). The methods used include *k-nearest neighbor* (e.g., Bao & Intille, 2004; Zhang, Rowlands, Murray, & Hurst, 2012), *hidden Markov models* (e.g., Pober, Staudenmayer, Raphael, & Freedson, 2006), and *artificial neural networks* (e.g., Hagenbuchner, Cliff, Trost, van Tuc, & Peoples, 2015; Staudenmayer, Pober, Crouter, Bassett, & Freedson, 2009). The “state-of-the-art” method that has proven effective for such tasks is the use of *support vector machines* (SVMs; e.g., He & Jin, 2009). This is the method we used to classify the acceleration data in the present empirical study.

Despite all the advantages of conducting one’s own reference measurements with each subject of a study, they do have one major disadvantage: They are time-consuming, and can thus also be rather expensive in studies with large samples. Every subject has to be supervised individually and guided to conduct the activities according to the protocol. In the present study, reference-pattern-based classification was conducted for children. This made it possible to try out a—to our knowledge—entirely new approach in which the reference data were collected for whole school classes of children at a time, during a single physical education lesson. If this approach turns out to

be successful, it would create great opportunities for the measurement of the PA of children in larger studies. The possible sample size would be much less restricted by a limited time for the reference measurements. In our study, we were able to conduct the reference measurements for 70 children in a total time of only four and a half hours. Of course, this approach also bears potential problems regarding the accuracy of the reference measurements. When conducting these measurements with only one person at a time, it is relatively simple to make sure that the activities are executed as required. It is much harder to exert this control over a whole class of 20 to 30 children. Consequently, it would not be surprising, should the acquired data be less precise than data acquired in individual reference measurements. The main goal of the present study was to investigate whether data collected in this way would still allow for classification with reasonable precision. Additionally, the generalizability of the estimated models is of interest. *Generalizability* denotes the ability of a model to predict activities from the data of different children. This is important when some children miss the physical education lesson in which the reference measures are conducted. A good generalizability of the models would allow for a classification even for the data of the children for whom no reference data are available.

The usefulness of the approach presented here can be assessed better when the obtained results are compared to those that would be obtained if data were analyzed with the traditional approach of using counts and cut points. Thus, we also translated the data into counts and classified them on the basis of the cut points provided by the ActiLife software (Wyatt, 2012, Version 6.9.0). These cut points are based on the recommendations from calibration studies, specifically for children (Butte et al., 2012; Evenson et al., 2008; Freedson et al., 2005; Mattocks et al., 2007; Pulsford et al., 2011; Puyau et al., 2002). Most research that uses accelerometers to measure PA in children uses one of these sets of cut points.

We expected our approach to yield better results than a classification based on counts and commonly used cut points, for two reasons. First, transforming raw acceleration data into counts inevitably leads to a loss of information. In the case that this information is more than just random noise, its loss should lead to less accurate classifications. Second, commonly used cut points are always derived for, and applied to, a whole sample and not individual subjects. Analyzing the data of single subjects, thereby taking individual differences in movements into account, should also lead to better results.

Most importantly, we expected our approach to conducting reference measurements in groups to yield a good trade-off between classification accuracy and time-efficiency. This means that, given the inherent time-efficiency of our approach, it should still allow for a reasonably accurate classification of accelerometer data, thereby being a feasible alternative to using predefined cut points.

Materials and method

Design and subjects

The present study was part of the FLUX (Assessment of Cognitive Performance FLUctuations in the School ConteXt) project, which aims at investigating daily fluctuations in children's cognitive performance in the school context and their potential correlates. A total of 110 children received a smartphone on which they solved working memory tasks and answered self-report questionnaires several times a day for four weeks. Additionally, 80 of these children wore accelerometers for the time of the study, which was approved by a local ethics committee. For their participation, the subjects received a reward in the form of money or a gift coupon. The present article reports the results of PA reference measurements conducted during the pretesting of the FLUX study. Out of the 80 children with an ActiGraph, 70 (43 boys and 27 girls) were present in the physical education lesson during which the reference measurements were conducted. The children were in third or fourth grade, with their ages ranging from 8 to 11 years ($M = 9.77$, $SD = 0.62$ years). In total, three third grade and three fourth grade classes took part in the study. The average size of the classes was 22.3 ($SD = 1.4$). All children in the classes performed the activities for the reference measurements, but only those whose parents had signed an informed consent wore an ActiGraph at the time. On average, 11.7 ($SD = 2.6$) children per class wore an ActiGraph attached to the waist on their nondominant side. The measurements were done during a physical education lesson by three trained staff members. Each lesson lasted 45 min, resulting in a total measurement time of 4.5 h in six classes.

Accelerometers

In the present study, we used the ActiGraph GT3X+ (ActiGraph, LLC, Fort Walton Beach, FL). The GT3X+ is a triaxial accelerometer that measures acceleration in a range from -6 to $+6$ g. As is usual for children, the devices were worn on the waist (Strath et al., 2012). The sampling rate was set to 30 Hz. This sampling rate was chosen because the proposed method was developed for use in a long-term study. A higher sampling rate would require too much maintenance and coordination with the subjects, due to limited memory space and battery life.

Reference measurements

We based the decision of which activities should be conducted in the reference measurements on the protocol recommended by Foerster and Fahrenberg (2000). This protocol was adopted according to specific requirements and limitations of our study design, ensuing from the fact that the reference measurements

were done for all children simultaneously and during a physical education lesson. For example, climbing stairs could not be included, due to a lack of stairs in the gym of the school. Taking these limitations into account, while trying to stay as close as possible to the protocol of Foerster and Fahrenberg (2000), we came up with a protocol of six activities: lying down, sitting, standing, slow walking, fast walking, and running.

Ideally, each of these activities should be conducted for 90 s. The only exception was running. Since it may be hard for children to run for 90 s straight, they were allowed to stop whenever they felt they could not keep their pace much longer. In that case, the children were instructed to run to the middle of the gym and high-five one of the research assistants, at which point the time was stopped for that child. This assured the availability of detailed information about the running time of each child. The time was taken from a stopwatch, the starting time of which was synchronized with the starting time of the ActiGraph measurements. Because we carefully recorded the start and stop times of each activity, it was later possible to assign the collected reference data to the activities.

