

Investigating optimal accelerometer placement for energy expenditure prediction in children using a machine learning approach

K A Mackintosh¹, A H K Montoye^{2,3}, K A Pfeiffer⁴ and M A McNarry¹

¹Swansea University, Swansea, Wales, UK, ² Ball State University, Muncie, Indiana, USA, ³Alma College, Alma Michigan, USA, ⁴Michigan State University, East Lansing, Michigan, USA.

Email: k.mackintosh@swansea.ac.uk

Abstract. Accurate measurement of energy expenditure (EE) is imperative for identifying and targeting health-associated implications. Whilst numerous accelerometer-based regression equations to predict EE have been developed, there remains little consensus regarding optimal accelerometer placement. Therefore, the purpose of the present study was to validate and compare artificial neural networks (ANNs) developed from accelerometers worn on various anatomical positions, and combinations thereof, to predict EE.

Twenty-seven children (15 boys; 10.8±1.1 years) participated in an incremental treadmill test and 30-minute exergaming session wearing a portable gas analyser and nine ActiGraph GT3X+ accelerometers (chest and left and right wrists, hips, knees, and ankles). Age and sex-specific resting EE equations (Schofield) were used to estimate METs from the oxygen uptake measures. Using all the data from both exergames, incremental treadmill test and the transition period in between, ANNs were created and tested separately for each accelerometer and for combinations of two or more using a leave-one-out approach to predict EE compared to measured EE. Six features (mean and variance of the three accelerometer axes) were extracted within each 15-second window as inputs in the ANN. Correlations and root mean square error (RMSE) were calculated to evaluate prediction accuracy of each ANN, and repeated measures ANOVA was used to statistically compare accuracy of the ANNs.

All single-accelerator ANNs and combinations of two-, three-, and four-accelerometers performed equally ($r=0.77-0.82$), demonstrating higher correlations than the 9-accelerator ANN ($r=0.69$) or the Freedson linear regression equation ($r=0.75$). RMSE did not differ between single-accelerator ANNs or combination of two, three, or four accelerometers (1.21-1.31

Using machine learning to predict energy expenditure

METs), demonstrating lower RMSEs than the 9-accelerometer ANN (1.46 METs) or Freedson equation (1.74 METs).

These findings provide preliminary evidence that ANNs developed from single accelerometers mounted on various anatomical positions demonstrate equivalency in the accuracy to predict EE in a semi-structured setting, supporting the use of ANNs in improving EE prediction accuracy compared with linear regression.

Keywords: Artificial Neural Network, physical activity measurement, youth, METS, accelerometer, placement

1. Introduction

Higher levels of physical activity (PA) and lower levels of sedentary behaviour are associated with improved physiological and psychosocial health in children (Jansen et al., 2011). However, the majority of children in Europe are not sufficiently active (Riddoch et al., 2007), with only 5.1% estimated to meet current government guidelines of at least 60 minutes daily of moderate-to-vigorous physical activity (MVPA), defined as any activity eliciting an energy expenditure of at least 4.0 metabolic equivalents (METs; Department of Health, 2011). Fundamental to these low PA levels and, pivotally, their resolution, is the accurate measurement of PA and thus prediction of energy expenditure.

Early models of energy prediction generally utilised accelerometer counts as an independent variable to derive linear regression equations (Freedson, Pober, & Janz, 2005), though estimates have shown poorer accuracy during free-living than under laboratory-based conditions (Hendelman, Miller, Bagget, Debold, & Freedson, 2000; Swartz et al., 2000). Furthermore, the applicability of such approaches in children has been questioned given their highly sporadic nature of movement (Bailey et al., 1995; Baquet, Stratton, Van Praagh, & Berthoin, 2007). Significant improvements in energy expenditure prediction have been elicited through the utilisation of novel computational analysis techniques, such as machine learning (Freedson, Lyden, Kozey-Keadle, & Staudenmayer, 2011; Trost, Wong, Pfeiffer, & Zheng, 2012b). Central to the application of these techniques and, thus, the accurate prediction of energy expenditure, is the identification of the optimal accelerometer placement. Specifically, the right hip has conventionally been used as it is close to the centre of mass and has been shown to yield higher accuracy for energy expenditure prediction than other placements when employing traditional regression techniques (Puyau, Adolph, Vohra, & Butte, 2002; Swartz et al., 2000), although other studies have shown contradictory findings (Chandler, Brazendale, Beets, & Mealing, 2016; Crouter, Flynn, & Bassett, 2015). However, recent research utilising more sophisticated machine learning techniques have demonstrated a high accuracy of energy expenditure prediction from accelerometers located at several body locations (Ellis et al., 2014; Montoye, Mudd, Biswas, & Pfeiffer, 2015). Such findings therefore promote the re-evaluation of optimal accelerometer placement sites, especially given participant compliance issues with hip-mounted accelerometers (Colley, Connor Gorber, & Tremblay, 2010; Trost, McIver, & Pate, 2005) and the greater acceptability and user-friendliness of accelerometer placements such as the wrist (Ekblom, Nyberg, Bak, Ekelund, & Marcus, 2012).

In addition to the controversy regarding the optimal position, it has been postulated that amalgamating the data from multiple monitoring devices could improve estimated energy expenditure; accelerometers positioned at different body locations are likely to provide nontrivial, complementary information (He et al., 2014). Although it remains to be elucidated how this information can be combined

to enhance energy expenditure prediction, the use of multiple accelerometers for activity detection remains equivocal. Specifically, whilst He *et al* (2014) reported a modest improvement in the prediction accuracy of specific lifestyle activities when integrating information from three locations (hip and wrists), Cleland *et al* (2013) found no significant improvement in activity detection when two or more accelerometers were used. Additionally, in successive studies, Montoye *et al* (2014) found improved energy expenditure prediction accuracy with a three-piece accelerometer system compared to a hip-mounted accelerometer in a laboratory setting but minimal improvement in a semi-structured setting (Montoye, Dong, Biswas, & Pfeiffer, 2016). Interpretation of these findings with regard to the potential utility of multiple accelerometers may, however, at least in part, be limited by the structured activities incorporated and a failure to consider all possible combinations of accelerometer placements. Indeed, it may be hypothesized that multiple accelerometers would provide greater benefits in terms of energy expenditure prediction within a free-living setting (Cleland *et al.*, 2013); any such enhancements would, however, need to be balanced against the increased participant burden and risk of non-compliance.

