Calibration and validation of the Youth Activity Profile as a physical activity and sedentary behaviour surveillance tool for English youth

Stuart J Fairclough (stuart.fairclough@edgehill.ac.uk)
Edge Hill University https://orcid.org/0000-0001-8358-1979
Danielle L Christian
University of Central Lancashire
Pedro F Saint-Maurice
National Institutes of Health
Paul R Hibbing
University of Tennessee Knoxville
Robert J Noonan
University of Liverpool
Greg J Welk
Iowa State University
Philip M Dixon
Iowa State University
Lynne M Boddy
Liverpool John Moores University

Research article

Keywords: Calibration, validation, physical activity, self-report, accelerometer, sedentary behaviour, children, adolescents, measurement

Posted Date: June 19th, 2019

DOI: https://doi.org/10.21203/rs.2.10400/v1

Version of Record: A version of this preprint was published at International Journal of Environmental Research and Public Health on October 2nd, 2019. See the published version at https://doi.org/10.3390/ijerph16193711.

Abstract

Background Calibration algorithms applied to the Youth Activity Profile (YAP) self-report questionnaire in the US have accurately estimated moderate-to-vigorous physical activity (MVPA) and sedentary behaviour (SB). However, the efficacy of the calibration algorithms may vary when applied to different populations. We aimed to: (1) assess the accuracy of US-generated YAP calibration algorithms for MVPA and SB with English youth, (2) validate English-specific YAP calibration algorithms, (3) examine their potential surveillance utility to assess compliance to MVPA guidelines. Methods Four primary schools and five secondary schools were recruited. Four-hundred-and-two participants (aged 9-16 years; 212 boys) wore SenseWear Armband Mini devices (SWA) for eight days and completed the YAP on the eighth day. For aim (1) the original US calibration algorithms were applied to the YAP scores, which were matched to SWA-estimated in-school, out-of-school, and weekend MVPA and out-of-school SB data. For aim (2) new calibration algorithms for the equivalent time-segments were generated from the English YAP data using quantile regression. The algorithms were applied in an independent cross-validation sample, and individual- and group-level agreement were assessed using bias, mean absolute percent error (MAPE) and equivalency tests, respectively. For aim (3) the utility of the English YAP algorithms to assess compliance to MVPA guidelines was examined using kappa, sensitivity, and specificity. Results Agreement between the US calibration algorithms and SWA estimates of MVPA and SB was poor. Group-level MAPE for the English YAP-estimates of in-school, out-of-school, and weekend MVPA ranged from 3.6% to 17.3%. Bias for these estimates were 17.2 (34.4), 31.6 (14.0), and -4.9 (3.6) min·week-1, respectively. Out-of-school SB was over-predicted by 109.2 (11.8) min·week-1 (MAPE=11.8%). Predicted YAP values were within 15%-20% equivalence of the SWA estimates. Classification accuracy of the English YAP MVPA estimates for compliance to 60 min·day-1 and 30 min·school-day-1 MVPA recommendations were 91%/37% and 89%/57% sensitivity/specificity, respectively. Conclusions The English YAP generated robust group-level estimates of MVPA and SB and has potential for surveillance to monitor compliance with MVPA guidelines. The YAP's accuracy may be further improved through research work with more representative UK samples to enhance the calibration process and to refine the resultant algorithms.

Background

Physical inactivity and sedentary behaviour (SB) are increasingly prevalent among children and young people (1), and are associated with undesirable health and wellbeing outcomes (2, 3). To further understand physical activity (PA) and SB in youth it is critical to develop more effective ways to assess these complex behaviours. Accelerometry-based devices have been shown to provide reasonable estimates of both PA (4, 5) and SB (6, 7) and are widely used in various research applications. Accelerometers rank first in validity among field-based measures of PA/SB and for that reason their use in large-scale studies is becoming more common. However accelerometers are often inaccessible outside of well-funded research studies (8-10), or are only used with modestly sized samples due to limited device availability to research teams (11-13). Moreover, accelerometers are not routinely used by PA practitioners, PA and health promotion organisations, or schools because they are costly and labour-intensive (14), and often require high levels of data processing expertise. Hence, technological improvements of accelerometers and development of built-in algorithms to assess PA and SB do not translate in direct benefits to health and education professionals, which limits the application of these devices beyond the research community. Accelerometers also do not provide any information on the types of PA and SB being undertaken and where these behaviours take place, which is often integral to understanding the context underpinning them (15).

Self-report questionnaires of PA and SB overcome some of the limitations of accelerometry, by providing important information about context and setting (16). Moreover, self-report questionnaires reduce participant burden, and are more affordable than accelerometers when used at scale (14). Recent systematic reviews identified 89 (17) and 46 (18) different self-report questionnaires to assess youth PA and SB, respectively. Methodological quality of these studies was generally low, and none of the questionnaires demonstrated acceptable reliability and validity, with content validity noted as being particularly weak (17, 18). A variety of surveys are used in school and public health surveillance to capture information about PA and SB patterns in youth. In England, no youth PA and SB questionnaire exists that is calibrated to produce acceptable PA and SB estimates, and which is designed to be used easily across different levels of expertise. The Health Survey for England asks youth to recall the frequency and duration of PA over the last seven days to capture in-school and out-of-school sports/activities. However, this survey under and over estimates MVPA in younger and older children, respectively (19). The recently introduced Sport England Active Lives Survey for Children and Young People includes a 7-day recall list of in- and out-of-school activities, when children did these, and for how long (20), but the validity of this instrument has not been established. Both questionnaires estimate PA based on participants' raw scores and are therefore subject to un-addressed systematic and random error, which are likely to result in estimations of PA that lack equivalence with device-based estimates. Additionally, neither of these surveys assess SB. Thus, there is a need for a more accurate, self-report methodology to assess both PA and SB in youth that is suitable for youth in England and the wider UK

