Journal for the Measurement of Physical Behaviour, 2019, 2, 94-102
https://doi.org/10.1123/jmpb.2018-0032

Human Kinetics
ORIGINAL RESEARCH

Accelerometry-Based Prediction of Energy Expenditure in Preschoolers

Berit Steenbock
Leibniz Institute for Prevention Research and Epidemiology – BIPS and University of Bremen

Marvin N. Wright, Norman Wirsik, and Mirko Brandes
Leibniz Institute for Prevention Research and Epidemiology – BIPS

Purpose: Study purposes were to develop energy expenditure (EE) prediction models from raw accelerometer data and to investigate the performance of three different accelerometers on five different wear positions in preschoolers. **Methods**: Fourty-one children (54% boys; 3–6.3 years) wore two Actigraph GT3X (left and right hip), three GENEActiv (right hip, left and right wrist), and one activPAL (right thigh) while completing a semi-structured protocol of 10 age-appropriate activities. Participants wore a portable indirect calorimeter to estimate EE. Utilized models to estimate EE included a linear model (LM), a mixed linear model (MLM), a random forest model (RF), and an artificial neural network model (ANN). For each accelerometer, model, and wear position, we assessed prediction accuracy via leave-one-out cross-validation and calculated the root-mean-squared-error (RMSE). **Results**: Mean RMSE ranged from 2.56–2.76 kJ/min for the RF, 2.72–3.08 kJ/min for the ANN, 2.83–2.94 kJ/min for the LM, and 2.81–2.92 kJ/min for the MLM. The GENEActive obtained mean RMSE of 2.56 kJ/min (left and right wrist) and 2.73 kJ/min (right hip). Predicting EE using the GT3X on the left and right hip obtained mean RMSE of 2.60 and 2.74 kJ/min. The activPAL obtained a mean RMSE of 2.76 kJ/min. **Conclusion**: These results demonstrate good prediction accuracy for recent accelerometers on different wear positions in preschoolers. The RF and ANN were equally accurate in EE prediction compared with (mixed) linear models. The RF seems to be a viable alternative to linear and ANN models for EE prediction in young children in a semi-structured setting.

Keywords: accelerometer, children, linear mixed model, machine learning, physical activity, validation

It is generally agreed that regular physical activity (PA) is related to important health outcomes in children (e.g., cardiometabolic and psychosocial health; Knaeps et al., 2018; Reddon, Meyre, & Cairney, 2017; Shoup, Gattshall, Dandamudi, & Estabrooks, 2008; Skrede et al., 2017; Wafa et al., 2016). PA is defined as any bodily movement produced by skeletal muscles that results in energy expenditure (EE) (Caspersen, Powell, & Christenson, 1985). In order to monitor children's PA, analyze associations between PA and health outcomes, and evaluate the effectiveness of interventions promoting PA among children, valid measures of children's PA and EE are needed (Lamonte & Ainsworth, 2001). In recent years, accelerometers have gained considerable popularity as an objective measure of sedentary behaviors, PA and other outcomes, such as EE. They detect accelerations of the body and enable an estimation of intensity, frequency, duration, and type of movement (Hills, Mokhtar, & Byrne, 2014; Skotte, Korshoj, Kristiansen, Hanisch, & Holtermann, 2014). Accelerometers have several advantages over traditional questionnaire-based measures of PA, including superior reliability and validity, and are increasingly being used in studies with very young children (Hills et al., 2014). However, traditional linear model equations developed for activity count-based data do not provide accurate estimates of EE in preschoolers (Janssen et al., 2013; Reilly et al., 2006).

Because the relationships between accelerometer output and EE differ in preschoolers compared with older children, prediction equations require development and validation in this specific age group (Butte et al., 2014). Considerable progress has been made in predicting EE for adults and older children (Jimmy, Seiler, & Maeder, 2013; Montoye, Begum, Henning, & Pfeiffer, 2017; Montoye, Mudd, Biswas, & Pfeiffer, 2015) whereas several methodological questions concerning the use of accelerometry in young children remain open. In their recently published review that provides age-specific practical considerations on accelerometer data collection (e.g., device placement) and processing criteria (e.g., epoch length, cut-points, and algorithms), Migueles and colleagues (2017) observed a lack of calibration and validation studies for preschoolers that address important processing criteria (such as EE algorithms for wrist- and hip-worn accelerometers). However, as studies included in the review were restricted to those applying the latest version of the Actigraph device (GT3X), no practical consideration about device selection in this age group could be drawn. Besides this, most of the research on EE prediction that has been done in preschoolers is limited by the use of direct observation as the criterion measurement and the assignment of fixed metabolic equivalents (METs) to activities and accelerometer output (Davies et al., 2012; De Decker et al., 2013; Hagenbuchner, Cliff, Trost, Van Tuc, & Peoples, 2015). Additionally, the use of highly structured protocols under laboratory settings has been found to overestimate EE in children, which limits the transfer to free-living behaviors (Nilsson et al., 2008).

Recent studies in older children and adults show improvements in EE prediction using non-linear models (Mackintosh,

Steenbock, Wright, Wirsik, and Brandes are with the Leibniz Institute for Prevention Research and Epidemiology–BIPS, Bremen, Germany. Steenbock is with the Center of Clinical Psychology and Rehabilitation, University of Bremen, Bremen, Germany. Steenbock (steenbock@leibniz-bips.de) is corresponding author.

Montoye, Pfeiffer, & McNarry, 2016; Montoye et al., 2017). Preliminary evidence demonstrates that machine learning models outperform simplified regression models in preschoolers (Chowdhury et al., 2018).

This study had three purposes related to accelerometer use in preschoolers: (1) to provide energy expenditure (EE) prediction models from raw accelerometry data established against indirect calorimetry, (2) to compare two linear and two machine learning models, and (3) to compare accuracy of different accelerometers placed on the hips, thigh, and wrists.

Methods

Study Participants

To recruit 3- to 6-year-old children from four daycare facilities (kindergartens) in Lower Saxony, we first contacted the administration of the kindergartens. In the next step, we explained the aim and the procedures of the study to the administration as well as to parent representatives during a pre-arranged meeting in each kindergarten. Thereafter, study information material was distributed to all parents. In the course of regular kindergarten morning circle time, we demonstrated the indirect calorimetry to all children. Children with written parental informed consent were allocated to a schedule. On the measurement day, we asked for verbal assent from each child and those refusing to take part in the study were excluded. The local ethics committee (University of Bremen, Bremen) approved the procedures of the study (ethics committee running title: accelerometry in preschoolers [ACCIPS]).