Therefore, the purpose of the present study was to investigate children's predicted energy expenditure through the implementation of machine learning methods of analysing accelerometer data in a semi-structured setting. Specifically, the present study sought to validate and compare artificial neural networks (ANNs) developed from data from accelerometers worn on nine anatomical positions, and combinations thereof, and to compare and contrast accuracies to Freedson energy expenditure prediction equations (Freedson *et al.*, 2005).

2. Methods

2.1. Participants

In total, 27 children (15 boys, 10.8 ± 1.0 years) were recruited via a local primary school to participate in this study. All participants were familiar with playing active video games and were asked to attend the laboratory in a rested state, at least two hours postprandial and to have avoided strenuous exercise and caffeine in the preceding 24 hours. The local research ethics committee granted ethical approval for this study; written informed parental consent and participant assent were obtained prior to data collection.

2.2. Procedure

Participants attended the laboratory on one occasion during which anthropometric measures and peak $\dot{V}O_2$ were assessed, in addition to playing exergames, which were selected to represent an optimal trade-off between the sporadic nature of children's movement and a controlled laboratory setting. Specifically,

stature and sitting stature (Holtain, Crymych, Dyfed, UK) were measured to the nearest 0.01 m and body mass (Seca, Hamburg, Germany) to the nearest 0.1 kg. Age was calculated from the date of assessment, and maturity was estimated using the methods described by Mirwald *et al* (2002) and expressed as the estimated time in years from the age at peak height velocity (APHV). Participants subsequently completed two exergames (River Rush and Reflex Ridge; Kinect Adventures!, Xbox 360), in a randomly assigned order, for 15 minutes each. The difficulty levels for the exergames were set by the researchers and standardized throughout testing for consistency. Finally, following a 15-minute rest, participants completed a continuous, incremental treadmill test to volitional exhaustion. Due to the variation in biological age of the participants, the speeds utilized during the test were individually calibrated utilising Froude numbers (Fr), as described by Hopkins *et al* (2010). The protocol required participants to complete 2-minute stages, beginning with a walking speed equivalent to Fr 0.25 and subsequently increasing to the equivalent of Fr 0.5 (Walk/run transition) after which successive increments were determined by the difference in the speed for stages 1 and 2 ($\sim 2 \text{ km}\cdot\text{hr}^{-1}$) until volitional exhaustion.

2.3. Equipment

Throughout both exergames, the incremental treadmill test, and in the rest time between activities, gas exchange variables were measured on a breath-by-breath basis (MetaMax 3B, Cortex, Biophysik, Leipzig, Germany). Prior to testing each participant, the gas analysers were calibrated using gases of known concentration and the turbine volume transducer was calibrated using a 3-litre syringe (Hans Rudolph, Kansas City, MO). The delays in the capillary gas transit and analyser rise time were accounted for relative to the volume signal, thereby time-aligning the concentration and volume signals.

Furthermore, during both exergames and the incremental treadmill test, participants were fitted with nine tri-axial accelerometers (Actigraph wGT3X+, Florida, USA), set at a measurement frequency of 100 Hz. Specifically, participants were asked to wear an accelerometer on the lateral plane of each ankle, knee, hip, wrist and centre of the chest which were fitted using self-adhering bandages at each location to minimize movement artefacts.

2.4. Data processing

All data collected during the two exergames, the incremental treadmill test, and the transition times between these activities were included in analysis. Breath-by-breath oxygen consumption (VO_2) data were reintegrated to 15-s windows for analysis and converted to METs in each 15-s window. In order to calculate METs, an estimate of daily resting energy expenditure was calculated for each participant using the Schofield prediction equations (Schofield, 1985), which are sex- and age-specific. Once daily resting energy expenditure was estimated, it was converted to resting relative VO_2 , and MET values for each 15-s

window were calculated by dividing the measured relative VO_2 by the Schofield-estimated resting VO_2 . The raw ActiGraph data were reintegrated to counts, similar to the work of Trost *et al* (2012b) and Staudenmayer *et al* (2009), in 1-s epochs (windows) in each of the three axes for feature generation and development of the machine learning models. Additionally, raw ActiGraph data from the right hip ActiGraph were reintegrated into 15-s windows for the vertical accelerometer axis only for use in the Freedson MET prediction equation. Occasionally, when transitioning between activities, participants would request to remove the metabolic analyzer for a water break or in the walk from one activity to the next. In these cases, the VO_2 registered by the metabolic analyzer would be 0 or a value close to 0. Given that McMurray *et al.* (2015) recently reported standard deviations for sedentary behaviours of 0.2-0.3 METs, using the commonly applied principle of removing any data more than two standard deviations, any windows with $\text{EE} < 0.5$ corrected METs were removed from analysis due to high likelihood that the analyzer had been removed for part or all of a given window. Using this threshold, approximately 1.4% of collected data windows were removed.

ANNs, a specific type of machine learning model, were chosen for use in this study because they have previously shown high accuracy for prediction of energy expenditure in children (Troost *et al.*, 2012b) and adults (Staudenmayer *et al.*, 2009) using accelerometer count data. Further description of the theoretical basis and structure of ANNs can be found in the work of Trost *et al* (2012b) and Staudenmayer *et al* (2009) as well as Montoye *et al* (2015). The 1-s accelerometer data from each axis were used to extract features in 15-s windows, and these extracted features were used as inputs into the ANNs. Originally, two different sets of features were calculated. The first set included the 10th, 25th, 50th, 75th, and 90th percentiles and the covariance of the accelerometer data in each 15-s window; this feature set is very similar to those previously used and validated in Trost *et al* (2012b) and Staudenmayer *et al* (2009). The second feature set was simpler, consisting of mean and variance of the accelerometer counts in each 15-s window, which has been previously used by members of our research group (Dong, Biswas, Montoye, & Pfeiffer, 2013; Montoye *et al.*, 2014; Montoye *et al.*, 2015). Preliminary analyses revealed no difference in predictive accuracy by ANNs developed from each feature set; therefore, the simpler feature set, mean and variance, was used in further analyses.

