Accelerometry-based prediction of energy expenditure in preschoolers

Berit Steenbock, Marvin N. Wright, Norman Wirsik, Mirko Brandes

This is the author's final accepted version. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

Accelerometry based prediction of energy expenditure in preschoolers

Date of submission: 24.07.2018

Berit Steenbock[1,2], Marvin N. Wright[1], Norman Wirsik[1], Mirko Brandes[1]

[1]Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

[2]Center of Clinical Psychology and Rehabilitation, University of Bremen, Bremen, Germany

Corresponding author

Mirko Brandes, Ph.D.

Leibniz Institute for Prevention Research and Epidemiology – BIPS

Achterstrasse 30

28359 Bremen

Germany

Phone: +49 421 218 56905

Fax: +49 421 218 56941

brandes@leibniz-bips.de

Berit Steenbock, M.Sc.

Center of Clinical Psychology and Rehabilitation, University of Bremen

Grazer Strasse 6

28359 Bremen

Germany

steenbock@uni-bremen.de

Marvin N. Wright, Ph.D.

Leibniz Institute for Prevention Research and Epidemiology – BIPS

Achterstrasse 30

28359 Bremen

Germany

wright@leibniz-bips.de

Norman Wirsik, Ph.D.

Leibniz Institute for Prevention Research and Epidemiology – BIPS

Achterstrasse 30

28359 Bremen

Germany

wirsik@leibniz-bips.de

1   **Accelerometry based prediction of energy expenditure in preschoolers**

2   Date of submission: 24.07.2018

3

4   **Abstract**

5   **Purpose**

6   Study purposes were to develop energy expenditure (EE) prediction models from raw

7   accelerometer data and to investigate the performance of three different accelerometers on

8   five different wear positions in preschoolers.

9   **Methods**

10   Fourty-one children (54% boys; 3-6.3 years) wore two Actigraph GT3X (left and right hip),

11   three GENEActiv (right hip, left and right wrist), and one activPAL (right thigh) while

12   completing a semi-structured protocol of 10 age-appropriate activities. Participants wore a

13   portable indirect calorimeter to estimate EE. Utilized models to estimate EE included a linear

14   model (LM), a mixed linear model (MLM), a random forest model (RF), and an artificial

15   neural network model (ANN). For each accelerometer, model and wear position, we assessed

16   prediction accuracy via leave-one-out cross-validation and calculated the root-mean-squared-

17   error (RMSE).

18   **Results**

19   Mean RMSE ranged from 2.56–2.76 kJ/min for the RF, from 2.72-3.08 kJ/min for the ANN,

20   from 2.83–2.94 kJ/min for the LM, and from 2.81-2.92 kJ/min for the MLM. The

21   GENEActive obtained mean RMSE of 2.56 kJ/min (left and right wrist) and 2.73 kJ/min

22   (right hip). Predicting EE using the GT3X on the left and right hip obtained mean RMSE of

23   2.60 and 2.74 kJ/min.The activPAL obtained a mean RMSE of 2.76 kJ/min.

24   **Conclusion**

25   These results demonstrate good prediction accuracy for recent accelerometers on different

26   wear positions in preschoolers. The RF and ANN were equally accurate in EE prediction

27 compared with (mixed) linear models. The RF seems to be a viable alternative to linear and

28 ANN models for EE prediction in young children in a semi-structured setting.

29 **Key words** accelerometer, physical activity, validation, linear mixed model, machine

30 learning, children

31   **Abbreviations**

32   absEE absolute energy expenditure

33   AEE   Activity-related energy expenditure

34   ANN   Artificial neural network

35   EE   Energy expenditure

36   BMI   Body mass index

37   LM   Linear model

38   MET   Metabolic equivalent

39   MIN   Minute

40   MLM   Mixed linear model

41   PA   Physical activity

42   REE   Resting energy expenditure

43   relEE relative energy expenditure

44   RF   Random forest

45   RMR   Resting metabolic rate

46   RMSE Root-mean-square error

47

Introduction

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

73 data collection (e.g. device placement) and processing criteria (e.g. epoch length, cut-points,

74 and algorithms), Migueles and colleagues (2017) observed a lack of calibration and validation

75 studies for preschoolers that address important processing criteria (such as EE algorithms for

76 wrist- and hip-worn accelerometers). However, as studies included in the review were

77 restricted to those applying the latest version of the Actigraph device (GT3X), no practical

78 consideration about device selection in this age group could be drawn. Besides this, most of

79 the research on EE prediction that has been done in preschoolers is limited by the use of direct

80 observation as the criterion measurement and the assignment of fixed metabolic equivalents

81 (METs) to activities and accelerometer output (Davies et al., 2012; De Decker et al., 2013;

82 Hagenbuchner, Cliff, Trost, Van Tuc, & Peoples, 2015). Additionally, the use of highly

83 structured protocols under laboratory settings has been found to overestimate EE in children,

84 which limits the transfer to free-living behaviors (Nilsson et al., 2008).

85 Recent studies in older children and adults show improvements in EE prediction using non-

86 linear models (Mackintosh, Montoye, Pfeiffer, & McNarry, 2016; Montoye et al., 2017).