Nonwear time

The described approach has the problem that it forces the acceleration data to be classified into one of the activity categories of the reference measurements. Periods of time in which the accelerometer was not worn can thus not be identified, which is a serious problem when one wants to take this approach to real-life studies. We addressed this problem by introducing the accelerometer data of not-worn devices into the analyses. A recording ActiGraph was placed on a table in three different positions for 90 s each. Those three positions were the most likely positions for a device to lie on a flat surface, due to the shape of the device. The data recorded in this time were added to the classification. In this way, we attempted to find the patterns of acceleration on the different axes that would indicate that a device was not worn.

Feature extraction

Probably just as important as the classification technique was the proper preprocessing of the data. Usually, data are collapsed into time frames, and certain key values are extracted per frame for the classification of the data. We classified the data in nonoverlapping frames of 2.5 s. With a sampling rate of 30 Hz, this meant that each frame consisted of 75 acceleration measurements on each of the three axes. This information was further processed on the basis of techniques that have proven successful in previous studies (e.g., Bao & Intille, 2004; Ravi, Dandekar, Mysore, & Littman, 2005; Wu, Pan, Zhang, Qi, & Li, 2009). More specifically, for each of the three axes, the mean, variance, and energy, as well as the

correlations between the raw acceleration data of the three axes, were computed. We did not preprocess (e.g., square or take absolute values) the mean acceleration values to take out influences of gravity. Information about how gravity works on the different axes allows for inferences on the inclination of the device, and thus the posture of a person. This information is useful for distinguishing between the different sedentary activities—that is, lying down, sitting, and standing—as well as recognizing nonwear times. Different activities differ in the magnitudes of the acceleration values, and thus in their possible range. To distinguish between physical activities, the variance of the acceleration value is also useful. However, there can be an overlap in the variance values if a person is walking or running slowly (Chung, Purwar, & Sharma, 2008). In this case, additional information that can help distinguish between activities is the energy in the frequency domain. To calculate the energy, the data first have to be Fourier-transformed, which was done by a discrete *fast Fourier transformation* (FFT). Energy (i.e., spectral energy; cf. Bao & Intille, 2004; Murugappan, Murugappan, & Gerard; 2014) is calculated as the sum of the squared discrete FFT component magnitudes of the signal over the entire frequency range. Furthermore, the sum is divided by the length of the frame for normalization. The correlation between the acceleration data of the axes can further improve the recognition of activities (Bao & Intille, 2004). With these four features for each axis, a total of 12 values described the data in each frame. These 12 features were used to train the SVMs on the reference data and, accordingly, to predict activities from the data.

Support vector machines

SVMs were introduced by Boser, Guyon, and Vapnik (1992) and further developed to the version used here by Cortes and Vapnik (1995). SVMs can be used for classification based on high-dimensional data and to find patterns in nonlinear information. They are thus perfectly suited for the present task to categorize acceleration data into the activities in the reference measures. Since SVMs were developed for binary classification—that is, for distinctions between two classes—a one-against-one classification method was used. In this method, every possible pair of activities was compared and the data were classified as reflecting the activity that was chosen in most of these comparisons. It has been shown that this method produces good results for practical use in categorizing data into multiple classes (Hsu & Lin, 2002). SVMs separate different classes of data by a hyperplane. This is done by maximizing the margin between the closest points of the classes (see Fig. 1), which are called *support vectors*. The middle of the margin is the hyperplane that optimally separates the two classes. The problem with data like these is that they are often not linearly separable. SVMs circumvent this problem by projecting the data points into a higher-dimensional space, in

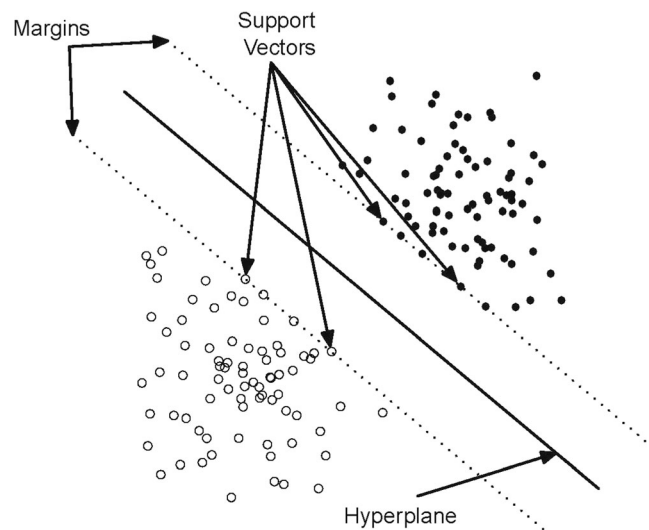


Fig. 1 Schematic representation of support vector machine (SVM) classifications in two dimensions. The conceptually most important aspects of SVMs are labeled (see the text for details)

which the points become linearly separable, corresponding to a nonlinear separation in the observed data space. This makes SVMs very flexible, and thus well-suited for very complex, nonlinear classification problems.

We conducted SVMs with the “svm” function from the e1071 package (Version 1.6-7; Dimitriadou, Hornik, Leisch, Meyer, & Weingessel, 2005) for the open-source statistical software R (R Development Core Team, 2012). Regarding the specific settings of the function, we followed the recommendations of the authors (Meyer, 2012). This means that we used C -classification with the radial basis function kernel. The C and γ parameters were determined using a grid search over all reasonable parameters, which was automatically done by the “tune.svm” function. The so-called soft-margin parameter C determines the punishment for data points on the wrong side of the hyperplane. This is important for cases in which no hyperplane can be found that separates all cases of the classes. Possible values for C ranged from 1 to 1,000, whereas possible values for γ ranged from .0001 to 10. The γ parameter determines the flexibility of the SVM in fitting the data. We constrained this parameter to be no larger than 10 to prevent overfitting (Ben-Hur & Weston, 2010). For each subject, a unique pair of C and γ was obtained. The “svm” function automatically normalizes the input data, which is important to obtain a reasonable accuracy in the classification (Ben-Hur & Weston, 2010).