Compared to accelerometers, there have been fewer efforts to improve the validity of self-report questionnaires to provide estimates of time spent in PA and SB. One approach is regression-based calibration of self-report data against accelerometer-derived PA and SB estimates. This method has shown promise in estimating whole-day (21) and context-specific estimates of youth PA (22). From these studies, Saint-Maurice and Welk developed the Youth Activity Profile (YAP) and demonstrated that coded self-report responses from US youth could be calibrated to provide more accurate estimates of group-level moderate-to-vigorous PA (MVPA) and SB, than self-reported estimates alone (23). Differences between calibrated YAP estimates and MVPA using SenseWear Armbands (SWA) during school, out-of-school, and at the weekend were -15.6, 3.4, and -21.7 minutes per week, respectively. Furthermore, calibrated YAP estimates of out-of-school SB time underpredicted SWA estimates by just 49.7 minutes per week (23). In a subsequent study, group level YAP MVPA and SB estimates were within 10%-20% of values obtained from ActiGraph GT3X+ accelerometers (11). The YAP has been administered by school teachers without any specialist PA/SB measurement knowledge in over 1000 US schools through the NFL

PLAY 60 FITNESSGRAM Partnership Project (24), which demonstrates its utility as a cost-effective population-level PA and SB measurement tool. These findings highlight the value in developing low-cost, standardised, and scalable self-report questionnaires and associated analytical techniques, that can produce estimates of PA and SB that are equivalent to those produced by accelerometers (25). Saint-Maurice and Welk (2015) advocated the testing and potential refinement of the YAP algorithms on independent samples beyond their original study (23). Reflecting this recommendation, this study aimed to (1) assess the predictive accuracy of applying US-generated YAP calibration algorithms for PA and SB in a sample of English youth, (2) develop and validate English-specific YAP calibration algorithms, and (3) examine their potential surveillance utility to assess compliance to PA guidelines.

Methods

Participants and settings

Eleven schools (five primary and six secondary) from northwest England were informed about the study. Nine of the 11 schools (four primary and five secondary) were recruited and students from randomly selected classes in Year (i.e., Grade) 5 (primary school stage; 10.2 ± 0.3 years), Year 8 (secondary school stage; 13.2 ± 0.3 years) and Year 10 (15.2 ± 0.3 years) were invited to participate (N = 409; 212 boys). Informed parental consent and child assent were obtained from 402 participants (209 boys; 98% recruitment rate) who were each assigned to a unique ID code. Each school received a £300 financial incentive for participating, and each participant received a £10 shopping voucher following completion of data collection, which took place between March and July 2017. The study received ethical approval from the Liverpool John Moores University Research Ethics Committee (#14/SPS/012).

Measures

Anthropometric measures

Height was measured to the nearest 0.1cm using a portable stadiometer (Seca 213 height measure, Seca UK, Birmingham, UK). Body mass was measured to the nearest 0.1kg using digital scales (Seca 877 digital scales, Seca UK, Birmingham, UK). Waist circumference was measured to the nearest 0.1cm using a non-elastic measuring tape, which was positioned around the mid-section of the waist, over the participants' school shirts. All anthropometric measures were administered and recorded by pairs of trained researchers in accordance with standardised procedures (26). Height and weight data were converted to body mass index (BMI), which were subsequently used to categorise participants into weight status classifications using IOTF BMI cutpoints (11).

Socioeconomic status

Socioeconomic status was assessed using the 2015 English Indices of Multiple Deprivation (IMD) raw scores (27), which were derived from individual participants' postcode entries. IMD scores range from 0.5 to 92.6, and are composed of seven domains of deprivation (income, employment, health, education, access to services, living environment and crime) with higher aggregated scores representing higher degrees of deprivation.

Device-measured physical activity and sedentary time

PA was measured over eight days using a SenseWear Armband Mini (SWA) (Bodymedia, Inc., Pittsburgh, PA). The SWA is a multi-sensor device that detects and records movement and physiological responses at 60-second epochs (default setting), to provide accurate estimates of energy expenditure. A key advantage of the SWA is that it automatically detects non-wear time (23). The SWA is validated for use with children (28, 29) and was used in the original US YAP calibration study (23). Participants were instructed to wear the SWA on the back of the upper arm in direct contact with the skin. They were asked to only remove the device for water-based activities, such as bathing or swimming.

Self-reported physical activity and sedentary time

Self-reported PA and SB data were collected using the YAP (23). The YAP is an online 15-item questionnaire comprised of three sections (school day, out-of-school, and sedentary behaviour), with five questions per section. Participants are asked to recall their PA and SB over the past 7-days during context-specific time segments. For example, the school day questions ask on how many days participants undertook active travel to and from school, and their activity levels during break time, lunch time, and PE. The out-of-school segment refers to activity levels before school, immediately after-school, evening, and across both Saturday and Sunday. The SB section asks about time spent watching TV, playing video games, using a mobile phone, a computer/tablet, and overall SB. All questions are structured using a 1-5 Likert scale (e.g., for active travel to school, a score of 1 indicates 0 days per week of active travel, whereas a score of 5 indicates 4-5 days per week). The questions referring to break time, lunch time, and PE also include the option for participants to indicate that these PA opportunities did not occur during the previous week. In such instances, a score of 0 is assigned.

Prior to the study commencing, the YAP was minimally amended by the research team to make the clarity, language, and terminology more appropriate for English youth (e.g., the word 'recess' was replaced with 'break time', 'cell phone' was replaced with 'mobile phone', etc). Through this process the fundamental content and meaning of the YAP questions were unaltered. Differences between the two YAP versions are highlighted in Additional file 1. Participants completed the YAP using desktop PCs or iPads in a classroom eight days after receiving the SWA. This was to ensure the seven-day recall of the YAP temporally matched the collected SWA data. All participants received the same instructions on how to complete the YAP from a prepared script. Research staff were on hand throughout to assist with any further questions. On completion of the YAP, researchers used recall 'probing' questions as a quality assurance mechanism to improve the accuracy of responses (23). These probes were specifically developed for the YAP calibration and are not part of the YAP, nor are recommended for field applications when using the tool (23). The English YAP version used in this study is provided as Additional file 2.