87 Preliminary evidence demonstrates that machine learning models outperform simplified

88 regression models in preschoolers (Chowdhury et al., 2018).

89 This study had three purposes related to accelerometer use in preschoolers: (1) to provide

90 energy expenditure (EE) prediction models from raw accelerometry data established against

91 indirect calorimetry, (2) to compare two linear and two machine learning models, and (3) to

92 compare accuracy of different accelerometers placed on the hips, thigh, and wrists.

93 **Methods**

94 *Study participants*

95 To recruit three to six years old children from four daycare facilities (kindergartens) in Lower

96 Saxony, we first contacted the administration of the kindergartens. In the next step, we

97 explained the aim and the procedures of the study to the administration as well as to parent

98 representatives during a pre-arranged meeting in each kindergarten. Thereafter, study

99 information material was distributed to all parents. In the course of regular kindergarten

100 morning circle time, we demonstrated the indirect calorimetry to all children. Children with

101 written parental informed consent were allocated to a schedule. On the measurement day, we

102 asked for verbal assent from each child and those refusing to take part in the study were

103 excluded. The local ethics committee (University of Bremen, Bremen) approved the

104 procedures of the study (ethics committee running title: accelerometry in preschoolers

105 (ACCIPS)).

106 *Study design*

107 The measurements took place from October to November 2016 and were conducted at each

108 respective kindergarten, using the available indoor and outdoor infrastructure. The children

109 had to fast for at least two hours before measurement, but were allowed to drink water. The

110 measurement was restricted to two children per morning, measured one after the other, to

111 avoid interference with lunchtime. A total of 75 min were allocated for the measurement

112 protocol of one child. This included the assessment of anthropometrics, handedness, mounting

113 and demounting the devices, measurement of resting metabolic rate (RMR), performance of

114 up to nine in- and outdoor activities, changing rooms within the kindergarten, and dressing

115 and undressing for outdoor activities (Table 1). We measured RMR during 10 min of supine

116 rest in a quiet, dimly lit room. The child lay on a mattress and was provided with a blanket. A

117 short learning story was shown on a tablet to keep the child calm for the measurement of

118 RMR. We decided on a measurement period of 10 minutes as Borges et al. (2016) showed

119 that this is an appropriate length of time to achieve steady state conditions and delineate an

120 optimum abbreviated period to estimate RMR by indirect calorimetry.

121 ***** Table 1 here *****

122 *Activities*

123 The child was free to choose up to three friends to play with. Five activities were

124 predetermined and ranged from light (e.g. drawing) to vigorous intensities (e.g. jogging). An

125 additional four activities to be chosen independently with respect to the resources of the

126 daycare facility and the child's individual preferences were offered to each child. Examples

127 are given in Table 1. Starting with the first predetermined activity 'drawing', the child was

128 asked to sit down for five minutes at a table in a group room and to choose one out of five

129 colouring pages with different motifs to crayon. All subsequent activities were performed

130 continuously for at least three minutes. Additionally, the child could choose two indoor

131 physical activities (e.g. playing with toy cars, toy blocks, hide and seek, playing with dolls).

132 For the performance of the outdoor activities, a well-defined and plain area was chosen. On

133 the outdoors, the child started with a predetermined activity, which was 'tag'. Afterwards, the

134 child chose two other outdoor activities, depending on the equipment of each of the four local

135 daycare facilities (e.g. swing, tricycle, monkey bars, or scooter). Finally, the child performed

136 the three activities scheduled last, which were walking at normal speed, walking fast, and

137 jogging. During these activities, the researcher accompanied the child to ensure continuous

138 walking and jogging, while the child set the pace. For fast walking, the researcher encouraged

139 the child by telling an imaginary journey of a holiday trip, including catching the tram, train

140 and plane to get to a holiday location. If the child changed to jogging during the fast walking

141 task, the researcher immediately slowed the child down to fast walking. If the child refused to

142 perform an activity continuously due to motivational reasons, the activity was stopped and the

143 child was asked to continue with the next activity. The protocol was completely aborted if the

144 child was exhausted or refused to go on any more.

145 *Accelerometers*

146 The children were equipped with six accelerometers of three different brands. One

147 accelerometer was placed at the left and the right wrist (GENEActiv, ActivInsight Ltd,

148 Kimbolton, UK), one accelerometer at the left hip (GT3X Actigraph, Pensacola, Florida,

149 USA), two accelerometers at the right hip (GENEActiv, GT3X), and one accelerometer at the

150 right thigh (activPAL, PAL Technologies Ltd, Glasgow, Scotland, UK). Accelerometers at

151 the wrist were firmly mounted with ¾ inch Tyvek security bracelets, which were used instead

152 of the standard bracelets provided by the manufacturer. An elastic belt was used to fix the hip

153 worn monitors close to the lateral hip bone, and an adhesive pad was used to fix the activPAL

154 to the skin on the front of the right thigh. All accelerometers were operated with the highest

155 possible resolution, which was 100Hz in the GENEActiv and the GT3X, and 20Hz in the

156 activPAL. Before each measurement we initialized and synchronized the accelerometers with

157 the computer running the indirect calorimetry. After completion of the protocol we

158 downloaded accelerometer data from all devices using dedicated software from the

159 manufacturers.