Individual versus general models

An SVM model was estimated for each child individually, using tenfold cross-validation. In this process, a model is trained on 90 % of the data of a child in order to predict the activities corresponding to the remaining 10 % of the data.

This process is repeated ten times, so that each data point (i.e., frame of 2.5 s) is predicted from training once. The classification accuracy, defined as the percentage of correctly classified frames, is calculated for each model and averaged over the ten repetitions. Additionally, the generalizability of the reference data to different children was assessed with a leave-one-out cross-validation. This means that, for each child, an SVM model was trained with the combined data of all other children. This model was then used to classify the activities of the child whose data were left out in the training process. Again, the percentage of correct classifications was used as a measure of accuracy.

Comparison to ActiLife results

The classification accuracies that were obtained in the described way were then compared to the classification accuracies of the standard procedure (i.e., translating the data into counts and classifying them with the use of cut points). Counts were calculated for 1-s intervals. The cut points we used were the ones that are recommended by the manufacturer of the ActiGraph devices (Butte et al., 2014; Evenson et al., 2008; Freedson et al., 2005; Mattocks et al., 2007; Pulsford et al., 2011; Puyau et al., 2002) and are readily available in the ActiLife software. Existing cut points for children usually classify activities as either “sedentary,” “light,” “moderate,” or “vigorous” (Butte et al., 2014; Evenson et al., 2008; Mattocks et al., 2007; Pulsford et al., 2011; Puyau et al., 2002), and optionally with an additional “very vigorous” category (Freedson et al., 2005).

For the reasons discussed in the introduction, it is difficult to assign the activities used (walking slow, walking fast, running) to the different categories (light, moderate, vigorous) in a definitive way (cf. Janssen et al., 2013). Most importantly, whether slow walking should be considered light or moderate PA cannot be definitively decided, because different studies have found different MET values for this activity. To guarantee a fair comparison, analyses that include a distinction between different (intensities of) PAs were conducted twice (i.e., once with slow walking considered as light PA, and once with it considered as moderate PA).

Research describing the activities of subjects or relating the amounts of activity to other variables usually describes PA in terms of the time that is spent in MVPA. In line with this, the WHO expresses recommendations about the amount of PA in terms of time per day that should be spent in MVPA. Accordingly, we first classified the data into time spent in MVPA and inactive time (including sedentary behavior and nonwear time), using the cut points that are provided by the ActiLife software (i.e., Comparison 1). The accuracy of these classifications was then compared to the accuracy of our approach based on using SVMs to classify raw data. This was done twice. In one analysis, slow walking was considered

moderate PA, and thus part of MVPA, and in another analysis it was considered *light* PA, and thus not part of MVPA.

The measurement accuracy of ActiGraphs does also allow for a classification of activities that is more detailed than just the distinction between sedentary behavior and MVPA. For the next comparison (2), the data were classified as either *sedentary behavior* (or *inactivity*), *light PA*, *moderate PA*, or *vigorous PA*. As before, *slow walking* was considered light PA in one analysis and moderate PA in the other. This means that in the second analysis, there was no light PA category. Since these analyses were only concerned with the ability to distinguish between different activities, all nonactive categories were considered as one.

Another advantage of the ActiGraph is its ability to distinguish between different postures, and thus between different types of sedentary (or inactive) behavior. Specifically, the angle of the device is calculated from the acceleration values on the three axes. From this, the posture of the person wearing the device is deduced. The ActiLife software is able to differentiate between standing, sitting, lying down, and nonwear—that is, times when the device is not worn. This feature is particularly important for real-life research, because it allows for an assessment of the times when an inactive device is not due to an inactive subject, but rather to the device not being worn. The next analysis (3) thus addressed the ability to discriminate between the three inactive behaviors (i.e., standing, sitting, and lying down) and nonwear time. Whether standing should be seen as sedentary behavior is still up for debate. Since the MET of standing quietly is slightly higher than that of sitting or lying down (1.5 vs. 1.2–1.4; Ridley et al., 2008), it can be argued that standing should not be considered *sedentary*, but rather *inactive behavior*. However, since this distinction was not reflected in the cut points used here, we considered standing as sedentary behavior.

For all intents and purposes, the analyses described can give a good insight into the performance differences of the two analysis techniques. Practically relevant distinctions are covered, the main ones being the distinctions between different (intensities of) activity and between nonactive behavior and times that the device is not worn. For the sake of completeness, the abilities to distinguish between all categories (lying down, sitting, standing, walking slowly, walking fast, running, and nonwear) were finally compared between the two analysis techniques (Comparison 4). It is important to keep in mind that the SVMs can differentiate between nine categories, since nonwear data were recorded in three different ways, but in the classical approach using the angle of the device, only one nonwear category can be identified. This means that differentiation can only be done between a total of seven categories.

For any performance differences found in the previous analyses, one might argue that they are not due to a difference in the analysis techniques, or to the fact that transferring raw

data into counts reduces the amount of available information, but rather to the fact that the same cut points are used for the whole sample, whereas each subject has a distinct SVM model. If this was the reason for performance differences, they could be eliminated by calculating a separate SVM for each subject, but using counts instead of the raw data. This is what we did in the last analysis. Since counts can only be used to differentiate between activities and not between postures, only the three active categories were considered in this analysis. For better results, the classification was done with the counts from the x -axis, from all three axes, and from vector magnitudes, respectively. Any remaining performance differences in this comparison would strongly indicate the usefulness of using raw data instead of counts.

Results

Classification with raw data and SVMs

The classification of physical activities achieved high accuracy. Table 1 displays the activities predicted by the individual models against the actual activities. The table shows the sums of all classifications of all 70 children. With individual models for each child, the data could be classified into nine categories with an average accuracy of 96.9 %; with the general models, the data of each child could be classified into nine categories with an average accuracy of 87.5 % (see Table 2).

It can be seen that even misclassified activities usually resemble the actual activity. Most misclassifications happened between slow and fast walking. Almost never did the predicted activity deviate strongly from the actual activity. For example, lying down and sitting were only very rarely classified as walking or running. Since there were virtually no misclassifications of the nonwear time, collapsing all three nonwear time categories into one did not have a noticeable influence on these results.