ANNs were created using a leave-one-out approach. For this approach, data from all but one participant were used to develop (train) the ANNs. For the training phase, both the input features and corrected MET values were used to tune the weights of ANN parameters. Once developed, the ANNs were tested for predictive accuracy on the participant left out of the ANN training. This approach was iterative, repeated 27 times so that each participant's data were used as the test data once. ANNs were created and tested separately for each of the 9 accelerometers in order to determine optimal accelerometer

Using machine learning to predict energy expenditure

placement for prediction of energy expenditure. Additionally, combinations of two or more accelerometers were used in ANN creation and testing to determine if using additional accelerometers would improve energy expenditure prediction accuracy. For simplicity in reporting, accuracy of all 9 single accelerometers is reported, but only the highest accuracy achieved with any combination of 2, 3, 4, and all 9 accelerometers is reported. Combinations of 5-8 accelerometers are not reported as it became apparent that addition of more accelerometers did not yield additional predictive accuracy over using only one or two accelerometer placements. Each ANN was saved as a RDA file and can be downloaded for use from the following link <https://sites.google.com/site/alexmontoye/machine-learning-model-code> and clicking the “ANN 9 accelerometers in children.zip” link.

In order to compare energy expenditure prediction accuracy of the ANNs developed in this study to commonly used accelerometer data analysis methods, the Freedson child MET prediction equation developed for children was used to predict energy expenditure using data from the right hip accelerometer (Freedson et al., 2005). As noted previously, the vertical axis data from the right hip accelerometer were reintegrated into 15-s windows, and the number of counts was multiplied by 4 before being input into the prediction equation (since the Freedson equation was developed for data collected in 60-s windows). The Freedson equation is as follows:

$$(1) \text{ METs} = 2.757 + (0.0015 \cdot \text{counts} \cdot \text{min}^{-1}) - (0.08957 \cdot \text{age} [\text{yrs}]) - (0.000038 \cdot \text{counts} \cdot \text{min}^{-1} \cdot \text{age})$$

2.5. Data analysis

Correlations with measured energy expenditure, root mean square error (RMSE), and bias statistics were calculated for each iteration of the leave-one-out approach using Microsoft Excel (Microsoft Corp, Redmond, WA) in order to determine performance of the ANNs and of the Freedson MET prediction equation. Due to negative skew of the correlations, a Fisher Z transformation was conducted to normalize the distribution prior to conducting statistical analyses. In order to compare accuracy of the ANNs and Freedson MET prediction equation, repeated measured analysis of variance tests were conducted separately for transformed correlations, RMSE, and bias. A least significant difference post-hoc test, equivalent to no correction, was used for pairwise comparisons when the overall test was statistically significant. A p-value of $p < 0.05$ was used to determine statistical significance. Residual plots were also created to better assess bias at different intensities of PA. Statistical comparisons were conducted using SPSS (v. 22.0, IBM Corp., Armonk, NY).

3. Results

Anthropometric characteristics of the sample can be seen in Table 1. Correlations for predicted vs. measured (criterion) energy expenditure can be seen in figure 1. Correlations with measured energy expenditure for single-accelerometer placements fell in a narrow range ($r=0.77-0.81$), although some minor differences were noted. The ANNs for the right and left hips, right ankle, and chest placements ($r=0.81$) had slightly but significantly higher correlations than the left and right wrist accelerometers ($r=0.77-0.78$). The best 2- (right wrist-right knee), 3- (chest - right hip - right ankle) and 4-accelerometer combinations (chest - right knee - right wrist - right hip) achieved correlations of $r=0.80$, $r=0.81$, and $r=0.82$, respectively, none of which were significantly different from each other or from the right ankle or chest accelerometer placements. The ANN created from all 9 accelerometers had a significantly lower correlation with predicted energy expenditure ($r = 0.69$) than any single accelerometer or combination of 2, 3, or 4 accelerometers. With a correlation of $r=0.75$, the Freedson MET prediction equation had a significantly higher correlation with measured energy expenditure than the 9-accelerometer ANN but a significantly lower correlation than all ANNs for all single accelerometers and the best 2-, 3-, and 4-accelerometer combinations.

Table 1. Participant anthropometric and peak exercise responses.

	Total (n=27)	Boys (n=15)	Girls (n=12)
Age (years)	10.8 ± 1.1	10.8 ± 1.1	10.8 ± 1.3
Stature (m)	1.45 ± 0.07	1.46 ± 0.08	1.45 ± 0.07
Body mass (kg)	38.7 ± 8.6	39.8 ± 8.4	37.2 ± 9.1
Maturity offset (years)	-1.9 ± 1.0	-2.4 ± 0.8	-1.1 ± 1.0

Means ± SD.