160 *Physiological measures*

161 A portable, open-circuit indirect calorimetry system (MetaMax3b, Cortex Biophysics,

162 Leipzig, Germany) was used as the criterion measure. The MetaMax3b has proven reliability

163 in a study with adolescents but was found to slightly overestimate VO_2 during moderate and

164 vigorous exercise (Macfarlane & Wong, 2012). It is highly comparable to other common

165 devices (Mirko Brandes, Klein, Ginsel, & Heitmann, 2015). The system was mounted to the

166 children using a pediatric harness. In very small children, additional adhesive tape was used to

167 adjust the harness to the child, so that the system did not interfere with the activities. A

168 facemask was secured over the child´s nose and mouth with an adjustable nylon harness. A

169 bidirectional turbine, inserted to the facemask, measured the volume of inspired and expired

170 air. A sample tube, connected to the turbine, retrieved expired air samples breath-by-breath.

171 Air samples were analysed for oxygen uptake and carbon dioxide production within the

172 sensor unit of the system. Stable and dry weather conditions enabled sound outdoor

173 measurements. Data were transferred via telemetry to a laptop and available in real-time

174 (MetaSoft 3, Cortex Biophysics, Leipzig, Germany). The laptop operator used the marker

175 function of the software to identify the beginning and the end of each activity. Each morning,

176 prior to the first measurement, we calibrated the indirect calorimetry system according to the

177 manufacturer recommended procedure.

178 *Data reduction and modelling*

179 Oxygen consumption and carbon dioxide production were measured continuously and

180 converted to absolute (kJ/min) and relative (J/min/kg) EE using the equations by Weir (1949).

181 RMR was calculated as the minimum of a rolling one-minute mean during supine rest.

182 Metabolic equivalents (METs) were calculated by dividing the relative oxygen consumption

183 (expressed in ml/min/kg) by 3.5.

184 As done by Montoye et al. (2017), for each individual and device, the following 10 summary

185 statistics were calculated in 30s non-overlapping windows for each axis (X, Y, Z) of the raw

186 acceleration data: mean, standard deviation, minimum, maximum, lag one autocorrelation,

187 percentiles (10%, 25%, 50%, 75%, 90%), resulting in 30 summary statistics for each time

188 window. No filtering of the raw accelerometer data was conducted.

189 Predictive models were created separately for each accelerometer and placement, resulting in

190 24 (4x6) models developed and tested. We built four predictive models: (1) a linear regression

191 model, (2) a linear mixed model, (3) a random forest, and (4) an artificial neural network

192 model. Linear regression models (LM) and linear mixed models (MLM) were created

193 separately for the outcome variables absolute EE (absEE), relative EE (relEE), and METs and

194   included the above mentioned 30 summary statistics as independent variables. The MLM

195   accounts for repeated measurements to compute EE prediction.

196   Random forests (RF) are ensembles of classification or regression trees. Each tree is grown on

197   a bootstrap sample of the data to create an ensemble of diverse trees, modelling different

198   aspects of the data. Further randomization is introduced by drawing a random subset of

199   variables at every split. We used regression RFs to predict EE (METs, absEE or relEE) with

200   raw acceleration data, based on the 30 summary statistics.

201   Artificial neural networks (ANNs), inspired by biological neural networks, are widely used to

202   model complex relationships between inputs and outputs. In contrast to the LM and MLM,

203   they do not assume a specific type of relationship between the outcome and covariates and

204   allow for non-linear dependencies. In this study, ANNs were used to approximate the

205   functional relationship between the response (METs, absEE or relEE) and the 30 summary

206   statistics as covariates.

207   All analyses were performed in R (R Core Team, 2017) with the add-on packages 'lme4'

208   (Bates, Mächler, Bolker, & Walker, 2015) for MLM, 'ranger' (Wright & Ziegler, 2017) for

209   RF and 'nnet' (Venables & Ripley, 2002) for ANN. All model fitting, cross validation,

210   parameter tuning and evaluation was performed with the 'mlr' (Bischl et al., 2016) package.

211   In the RF, we grew 500 trees and tuned the parameter 'mtry' using model-based optimization

212   (Bischl et al., 2017). For the ANN we chose a single hidden layer with 15 units and tuned the

213   regularization parameter 'decay' to avoid overfitting. Code for using the developed models

214   can be found at the following link: https://github.com/bips-hb/EE_prediction .

215   *Statistical analysis*

216   Separately for each of the models, accelerometers, and positions, we calculated the root-

217   mean-squared-error (RMSE) to test EE prediction accuracy. We tested all models using a

218   leave-one-out cross-validation.

219

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.

*** Table 2 here ***

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 $kJ*m^{-1}$ to 16.3 ± 4.6 $kJ*m^{-1}$ (Table 3).

*** Table 3 here ***

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.

*** Figure 1 here ***

Boxplots for RMSE for all accelerometers, locations and models are shown in Figure 2, 3 and 4 for absEE, relEE and METs, respectively.