Comparison to ActiLife classifications

The proposed method for classifying physical activity proposed in this article was compared to the traditional approach (i.e., converting the raw data into counts and classifying them with predefined cut points). As we explained before, this comparison should not be done in one step. To achieve a fair and thorough comparison, several steps, as detailed in the Method section, are necessary.

It can be seen in Table 2 that the proposed method outperforms the traditional one on all comparisons. The difference in accuracy was largest when the data were classified into all available categories (4). Our achieved classification accuracy of 96.9 % was much higher than the accuracies of approximately 57 % to 62 % that result from classifications with the traditional approach. From Comparisons 2 and 3, we can see that this discrepancy was due to two reasons. First, our approach was able to differentiate between activities with an accuracy of 95.9 %, and thus was better at distinguishing activities than were counts and cut points (2). This was true regardless of whether slow walking was considered light PA (accuracy around 80 % to 84 %) or moderate PA (accuracy around 87 % to 91 %). Second, our approach could distinguish active behavior, inactive behavior, and nonwear time (3) with an accuracy of 96.3 %. Thus, it was better able to tell inactive behavior from nonwear time than were inferences drawn from the inclination of the device (accuracy around 82 % to 83 %). Since all active behaviors were treated as one, this difference was not due to differences in the abilities to tell apart activities. Rather, it was either due to differing abilities to tell inactive behavior from nonwear time, or differing abilities to generally distinguish activity from nonactivity. From Comparison 1, we can see that the general abilities to distinguish activity from nonactivity did not differ greatly between the two techniques. This means that the differences in Comparison 3 were due to differences in the ability to tell inactive behavior from nonwear time, which is a

Table 1 Predicted and actual activities, summed over all 70 subjects: Predictions by support vector machines, based on raw data

		True Activities								
		Sit	Stand	Lie	Walk	Fast walk	Run	Nonwear 1	Nonwear 2	Nonwear 3
Predicted activities	Sit	2,274	6	0	1	0	2	0	0	0
	Stand	35	2,299	0	14	2	2	0	0	0
	Lie	1	0	2,304	0	0	0	0	0	0
	Walk	0	4	0	2,256	138	10	0	0	0
	Fast walk	0	1	0	38	2,159	18	0	0	0
	Run	0	0	0	1	11	2,062	0	0	0
	Nonwear 1	0	0	2	0	0	0	2,520	0	0
	Nonwear 2	0	0	4	0	0	0	0	2,520	0
	Nonwear 3	0	0	0	0	0	0	0	0	2,520

Table 2 Accuracies of activity predictions by raw data and SVMs and by counts and cut points

		Raw Data and SVMs		Count and Cut-Point Data From						
		Individual Models	General Models	Freedson	Puyau	Mattocks	Evenson	Pulsford	Butte (x-Axis)	Butte (Vector Magnitude)
(4) All categories	Slow walk as light PA	96.9	87.5	59.1	56.9	60.0	61.6	61.4	61.4	61.9
	Slow walk as moderate PA			67.9	64.5	68.0	67.9	67.6	67.0	68.9
(1) PA vs. inactivity		98.3		96.2	95.9	96.1	96.1	96.1	94.5	96.2
(2) Different activities and inactivity	Slow walk as light PA	95.9		81.4	79.7	82.3	83.9	83.6	83.0	84.2
	Slow walk as moderate PA			90.2	87.3	90.3	90.1	89.8	88.6	91.2
(3) PA vs. nonwear vs. sedentary		96.3		83.4	83.1	83.3	83.3	83.3	81.8	83.4

All numbers are the percentages of correctly classified frames.

very important distinction when real-life accelerometer data are to be analyzed.

Table 3 shows the accuracies of classifying counts with SVMs, as compared to the accuracies of classifying counts with cut points and classifying raw data with SVMs. It has to be kept in mind that this classification only contains a distinction between the three activities (i.e., slow walking, fast walking, and running). It can be seen that SVMs were able to classify counts more accurately than almost all cut points. At the same time, all classifications based on counts were much less accurate than the ones based on raw data. Interestingly, classifications based only on the counts from the *x*-axis were less accurate than both the classifications based on all three axes and the classifications based on vector magnitudes.

Discussion

Overall, the results demonstrate that SVMs can classify raw accelerometer data with great precision. More importantly, our new approach of using group assessment for reference data was time-efficient, and thus avoided a major obstacle for applying reference measurements in large-scale research. The accuracy with which activities could be classified was comparable to, or even higher than, that reported in similar studies using various classification techniques (e.g., 84 % in Bao & Intille, 2004; 89 % in Ermes, Pärkka, Mantyjarvi, & Korhonen, 2008; 96.8 % in Foerster & Fahrenberg, 2000; 73 %–91 % in Hagenbuchner et al., 2015; 70 %–80 % in Pober et al., 2006; >99 % in Ravi et al., 2005; 98.7 % in Zhang et al., 2012). These numbers are meant to give a rough impression of the results of other studies. However, due to methodological differences between the studies, accuracy values should be compared with care. Studies differ in various aspects. For example, in some studies each subject wore multiple devices (e.g., Bao & Intille, 2004; Ermes et al., 2008),

which is an advantage in terms of classification accuracy, but a disadvantage if a method is to be used in real-life research. Other sources of diversity include the performance of a wide range of different activities; different settings in which the measurements were done (e.g., laboratory vs. free-range), durations of the measurements, and numbers of subjects; and the use of many different methods for classification. A good and detailed review was done by Preece and colleagues (2009).

One of the assumptions that underlies all of these studies is that there are better ways to analyze accelerometer data than with the use of counts and cut points. Accordingly, as we mentioned, it has been recommended to use the raw acceleration data (e.g., Corder et al., 2008; Freedson et al., 2012; Freedson et al., 2005; John & Freedson, 2012). However, to the best of our knowledge, until now no study has directly compared the accuracies of classifications based on counts and raw data with the same dataset.¹ The present research contributes to this question by providing empirical evidence for the superiority of raw data over counts.

The results indicated that it is possible to collect reference data from many subjects at the same time. Although we only measured an average of about 12 children at a time, more than 22 children were present during the reference measurements, on average. If more of them had agreed to participate in our study, we could have gathered even more reference data without needing additional time. Furthermore, if reference measurements in groups can be done with children, it should also be possible with adults, since they tend to comply much better with instructions.