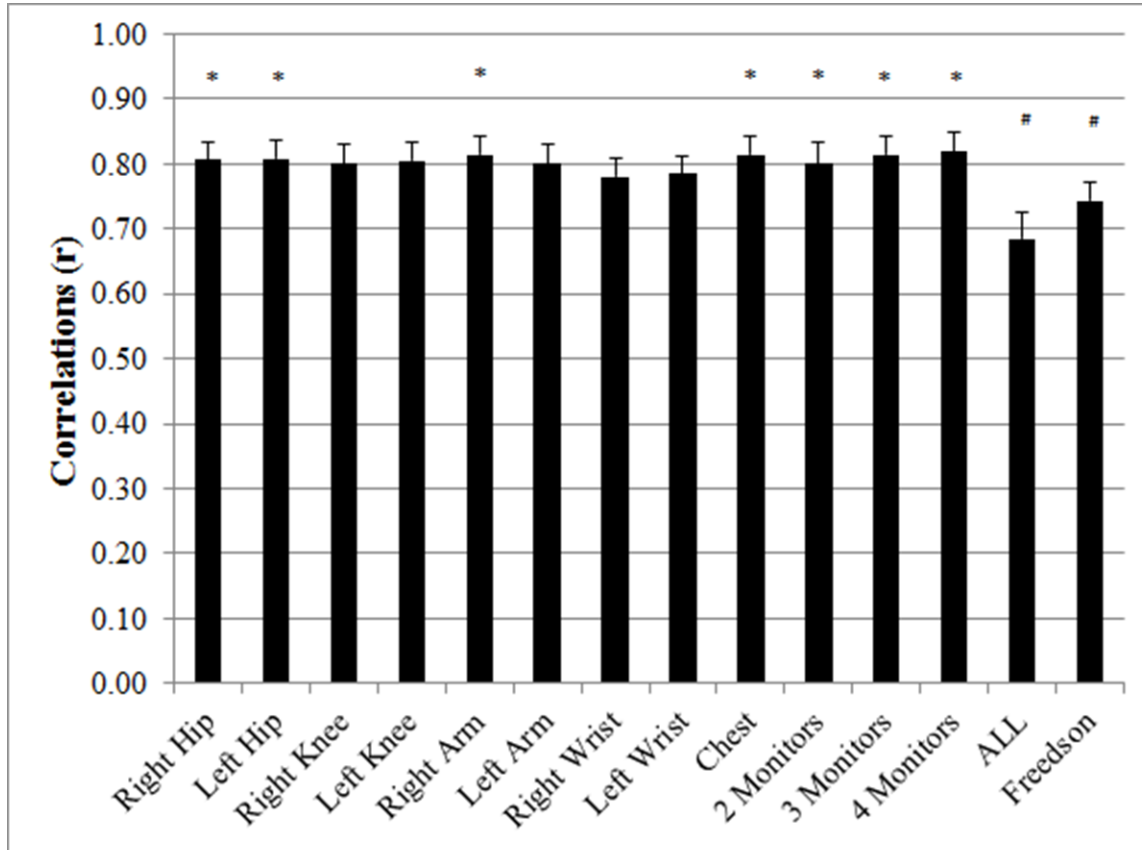


Figure 1. Correlations with criterion-measured METs.

Data are shown as mean (standard error [SE]). 2 Monitors = right wrist-right knee; 3 Monitors = chest-right hip-right ankle; 4 Monitors = chest-right knee-right wrist-right hip; ALL = all 9 accelerometers; Freedson = Freedson MET prediction equation.

* Indicates significant difference from RW and LW.

Indicates significant difference from all other accelerometers and accelerometer combinations.

In the RMSE analysis (figure 2), there were no significant differences among any of the single-accelerometer ANNs or the best 2-, 3-, and 4-accelerometer combinations (min-max 1.21-1.31 METs). With an RMSE of 1.46 METs, the 9-accelerometer ANN had significantly higher RMSE than any of the other single- or multiple-accelerometer ANNs tested. With an RMSE of 1.74 METs, the Freedson MET equation had higher RMSE than any of the single- or multiple-accelerometer ANNs tested.

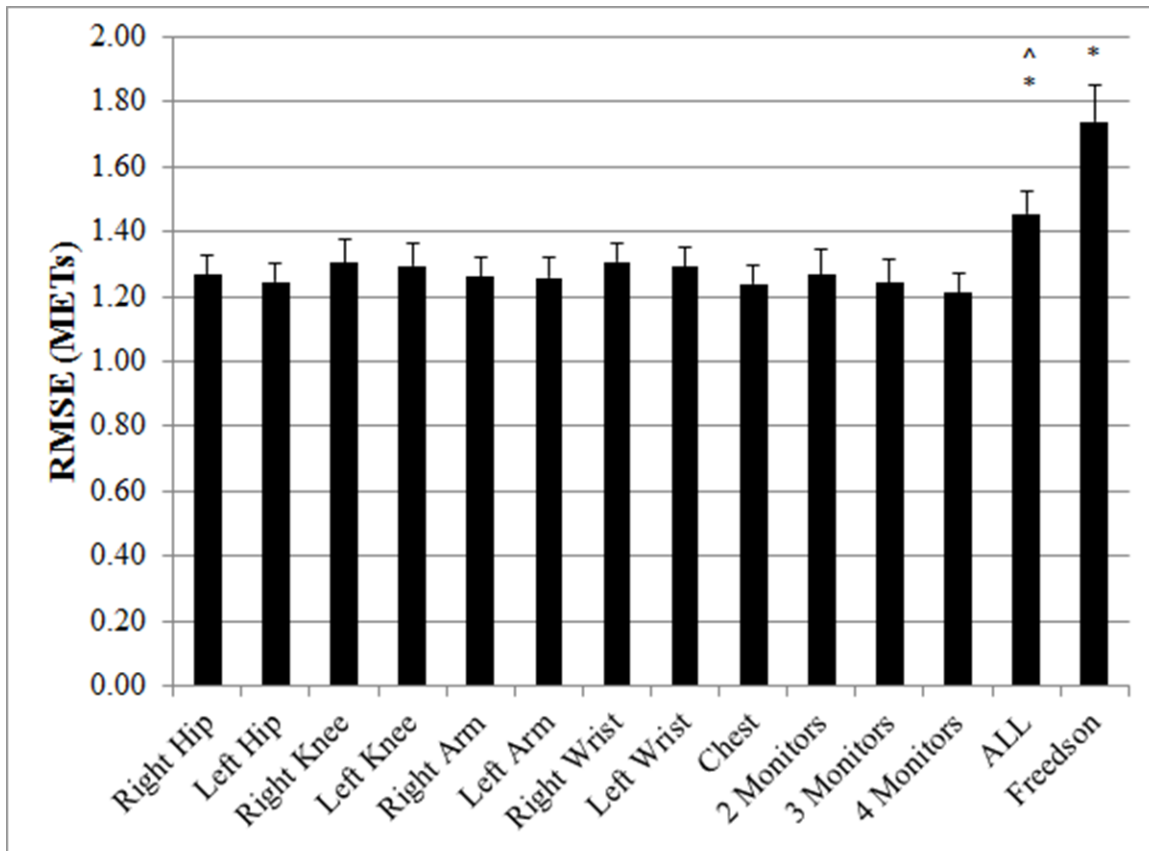


Figure 2. RMSE for predicted vs. criterion-measured METs.

*Indicates significant difference from all single accelerometers and combinations of 2, 3, and 4 accelerometers.

^Indicates significant difference between the 9 accelerometer ANN and Freedson equation
 Data are shown as mean (standard error [SE]). 2 Monitors = right wrist-right knee; 3 Monitors = chest-right hip-right ankle; 4 Monitors = chest-right knee-right wrist-right hip; ALL = all 9 accelerometers; Freedson = Freedson MET prediction equation.