*** Figure 2, 3 and 4 here ***

242   Mean RMSE ranged from 2.56–2.76 kJ/min for the RF, from 2.72-3.08 kJ/min for the ANN,

243   from 2.83–2.94 kJ/min for the LM, and from 2.81-2.92 kJ/min for the MLM (Table 4). A

244   comparison of the four models shows that the RF revealed slightly lower RMSE for absEE,

245   relEE and METs than the other three models. GENEActive devices obtained mean RMSE of

246   2.56 kJ/min (left and right wrist) and 2.73 kJ/min (right hip). Predicting EE using the GT3X

247   on the left and right hip obtained mean RMSE of 2.60 and 2.74 kJ/min, respectively. Use of

248   the thigh worn activPAL provided a mean RMSE of 2.76 kJ/min. Looking at wear positions,

249   EE prediction accuracy for devices worn at the hip obtained mean RMSE of 2.60-2.74 kJ/min.

250   Placement at the wrists obtained slightly lower mean RMSE (2.56 kJ/min), and placement at

251   the thigh slightly higher mean RMSE (2.76 kJ/min) (Table 4).

252

253   ***** Table 4 here *****

254
255
256   **Discussion**

257   The study's purpose was to develop and to provide EE prediction models from raw

258   accelerometry data, to compare two linear and two machine learning models, and to compare

259   EE prediction accuracy of different accelerometers placed on the hips, thigh, and wrists. To

260   our knowledge, this is the first study that provides a calibration and comparative validation of

261   multiple accelerometers in preschool children aged three to six years against indirect

262   calorimetry as the criterion measurement.

263   For adults, similar EE prediction accuracy has been reported for linear models vs. an ANN

264   model using different accelerometers (Montoye et al., 2017). Only for wrist-worn devices,

265   Montoye et al. (2017) report a significant improvement of the ANN over linear models. In

266   children, there is preliminary evidence that the application of machine learning improves EE

267   prediction accuracy compared to linear regression algorithms (Chowdhury et al., 2018;

268 Mackintosh et al., 2016). In our study, machine learning models and (mixed) linear models

269 were equally accurate in EE prediction irrespective of the device and wear position. This

270 could be due to the fact, that our models all included the same summary statistics as

271 independent variables whereas in the simplified approach by Chowdhury et al. (2018) only

272 three features, including mean(x), mean(y), mean(z), were extracted from each accelerometer

273 window. Thus our data shows preliminary evidence that application of a RF or ANN model

274 revealed only minor improvements in EE prediction compared to (mixed) linear models

275 should the models are built upon the same summary statistics as independent variables. We

276 can therefor recommend using linear as well as non-linear models for EE prediction on raw

277 accelerometry data in studies that aim at assessing young children's EE under free-living

278 conditions.

279 Furthermore, comparison of mean RMSE did not reveal meaningful differences between the

280 three different devices. All of them can therefore be recommended for use in young children.

281 Accelerometers are commonly worn on a waist belt, aligned with the right anterior axillary

282 line for the entire day up to 7 days to estimate habitual PA (Hills et al., 2014). To date, most

283 research has used the hip-worn GT3X, which has been calibrated and validated in a wide

284 range of populations (Borghese et al., 2017; Evenson et al., 2015; Johansson, Larisch,

285 Marcus, & Hagstromer, 2016). Compliance among young children has been shown to be

286 better when using wrist-worn accelerometers compared to hip-worn devices (Fairclough et al.,

287 2016). Therefore, our main aim was to test whether EE prediction of wrist-, hip- and thigh-

288 worn devices is equally accurate for age-appropriate activities in preschoolers. Our results

289 show good accuracy for all wear positions. This is in line with the findings of another study

290 with children in which accelerometers mounted on various anatomical positions demonstrated

291 equivalency in the accuracy to predict EE in a semi-structured setting (Mackintosh et al.,

292 2016).

293   In our study, we further observed only small differences regarding the predictive accuracy

294   between the dominant and non-dominant wrist. This also is in line with the findings of

295   another study that examined the classification accuracy of the GENEActiv with a cut-point

296   based approach for the assessment of PA intensities in preschoolers (Roscoe, James, &

297   Duncan, 2017) and may help to inform choice of wrist placement when using accelerometers

298   in field-based research. We assume good accuracy in predicting EE should the accelerometer

299   either be placed on the left or right wrist.

300   There are several strengths of our study which are in line with best practice recommendations

301   for calibration studies (Welk, 2005). We applied a semi-structured nature of the activity

302   protocol, numerous accelerometer positions, and a high-quality criterion measure for EE. Our

303   sample size of 41 children is in line with recent calibration and validation studies and the

304   included activities enabled their differentiation into light- to vigorous intensity activities quite

305   well, with the lowest activity-specific metabolic equivalents (AME) being observed in

306   drawing [mean (95% CI): 1.89 (1.7–2.1)], a moderate AME in playing with cars [3.1 (2.5–

307   3.6)] and walking [3.6 (3.3–4.0)], and the most vigorous AME in jogging [7.7 (6.7–8.7)] (M.