Study Design

The measurements took place from October to November 2016 and were conducted at each respective kindergarten, using the available indoor and outdoor infrastructure. The children had to fast for at least two hours before measurement, but were allowed to drink water. The measurement was restricted to two children per morning, measured one after the other, to avoid interference with lunchtime. A total of 75 min were allocated for the measurement protocol of one child. This included the assessment of anthropometrics, handedness, mounting and demounting the devices, measurement of resting metabolic rate (RMR), performance of up to nine in- and outdoor activities, changing rooms within the kindergarten, and dressing and undressing for outdoor activities (Table 1). We measured RMR during 10 min of supine rest in a quiet, dimly lit room. The child lay on a mattress and was provided with a blanket. A short learning story was shown on a tablet to keep the child calm for the measurement of RMR. We decided on a measurement period of 10 minutes as Borges et al. (2016) showed that this is an appropriate length of time to achieve steady state conditions and delineate an optimum abbreviated period to estimate RMR by indirect calorimetry.

Activities

The child was free to choose up to three friends to play with. Five activities were predetermined and ranged from light (e.g., drawing) to vigorous intensities (e.g., jogging). An additional four activities to be chosen independently with respect to the resources of the daycare facility and the child's individual preferences were offered to each child. Examples are given in Table 1. Starting with the first predetermined activity 'drawing', the child was asked to sit down for five minutes at a table in a group room and to choose one out of five coloring pages with different motifs to crayon. All subsequent activities were performed continuously for at least three minutes. Additionally, the child could choose two indoor physical activities (e.g., playing with toy cars, toy blocks, hide and seek, playing with dolls). For the performance of the outdoor activities, a well-defined and plain area was chosen. On the outdoors, the child started with a predetermined activity, which was 'tag'. Afterwards, the child chose two other outdoor activities, depending on the equipment of each of the four local daycare facilities (e.g., swing, tricycle, monkey bars, or scooter). Finally, the child performed the three activities scheduled last, which were walking at normal speed, walking fast, and jogging. During these activities, the researcher accompanied the child to ensure continuous walking and jogging, while the child set the pace. For fast walking, the researcher encouraged the child by telling an imaginary journey of a holiday trip, including catching the tram, train, and plane to get to a holiday location. If the child changed to jogging during the fast walking task, the researcher immediately slowed the child down to fast walking. If the child refused to perform an activity continuously due to motivational reasons, the activity was stopped and the child was asked to continue with the next activity. The protocol was completely aborted if the child was exhausted or refused to go on any more.

Table 1 Overview, Order, and Duration of Study Procedures

Activity/Task	Time (min)
Welcoming the child, changing to side room for preparation and resting metabolic rate (RMR) measurement	3
Mounting devices (oxygen analyzer, accelerometers), preparing RMR measurement	12
RMR measurement	10
Finishing RMR measurement, changing room for indoor activities	2
Drawing	5
Break/transition	1
Free choice activity indoors #1, such as building	3
Break/transition	1
Free choice activity indoors #2, such as playing with cars	3
Preparing and dressing for outdoor activities	5
Tag	3
Free choice activity outdoors #1, such as tricycling	3
Break/transition	1
Free choice activity outdoors #2, such as climbing	3
Break/transition	1
Regular Walking	3
Break	1
Fast walking	3
Break	1
Jogging	3
Undressing, demounting devices, say goodbye and return to group room/breakfast	8
Total	75

Accelerometers

The children were equipped with six accelerometers of three different brands. One accelerometer was placed at the left and the right wrist (GENEActiv, ActivInsight Ltd, Kimbolton, UK), one accelerometer at the left hip (GT3X Actigraph, Pensacola, Florida, USA), two accelerometers at the right hip (GENEActiv, GT3X), and one accelerometer at the right thigh (activPAL, PAL Technologies Ltd, Glasgow, Scotland, UK). Accelerometers at the wrist were firmly mounted with ¾ inch Tyvek security bracelets, which were used instead of the standard bracelets provided by the manufacturer. An elastic belt was used to fix the hip-worn monitors close to the lateral hip bone, and an adhesive pad was used to fix the activPAL to the skin on the front of the right thigh. All accelerometers were operated with the highest possible resolution, which was 100 Hz in the GENEActiv and the GT3X, and 20 Hz in the activPAL. Before each measurement we initialized and synchronized the accelerometers with the computer running the indirect calorimetry. After completion of the protocol we downloaded accelerometer data from all devices using dedicated software from the manufacturers.

Physiological Measures

A portable, open-circuit indirect calorimetry system (MetaMax3b, Cortex Biophysics, Leipzig, Germany) was used as the criterion measure. The MetaMax3b has proven reliability in a study with adolescents but was found to slightly overestimate VO_2 during moderate and vigorous exercise (Macfarlane & Wong, 2012). It is highly comparable to other common devices (Brandes, Klein, Ginsel, & Heitmann, 2015). The system was mounted to the children using a pediatric harness. In very small children, additional adhesive tape was used to adjust the harness to the child, so that the system did not interfere with the activities. A facemask was secured over the child´s nose and mouth with an adjustable nylon harness. A bidirectional turbine, inserted to the facemask, measured the volume of inspired and expired air. A sample tube, connected to the turbine, retrieved expired air samples breath-by-breath. Air samples were analyzed for oxygen uptake and carbon dioxide production within the sensor unit of the system. Stable and dry weather conditions enabled sound outdoor measurements. Data were transferred via telemetry to a laptop and available in real-time (MetaSoft 3, Cortex Biophysics, Leipzig, Germany). The laptop operator used the marker function of the software to identify the beginning and the end of each activity. Each morning, prior to the first measurement, we calibrated the indirect calorimetry system according to the manufacturer recommended procedure.

Data Reduction and Modeling

Oxygen consumption and carbon dioxide production were measured continuously and converted to absolute (kJ/min) and relative (J/min/kg) EE using the equations by Weir (1949). RMR

Table 2 Participant Characteristics

	All (N = 41)
Age (years)	4.8 ± 0.8 (3.0, 6.3)
Weight (kg)	20.5 ± 4.3 (12.8, 31.1)
Height (cm)	115 ± 9 (96, 130)
BMI (kg/m^2)	15.4 ± 2.1 (12.4, 22.3)
REE (kJ/min)	2.3 ± 0.5 (1.4, 3.4)

Note. Data are displayed as $M \pm SD$ (min, max). REE = resting energy expenditure.