Overall bias is displayed in figure 3. None of the ANNs or the Freedson MET prediction equation had significant bias for prediction of energy expenditure, although the Freedson equation trended toward underprediction. Figure 4 provides a more detailed view of bias, showing a residual plot for the chest accelerometer ANN and the Freedson MET prediction equation. The other 12 ANNs have similar residual plots to the chest, so only the chest is shown for simplicity. Despite no overall bias, all 13 ANNs tended to overestimate energy expenditure when the measured energy expenditure was low and underestimate energy expenditure when the measured energy expenditure was high. The Freedson MET prediction

equation had higher energy expenditure prediction accuracy when measured energy expenditure was low and underestimated energy expenditure when measured energy expenditure was high.

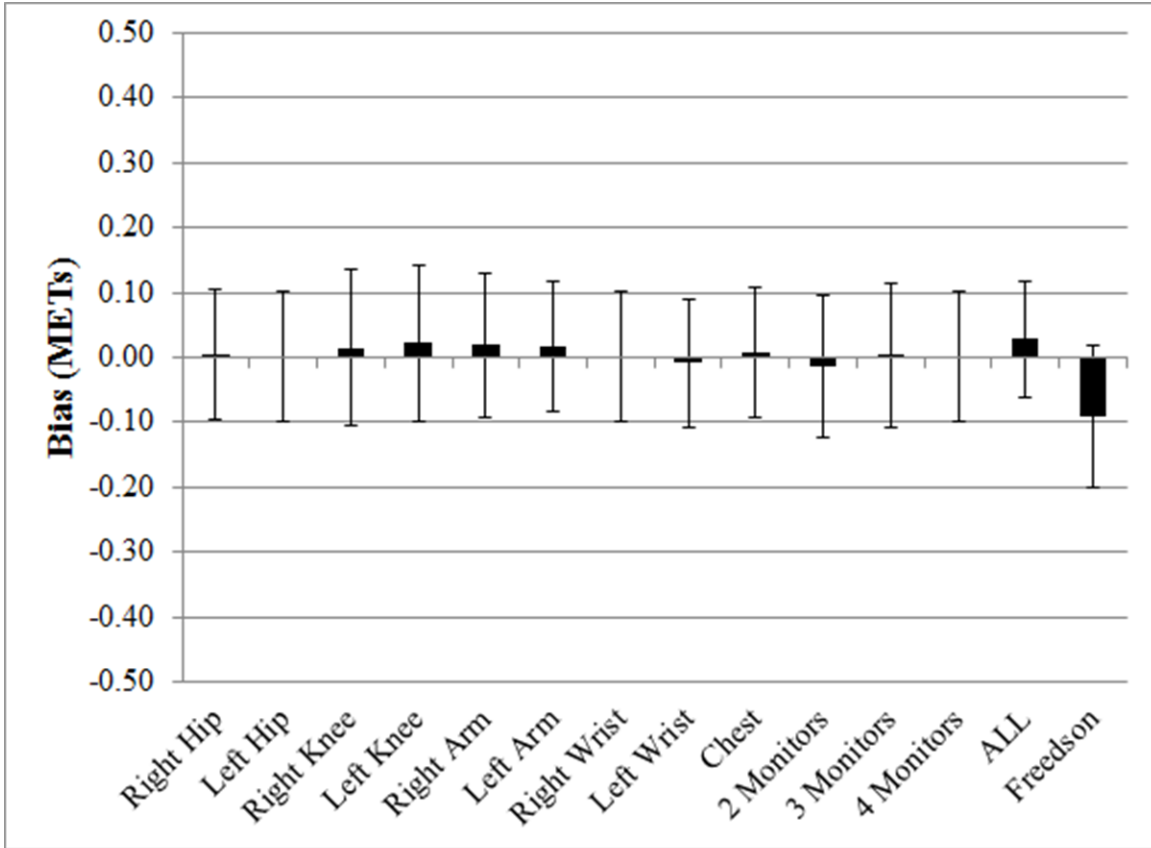


Figure 3. Overall bias for predicted vs. criterion-measured METs.

are shown as mean (standard error [SE]). 2 Monitors = right wrist-right knee; 3 Monitors = chest-right hip-right ankle; 4 Monitors = chest-right knee-right wrist-right hip; ALL = all 9 accelerometers; Freedson = Freedson MET prediction equation.

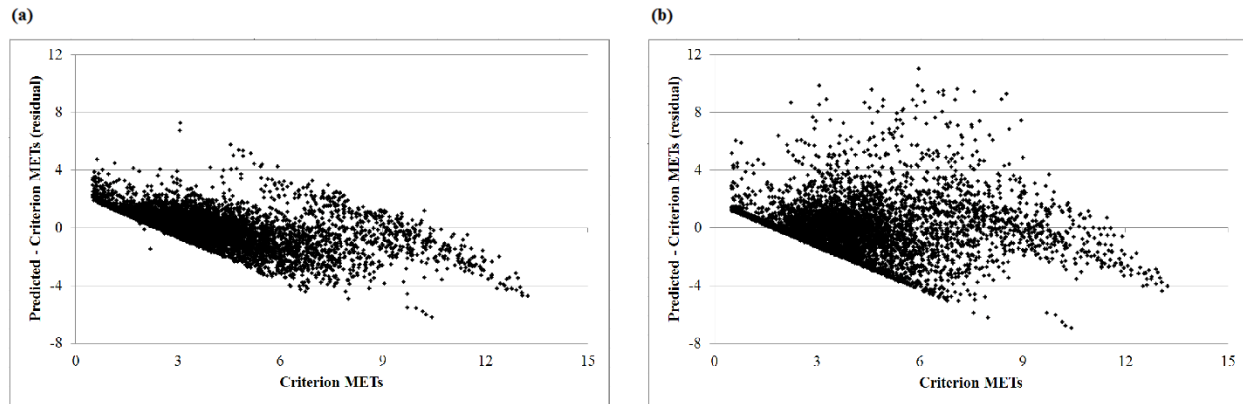


Figure 4. Residual plots of predicted vs. criterion-measured METs.

(a) Chest; (b) Freedson MET prediction equation.

4. Discussion

The purpose of the present study was to evaluate the relative accuracy of nine different accelerometer placements, and combinations thereof, on the estimation of energy expenditure derived from ANNs. The main findings were that neither anatomical location nor the combination of two or more accelerometers significantly influenced prediction accuracy. Furthermore, the Freedson MET equation was associated with significantly poorer prediction accuracy than all ANNs during semi-structured activities in children. These findings therefore demonstrate that accelerometer placement will minimally effect the accuracy of energy expenditure prediction and highlight the potential utility of ANNs to advance our understanding and interpretation of accelerometer traces derived from children.