308   Brandes, Steenbock, & Wirsik, 2018). The generalizability of our results is indeed limited by

309   the homogenous sample of apparently healthy Caucasian children. Therefore, our findings

310   cannot be generalized to other ethnicities or to children with a chronic condition or disease. A

311   further limitation of the study is that not all children completed the five predetermined

312   activities that stretch across all activity intensities (drawing, catching, walking at regular

313   speed, walking fast, and jogging). This was due to the young age and motivational reasons in

314   our target group.

315   **Conclusion**

316   These findings provide preliminary evidence that recent accelerometers mounted on five

317   different anatomical positions demonstrate equivalency in the accuracy to predict EE from

318 raw accelerometer data. Our models are publicly available and can be used for studies
319 assessing EE in preschoolers under free-living conditions. The RF, ANN and (mixed) linear
320 models were equally accurate in predicting EE irrespective of the device and wear position.
321 We therefor recommend utilizing linear as well as non-linear models for the estimation of EE
322 in preschoolers. We recommend a further investigation of RF models for EE prediction from
323 raw accelerometer data in young children as well as in other populations, which seems to be a
324 viable alternative to linear and ANN models.

325

Acknowledgements

331

Conflict of Interest

333 The authors declare that they have no conflicts of interest.

334

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.

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. 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 and Health, 15*(4), 233-238. 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 and Science in Sports and Exercise, 46*(6), 1216-1226. 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.

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 and Science in Sports and Exercise, 44*(4), 761-768.

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 and Science in Sports and Exercise, 45*(10), 2002-2011. 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. 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 and Science in Sports and Exercise, 48*(2), 245-253. 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.

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. doi: 10.3389/fnut.2014.00005

385 **Janssen, X., Cliff, D. P., Reilly, J. J., Hinkley, T., Jones, R. A., Batterham, M., . . . Okely, A. D. (2013).**
386 **Predictive validity and classification accuracy of ActiGraph energy expenditure equations**
387 **and cut-points in young children.** *PLoS One, 8*(11), e79124. doi:
388 **10.1371/journal.pone.0079124**

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

392 **Johansson, E., Larisch, L. M., Marcus, C., & Hagstromer, M. (2016). Calibration and Validation of a**
393 **Wrist- and Hip-Worn Actigraph Accelerometer in 4-Year-Old Children.** *PLoS One, 11*(9),
394 **e0162436. doi: 10.1371/journal.pone.0162436**

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

399 **Lamonte, M. J., & Ainsworth, B. E. (2001). Quantifying energy expenditure and physical activity in**
400 **the context of dose response.** *Medicine and Science in Sports and Exercise, 33*(6 Suppl),
401 **S370-378; discussion S419-320.**

402 **Macfarlane, D. J., & Wong, P. (2012). Validity, reliability and stability of the portable Cortex**
403 **Metamax 3B gas analysis system.** *European Journal of Applied Physiology, 112*(7), 2539-
404 **2547. doi: 10.1007/s00421-011-2230-7**

405 **Mackintosh, K. A., Montoye, A. H., Pfeiffer, K. A., & McNarry, M. A. (2016). Investigating optimal**
406 **accelerometer placement for energy expenditure prediction in children using a machine**
407 **learning approach.** *Physiological Measurement, 37*(10), 1728-1740. doi: 10.1088/0967-
408 **3334/37/10/1728**

409 **Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nystrom, C., Mora-Gonzalez, J., Lof, M., . .**
410 **. Ortega, F. B. (2017). Accelerometer Data Collection and Processing Criteria to Assess**
411 **Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations.**
412 *Sports Medicine, 47*(9), 1821-1845. doi: 10.1007/s40279-017-0716-0

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

416 **Montoye, A. H., Mudd, L. M., Biswas, S., & Pfeiffer, K. A. (2015). Energy Expenditure Prediction**
417 **Using Raw Accelerometer Data in Simulated Free Living.** *Medicine and Science in Sports*
418 *and Exercise, 47*(8), 1735-1746. doi: 10.1249/mss.0000000000000597

419 **Nilsson, A., Brage, S., Riddoch, C., Anderssen, S. A., Sardinha, L. B., Wedderkopp, N., . . . Ekelund, U.**
420 **(2008). Comparison of equations for predicting energy expenditure from accelerometer**
421 **counts in children.** *Scandinavian Journal of Medicine & Science in Sports, 18*(5), 643-650.
422 **doi: 10.1111/j.1600-0838.2007.00694.x**

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

425 **Reddon, H., Meyre, D., & Cairney, J. (2017). Physical Activity and Global Self-worth in a**
426 **Longitudinal Study of Children.** *Medicine and Science in Sports and Exercise, 49*(8), 1606-
427 **1613. doi: 10.1249/mss.0000000000001275**

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

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

434 **Shoup, J. A., Gattshall, M., Dandamudi, P., & Estabrooks, P. (2008). Physical activity, quality of life,**
435 **and weight status in overweight children.** *Quality of Life Research, 17*(3), 407-412. doi:
436 **10.1007/s11136-008-9312-y**

437   Skotte, J., Korshoj, M., Kristiansen, J., Hanisch, C., & Holtermann, A. (2014). Detection of physical
438   activity types using triaxial accelerometers. *Journal of Physical Activity and Health, 11*(1),
439   76-84. doi: 10.1123/jpah.2011-0347

440   Skrede, T., Stavnsbo, M., Aadland, E., Aadland, K. N., Anderssen, S. A., Resaland, G. K., & Ekelund,
441   U. (2017). Moderate-to-vigorous physical activity, but not sedentary time, predicts changes
442   in cardiometabolic risk factors in 10-y-old children: the Active Smarter Kids Study. *The*
443   *American Journal of Clinical Nutrition, 105*(6), 1391-1398. doi: 10.3945/ajcn.116.150540

444   Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (fourth ed.). New York:
445   Springer.