¹ Pober et al. (2006) did compare the use of cut points to more advanced methods. However, they used counts, and not raw data, for all analyses. Furthermore, for the analyses with cut points, they only reported an estimate of the overall time that a subject had spent at a certain activity level. As they stated themselves (see p. 1631), this means that different activities might consistently be confused with each other, making it hard to assess the actual accuracy of the classification.

Table 3 Accuracies of activity predictions by raw data and SVMs, counts and SVMs, and counts and cut points

		Raw Data and SVMs	Counts and SVMs			Counts and Cut-Point Data From						
			<i>x</i> -Axis	<i>x</i> -, <i>y</i> -, and <i>z</i> -Axis	Vector Magnitude	Freedson	Puyau	Mattocks	Evenson	Pulsford	Butte (<i>x</i> -Axis)	Butte (Vector Magnitude)
Three activities	Slow walk as light PA	92.7	70.9	79.0	71.2	52.8	46.3	55.8	61.2	60.4	63.8	62.0
	Slow walk as moderate PA					79.2	32.0	34.3	55.3	55.2	54.3	61.5

All numbers are the percentages of correctly classified frames.

In addition to being as accurate as other techniques that have used reference measurements, our approach was more accurate than the traditional approach of converting data to counts and classifying them with predefined cut points. Interestingly, these differences in performance were bigger when the data were classified into many as opposed to a few categories. When one only wants to tell any activity from nonactivity, it does not really matter which technique is used. However, if one wants to achieve a distinction between many different categories, the disadvantages of counts and cut points become more and more pronounced. In other words, if one wants to make full use of the advantages of greatly accurate devices, the use of counts is arguably not appropriate. Today, techniques such as the one presented here are able to utilize the full complexity and precision of modern accelerometer devices.

Some additional details can be inferred from the results. We can see that the movement patterns of children do indeed differ, which is reflected in the fact that the individual models classified activity more accurately than the general models. However, even the general models did a rather satisfactory job. For many applications, the general model is probably accurate enough to be used for classification for subjects that are not present during the reference measures. This makes our approach of measuring many subjects at the same time even more practical for use in larger studies with children, even when their ages show a considerable range. As long as children attend school, they can be measured in large groups at the same time. Also, even children who miss the reference measurements can still participate in a study, due to the reasonable generalizability of the patterns found in the data. These findings make the method employed here very well-suited for large-scale field research.

From Table 1, it can be seen that even misclassified activities are usually still in the right category (i.e., sedentary behavior, PA, or nonwear). For instance, fast walking is sometimes classified as running, whereas the classification of, for instance, lying down as running is scarce. Also, nonwear time was classified correctly in virtually all cases. On those rare occasions when nonwear was not detected, it was classified

as lying down, which is indeed the closest category to not wearing the device. These more detailed considerations further emphasize the usefulness of the chosen approach. To the best of our knowledge, the approach of collecting data that represent nonwear time and including these data in the analyses is new. The fact that nonwear time was correctly classified with great accuracy makes this approach promising for field studies, in which it is important to correctly identify periods in which a device was not worn.

The use of raw data also leads to much better detection of nonwear time than does the use of inclination for deducing posture. The latter method often leads to confusion between standing and nonwear time, as well as between sitting and lying down (see Table 4). The mix-up of standing and nonwear time can be explained by the fact that in one of the recorded nonwear time categories, the device was placed on a table “upright,” thus resembling the position the device would be in when someone who wore it at the waist was standing upright. Although this distinction was mostly lost when only the inclination was considered, the use of raw data made it possible to distinguish between those two perfectly. This is probably due to the fact that the raw data still contain information about slight movements that helps distinguish between a standing person and a not-worn device. An algorithm that only relies on the inclination does not have a way to detect such subtleties in the data. Accordingly, it is less suited for use in real-life research.

In real-life research, information about nonwear time is usually processed further (cf. Choi, Liu, Matthews, & Buchowski, 2011). A period of time is only considered nonwear time if the device is not moved for a certain period of time (e.g., 1 h), allowing for only very short interruptions of this lack of motion. This is done to separate nonwear time from times when a subject is just lying or sitting very still. In doing so, the accuracy of detecting nonwear time is increased. For reference measurements this cannot be done, because only a few minutes of data are available, so all available postprocessing techniques do not apply. This means that it could be argued that our comparison is not completely fair, because it does not take the complete calculation process of

Table 4 Predicted and actual intensities of activities, summed over all 70 subjects: Predictions by inclination with cut points, based on count data

		True Activities						
		Sit	Stand	Lie	Walk	Fast walk	Run	Nonwear
Predicted activities	Sit	3,230	2,125	0	178	38	10	369
	Stand	2,709	2,986	0	108	11	66	0
	Lie	209	80	4,317	512	210	165	387
	Walk	135	75	0	2,554	2,596	520	0
	Fast walk	74	18	1	1,127	3,244	1,416	0
	Run	11	19	0	28	754	4,515	0
	Nonwear	0	6,090	2	93	236	71	12,340

Cut points are taken from Butte et al. (2014), since they yielded the highest accuracy of all discussed sets of cut points.

nonwear time into account. However, we argue that all postprocessing can just as well be done with the information that is obtained by our suggested method. The only difference is the accuracy of the information that stands at the beginning of this process. Arguably, feeding more accurate information into a process will lead to more accurate results.

Comparison 2 gives further insight into possible sources of performance differences between the two methods. First, the classification accuracy rose from around 80 % to 90 % when slow walking was considered moderate instead of light PA. This shows that, for a considerable part, inaccuracies in the classification stem from the uncertainty of how to interpret slow walking. The ambiguity of this interpretation also becomes clear from Table 3. The cut points by Freedson et al. (2005), Puyau et al. (2002), and Mattocks et al. (2007) all led to a classification accuracy of around 50 % when slow walking was considered light PA. However, when it was considered moderate PA, the accuracy of the Freedson et al. (2005) cut points rose to almost 80 %, whereas the accuracy of the Puyau et al. (2002) and Mattocks et al. (2007) cut points fell to almost 30 %. This shows the differing underlying interpretations of slow walking when the cut points were calculated and emphasizes the huge impact that the choice for a certain set of cut points can have on the interpretation of accelerometer data. Our approach does not have the same problem, since slow walking will always be recognized as slow walking. Second, when interpreting slow walking as moderate PA, the accuracies of the classifications with cut points come close to the accuracy of the SVM. This shows that cut points are much less problematic when it comes to distinguishing between moderate and vigorous PA.