In the present study, accelerometers placed on all nine anatomical positions demonstrated a statistically similar ability to predict energy expenditure during exergaming, an incremental treadmill test, and the transitions between these activities in children. These findings have important implications in the design of future studies wishing to estimate energy expenditure during field based studies where accelerometer placement may be dictated by practicalities and adherence. However, it is important to consider that subtle differences may be of practical importance; when measured over a longer period, such subtle differences could culminate in substantially different total energy expenditure estimations, thus altering the interpretation of the findings. In agreement with conventional protocols, placement sites closest to the centre of mass tended to have slightly higher correlations with measured energy expenditure than placements on distal body locations. However, it has been suggested that such placement positions fail to detect upper body movements and thus engender significant measurement errors and PA intensity misclassifications (Chen & Bassett, 2005). Indeed, in a recent study by Ellis and colleagues (2014), they

found that whilst activity identification was relatively similar between a hip and wrist worn accelerometer, the former was associated with greater accuracy of EE prediction during locomotion. Consequently, alternative placement sites that may elicit improved accuracy have received increasing interest, with the wrist perhaps the most researched alternative. Despite this, the findings remain equivocal regarding the relative accuracy of energy expenditure prediction derived from the wrist and hip, with some reporting a greater or similar accuracy of the non-dominant wrist compared to the hip (Hildebrand, VT, Hansen, & Ekelund, 2014; Melanson & Freedson, 1995) while others, in agreement with the current findings, demonstrated by slightly lower correlations, found a poorer accuracy associated with the wrist (Puyau et al., 2002; Swartz et al., 2000). Nonetheless, wrist-worn accelerometers provide significant advantages with respect to compliance (Troost, Zheng, & Wong, 2012a), a consideration that needs to be balanced against potentially marginal losses in energy expenditure prediction accuracy. Specifically, whilst non-compliance and inadequate wear time have been longstanding problems with regard to hip-worn accelerometers (Belton, O'Brien, Wickel, & Issartel, 2013), recent studies have shown 70-80% compliance and median wear time of 21-22 hours per day using wrist-worn accelerometers (Rowlands et al., 2014). Of note, each of the nine single-accelerometer ANNs had significantly higher accuracy for predicting energy expenditure than a previously-developed, regression-based energy expenditure prediction equation (Freedson et al., 2005). This finding provides further evidence of the utility of moving toward use of machine learning or pattern recognition modelling techniques for improvement of energy expenditure estimation.

It has been postulated that amalgamating the data from multiple monitoring devices could improve estimated energy expenditure; accelerometers positioned at different body locations are likely to provide nontrivial, complementary information (He et al., 2014). However, contrary to expectation, the utilisation of multiple accelerometers provided only minimal improvements in prediction accuracy, signified by slightly higher correlations but no difference in RMSE. While several studies have investigated the use of multiple accelerometers for activity recognition in adults, reporting similar findings to those reported here (Cleland et al., 2013; He et al., 2014), there is limited research specifically considering energy expenditure prediction, despite its importance as an outcome variable in PA research (Welk, 2002). Montoye *et al* (2014) recently reported that a three piece accelerometer system significantly improved energy expenditure prediction accuracy during laboratory-based activities but these improvements were minimal in a semi-structured setting (Montoye et al., 2016). These discrepancies were largely attributed to the inclusion of transitional periods as well as non-steady-state data in the analysis, a methodology employed to increase the applicability of the ANNs developed to free-living settings. Specifically, PA, especially at higher intensities, is rarely performed in a sufficiently

sustained bout to achieve a steady state, thus predictive models must be able to handle non-steady-state data. Given the highly sporadic nature of children's PA (Bailey et al., 1995), with bouts typically lasting no longer than 6-s, the incorporation of non-steady-state data as done in the present study is especially pertinent to the development of appropriate energy expenditure prediction models for children. Therefore, whilst the accuracy associated with the present ANNs may be lower than could be derived during controlled laboratory exercises, the applicability of the present models is likely to be considerably higher. Furthermore, it is postulated that information contained within the raw signal may have been lost following the conversion to count-based data, resulting in a lack of improvement in energy expenditure prediction. Nonetheless, the reintegration to count-based data enabled comparisons to previous studies (Staudenmayer et al., 2009; Trost et al., 2012b). It is also anticipated that the implications associated with wearing multiple monitors for compliance outweigh such minimal improvements in prediction accuracy. However, further investigation into the use of multiple monitors is warranted given the ever-increasing vast array of technology available and the potential to embed sensors within every wearable items (i.e., clothing) with no added burden to participants.

This is the first study to consider a diverse range of dominant and non-dominant accelerometer placements, and the optimal combinations thereof, for the prediction of energy expenditure, demonstrating the highest accuracy with single accelerometers and when accelerometers on the upper and lower body were combined. This combination is largely in accord with the literature suggesting that both upper and lower body movements need to be accounted for (Chen & Bassett, 2005). Indeed, hip worn accelerometers have largely been validated during ambulatory movements, whilst many activities of daily living incorporate a substantial upper body contribution, limiting the applicability of those prediction algorithms developed solely using hip-worn accelerometers. The failure of greater numbers of accelerometers to improve the energy expenditure prediction accuracy may be attributable to over-fitting of the data and the relatively limited information used for data extraction in the development of ANNs. In reality, the decision to use single or multiple accelerometers will principally depend on the relative importance of measurement accuracy compared with participant and researcher burden. For population level studies, inherent inter-participant variability is likely to negate the small gains in accuracy achieved by using multiple accelerometers whereas for smaller studies or those researching a specific outcome (i.e. weight-loss interventions or posture), multiple accelerometers may account for greater variance in energy expenditure than a single accelerometer and minimize bias in energy expenditure prediction of specific types of activities. With regard to the latter, it is important to highlight that the present ANNs generally demonstrated a bias towards the mean, overestimating energy expenditure at low intensities and underestimating at higher intensities. Whilst commonly reported, this form of bias would need to be

considered if the ANNs developed were to be used in a true free-living setting and when interpreting time spent at the various exercise intensities. However, ActiGraph accelerometers, and the present data processing techniques, were not designed to measure sedentary behaviour. Nonetheless, given the ever increasing interest in understanding the relationship between sedentary behaviour and health, and indeed the contribution sedentary behaviours make to a child's day, the accurate estimation of sedentary pursuits is imperative. As such, future machine learning model development should include a broad range of energy expenditure representative of children's play and daily living.