Table 3 Energy Expenditure per Activity

Activity	Number of Children (n)	Energy Expenditure (kJ/min)
Ball throwing	1	11.2
Building	29	5.4 ± 1.3 (3.3, 9.3)
Playing with cars	17	6.6 ± 1.2 (4.6, 8.9)
Tag	34	16.3 ± 4.6 (8.1, 27.8)
Children's slide	1	11.2
Climbing	24	11.6 ± 3.0 (7.4, 19.9)
Playing with dolls	7	6.4 ± 1.1 (4.8, 8.2)
Drawing	40	4.2 ± 0.9 (2.2, 6.2)
Going upstairs	1	10.6
Hide	19	6.6 ± 1.4 (3.6, 9.0)
Hide (outside)	2	11.0 ± 3.2 (8.7, 13.2)
Jogging	11	16.0 ± 3.9 (6.1, 21.7)
Kitchen	1	5.0
Layback	1	8.2
RMR measurement (laying)	41	2.3 ± 0.5 (1.4, 3.4)
Rocking car	1	5.4
Rocking horse	1	4.9
Playing soccer	3	10.9 ± 5.0 (5.9, 15.8)
Swinging	7	8.0 ± 3.2 (5.3, 12.7)
Playing tennis	2	10.9 ± 4.2 (7.9, 13.8)
Trampoline	1	10.8
Tricycle	24	13.4 ± 3.8 (5.5, 21.9)
Regular walking	37	8.2 ± 2.4 (0.9, 14.3)
Walking fast	30	12.4 ± 3.5 (5.5, 17.8)
All	335	8.7 ± 5.2 (0.9, 27.8)

Note. Data are displayed as $M \pm SD$ (min, max).

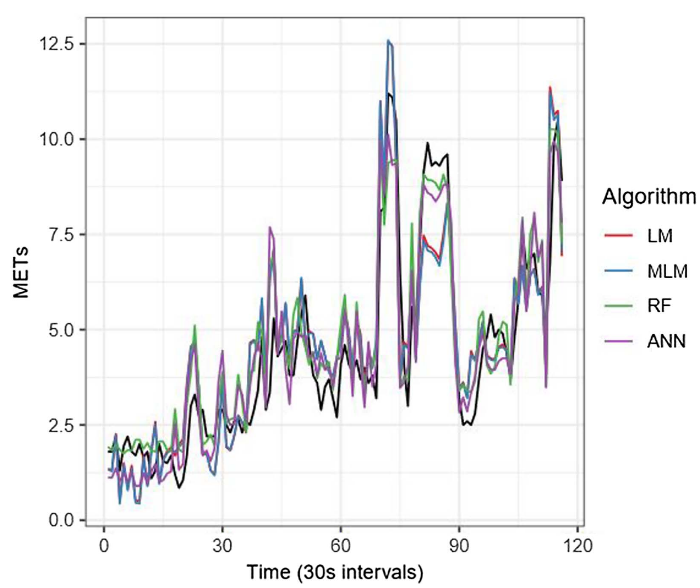

Figure 1 — Predicted METs from the four different models (GENEActiv, left wrist) and METs derived from the indirect calorimeter (black line) over the time course of the study protocol. LM = linear model; MLM = mixed linear model; RF = random forest; ANN = artificial neural network.

was calculated as the minimum of a rolling one-minute mean during supine rest. Metabolic equivalents (METs) were calculated by dividing the relative oxygen consumption (expressed in ml/min/kg) by 3.5.

As done by Montoye et al. (2017), for each individual and device, the following 10 summary statistics were calculated in 30-s non-overlapping windows for each axis (X, Y, Z) of the raw acceleration data—mean, standard deviation, minimum, maximum, lag one autocorrelation, percentiles (10%, 25%, 50%, 75%, 90%)—resulting in 30 summary statistics for each time window. No filtering of the raw accelerometer data was conducted.

Predictive models were created separately for each accelerometer and placement, resulting in 24 (4×6) models developed and tested. We built four predictive models: (1) a linear regression model, (2) a linear mixed model, (3) a random forest, and (4) an artificial neural network model. Linear regression models (LM) and linear mixed models (MLM) were created separately for the outcome variables absolute EE (absEE), relative EE (relEE), and METs and included the above mentioned 30 summary statistics as independent variables. The MLM accounts for repeated measurements to compute EE prediction.

Random forests (RF) are ensembles of classification or regression trees. Each tree is grown on a bootstrap sample of the data to create an ensemble of diverse trees, modelling different aspects of the data. Further randomization is introduced by drawing a random subset of variables at every split. We used regression RFs to predict EE (METs, absEE, or relEE) with raw acceleration data, based on the 30 summary statistics.

Artificial neural networks (ANNs), inspired by biological neural networks, are widely used to model complex relationships between inputs and outputs. In contrast to the LM and MLM, they do not assume a specific type of relationship between the outcome and covariates and allow for non-linear dependencies. In this study, ANNs were used to approximate the functional relationship between the response (METs, absEE, or relEE) and the 30 summary statistics as covariates.

All analyses were performed in R (R Core Team, 2017) with the add-on packages ‘lme4’ (Bates, Mächler, Bolker, & Walker, 2015) for MLM, ‘ranger’ (Wright & Ziegler, 2017) for RF, and ‘nnet’ (Venables & Ripley, 2002) for ANN. All model fitting, cross validation, parameter tuning, and evaluation was performed with the ‘mlr’ (Bischl et al., 2016) package. In the RF, we grew 500 trees