For most comparisons we have discussed, it could be argued that performance differences were due to the fact that our approach uses an individual model for each subject, whereas cut points are the same for all subjects. Behind the tackled question of whether it is better to use SVMs to classify raw data or cut points to classify counts lie two more basic questions. First, is it important to use an individual model for each

subject, rather than general cut points? Second, is it generally better to use raw data rather than counts? Two results relate to these questions. First, the general SVM models that were trained on the data of all children but one and were used to predict the activities of the remaining child could still classify the activities with an accuracy of almost 90 %, and were thus much more accurate than the use of cut points. This shows that, whereas individual models were always more accurate, the use of one and the same model for all subjects can still lead to satisfactory results. Second, as Table 3 shows, the individual SVM models using counts to differentiate between activities reached an accuracy between those of the other two approaches. This shows that with individual models, the use of counts can yield higher accuracies than their classification with cut points would suggest. However, the accuracies were still much lower than those obtained by using raw data. This shows that the transformation of raw data into counts eliminates meaningful information that can help to correctly classify accelerometer data. Using the counts of only the *x*-axis led to the least accurate results (cf. Howe, Staudenmayer, & Freedson, 2009). Interestingly, this is exactly the information that the great majority of cut points rely on. In sum, we can infer that the performance differences we have found are not only due to the fact that we used an individual model for each child, but also to the fact that we used raw data rather than counts.

It should be kept in mind that, on their own, the results presented here are somewhat limited by the fact that all data were gathered in the same setting (i.e., a physical education lesson at school). It would be useful to further validate the described method in a different study. This can be done in different ways. Data could be gathered in a range of settings (e.g., laboratory, school, home). The accuracy of the classifications could then be assessed by comparing them to either a carefully made protocol by the subjects or, given the problems with self-report measures of activity (e.g., Baumeister, Vohs, & Funder, 2009; Ebner-Priemer et al., 2006), a protocol by a researcher who is present in these settings. Although such a

study could give strong proof for the accuracy of the proposed method, it would require a very high time investment and/or very compliant subjects. Another, more straightforward way to validate the proposed method would be to assess its face validity when applied to real-life data. Although this does not allow for a precise assessment of its accuracy, it does provide information on its usability. In line with this, we were already able to successfully apply the described method to the real-life data collected in the FLUX study (see Kühnhausen, Leonhardt, Dirk, & Schmiedek, 2013), indicating that it does indeed generalize to, and is useable in, real-life research.

In conclusion, we have shown that researchers should seriously consider conducting reference measurements prior to an actual study. They are a requirement for obtaining individual models that allow for a much more accurate recognition of physical activities than existing general cut points do. Furthermore, reference measurements enable researchers to work independently from counts, which have proven to be inferior to raw data when it comes to differentiating between activities. Most importantly, we have shown that the high costs that have been associated with reference measurements so far do not have to be a reason not to conduct them. Reference measurements can be done in groups with relatively little effort, while still allowing for very accurate classifications of accelerometer data.

Author note This research was part of the FLUX project at the IDEa Center in Frankfurt am Main, Germany, which is funded by the Hessian Initiative for the Development of Scientific and Economic Excellence (LOEWE). We owe special thanks to Verena Diel, Tanja Koenen, Anja Leonhardt, Philipp Wiesemann, and a team of highly committed student assistants for their important roles in conducting the FLUX project. The authors declare that they have no competing or potential conflicts of interest.

References

- Ahn, S., & Fedewa, A. L. (2011). A meta-analysis of the relationship between children's physical activity and mental health. *Journal of Pediatric Psychology, 36*, 385–397. doi:10.1093/jpepsy/jsq107
- Aznar, S., Naylor, P. J., Silva, P., Pérez, M., Angulo, T., Laguna, M., & López-Chicharro, J. (2011). Patterns of physical activity in Spanish children: A descriptive pilot study. *Child: Care, Health and Development, 37*, 322–328. doi:10.1111/j.1365-2214.2010.01175.x
- Bao, L., & Intille, S. S. (2004). Activity recognition from user-annotated acceleration data. In T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, O. Nierstrasz, ... M. Y. Vardi (Eds.), *Proceedings of Pervasive Computing: Second International Conference, PERVASIVE 2004, Linz/Vienna, Austria, April 2004* (pp. 1–17). Berlin, Germany: Springer.
- Basterfield, L., Adamson, A. J., Frary, J. K., Parkinson, K. N., Pearce, M. S., & Reilly, J. J. (2011). Longitudinal study of physical activity and sedentary behavior in children. *Pediatrics, 127*, e24–e30. doi:10.1542/peds.2010-1935
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2009). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research. Purdue Symposium on Psychological Sciences* (pp. 12–27). Oxford, UK: Oxford University Press.
- Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. *Methods in Molecular Biology, 609*, 223–239. doi:10.1007/978-1-60327-241-4_13
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory COLT '92* (pp. 144–152). Pittsburgh, PA: ACM Press.
- Bussmann, J. B., Ebner-Priemer, U. W., & Fahrenberg, J. (2009). Ambulatory activity monitoring. *European Psychologist, 14*, 142–152. doi:10.1027/1016-9040.14.2.142
- Butte, N. F., Ekelund, U., & Westerterp, K. R. (2012). Assessing physical activity using wearable monitors: Measures of physical activity. *Medicine and Science in Sports and Exercise, 44*, S5–S12. doi:10.1249/MSS.0b013e3182399c0e
- Butte, N. F., Wong, W. W., Lee, J. S., Adolph, A. L., Puyau, M. R., & Zakeri, I. F. (2014). Prediction of energy expenditure and physical activity in preschoolers. *Medicine and Science in Sports and Exercise, 46*, 1216–1226. doi:10.1249/MSS.0000000000000209
- Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometer-based activity monitors: Current and future. *Medicine & Science in Sports & Exercise, 37*, S490. doi:10.1249/01.mss.0000185571.49104.82
- Choi, L., Liu, Z., Matthews, C. E., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and Science in Sports and Exercise, 43*, 357–364. doi:10.1249/MSS.0b013e3181ed61a3
- Chung, W.-Y., Purwar, A., & Sharma, A. (2008). Frequency domain approach for activity classification using accelerometer. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008*, 1120–1123. doi:10.1109/IEMBS.2008.4649357
- Corder, K., Ekelund, U., Steele, R. M., Wareham, N. J., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology, 105*, 977–987. doi:10.1152/jappphysiol.00094.2008
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*, 273–297. doi:10.1007/BF00994018
- Development Core Team, R. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2005). e1071: Misc functions of the Department of Statistics. Retrieved from <http://cran.r-project.org/web/packages/e1071/index.html>
- Ebner-Priemer, U. W., Kuo, J., Welch, S. S., Thielgen, T., Witte, S., Bohus, M., & Linehan, M. M. (2006). A valence-dependent group-specific recall bias of retrospective self-reports: A study of borderline personality disorder in everyday life. *Journal of Nervous and Mental Disease, 194*, 774–779. doi:10.1097/01.nmd.0000239900.46595.72
- Ermes, M., Pärkka, J., Mantyjarvi, J., & Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE Transactions on Information Technology in Biomedicine, 12*, 20–26. doi:10.1109/TITB.2007.899496
- Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal of Sports Sciences, 26*, 1557–1565. doi:10.1080/02640410802334196
- Foerster, F., & Fahrenberg, J. (2000). Motion pattern and posture: Correctly assessed by calibrated accelerometers. *Behavior*