It is postulated that raw accelerometry data may enhance the accuracy of the models, however, despite the presented study utilising count-based data, a machine learning approach significantly enhanced the energy expenditure prediction accuracy in comparison to the Freedson MET prediction equation. Specifically, the RMSE for any of the single-accelerometer ANNs and the best 2-, 3- and 4-accelerometer combinations in the present study ranged from 1.21-1.31 METs, in comparison to 1.74 METs for the Freedson MET equation. One possible explanation for this greater accuracy is the wider range of different movement patterns as well as the traditional locomotor treadmill activities in the present training set. However, it is pertinent to note that the current ANNs were developed and cross-validated on the same activities and would thus be expected to perform better than other models which were developed for different activities. Further validation of the developed ANNs in a more heterogeneous sample population and during a wider range of activities is warranted.

Importantly, none of the present ANNs demonstrated a significant bias in the prediction of energy expenditure, whereas, consistent with previous research (Crouter, Klowers, & Bassett, 2006; Rothney, Neumann, Beziat, & Chen, 2007; Staudenmayer et al., 2009), the Freedson *et al* regression method trended toward an overall under-estimation. This under-estimation is likely to be attributable to the equations being based on only three treadmill activities in a small sample. Given that the linear energy expenditure relationship tends to break down at higher running speeds (Cavagna, Thys, & Zamboni, 1976) and does not translate to non-locomotive activities, it is not surprising that the linear regression equation under-estimates free-living or vigorous treadmill activities (Crouter et al., 2006; Rothney et al., 2007; Staudenmayer et al., 2009). However, in alignment with the lower intensity activities (brisk walking to slow jogging), the Freedson MET prediction equation had higher energy expenditure prediction accuracy when measured energy expenditure was low, although the lack of specific sedentary behaviours in the present training set limits further interpretation. A similar relationship between prediction accuracy and exercise intensity was reported by Staudenmayer and colleagues (2009), resulting in a RMSE of 1.22 METs, a value not dissimilar to that reported here. However, it is pertinent to note that this study was conducted in adults, raising questions regarding the applicability of such ANNs to children

given developmental differences in stride length and frequency and differences in economy of movement and resting metabolic rate (Troost, 2007; Troost, Loprinzi, Moore, & Pfeiffer, 2011). More recently, Troost et al. (2012b) investigated the prediction of energy expenditure using an ANN in children completing activities ranging from sedentary to moderate-to-vigorous intensity household activities and games. This study reported a RMSE of 0.9-1.1 METs depending on the measurement window used; higher RMSEs were associated with the shorter measurement windows. The greater degree of accuracy in the study by Troost et al. (2012b) compared to the present study is likely to be attributable to their use of steady-state data and more controlled laboratory-based activities. The exergames in the present study were utilised to provide a more representative example of the highly sporadic nature of children's physical activity (Bailey et al., 1995).

The present study significantly advances our understanding of the use of ANNs in predicting children's energy expenditure during semi-structured activities. Nonetheless, it is pertinent to note certain methodological considerations which limit our interpretation. Specifically, the reliance on select activity modalities may limit generalisability of the present study to habitual physical activity, as may the omission of a thigh accelerometer to account for sedentary behaviours. Moreover, our sample size was relatively small and homogenous, limiting the applicability of our ANNs to other populations with disparate demographic characteristics. Furthermore, whilst indirect calorimetry is widely accepted as a criterion measure of energy expenditure, it is important to be cognizant that such measures inherently lag behind the actual energy expenditure elicited by an activity. Thus, the energy expenditure measured by indirect calorimetry may not represent a true reflection of the energy cost of non-steady-state activities.

In conclusion, the present findings suggest that neither the anatomical location of single accelerometers nor the combination of multiple anatomical locations significantly influences the accuracy of EE prediction using ANNs in children during semi-structured activities. Furthermore, data supports the use of ANN in improving prediction accuracy compared with simple regression. Such findings can advance our understanding of children's accelerometer traces, though raw accelerometer data from multiple accelerometers may be more important when assessing specific behaviours and postures. The study highlights the difficulties of prediction models, where activities in semi-structured protocols may not be truly representative as a training set for validation purposes. Future studies should seek to incorporate a greater variety activities and postures to ensure accurate prediction of energy expenditure in a true free-living setting.

Acknowledgments

The authors would like to thank David Barnett for his assistance with data collection and participant recruitment.

Conflicts of interest

The authors attest that they have no conflicts of interest to report.

References

- Bailey, R. B., Olson, J., Pepper, S. L., Porszasz, J., Barstow, T. J., & Cooper, D. M. (1995). The level and tempo of children's physical activities: an observational study. *Med Sci Sport Exerc*, *27*, 1033-1041.
- Baquet, G., Stratton, G., Van Praagh, E., & Berthoin, S. (2007). Improving physical activity assessment in prepubertal children with high-frequency accelerometry monitoring: A methodological issue. *Prev Med*, *44*, 143-147.
- Belton, S., O'Brien, W., Wickel, E. E., & Issartel, J. (2013). Patterns of noncompliance in adolescent field-based accelerometer research. *J Phys Act Health*, *10*, 1181-1185.
- Cavagna, G. A., Thys, H., & Zamboni, A. (1976). The sources of external work in level walking and running. *J Physiol*, *262*, 639-657.
- Chandler, J. L., Brazendale, K., Beets, M. W., & Mealing, B. A. (2016). Classification of physical activity intensities using a wrist-worn accelerometer in 8–12-year-old children. *Pediatr Obes*, *11*, 120-127.
- Chen, K. Y., & Bassett, D. R. (2005). The technology of accelerometry-based activity monitors: current and future. *Med Sci Sport Exer*, *37*, S490-500.
- Cleland, I., Kikhia, B., Nugent, C., Boytsov, A., Hallberg, J., Synnes, K., . . . Finlay, D. (2013). Optimal placement of accelerometers for the detection of everyday activities. *Sensors (Basel)*, *13*, 9183-9200.
- Colley, R., Connor Gorber, S., & Tremblay, M. S. (2010). Quality control and data reduction procedures for accelerometry-derived measures of physical activity. *Health Rep*, *21*, 63-69.
- Crouter, S. E., Flynn, J. I., & Bassett, D. R. (2015). Estimating Physical Activity in Youth Using a Wrist Accelerometer. *Medicine and Science in Sports and Exercise*, *47*, 944-951.
- Crouter, S. E., Klowers, K. G., & Bassett, D. R. (2006). A novel method for using accelerometer data to predict energy expenditure. *J Appl Physiol*, *100*, 1324-1331.
- Department of Health. (2011). Start Active, Stay Active: a report on physical activity for health from the four home countries' Chief Medical Officers. London, UK.