Study design

The study followed a similar protocol as that detailed in the original US YAP calibration study (23). Data were collected on a two-week cycle which consisted of two data collection visits to the schools. At the first visit, participants were provided with instructions on how to wear the SWA devices, which were distributed during this session. Anthropometric measures were also obtained. On the second data collection visit (8 days after the first visit), SWA devices were collected and the YAP was administered. Individual students' home postcodes, ethnicity, and sex were obtained via schools' information management systems. Schools also provided details of the previous week's school timetable schedule which included days and times for school start and end, recess, lunchtime, and physical education (PE) lessons.

Data Processing

Predictive ability of US algorithms with English sample (Aim 1)

Data were processed using an identical data processing routine to that used in the original US YAP calibration study (23). SWA data were downloaded using the BodyMedia SenseWear Professional Software v8.0. The SenseWear software automatically detected non-wear time and classified the data into PA or SB on a per minute basis, which is the default setting. Epochs spent in PA ≥4.0 METs were classified as MVPA, and epochs spent in activities ≤2.0 METs were classified as SB (30, 31). SWA MVPA and SB data were then temporally allocated to specific time segments which corresponded to the time segments integrated into the YAP questions (Table 1; e.g., the SWA data between 6pm and 10pm was classified as 'Evening'). The process of segmentation was conducted in R (32) using code specifically written for this purpose. This generated the number of minutes in MVPA and SB for each segment (e.g., break time, lunch) from the SWA.

Table 1. Time segments used in the YAP calibration (adapted from (23)).

The structure and questions of the YAP were designed to temporally link the SWA data to the recall responses. The first 10 questions captured a discrete time segment in which there were specific opportunities to be physically active (Table 1). Previous research has shown that the YAP segments capture 94.6% of the total MVPA that occurs throughout the day (23). Procedures for scoring the YAP values and converting them to estimates of MVPA and SB are presented in Additional file 3 and are briefly summarised here. After the YAP data were checked and cleaned, the segment-specific US YAP algorithms (23) were applied to the raw scores for questions 1 to 10 to generate percentage time in MVPA per question, and to the aggregated out-of-school SB score to generate percentage time in out-of-school SB. These percentage values were then multiplied by the duration of each segment (e.g., break time), which was determined by the YAP protocol and school-specific schedules (Table 1), to give the average minutes in MVPA per segment per day. These values were subsequently multiplied by the number of days per week per segment (e.g., 5 days for break time), which resulted in estimates of the average number of minutes in MVPA per week. The same procedure was used with the percentage out-of-school SB values, to produce absolute estimates of daily and weekly SB. The YAP data were then aggregated to reflect estimates of MVPA in-school (questions 1-5), out-of-school (questions 6-8), and at the weekend (questions 9-10), and SB out-of-school (questions 11-15). These data were temporally matched with the corresponding MVPA and SB estimates from the SWA to assess the predictive ability of the US YAP algorithms in an English population. This data segmentation process was conducted at the individual participant level according to the school day schedules provided by each school.

Following segmentation, individual records were screened for compliance to ensure quality and representativeness of the data. Compliance criteria were that the SWA had to be worn for ≥70% of the corresponding segment durations (e.g., break time) on at least three days (11), although the number of valid days required for SWA wear during PE classes and weekend days was set to 1. Participants with incomplete YAP data were also removed and not considered for further analysis. From the initial sample of 402 participants, after compliance checks 331 (82%; 170 boys) were included in the final analytical sample.

Agreement between the US YAP algorithm-predicted in-school, out-of-school, and weekend MVPA $min \cdot week^{-1}$ and out-of-school SB $min \cdot week^{-1}$, and corresponding estimates from the SWA were examined at the individual-level (i.e., variability in agreement between each individual participant's YAP and SWA estimates) and group-level (i.e., variability in agreement between aggregated YAP and SWA estimates for each school). The validity of each of the time segments was analysed by calculating the overall mean error or bias between observed and predicted values, and the mean absolute percentage error (MAPE). Group-level agreement was examined using the bias and MAPE.

Generationof English-specific YAP algorithms (Aim 2)

Calibration. The analytical sample data (n=331) were randomly allocated by school (stratified by school level) into calibration and cross-validation data sets. The calibration data (6 schools; 3 primary and 3 secondary) were used to generate the YAP prediction equations for in-school, out-of-school, and weekend MVPA and out-of-school SB. SWA and English YAP data were processed identically to the previous US YAP studies (11, 23). During calibration, daily percent time in SWA-derived MVPA and SB were treated as the dependent variables. School stage (primary or secondary school), sex, and the corresponding YAP segment composite scores were the independent variables. Quantile regression models (33) were fit separately for each time segment (in-school, out-of-school, and weekend) for MVPA and SB (out-of-school only). Preliminary evaluations considered different combinations of variables and their two-way interactions. Some models did not have unique solutions, and these were not considered further. Robustness of the final models was examined using root mean square error (RMSE). Calibration analyses were completed in R using the quantreg (34), tidyverse (35), and modelr packages (36).

Cross-validation. Data from the remaining three schools (1 primary and 2 secondary) were used to independently assess prediction accuracy of the English YAP algorithms from the calibration phase. Agreement was investigated by converting the YAP composite segment scores into weekly minutes of MVPA or SB using the algorithms developed in the calibration analyses. Individual-level agreement for each YAP segment was determined using bias, and MAPE. Bias and MAPE were calculated to explore group-level agreement. Equivalence testing was also applied with the cross-validation sample to examine whether 95% confidence intervals (CI) for YAP-predicted minutes of MVPA/SB were within a 10% range (equivalence zone) of estimates from the SWA. Where there was no evidence of equivalence at 10%, the equivalence zone was increased by 5% until equivalence was reached (i.e., 15%, 20%, etc).