446   Wafa, S. W., Shahril, M. R., Ahmad, A. B., Zainuddin, L. R., Ismail, K. F., Aung, M. M., & Mohd
447   Yusoff, N. A. (2016). Association between physical activity and health-related quality of life
448   in children: a cross-sectional study. *Health and Quality of Life Outcomes, 14*, 71. doi:
449   10.1186/s12955-016-0474-y

450   Weir, J. B. (1949). New methods for calculating metabolic rate with special reference to protein
451   metabolism. *The Journal of Physiology, 109*(1-2), 1-9.

452   Welk, G. J. (2005). Principles of design and analyses for the calibration of accelerometry-based
453   activity monitors. *Medicine and Science in Sports and Exercise, 37*(11 Suppl), S501-511.

454   Wright, M. N., & Ziegler, A. (2017). Ranger: a fast implementation of random forests for high
455   dimensional data in C++ and R. *Journal of Statistical Software, 77*(1), 1-17. doi:
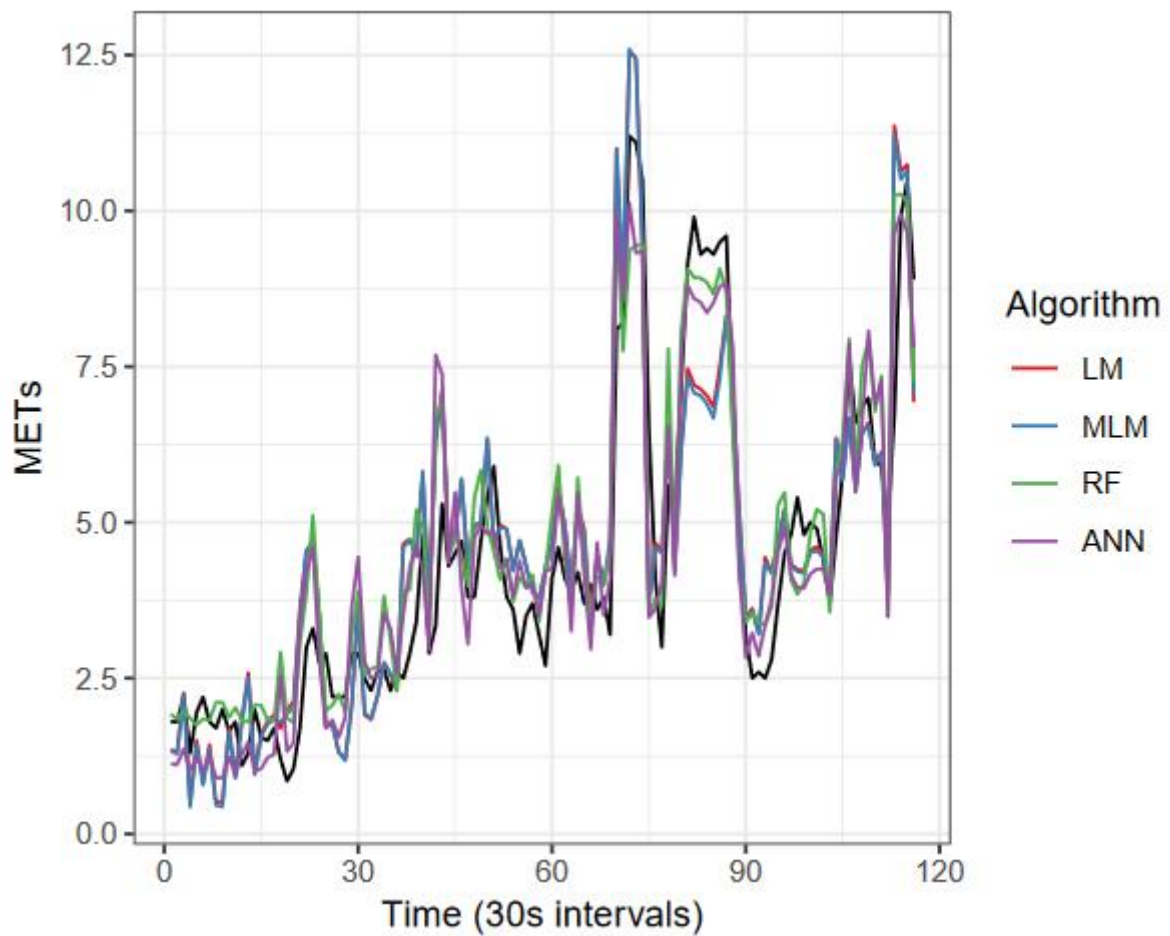456   10.18637/jss.v077.i01