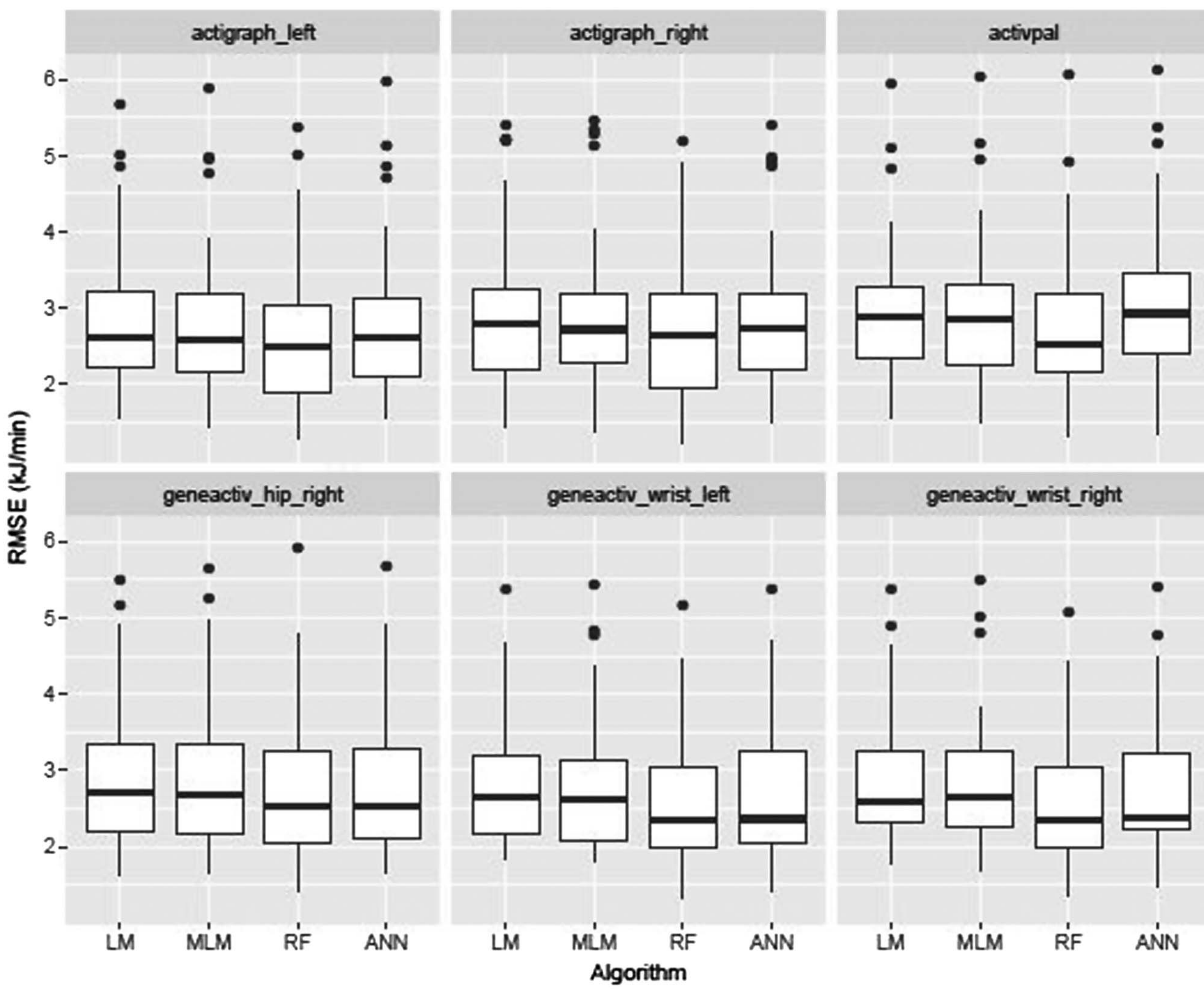

Figure 2 — absEE: Root mean square error (RMSE) for predicted versus measured absolute energy expenditure (absEE) for all accelerometers, locations, and models. Data are displayed in boxplots with median and interquartile range. Black dots represent outliers. LM = linear model; MLM = mixed linear model; RF = random forest; ANN = artificial neural network.

and tuned the parameter 'mtry' using model-based optimization (Bischl et al., 2017). For the ANN we chose a single hidden layer with 15 units and tuned the regularization parameter 'decay' to avoid overfitting. Code for using the developed models can be found at the following link: https://github.com/bips-hb/EE_prediction.

Statistical Analysis

Separately for each of the models, accelerometers, and positions, we calculated the root-mean-squared-error (RMSE) to test EE prediction accuracy. We tested all models using a leave-one-out cross-validation.

Results

We had written informed consent from parents of 62 children. Of these, 41 children completed the protocol. All of the participating children were Caucasians and were able to speak and understand German. Participant characteristics can be found in Table 2.

Reasons for non-participation were withdrawals of the consent on the measurement day (11), illness (5), not being fasted for at least two hours (2), withdrawn after the second activity (1), and missing data due to calorimeter failure (2). Children completed the protocol with a median of eight out of nine possible activities. Mean EE per activity ranged from 4.2 ± 0.9 kJ/min to 16.3 ± 4.6 kJ/min (Table 3).

Forty (98%) children had valid data for the GENEActiv on the left wrist and for the GT3X on the right hip. Thirty-nine (95%) children had valid data for the GENEActiv on the right wrist and for the GT3X on the left hip. Thirty-eight (93%) children had valid data for the GENEActiv on the right hip and for the activPAL. Predicted METs from the four different models (GENEActiv, left wrist) and METs derived from the indirect calorimeter of one child passing the protocol are exemplarily shown in Figure 1.

Boxplots for RMSE for all accelerometers, locations, and models are shown in Figures 2–4 for absEE, relEE, and METs, respectively.

Mean RMSE ranged from 2.56–2.76 kJ/min for the RF, from 2.72–3.08 kJ/min for the ANN, from 2.83–2.94 kJ/min for the LM, and from 2.81–2.92 kJ/min for the MLM (Table 4). A comparison