Potential surveillance utility of the English YAP algorithms (Aim 3)

To examine the potential of the English YAP algorithms for PA surveillance, the ability of the YAP to identify participants that met PA guidelines was assessed, using the SWA data as the criterion. Using the full analytical sample (n=331), average MVPA min·day^{-1} and average MVPA min·school day^{-1} were computed for YAP-predicted values and SWA values. Binary codes were used to classify participants according to whether they achieved at least an average of 60 min·day^{-1} MVPA, which reflects current PA recommendations in the UK (37) and internationally (38). The same method was used for an average of 30 min·school day^{-1} MVPA, which reflects recommendations in the UK (39) and US (40) that schools provide opportunities for youth to achieve at least 50% of the daily recommended PA during school time. Agreement between the proportion of participants achieving the respective 60 and 30 min MVPA recommendations according to the YAP and SWA were compared and classification accuracy of the YAP was evaluated using percent agreement, kappa, sensitivity, and specificity

Results

Descriptive Statistics

The descriptive characteristics of the participants who were included in the analyses are detailed in Table 2. There were no significant differences in age, BMI, and SES between participants included and not included in the analytical sample (p>.05; i.e., those who did not have complete YAP data and/or who not meet the SWA wear time criteria).

Table 2. Descriptive statistics of complete, calibration, and cross-validation samples. (Mean (SD) unless stated).

Predictive ability of US YAP algorithms with English sample (Aim 1)

Individual-level agreement between US YAP-predicted and SWA estimates was very poor with MAPE ranging from 59.0% to 93.6% (Additional file 4). The US YAP algorithms performed somewhat better at the group-level but agreement was still weak (Table 3). Bias ranged from -32.1 (116.8) min·week^{-1} for in-school MVPA to 445.7 (106.4) min·week^{-1} for out-of-school SB. Group-level MAPE was lowest for weekend MVPA (26.4%) and highest for out-of-school SB (51.0%).

Table 3. Group-level agreement between SWA and US YAP predicted MVPA and SB (n=9 schools).

Generation of English YAP algorithms (Aim 2)

Calibration

For the calibration analyses 200 participants had valid YAP and SWA data for at least one of the YAP segments of the week. In the final models the predictors of MVPA and SB were school level, sex, and the interaction between the segment YAP score and school level (Table 4). RMSE was 12.1, 9.6%, 8.5%, and 15.3% for in-school, out-of-school, and weekend MVPA, and out-of-school SB, respectively. More detailed summaries of the calibration models and resultant level- and sex-specific algorithms are presented in Additional file 5.

Table 4. Regression coefficients (SE) for in-school, out-of-school, weekend, and sedentary behaviour YAP segments

Cross-validation

From the three cross-validation schools, there were 129 participants with valid YAP and SWA data for at least one YAP segment of the week. Results are presented separately for each of the four segments. Individual- and group-level results are summarised in Additional file 6. Group-level results are presented in Table 5 (min·week^{-1} in MVPA and SB for cross-validation sample), Figure 1 (plots of YAP-predicted and SWA-estimated MVPA and SB for all schools), and Additional file 7 (percent of segment time in MVPA and SB for cross-validation sample).

In-school MVPA. The bias for predicted in-school MVPA min·week^{-1} and SWA-derived MVPA was 17.4 (88.0) min·week^{-1}, and the MAPE was 70.8%. Group-level estimates of MVPA from the YAP and SWA differed by 17.2 (34.4) min·week^{-1} (approximately 3.4 min·day^{-1}), and group-level MAPE was 17.3%. The upper and lower limits of the 95% CI for in-school YAP MVPA (95% CI=241.0, 282.6 min·week^{-1}) were within 20% of SWA-estimated MVPA (20% zone=195.5, 293.3 min·week^{-1}) (Table 5).

Out-of-school MVPA. Individual-level out-of-school MVPA predicted by the YAP over-estimated MVPA from the SWA by 30.6 (153.7) min·week^{-1} and the MAPE was 83.9%. The YAP over-estimated group-level MVPA compared to SWA estimates by 31.6 (28.3) min·week^{-1} (6.3 min·day^{-1}), and the MAPE was 14.0%. YAP-predicted and SWA-estimated out-of-school MVPA reached agreement when the equivalence zone was set at 20% (20% zone=232.5, 348.7 min·week^{-1}).

Table 5. Cross-validation sample group-level estimates of MVPA and SB, bias, and MAPE.

FIGURE 1 HERE

TITLE: Figure 1. Group-level agreement between SWA estimates of MVPA and SB and predicted MVPA and SB minutes using the English YAP prediction algorithms.

LEGEND: Panel A: in-school MVPA; Panel b: out-of-school MVPA; Panel C: weekend MVPA; Panel D: out-of-school SB. Each data point represents one school; red dots: the calibration sample schools used for generating the English YAP equations, blue triangles: the cross-validation sample schools used to assess prediction accuracy.

Weekend MVPA. Individual-level bias for weekend MVPA between the YAP and SWA was -5.3 (139.4) min·week^{-1}, and the individual-level MAPE was very high (199.6%). At the group-level, the YAP under-predicted weekend SWA MVPA by 4.9 (13.2) min·week^{-1} (-2.5 min·day^{-1}), and the MAPE was very low (3.6%) (Table 5). The 95% CI for the weekend MVPA YAP values (203.4, 245.8 min·week^{-1}) were within the 15% equivalence zone (195.1, 263.9 min·week^{-1}).