- Research Methods, Instruments, & Computers*, 32, 450–457. doi:10.3758/BF03200815
- Freedson, P., Bowles, H. R., Troiano, R., & Haskell, W. (2012). Assessment of physical activity using wearable monitors: Recommendations for monitor calibration and use in the field. *Medicine & Science in Sports & Exercise*, 44(Suppl. 1), 1–4. doi:10.1249/MSS.0b013e3182399b7e
- Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise*, 37, 523. doi:10.1249/01.mss.0000185658.28284.ba
- Guinhouya, C. B., Hubert, H., Soubrier, S., Vilhelm, C., Lemdani, M., & Durocher, A. (2006). Moderate-to-vigorous physical activity among children: Discrepancies in accelerometry-based cut-off points. *Obesity*, 14, 774–777. doi:10.1038/oby.2006.89
- Hagenbuchner, M., Cliff, D. P., Trost, S. G., van Tuc, N., & Peoples, G. E. (2015). Prediction of activity type in preschool children using machine learning techniques. *Journal of Science and Medicine in Sport*, 18, 426–431. doi:10.1016/j.jsams.2014.06.003
- Harrell, J. S., McMurray, R. G., Baggett, C. D., Pennell, M. L., Pearce, P. F., & Bangdiwala, S. I. (2005). Energy costs of physical activities in children and adolescents. *Medicine & Science in Sports & Exercise*, 37, 329–336. doi:10.1249/01.MSS.0000153115.33762.3F
- He, Z., & Jin, L. (2009). Activity recognition from acceleration data based on discrete cosine transform and SVM. In *Proceedings of the 2009 I.E. International Conference on Systems, Man and Cybernetics* (pp. 5041–5044). Piscataway, NJ: IEEE Press. doi:10.1109/ICSMC.2009.5346042
- Howe, C. A., Staudenmayer, J. W., & Freedson, P. S. (2009). Accelerometer prediction of energy expenditure: Vector magnitude versus vertical axis. *Medicine and Science in Sports and Exercise*, 41, 2199–2206. doi:10.1249/MSS.0b013e3181aa3a0e
- Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13, 415–425. doi:10.1109/72.991427
- Janssen, X., Cliff, D. P., Reilly, J. J., Hinkley, T., Jones, R. A., Batterham, M., & Okely, A. D. (2013). Predictive validity and classification accuracy of ActiGraph energy expenditure equations and cut-points in young children. *PLoS ONE*, 8, e79124. doi:10.1371/journal.pone.0079124
- John, D., & Freedson, P. (2012). ActiGraph and actual physical activity monitors: A peek under the hood. *Medicine and Science in Sports and Exercise*, 44, S86–S89. doi:10.1249/MSS.0b013e3182399f5e
- John, D., Miller, R., Kozey-Keadle, S., Caldwell, G., & Freedson, P. (2012). Biomechanical examination of the “plateau phenomenon” in ActiGraph vertical activity counts. *Physiological Measurement*, 33, 219–230. doi:10.1088/0967-3334/33/2/219
- King, A. C., Parkinson, K. N., Adamson, A. J., Murray, L., Besson, H., Reilly, J. J., & Basterfield, L. (2011). Correlates of objectively measured physical activity and sedentary behaviour in English children. *European Journal of Public Health*, 21, 424–431. doi:10.1093/eurpub/ckq104
- Kühnhausen, J., Leonhardt, A., Dirk, J., & Schmiedek, F. (2013). Physical activity and affect in elementary school children’s daily lives. *Frontiers in Psychology*, 4(456), 1–8. doi:10.3389/fpsyg.2013.00456
- Lee, I.-M., & Shiroma, E. J. (2014). Using accelerometers to measure physical activity in large-scale epidemiological studies: Issues and challenges. *British Journal of Sports Medicine*, 48, 197–201. doi:10.1136/bjsports-2013-093154
- Mattocks, C., Leary, S., Ness, A., Deere, K., Saunders, J., Tilling, K., & Riddoche, C. (2007). Calibration of an accelerometer during free-living activities in children. *International Journal of Pediatric Obesity*, 2, 218–226. doi:10.1080/17477160701408809
- Meyer, D. (2012). Support vector machines: The interface to libsvm in package e1071. Retrieved from <http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>
- Murugappan, M., Murugappan, S., & Gerard, C. (2014). Wireless EEG signals based neuromarketing system using Fast Fourier Transform (FFT). In *IEEE 10th International Colloquium on Signal Processing and Its Applications (CSPA)* (pp. 25–30). Piscataway, NJ: IEEE Press. doi:10.1109/CSPA.2014.6805714
- Pasco, J. A., Jacka, F. N., Williams, L. J., Brennan, S. L., Leslie, E., & Berk, M. (2011). Don’t worry, be active: Positive affect and habitual physical activity. *Australian and New Zealand Journal of Psychiatry*, 45, 1047–1052. doi:10.3109/00048674.2011.621063
- Peach, D., Van Hoomissen, J., & Callender, H. L. (2014). Exploring the ActiLife® filtration algorithm: Converting raw acceleration data to counts. *Physiological Measurement*, 35, 2359–2367. doi:10.1088/0967-3334/35/12/2359
- Ploughman, M. (2008). Exercise is brain food: The effects of physical activity on cognitive function. *Developmental Neurorehabilitation*, 11, 236–240. doi:10.1080/17518420801997007
- Pober, D. M., Staudenmayer, J., Raphael, C., & Freedson, P. S. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and Science in Sports and Exercise*, 38, 1626–1634. doi:10.1249/01.mss.0000227542.43669.45
- Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., Howard, D., Meijer, K., & Crompton, R. (2009). Activity identification using body-mounted sensors—A review of classification techniques. *Physiological Measurement*, 30, R1–R33. doi:10.1088/0967-3334/30/4/R01
- Pulsford, R. M., Cortina-Borja, M., Rich, C., Kinnafick, F.-E., Dezateux, C., & Griffiths, L. J. (2011). Actigraph accelerometer-defined boundaries for sedentary behaviour and physical activity intensities in 7 year old children. *PLoS ONE*, 6, e21822. doi:10.1371/journal.pone.0021822
- Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obesity Research*, 10, 150–157. doi:10.1038/oby.2002.24
- Ravi, N., Dandekar, N., Mysore, P., & Littman, M. L. (2005). Activity recognition from accelerometer. In American Association for Artificial Intelligence (Ed.), *IAAI '05 Proceedings of the 17th conference on Innovative applications of artificial intelligence* (Vol. 3, pp. 1541–1546). American Association for Artificial Intelligence.
- Reed, J., & Buck, S. (2009). The effect of regular aerobic exercise on positive-activated affect: A meta-analysis. *Psychology of Sport and Exercise*, 10, 581–594. doi:10.1016/j.psychsport.2009.05.009
- Reilly, J. J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J. Y. (2008). Objective measurement of physical activity and sedentary behaviour: Review with new data. *Archives of Disease in Childhood*, 93, 614–619. doi:10.1136/adc.2007.133272
- Ridley, K., Ainsworth, B. E., & Olds, T. S. (2008). Development of a compendium of energy expenditures for youth. *International Journal of Behavioral Nutrition and Physical Activity*, 5, 45. doi:10.1186/1479-5868-5-45
- Rothney, M. P., Apker, G. A., Song, Y., & Chen, K. Y. (2008). Comparing the performance of three generations of ActiGraph accelerometers. *Journal of Applied Physiology*, 105, 1091–1097. doi:10.1152/jappphysiol.90641.2008
- Sherar, L. B., Griew, P., Eslinger, D. W., Cooper, A. R., Ekelund, U., Judge, K., & Riddoch, C. (2011). International children’s accelerometry database (ICAD): Design and methods. *BMC Public Health*, 11, 485. doi:10.1186/1471-2458-11-485
- Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: A meta-analysis. *Pediatric Exercise Science*, 15, 243–256.
- Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., & Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*, 107, 1300–1307. doi:10.1152/jappphysiol.00465.2009
- Strath, S. J., Pfeiffer, K. A., & Whitt-Glover, M. C. (2012). Accelerometer use with children, older adults, and adults with functional