457

458

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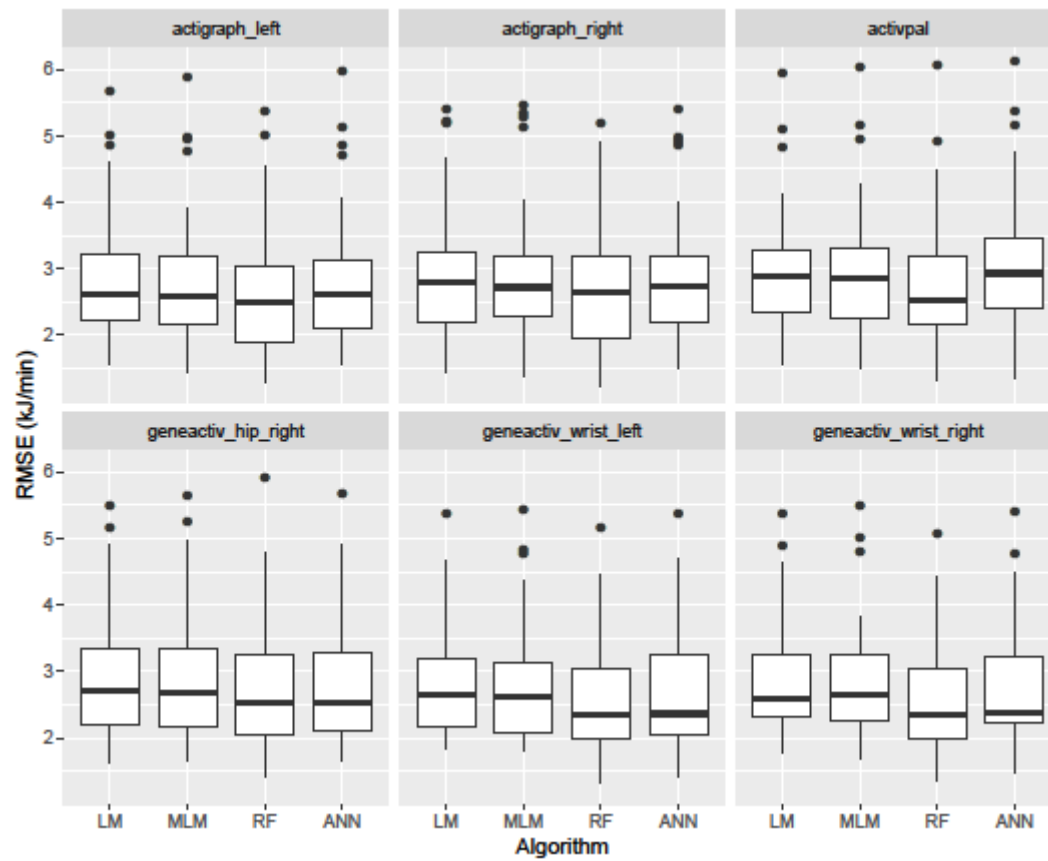
Figure 2 absEE: Root mean square error (RMSE) for predicted versus measured absolute energy expenditure (absEE) for all accelerometers, locations and models. Data are displayed in boxplots with median and interquartile range. Black dots represent outliers. LM: linear model, MLM: mixed linear model, RF: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