Out-of-school SB. At the individual-level, out-of-school time in SB was over-estimated by the YAP by 109.3 (262.1) min·week^{-1} relative to the SWA, and MAPE was 50.6%. Group-level YAP-predicted out-of-school time spent in SB was 109.2 (20.5) min·week^{-1} (21.8 min·day^{-1}) higher than the SWA estimates. Group-level MAPE for SB was 11.8% for min·week^{-1}. The upper and lower limits of the 95% CI for YAP-predicted SB (1120.5, 1196.9 min·week^{-1}) were equivalent at 15% (897.9, 1214.9 min·week^{-1}).

As the final calibration algorithms were specific to group-level and sex, additional sub-group agreement analyses were performed separately for all primary and secondary schools (Table 6), and for all boys and girls (Table 7). For primary schools, bias ranged from 12.9 (10.6) min·week^{-1} (in-school MVPA) to 47.2 (91.5) min·week^{-1} for SB. Primary school MAPE was lowest and highest for in-school MVPA (4.2%) and out-of-school MVPA (13.4%), respectively. Bias was lowest for weekend MVPA (-4.5 (28.8 min·week^{-1})

among secondary schools, and highest for SB (118.4 (50.0 min·week^{-1}). Conversely, MAPE was lowest for SB (8.0%) and highest for in-school MVPA (15.6%). Between-segment differences in bias were similar for primary and secondary schools. These ranged from -2.5 (51.6) min·week^{-1} (boys) and -3.9 (71.1 min·week^{-1} (girls) for out-of-school MVPA, to 89.7 (78.2) min·week^{-1} (boys) and 61.2 (84.1 min·week^{-1} (girls) for out-of-school SB. MAPE ranged from 10.3% (in-school MVPA) to 17.5% (weekend MVPA) for boys, and from 9.7% (out-of-school SB) to 16.0% (out-of-school MVPA) for girls. These additional group-level analyses are also presented in Additional file 6, with percent of segment time as the outcome.

Table 6. Group-level estimates of MVPA and ST, relative bias, and MAPE for primary and secondary stages.

Table 7. Group-level estimates of MVPA and ST, relative bias, and MAPE for boys and girls.

Classification accuracy of the YAP (Aim 3)

Sixty-min·day^{-1} MVPA was achieved by 81% of the participants according to the SWA. YAP-predicted estimates of daily MVPA indicated that the recommendation was met by 85.8% of participants. Agreement was 80.7% and the kappa value was 0.31 (fair agreement). Sensitivity and specificity were 91% and 37%, respectively. The school day 30 min·day^{-1} MVPA recommendation was achieved by 77.6% and 79.2% of participants, according to SWA and YAP-predicted estimates, respectively. Percent agreement and kappa values were 82.2% and 0.47, respectively (moderate agreement). Classification accuracy was 89% sensitivity and 57% specificity.

Discussion

This study aimed to examine the predictive accuracy of the US YAP algorithms for MVPA and SB with a sample of English youth, and to calibrate and test the validity and predictive utility of new English YAP algorithms. We found that the US YAP algorithms poorly predicted SWA estimates of MVPA and SB in English youth. Group-level predictions of in-school, out-of-school, and weekend MVPA, and out-of-school SB from the English YAP algorithms were promising, and the YAP demonstrated potential as a surveillance tool to identify prevalence of compliance to youth PA guidelines.

Aim 1

There was poor individual- and group-level agreement between the US YAP estimates of MVPA and SB and estimates from the SWA. The recommendation of Saint-Maurice and Welk that the US YAP algorithms be tested and refined on independent samples (23) is therefore well-founded. Moreover, it reinforces the notion that the content of self-report PA questionnaires is population- and context-specific, and that questionnaires developed and validated in one population cannot automatically be assumed to be suitable for youth elsewhere with different PA contexts and routines (41). Cross-cultural differences exist relating to individuals as well as the contexts and settings in which they live (e.g., school day schedules, sports practices, home routines, etc) (42). Although the development of the US YAP algorithms used the same data processing steps and a similarly aged-sample as in the present study, critically, the algorithms reflected the school schedules and out-of-school routines of the US youth, which differed to those in our English sample. These differences were reflected in the high individual-level and group-level MAPE values, for in-school MVPA and out-of-school SB in particular, when the US YAP algorithms were applied to the English SWA data.

Aim 2

When the estimates from the English YAP algorithms were compared to those from the SWA, individual-level agreement was poor. This is consistent with what has been observed previously for the YAP (11, 23), for other self-report instruments (17, 18), and also for most calibration equations developed for accelerometers (43). The high individual-level MAPE values exceeded those observed when the YAP was compared with MVPA and SB estimates from 12-17 year olds wearing wrist-mounted ActiGraph GT3X+ accelerometers (11). These results may have been influenced by a relatively high number of small observed SWA-derived MVPA and SB estimates, which when combined with moderate absolute error values result in very large MAPE values (44). The data clearly indicate that the YAP estimates of MVPA and SB lack predictive accuracy at the level of the student, and therefore should not be used to inform or evaluate individual-level intervention or PA prescription.