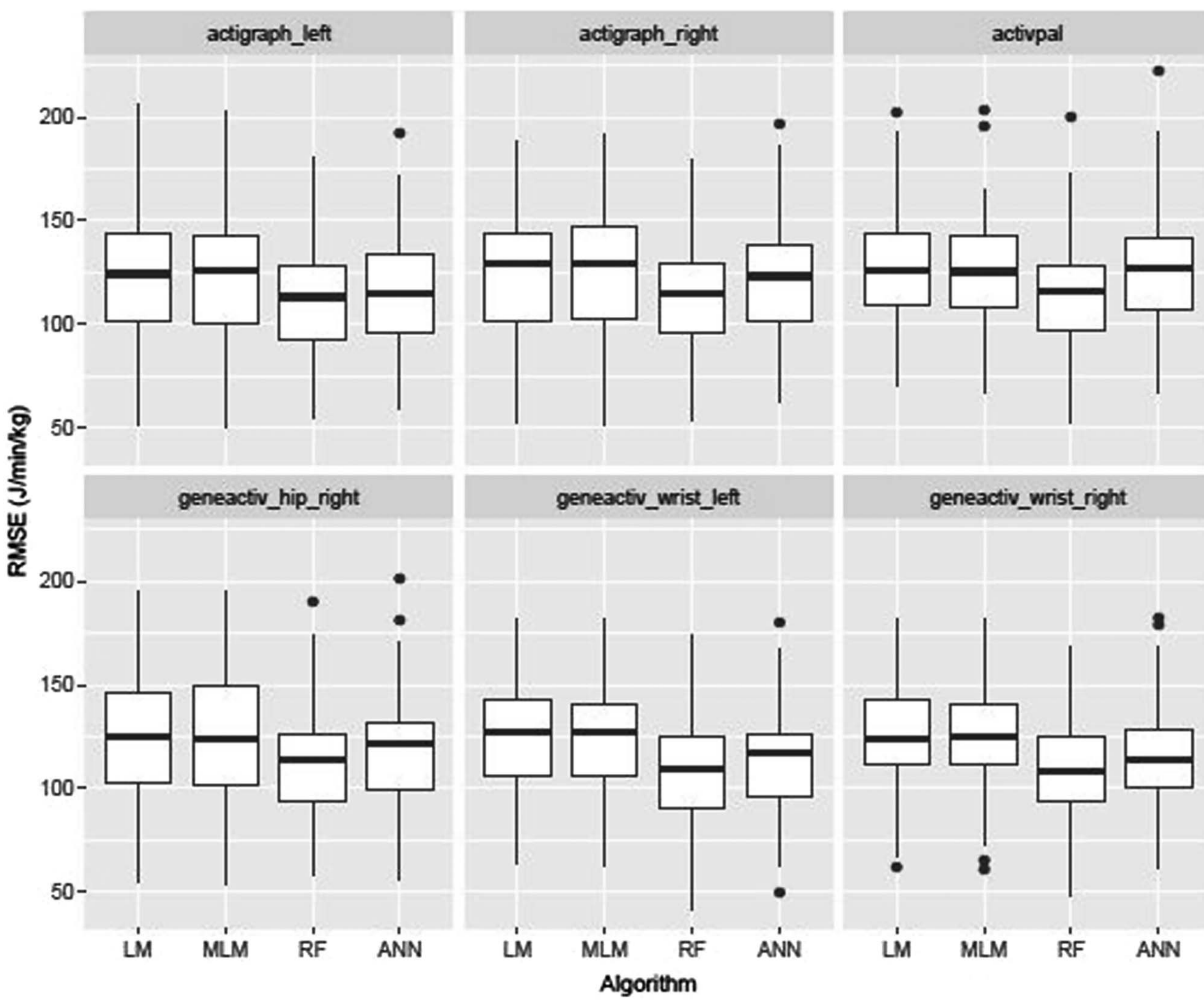

Figure 3 — relEE: Root mean square error (RMSE) for predicted versus measured relative energy expenditure (relEE) for all accelerometers, locations, and models. Data are displayed in boxplots with median and interquartile range. Black dots represent outliers. LM = linear model; MLM = mixed linear model; RF = random forest; ANN = artificial neural network.

of the four models shows that the RF revealed slightly lower RMSE for absEE, relEE and METs than the other three models. GENEActive devices obtained mean RMSE of 2.56 kJ/min (left and right wrist) and 2.73 kJ/min (right hip). Predicting EE using the GT3X on the left and right hip obtained mean RMSE of 2.60 and 2.74 kJ/min, respectively. Use of the thigh worn activPAL provided a mean RMSE of 2.76 kJ/min. Looking at wear positions, EE prediction accuracy for devices worn at the hip obtained mean RMSE of 2.60–2.74 kJ/min. Placement at the wrists obtained slightly lower mean RMSE (2.56 kJ/min), and placement at the thigh slightly higher mean RMSE (2.76 kJ/min) (Table 4).

Discussion

The study's purpose was to develop and to provide EE prediction models from raw accelerometry data, to compare two linear and two machine learning models, and to compare EE prediction accuracy of different accelerometers placed on the hips, thigh, and wrists. To our knowledge, this is the first study that provides a calibration and comparative validation of multiple accelerometers in preschool children aged three to six years against indirect calorimetry as the criterion measurement.

For adults, similar EE prediction accuracy has been reported for linear models vs. an ANN model using different accelerometers (Montoye et al., 2017). Only for wrist-worn devices, Montoye et al. (2017) report a significant improvement of the ANN over linear models. In children, there is preliminary evidence that the application of machine learning improves EE prediction accuracy compared to linear regression algorithms (Chowdhury et al., 2018; Mackintosh et al., 2016). In our study, machine learning models and (mixed) linear models were equally accurate in EE prediction irrespective of the device and wear position. This could be due to the fact, that our models all included the same summary statistics as independent variables whereas in the simplified approach by Chowdhury et al. (2018) only three features, including mean (x), mean (y), mean (z), were extracted from each accelerometer window. Thus our data shows preliminary evidence that application of an RF or ANN model leads to only minor improvements in EE prediction compared to (mixed) linear models if the models are built upon the same summary statistics as independent variables. We can therefore recommend using linear as well as