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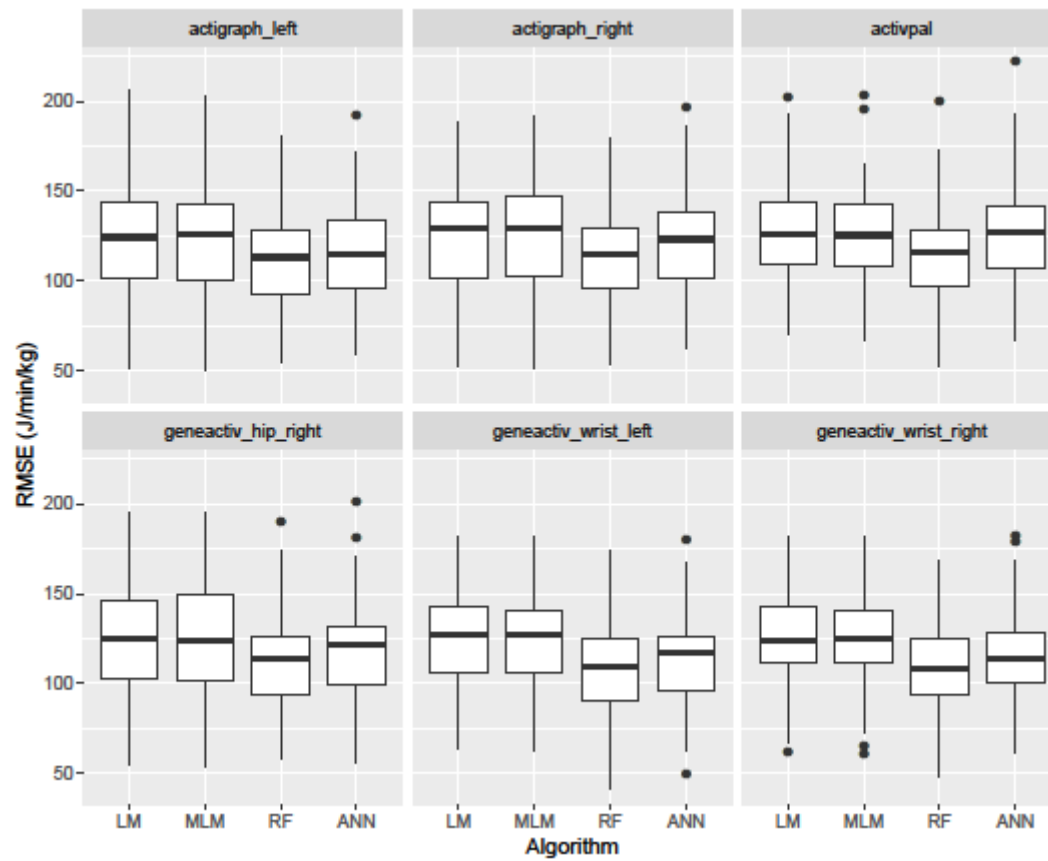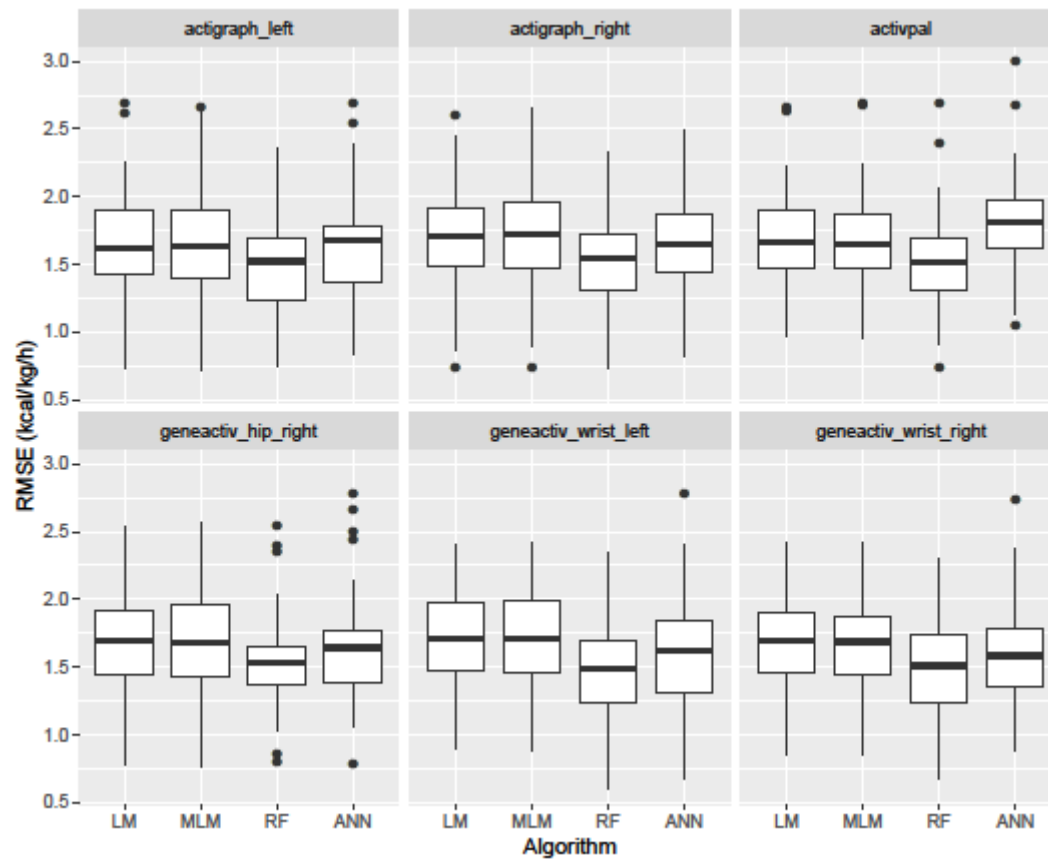
Figure 4 METs: Root mean square error (RMSE) for predicted versus measured metabolic equivalents (METs) for all accelerometers, locations and models. Data are displayed in boxplots with median and interquartile range. Black dots represent outliers. LM: linear model, MLM: mixed linear model, RF: random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

Table 1 Overview, order and duration of study procedures

Activity/task	Time (min)
Welcoming the child, changing to side room for preparation and RMR measurement	3
Mounting devices (oxygen analyser, 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

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*m^{-1})	2.3 ± 0.5 (1.4, 3.4)

Note: 1 Data are displayed as mean ± standard deviation (min, max)

Table 3 Energy expenditure per activity

Activity	Number of children (n)	Energy expenditure (kJ*m^{-1})
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: 2 Data are displayed as mean ± standard deviation (min, max)

Table 4 Overview of energy expenditure prediction accuracy of the different models, accelerometers and wear positions

Mean results (RMSE)						
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	1.70	1.69	1.69 (0.42)	1.69 (0.38)	1.69 (0.38)

	(0.41)	(0.40)	(0.39)			
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)

LM: linear model. MLM: mixed linear model. RF: random forest. ANN: artificial neural network. RMSE: Root-mean-square error