Group-level bias and error between English YAP-predicted MVPA and SB and estimates from the SWA were substantially better than those from the individual-level analyses. The degree of bias between in-school YAP and SWA MVPA was comparable with previous YAP validations (11, 23). The MAPE of 17.1% could not be directly compared to the previous YAP studies which employed tests of equivalence to determine group-level agreement, but the YAP-predicted in-school MVPA was equivalent to SWA-estimates at 20% equivalence. This is the same as that reported in the original US YAP calibration study using SWA devices (23), but higher than the 15% equivalence observed in a subsequent study using the ActiGraph GT3X+ (11). The out-of-school MVPA bias of 31.6 min·week^{-1} was much higher than the -3.0 min·week^{-1} (11) and 3.4 min·week^{-1} (23) previously reported for US youth, though the 10.8% MAPE indicated lower error than for our in-school MVPA predictions. YAP and SWA out-of-school MVPA were deemed equivalent at 20%, whereas Saint-Maurice and colleagues reported 15% (23) and 10% equivalence (11) in their US studies. Weekend MVPA was predicted to within -4.9 min min·week^{-1} of the SWA estimates and MAPE was very low (3.6%). This degree of bias is substantially less than the -21.7 min·week^{-1} and -17.8 min·week^{-1} observed by Saint-Maurice and Welk (23) and Saint-Maurice et al. (11), respectively. Further, the 15% equivalence between weekend MVPA from the YAP and SWA was superior to the US studies (30% (23) and 20% (11)). Conversely, bias and equivalence for out-of-school SB predicted by the YAP was higher than Saint-Maurice and colleagues reported (11, 23), although the 11.8% MAPE that we observed was lower than for our in-school and out-of-school segments.

A key difference between the English and US YAP estimates was over-prediction of in-school and out-of-school MVPA, and SB. The English YAP algorithms were based on the specific school schedules and daily routines of the participants in the calibration sample, although the same data processing methods as the previous YAP studies were used (11, 23). It is established that PA recall methods are subject to various sources of measurement error (45), and in this study the over-predictions may have reflected such factors related to the data collection protocol and the participants themselves. For example, though the research assistants received the same training there may have been variation in how and to what extent they used the probing questions to check participants' YAP recall accuracy. Moreover, differences in literacy and cognitive understanding of the YAP questions likely varied among the participants, and particularly in the primary school group. Even though the probing questions were employed, variations in how the YAP questions were interpreted would have contributed to measurement and processing errors (45), which would have affected the resultant algorithms. The participants in our study were also relatively active and any expected over-estimations of their PA (46) may have been exacerbated when recalling their MVPA behaviours. YAP-predicted SB was also overestimated but the 11.8% MAPE and 15% equivalence were promising. SBs tend to be more stable than active behaviours (47). However, the YAP overestimated SB by almost 22 min·day^{-1}, which may reflect that engagement in SBs, like TV viewing, or gaming occurred sporadically rather than as part of a set structure and routine (48). As a result, it is likely that recall of the specific SBs included in the YAP was challenging for some participants (49, 50).

Agreement between YAP-predicted and SWA-estimated MVPA was greatest for the weekend YAP segment. Weekends often involve greater choices and time for recreational activities, but weekend schedules can also reflect regularly occurring activities such as household chores, caring for siblings, and sports practices, etc. Although not as structured as the school day, weekends can represent familiar and routinised contexts for some youth. Moreover, the YAP uses two of the 15 questions to ask about PA during whole weekend days. Longer recall periods (i.e., a full day) are hypothesised to inhibit accurate recall of PA behaviours (51). It is though possible that only focusing on two specific days, rather than five, reduced recall burden and facilitated more accurate responses, which contributed to the low observed error for weekend MVPA.

Aim 3

The English YAP algorithms demonstrated their utility to evaluate compliance with health-related PA recommendations. Sensitivity values associated with the 60 and 30 min MVPA·day^{-1} recommendations indicated that 91% and 89% of youth who did not achieve the respective recommended daily MVPA would be correctly identified as 'inactive' based on the English YAP algorithm predicted MVPA. These sensitivity values are superior to those reported in the original US YAP calibration study (23), and reinforce the utility of the English YAP algorithms for identifying youth that do not meet PA guidelines. Saint-Maurice and

Welk also reported specificity values of 69% (60 min MVPA·day^{-1}) and 61% (30 min MVPA·school-day^{-1}) (23), compared to 37% and 57%, respectively in the present study. This suggests that the English YAP MVPA algorithms were less able to accurately classify youth who did achieve PA guidelines, as 'active'. Hence, further refinement of the YAP MVPA algorithms is needed to improve this classification accuracy. However, such misclassification is unlikely to cause harm if already active youth were targeted and/or offered additional opportunities to take part in PA programmes and interventions.

Strengths and limitations

The calibrated YAP estimates of MVPA and SB have great potential utility for future research and PA promotion, as existing calibrated self-report instruments for English youth are not available. Strengths of the study included (1) use of a proven, rigorous YAP protocol and methodology; (2) use of manageable group sizes for data collection which allowed use of recall probes to enhance the participants' recall accuracy; (3) recording detailed timetable information from each school to accurately determine each participant's schedule during the week when they wore the SWA, so as to enhance the degree of temporal precision required for the calibration analyses; (4) use of an independent sample for the cross-validation analyses, and (5) the choice of the SWA as the device-based measure, which has previously demonstrated superior agreement with criterion measures of free-living energy expenditure than other research-grade and consumer activity monitors (52, 53). There are also limitations which require consideration. Schools were not selected at random and so a degree of sampling bias in favour of more active participants may have been evident. Data were collected in the spring and summer months which may have reflected the relatively high estimates of MVPA. Therefore, the English YAP algorithms do not account for seasonal variation in the participants' PA and SB. The YAP content means that it can only be used to predict MVPA and SB during school-term time and not during vacation periods, and all modes of MVPA and SB may not be captured. However, schools in England are in session for around 39 weeks of the year so typical activity would be captured by the YAP. The YAP-predicted MVPA and SB estimates demonstrated good group-level agreement, but like values from all PA measurement tools, they cannot be considered exact values reflecting individual-level activity behaviours. Moreover, the calibration algorithms are based on MVPA and SB estimates from the SWA as the field-based criterion measure. There are inherent differences in how MVPA and SB are calculated by the SWA compared to accelerometer-only devices, and this limits comparability with data from other studies. Further, like all PA measurement instruments, the SWA is subject to measurement error which we could not control, and which may have attenuated the effects of the analyses (11). Incorporating measurement error modelling against a criterion measure has been shown help reduce the effects of measurement error and improve the precision of PA estimates from self-report questionnaires (45). A true criterion measure of free-living PA and SB requires accurate ground-truth measurement (e.g., wearable cameras) to label activity behaviours (54) (although unsupervised machine learning methods are now emerging, which may remove the need for criterion measures (55)). However, these approaches are yet to be feasible in large samples, and therefore, currently offer limited value for the calibration of self-report questionnaires.