- Dong, B., Biswas, S., Montoye, A., & Pfeiffer, K. (2013). Comparing metabolic energy expenditure estimation using wearable multi-sensor network and single accelerometer. *Conf Proc IEEE Eng Med Biol Soc, 2013*, 2866-2869.
- Eklom, O., Nyberg, G., Bak, E. E., Ekelund, U., & Marcus, C. (2012). Validity and comparability of a wrist-worn accelerometer in children. *J Phys Act Health, 9*, 389-393.
- Ellis, K., Kerr, J., Godbole, S., Lanckriet, G., Wing, D., & Marshall, S. (2014). A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. *Physiol Meas, 35*, 2191-2203.
- Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of Accelerometer Output for Children. *Medicine and Science in Sports and Exercise, 37*, S523-S530.
- Freedson, P. S., Lyden, K., Kozey-Keadle, S., & Staudenmayer, J. (2011). Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. *J Appl Physiol, 111*, 1804-1812.
- He, B., Bai, J., Zipunnikov, V. V., Koster, A., Caserotti, P., Lange-Maia, B., . . . Crainiceanu, C. M. (2014). Predicting human movement with multiple accelerometers using movelets. *Med Sci Sports Exerc, 46*, 1859-1866.
- Hendelman, D., Miller, K., Bagget, C., Debold, E., & Freedson, P. S. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sport Exerc, 32*, S442-S449.
- Hildebrand, M., VT, V. A. N. H., Hansen, B. H., & Ekelund, U. (2014). Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc, 46*, 1816-1824.
- Hopkins, N., Stratton, G., Maia, J., Tinken, T. M., Graves, L. E., Cable, T. N., & Green, D. J. (2010). Heritability of Arterial Function, Fitness, and Physical Activity in Youth: A Study of Monozygotic and Dizygotic Twins. *J Pediatr, 157*, 943-948.
- Jansen, W., Borsboom, G., Meima, A., Zwanenburg, E. J.-V., Mackenbach, J. P., Raat, H., & Brug, J. (2011). Effectiveness of a primary school-based intervention to reduce overweight. *Int J Pediatr Obes, 6*, e70-e77.
- McMurray, R. G., Butte, N. F., Crouter, S. E., Trost, S. G., Pfeiffer, K. A., Bassett, D. R., . . . For the, C. D. C. N. C. I. N. R. G. o. E. E. i. C. (2015). Exploring Metrics to Express Energy Expenditure of Physical Activity in Youth. *PLoS ONE, 10*, e0130869.
- Melanson, E. L., Jr., & Freedson, P. S. (1995). Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. *Med Sci Sports Exerc, 27*, 934-940.
- Mirwald, R. L., Baxter-Jones, A. D. G., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Med Sci Sport Exerc, 34*, 689-694.

Using machine learning to predict energy expenditure

- Montoye, A., Dong, B., Biswas, S., & Pfeiffer, K. (2014). Use of a wireless network of accelerometers for improved measurement of human energy expenditure. *Electronics*, *3*, 205-220.
- Montoye, A. H., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2015). Energy Expenditure Prediction Using Raw Accelerometer Data in Simulated Free Living. *Med Sci Sports Exerc*, *47*, 1735-1746.
- Montoye, A. H. K., Dong, B., Biswas, S., & Pfeiffer, K. A. (2016). Validation of a wireless accelerometer network for energy expenditure measurement. *Journal Of Sports Sciences*, 1-10.
- Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obes Res*, *10*, 150-157.
- Riddoch, C., Mattocks, C., Deere, K., Saunders, J., Kirkby, J., Tilling, K., . . . Ness, A. (2007). Objective measurement of levels and patterns of physical activity. *Arch Dis Child*, *92*, 963-969.
- Rothney, M. P., Neumann, M., Beziat, A., & Chen, K. Y. (2007). An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol (1985)*, *103*, 1419-1427.
- Rowlands, A. V., Rennie, K., Kozarski, R., Stanley, R. M., Eston, R. G., Parfitt, G. C., & Olds, T. S. (2014). Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc*, *46*, 2308-2316.
- Schofield, W. N. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr*, *39 Suppl 1*, 5-41.
- 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. *J Appl Physiol (1985)*, *107*, 1300-1307.
- Swartz, A. M., Strath, S. J., Bassett, D. R., Jr., O'Brien, W. L., King, G. A., & Ainsworth, B. E. (2000). Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Med Sci Sports Exerc*, *32*, S450-456.
- Trost, S., Zheng, Y., & Wong, W. (2012a). Machine learning for activity recognition: hip versus wrist data. *Physiol Meas*, *35*, 2183-2189.
- Trost, S. G. (2007). Measurement of Physical Activity in Children and Adolescents. *Am J Lifestyle Med*, *1*, 299-314.
- Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2011). Comparison of Accelerometer Cut Points for Predicting Activity Intensity in Youth. *Med Sci Sport Exer*, *43*, 1360-1368.
- Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc*, *37*, S531-543.

Using machine learning to predict energy expenditure

- Trost, S. G., Wong, W. K., Pfeiffer, K. A., & Zheng, Y. (2012b). Artificial neural networks to predict activity type and energy expenditure in youth. *Medicine and Science in Sports and Exercise*, *44*, 1801-1809.
- Welk, G. J. (2002). Use of Accelerometry-Based Activity Monitors to Assess Physical Activity. . In G. J. Welk (Ed.), *Physical Activity Assessments for Health-Related Research* (pp. 125-142). Champaign, IL: Human Kinetics.