- limitations. *Medicine and Science in Sports and Exercise*, 44, S77–S85. doi:10.1249/MSS.0b013e3182399eb1
- Strong, W. B., Malina, R. M., Blimkie, C. J. R., Daniels, S. R., Dishman, R. K., Gutin, B., & Trudeau, F. (2005). Evidence based physical activity for school-age youth. *Journal of Pediatrics*, 146, 732–737. doi:10.1016/j.jpeds.2005.01.055
- Tomporowski, P. D., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Psychology Review*, 20, 111–131. doi:10.1007/s10648-007-9057-0
- Trost, S. G., Blair, S. N., & Khan, K. M. (2014). Physical inactivity remains the greatest public health problem of the 21st century: Evidence, improved methods and solutions using the “7 investments that work” as a framework. *British Journal of Sports Medicine*, 48, 169–170. doi:10.1136/bjsports-2013-093372
- Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine and Science in Sports and Exercise*, 43, 1360–1368. doi:10.1249/MSS.0b013e318206476e
- Wijndaele, K., Westgate, K., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F., & Granat, M. H. (2015). Utilization and harmonization of adult accelerometry data: Review and expert consensus. *Medicine and Science in Sports and Exercise*, 47, 2129–2139. doi:10.1249/MSS.0000000000000661
- Wong, W. W., Ortiz, C. L., Lathan, D., Moore, L. A., Konzelmann, K. L., Adolph, A. L., & Butte, N. F. (2012). Underserved minority children are not meeting the US public health recommendation for moderate-vigorous physical activity. *Journal of Obesity & Weight Loss Therapy*, 2, 1000132. doi:10.4172/2165-7904.1000132
- Wu, J., Pan, G., Zhang, D., Qi, G., & Li, S. (2009). Gesture recognition with a 3-D accelerometer. In D. Zhang, M. Portmann, A.-H. Tan, & J. Indulska (Eds.), *Proceedings of Ubiquitous Intelligence and Computing: 6th International Conference* (pp. 25–38). Berlin, Germany: Springer.
- Wyatt, J. (2012). *Actilife 6 user's manual*. Pensacola, FL: ActiGraph.
- Zhang, S., Rowlands, A. V., Murray, P., & Hurst, T. L. (2012). Physical activity classification using the GENE wrist-worn accelerometer. *Medicine and Science in Sports and Exercise*, 44, 742–748. doi:10.1249/MSS.0b013e31823bf95c