Lastly, the SWA uses a default 60-second epoch setting to record data, which may not have fully captured intermittent bouts of higher intensity PA that are characteristic of school-aged youth (56), however, this monitor has been shown to provide accurate estimates of PA in this population (28, 29).

Conclusions

Poor agreement was observed in MVPA and SB derived from the US YAP algorithms and SWA worn by the English sample. YAP algorithms developed using the English sample data resulted in MVPA and SB estimates that had promising group-level agreement with the lowest error observed for weekend MVPA and out-of-school SB. The YAP has potential as a surveillance tool to monitor compliance with youth PA guidelines, but more refinement is needed to improve its classification accuracy. The group-level YAP estimates of MVPA indicate that the YAP is a promising self-report questionnaire for use with English youth, and potentially with samples from other countries in the UK. The YAP is a cost-effective, easy to implement instrument that can be used at scale and implemented by researchers and practitioners, to provide meaningful group-level estimates of MVPA and SB. Further refinement of the YAP algorithms with a more representative UK sample and employing replicate measurement error modelling procedures to enhance the precision of the calibration algorithms is advocated.

List Of Abbreviations

BMI - Body mass index

CI – Confidence interval

IMD – Indices of Multiple Deprivation

IOTF – International Obesity Task Force

MAPE – mean absolute percent error

MVPA - Moderate-to-vigorous physical activity

PA – Physical activity

RMSE – Root mean square error

SB – Sedentary behaviour

SWA – Sensewear Armband

UK – United Kingdom

US – United States

Declarations

Ethics approval and consent to participate. Ethical approval was granted by Liverpool John Moores University Research Ethics Committee (#14/SPS/012). Informed parental/carer consent and informed child assent was obtained from all participants.

Consent for publication. Not applicable.

Availability of data and material. All data generated or analysed during this study are deposited on the Open Science Framework and can be accessed from https://osf.io/zwy6j/

Competing interests. The authors declare that they have no competing interests.

Funding. This study was funded by the Youth Sport Trust and Edge Hill University. The funders had no involvement in the study design, data collection, analysis, and interpretation, or writing of the manuscript.

Authors' contributions. SJF conceived the study with LMB, designed the protocol, performed the analyses, and wrote the manuscript. DLC collected, cleaned, and prepared the data, and drafted the manuscript. PFS-M advised on the data analysis and protocol and drafted the manuscript. PRH performed the data segmentation procedures and drafted the manuscript. RJN drafted the manuscript. GJW advised on the data analysis and protocol and drafted the manuscript. PD performed the analysis and drafted the manuscript. LMB designed the protocol, advised on the analysis, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements. The authors wish to thank the children and teachers for taking part in the study and the field-research team for assistance with data collection.

Additional File Legend

Additional file 1. Amendments to original YAP wording.pdf

As per file title

Additional file 2. YAP.pdf

The wording of the YAP questionnaire

Additional file 3. Scoring & conversion of raw YAP values.pdf

Step-by-step process for YAP scoring and conversion to MVPA and SB estimates

Additional file 4. Individual-level results US YAP vs SWA.pdf

US YAP predicted and SWA MVPA and SB estimates, plus bias, and MAPE

Additional file 5. English YAP calibration models & equations.pdf

Summary of calibration models and resultant equations for use with primary and secondary school boys and girls

Additional file 6. Individual-level cross-validation results minutes & percent segment times.pdf

Individual-level YAP predicted and SWA MVPA and SB estimates, plus bias, and MAPE with min/week and percent of segment time as the outcomes

Additional file 7. Group-level cross-validation results percent segment time.pdf

Group-level YAP predicted and SWA MVPA and SB estimates, plus bias, and MAPE with percent of segment time as the outcome

References

1. Cooper A, Goodman A, Page A, Sherar L, Esliger D, van Sluijs E, et al. Objectively measured physical activity and sedentary time in youth: the International children's accelerometry database (ICAD). Int J Behav Nutr Phys Act. 2015;12(1):113.

2. Poitras VJ, Gray CE, Borghese MM, Carson V, Chaput J-P, Janssen I, et al. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. Appl Physiol, Nutr, Metab. 2016;41:S197-S239.

3. Carson V, Hunter S, Kuzik N, Gray CE, Poitras VJ, Chaput J-P, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. Appl Physiol, Nutr, Metab. 2016;41:S240-S65.

4. Crouter SE, Horton M, Bassett DR. Validity of ActiGraph child-specific equations during various physical activities. Med Sci Sports Exerc. 2013;45:1403-9.

5. Hildebrand M, Van Hees VT, Hansen BH, Ekelund U. Age-group comparibility of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc. 2014;46:1816-24.

6. Kim Y, Barry VW, Kang M. Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. Meas Phys Educ Exerc Sci. 2015;19:125-37.

7. Hurter L, Fairclough SJ, Knowles Z, Porcellato L, Cooper-Ryan A, Boddy L. Establishing raw acceleration thresholds to classify sedentary and stationary behaviour in children. Children. 2018;5:172.

8. Harrington DM, Davies MJ, Bodicoat DH, Charles JM, Chudasama YV, Gorely T, et al. Effectiveness of the ‘Girls Active’ school-based physical activity programme: A cluster randomised controlled trial.

International Journal of Behavioral Nutrition and Physical Activity. 2018;15(1):40.