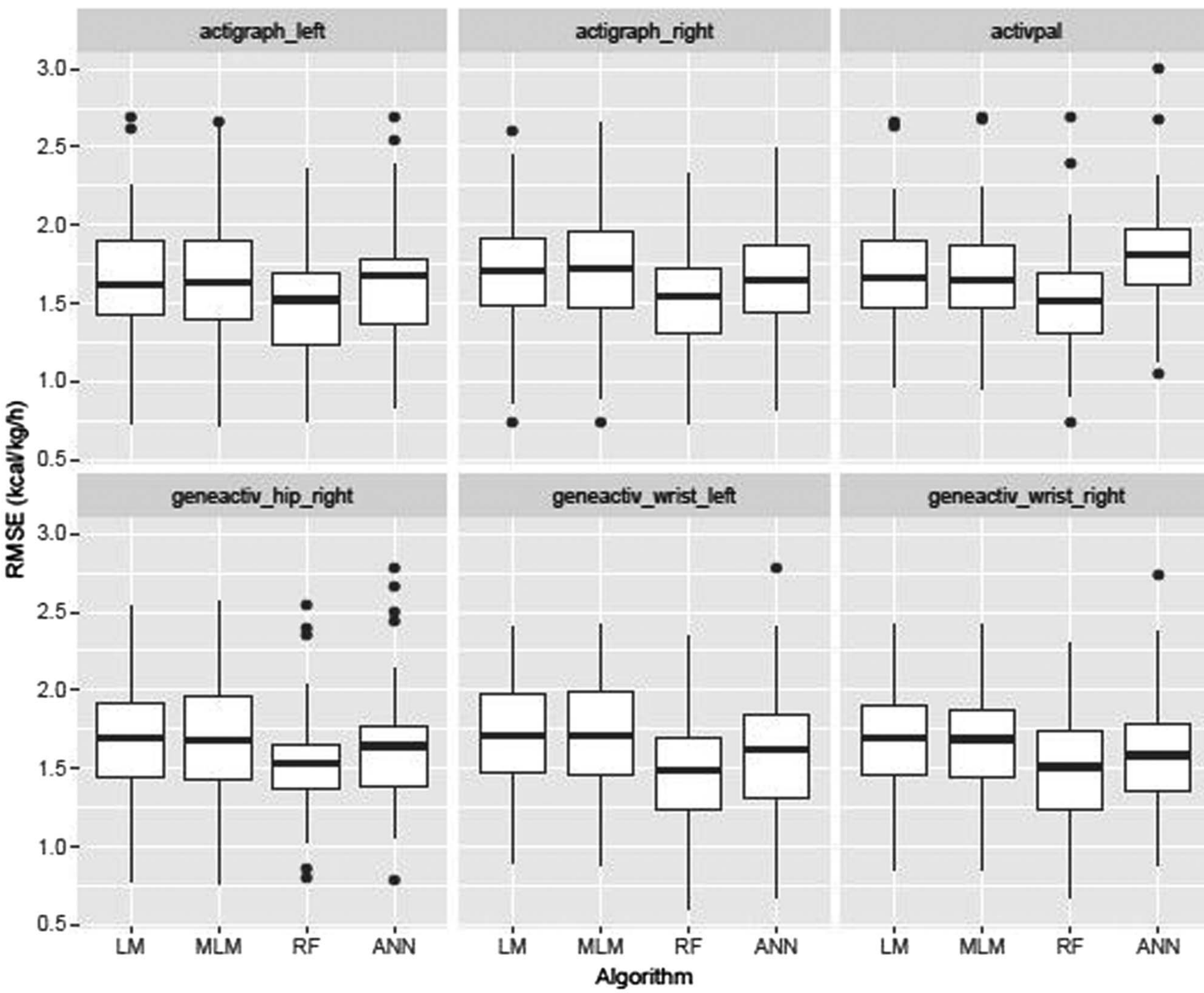

Figure 4 — METs: Root mean square error (RMSE) for predicted versus measured metabolic equivalents (METs) for all accelerometers, locations, and models. Data are displayed in boxplots with median and interquartile range. Black dots represent outliers. LM = linear model; MLM = mixed linear model; RF = random forest; ANN = artificial neural network.

non-linear models for EE prediction on raw accelerometry data in studies that aim at assessing young children's EE under free-living conditions. Furthermore, comparison of mean RMSE did not reveal meaningful differences between the three different devices. All of them can therefore be recommended for use in young children.

Accelerometers are commonly worn on a waist belt, aligned with the right anterior axillary line for the entire day up to seven days to estimate habitual PA (Hills et al., 2014). To date, most research has used the hip-worn GT3X, which has been calibrated and validated in a wide range of populations (Borghese et al., 2017; Evenson et al., 2015; Johansson, Larisch, Marcus, & Hagstromer, 2016). Compliance among young children has been shown to be better when using wrist-worn accelerometers compared to hip-worn devices (Fairclough et al., 2016). Therefore, our main aim was to test whether EE prediction of wrist-, hip- and thigh-worn devices is equally accurate for age-appropriate activities in preschoolers. Our results show good accuracy for all wear positions. This is in line with the findings of another study with children in which accelerometers mounted on various anatomical positions demonstrated equivalency in the accuracy to predict EE in a semi-structured setting (Mackintosh et al., 2016).

In our study, we further observed only small differences regarding the predictive accuracy between the dominant and non-dominant wrist. This also is in line with the findings of another study that examined the classification accuracy of the GENEActiv with a cut-point based approach for the assessment of PA intensities in preschoolers (Roscoe, James, & Duncan, 2017) and may help to inform choice of wrist placement when using accelerometers in field-based research. We assume good accuracy in predicting EE should the accelerometer either be placed on the left or right wrist.

There are several strengths of our study which are in line with best practice recommendations for calibration studies (Welk, 2005). We applied a semi-structured nature of the activity protocol, numerous accelerometer positions, and a high-quality criterion measure for EE. Our sample size of 41 children is in line with recent calibration and validation studies and the included activities enabled their differentiation into light- to vigorous intensity activities quite well, with the lowest activity-specific metabolic equivalents (AME) being observed in drawing [mean (95% CI): 1.89 (1.7–2.1)], a moderate AME in playing with cars [3.1 (2.5–3.6)] and walking [3.6 (3.3–4.0)], and the most vigorous AME in jogging [7.7 (6.7–8.7)] (Brandes, Steenbock, & Wirsik, 2018). The generalizability of our results is indeed limited by the homogenous sample of apparently healthy Caucasian children. Therefore, our findings cannot be generalized to other ethnicities or to children with a chronic condition or disease. A further limitation of the study is that not all children completed the five predetermined activities that stretch across all activity intensities (drawing, catching, walking at regular speed, walking fast, and jogging). This was due to the young age and motivational reasons in our target group.

Conclusion

These findings provide preliminary evidence that recent accelerometers mounted on five different anatomical positions demonstrate equivalency in the accuracy to predict EE from raw accelerometer data. Our models are publicly available and can be used for studies assessing EE in preschoolers under free-living conditions. The RF, ANN, and (mixed) linear models were equally accurate in predicting EE irrespective of the device and wear position. We therefore recommend utilizing linear as well as non-linear models for the estimation of EE in preschoolers. We recommend a further investigation of RF models for EE prediction from raw accelerometer data in young children as well as in other populations, which seems to be a viable alternative to linear and ANN models.

Table 4 Overview of Energy Expenditure Prediction Accuracy of the Different Models, Accelerometers, and Wear Positions

Mean Results: RMSE (SD)						
Absolute energy expenditure (kJ/min)						
Algorithm	GT3X left hip	GT3X right hip	activPAL	GENEActiv right hip	GENEActiv left wrist	GENEActiv right wrist
LM	2.84 (0.94)	2.91 (0.95)	2.94 (0.91)	2.89 (0.95)	2.83 (0.86)	2.85 (0.83)
MLM	2.81 (0.99)	2.91 (0.99)	2.92 (0.95)	2.90 (0.98)	2.83 (0.89)	2.83 (0.87)
RF	2.60 (0.97)	2.74 (0.96)	2.76 (0.94)	2.73 (1.00)	2.56 (0.83)	2.56 (0.83)
ANN	2.78 (1.01)	2.86 (0.95)	3.08 (1.00)	2.83 (0.98)	2.72 (0.91)	2.74 (0.88)
Relative energy expenditure (J/min/kg)						
Algorithm	GT3X left hip	GT3X right hip	activPAL	GENEActiv right hip	GENEActiv left wrist	GENEActiv right wrist
LM	123.64 (31.82)	124.91 (30.58)	126.22 (29.46)	123.79 (31.28)	125.21 (27.38)	125.60 (28.52)
MLM	123.58 (31.94)	125.40 (31.50)	125.96 (29.86)	124.70 (32.95)	124.91 (27.58)	125.42 (28.74)
RF	112.32 (28.40)	115.56 (27.35)	115.61 (27.93)	112.57 (28.83)	108.64 (26.33)	109.34 (26.98)
ANN	116.89 (28.41)	121.51 (29.57)	125.98 (31.50)	118.04 (29.80)	114.01 (27.88)	115.91 (28.55)
Metabolic equivalents (METs)						
Algorithm	GT3X left hip	GT3X right hip	activPAL	GENEActiv right hip	GENEActiv left wrist	GENEActiv right wrist
LM	1.67 (0.41)	1.70 (0.39)	1.69 (0.38)	1.68 (0.39)	1.70 (0.37)	1.69 (0.38)
MLM	1.67 (0.41)	1.70 (0.40)	1.69 (0.39)	1.69 (0.42)	1.69 (0.38)	1.69 (0.38)
RF	1.52 (0.38)	1.56 (0.36)	1.56 (0.38)	1.53 (0.38)	1.47 (0.36)	1.48 (0.37)
ANN	1.63 (0.42)	1.66 (0.39)	1.81 (0.40)	1.66 (0.43)	1.61 (0.44)	1.62 (0.40)

Note. LM = linear model; MLM = mixed linear model; RF = random forest; ANN = artificial neural network; RMSE = root-mean-square error.

Acknowledgments

The study was supported by an internal innovations grant of the Leibniz Institute for Prevention Research and Epidemiology (BIPS). We thank Johanna Sophie Lubasch und Marlena Böning for their assistance during data collection and all daycare facilities, daycare staff, parents and children for their support and participation in the study. The authors declare that they have no conflicts of interest.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., ... Jones, Z.M. (2016). mlr: Machine Learning in R. *The Journal of Machine Learning Research, 17*(170), 1–5.

Bischl, B., Richter, J., Bossek, J., Horn, D., Thomas, J., & Lang, M. (2017). *mlrMBO: A Modular Framework for Model-Based Optimization of expensive black-box functions*. Retrieved from http://arxiv.org/abs/1703.03373

Borges, J.H., Langer, R.D., Cirolini, V.X., Pascoa, M.A., Guerra-Junior, G., & Goncalves, E.M. (2016). Minimum time to achieve the steady state and optimum abbreviated period to estimate the resting energy expenditure by indirect calorimetry in healthy young adults. *Nutrition in Clinical Practice, 31*(3), 349–354. PubMed ID: 26888859

Borghese, M.M., Tremblay, M.S., LeBlanc, A.G., Leduc, G., Boyer, C., & Chaput, J.P. (2017). Comparison of ActiGraph GT3X+ and actical accelerometer data in 9–11-year-old Canadian children. *Journal of Sports Sciences, 35*(6), 517–524. PubMed ID: 27103499 doi:10.1080/02640414.2016.1175653

Brandes, M., Klein, H., Ginsel, S., & Heitmann, A. (2015). Comparability of three mobile respiratory gas analyzers. *Sportwissenschaft, 45*(4), 168–172. doi:10.1007/s12662-015-0368-x

Brandes, M., Steenbock, B., & Wirsik, N. (2018). Energy cost of common physical activities in preschoolers. *Journal of Physical Activity & Health, 15*(4), 233–238. PubMed ID: 29202634 doi:10.1123/jpah.2017-0348

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 & Science in Sports & Exercise, 46*(6), 1216–1226. PubMed ID: 24195866 doi:10.1249/mss.0000000000000209

Caspersen, C.J., Powell, K.E., & Christenson, G.M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Rep, 100*(2), 126–131. PubMed ID: 3920711

Chowdhury, A.K., Tjondronegoro, D., Zhang, J., Hagenbuchner, M., Cliff, D., & Trost, S.G. (2018). *Deep learning for energy expenditure prediction in pre-school children*. Paper presented at the IEEE Conference on Biomedical and Health Informatics, Las Vegas, NA.

Davies, G., Reilly, J., McGowan, A., Dall, P., Granat, M., & Paton, J. (2012). Validity, practical utility, and reliability of the activPAL in preschool children. *Medicine & Science in Sports & Exercise, 44*(4), 761–768. PubMed ID: 21983077

De Decker, E., De Craemer, M., Santos-Lozano, A., Van Cauwenberghe, E., De Bourdeaudhuij, I., & Cardon, G. (2013). Validity of the ActivPAL and the ActiGraph monitors in preschoolers. *Medicine & Science in Sports & Exercise, 45*(10), 2002–2011. PubMed ID: 23524516 doi:10.1249/MSS.0b013e318292c575

Evenson, K.R., Wen, F., Herring, A.H., Di, C., LaMonte, M.J., Tinker, L.F., ... Buchner, D.M. (2015). Calibrating physical activity intensity for hip-worn accelerometry in women age 60 to 91 years: The Women's Health Initiative OPACH Calibration Study. *Preventive Medicine Reports, 2*, 750–756. PubMed ID: 26527313 doi:10.1016/j.pmedr.2015.08.021

Fairclough, S.J., Noonan, R., Rowlands, A.V., Van Hees, V., Knowles, Z., & Boddy, L.M. (2016). Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. *Medicine & Science in Sports & Exercise, 48*(2), 245–253. PubMed ID: 26375253 doi:10.1249/mss.0000000000000771

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*(4), 426–431. PubMed ID: 25088983

Hills, A.P., Mokhtar, N., & Byrne, N.M. (2014). Assessment of physical activity and energy expenditure: An overview of objective measures. *Frontiers in Nutrition, 1*, 5. PubMed ID: 25988109 doi:10.3389/fnut.2014.00005

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*(11), e79124. PubMed ID: 24244433 doi:10.1371/journal.pone.0079124

Jimmy, G., Seiler, R., & Maeder, U. (2013). Development and validation of energy expenditure prediction models based on GT3X accelerometer data in 5- to 9-year-old children. *Journal of Physical Activity & Health, 10*(7), 1057–1067. PubMed ID: 23136371

Johansson, E., Larisch, L.M., Marcus, C., & Hagstromer, M. (2016). Calibration and validation of a wrist- and hip-worn actigraph accelerometer in 4-year-old children. *PLoS One, 11*(9), e0162436. PubMed ID: 27617962 doi:10.1371/journal.pone.0162436

Knaeps, S., Bourgois, J.G., Charlier, R., Mertens, E., Lefevre, J., & Wijndaele, K. (2018). Ten-year change in sedentary behaviour, moderate-to-vigorous physical activity, cardiorespiratory fitness and cardiometabolic risk: Independent associations and mediation analysis. *British Journal of Sports Medicine, 52*(16), 1063–1068. PubMed ID: 27491779 doi:10.1136/bjsports-2016-096083

Lamonte, M.J., & Ainsworth, B.E. (2001). Quantifying energy expenditure and physical activity in the context of dose response. *Medicine & Science in Sports & Exercise, 33*(Suppl. 6), S370–S378. discussion S419–S320. PubMed ID: 11427762

Macfarlane, D.J., & Wong, P. (2012). Validity, reliability and stability of the portable cortex metamax 3B gas analysis system. *European Journal of Applied Physiology, 112*(7), 2539–2547. PubMed ID: 22075643 doi:10.1007/s00421-011-2230-7

Mackintosh, K.A., Montoye, A.H., Pfeiffer, K.A., & McNarry, M.A. (2016). Investigating optimal accelerometer placement for energy expenditure prediction in children using a machine learning approach. *Physiological Measurement, 37*(10), 1728–1740. PubMed ID: 27653339 doi:10.1088/0967-3334/37/10/1728

Migueles, J.H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., ... Ortega, F.B. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: A systematic review and practical considerations. *Sports Medicine, 47*(9), 1821–1845. PubMed ID: 28303543 doi:10.1007/s40279-017-0716-0

Montoye, A.H., Begum, M., Henning, Z., & Pfeiffer, K.A. (2017). Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiological Measurement, 38*(2), 343–357. PubMed ID: 28107205 doi:10.1088/1361-6579/38/2/343

Montoye, A.H., Mudd, L.M., Biswas, S., & Pfeiffer, K.A. (2015). Energy expenditure prediction using raw accelerometer data in simulated free living. *Medicine & Science in Sports & Exercise, 47*(8), 1735–1746. PubMed ID: 25494392 doi:10.1249/mss.0000000000000597

Nilsson, A., Brage, S., Riddoch, C., Anderssen, S.A., Sardinha, L.B., Wedderkopp, N., . . . Ekelund, U. (2008). Comparison of equations for predicting energy expenditure from accelerometer counts in children. *Scandinavian Journal of Medicine & Science in Sports, 18*(5), 643–650. PubMed ID: 18208433 doi:10.1111/j.1600-0838.2007.00694.x

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reddon, H., Meyre, D., & Cairney, J. (2017). Physical activity and global self-worth in a longitudinal study of children. *Medicine & Science in Sports & Exercise, 49*(8), 1606–1613. PubMed ID: 28319588 doi:10.1249/mss.0000000000001275

Reilly, J.J., Kelly, L.A., Montgomery, C., Jackson, D.M., Slater, C., Grant, S., & Paton, J.Y. (2006). Validation of Actigraph accelerometer estimates of total energy expenditure in young children. *International Journal of Pediatric Obesity, 1*(3), 161–167. PubMed ID: 17899634

Roscoe, C.M.P., James, R.S., & Duncan, M.J. (2017). Calibration of GENEActiv accelerometer wrist cut-points for the assessment of physical activity intensity of preschool aged children. *European Journal of Pediatrics, 176*(8), 1093–1098. PubMed ID: 28674825 doi:10.1007/s00431-017-2948-2

Shoup, J.A., Gattshall, M., Dandamudi, P., & Estabrooks, P. (2008). Physical activity, quality of life, and weight status in overweight children. *Quality of Life Research, 17*(3), 407–412. PubMed ID: 18293100 doi:10.1007/s11136-008-9312-y

Skotte, J., Korshoj, M., Kristiansen, J., Hanisch, C., & Holtermann, A. (2014). Detection of physical activity types using triaxial accelerometers. *Journal of Physical Activity & Health, 11*(1), 76–84. PubMed ID: 23249722 doi:10.1123/jpah.2011-0347

Skrede, T., Stavnsbo, M., Aadland, E., Aadland, K.N., Anderssen, S.A., Resaland, G.K., & Ekelund, U. (2017). Moderate-to-vigorous physical activity, but not sedentary time, predicts changes in cardiometabolic risk factors in 10-y-old children: The active smarter kids study. *The American Journal of Clinical Nutrition, 105*(6), 1391–1398. PubMed ID: 28381476 doi:10.3945/ajcn.116.150540

Venables, W.N., & Ripley, B.D. (2002). *Modern Applied Statistics with S* (4th ed.). New York, NY: Springer.

Wafa, S.W., Shahril, M.R., Ahmad, A.B., Zainuddin, L.R., Ismail, K.F., Aung, M.M., & Mohd Yusoff, N.A. (2016). Association between physical activity and health-related quality of life in children: A cross-sectional study. *Health and Quality of Life Outcomes, 14*, 71. PubMed ID: 27146199 doi:10.1186/s12955-016-0474-y

Weir, J.B. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of Physiology, 109*(1–2), 1–9. PubMed ID: 15394301

Welk, G.J. (2005). Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S501–S511. PubMed ID: 16294113

Wright, M.N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1–17. doi:10.18637/jss.v077.i01