9. Jago R, Solomon-Moore E, Macdonald-Wallis C, Sebire SJ, Thompson JL, Lawlor DA. Change in children's physical activity and sedentary time between Year 1 and Year 4 of primary school in the B-PROACT1V cohort. Int J Behav Nutr Phys Act. 2017;14:33.

10. Wilkie HJ, Standage M, Gillison FB, Cumming SP, Katzmarzyk PT. Multiple lifestyle behaviours and overweight and obesity among children aged 9–11 years: results from the UK site of the International Study of Childhood Obesity, Lifestyle and the Environment. BMJ Open. 2016;6(2)e10677.

11. Saint-Maurice PF, Kim Y, Hibbing P, Oh AY, Perna FM, Welk GJ. Calibration and Validation of the Youth Activity Profile: The FLASHE Study. Am J Prev Med. 2017;52:880-7.

12. Fairclough SJ, Dumuid D, Taylor S, Curry W, McGrane B, Stratton G, et al. Fitness, fatness and the reallocation of time between children's daily movement behaviours: an analysis of compositional data. Int J Behav Nutr Phys Act. 2017;14(1):64.

13. Noonan RJ, Boddy LM, Kim Y, Knowles ZR, Fairclough SJ. Comparison of children's free-living physical activity derived from wrist and hip raw accelerations during the segmented week. J Sports Sci. 2017;35:2067-72.

14. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport. 2000;71:S59-S73.

15. Koorts H, Mattocks C, Ness AR, Deere K, Blair SN, Pate RR, et al. The association between the type, context, and levels of physical activity amongst adolescents. J Phys Act Health. 2011;8:1057-65.

16. Biddle SJ, Gorely T, Pearson N, Bull FC. An assessment of self-reported physical activity instruments in young people for population surveillance: Project ALPHA. Int J Behav Nutr Phys Act. 2011;8(1):1.

17. Hidding LM, Chinapaw MJM, van Poppel MNM, Mokkink LB, Altenburg TM. An updated systematic review of childhood physical activity questionnaires. Sports Med. 2018.48:2797-42

18. Hidding LM, Altenburg TM, Mokkink LB, Terwee CB, Chinapaw MJM. Systematic review of childhood sedentary behavior questionnaires: what do we know and what is next? Sports Med. 2017;47:677-99.

19. Craig R, Mindell J, Hirani V. Health survey for England 2008 Volume 1: physical activity and fitness. London: The Health and Social Care Information Centre; 2009.

20. Sport England. Active Lives: Children and Young People 2018. https://www.sportengland.org/research/active-lives-children-and-young-people (2018). Accessed 1 Dec 2018.

21. Saint-Maurice PF, Welk GJ, Beyler N, Bartee R, Heelan K. Calibration of self-report tools for physical activity research: the Physical Activity Questionnaire (PAQ). BMC Public Health. 2014;14(1):461.

22. Saint-Maurice PF, Welk GJ, Bartee RT, Heelan K. Calibration of context-specific survey items to assess youth physical activity behaviour. J Sports Sci. 2017;35:866-72.

23. Saint-Maurice PF, Welk GJ. Validity and calibration of the Youth Activity Profile. PLoS ONE. 2015;10(12):e0143949.

24. Welk GJ, Bai Y, Saint-Maurice PF, Allums-Featherston K, Candelaria N. Design and evaluation of the NFL PLAY 60 FITNESSGRAM® Partnership Project. Res Q Exerc Sport. 2016;87:1-13.

25. Hallal P, Andersen L, Bull F, Guthold R, Haskell W, Ekelund U. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet. 2012;380:247 - 57.

26. Lohman TG, Roche AFM, Martorell R. Anthropometric standardization reference manual. Illinois: Champaign, IL: Human Kinetics; 1991.

27. Department for Communities and Local Government. The English Indices of Deprivation. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (2015). Accessed 1 Dec 2017.

28. Calabro MA, Welk GJ, Eisenmann JC. Validation of the SenseWear Pro Armband algorithms in children. Med Sci Sports Exerc. 2009;41:1714-20.

29. Arvidsson D, Slinde F, Larsson S, Hulthen L. Energy cost in children assessed by multisensor activity monitors. Med Sci Sports Exerc. 2009;41:603-11.

30. Saint-Maurice PF, Kim Y, Welk GJ, Gaesser GA. Kids are not little adults: what MET threshold captures sedentary behavior in children? Eur J Appl Physiol. 2016;116:29-38.

31. Welk GJ, Morrow JR, Jr., Saint-Maurice PF. Measures registry user guide: individual physical activity. Washington DC: National Collaborative on Childhood Obesity Research, 2017.

32. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

33. Chandler J, Beets M, Saint-Maurice PF, Weaver R, Cliff D, Drenowatz C, et al. Wrist-based accelerometer cut-points to identify sedentary time in 5–11-year-old children. Children. 2018;5(10):137.

34. Koenker R. Quantile regression in R: a vignette. https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf. (2017). Accessed 1 April 2018.

35. Wickham H. tidyverse. https://cloud.r-project.org/web/packages/tidyverse/index.html. (2017). Accessed 1 April 2018.

36. Wickham H. modelr. https://cran.r-project.org/web/packages/modelr/index.html. (2019). Accessed 1 April 2018.

37. Chief Medical Officers. Start active, stay active. A report on physical activity for health from the four home countries. London: DH; 2011.

38. WHO. Global recommendations on physical activity for health. Geneva: WHO; 2010.

39. Department of Health. Childhood obesity. A plan for action. London: DH; 2016.

40. Institute of Medicine. Educating the student body. Taking physical activity and physical education to school. Washington DC: Institute of Medicine; 2013.

41. Terwee CB, Mokkink LB, van Poppel MNM, Chinapaw MJM, van Mechelen W, de Vet HCW. Qualitative attributes and measurement properties of physical activity questionnaires. Sports Med. 2010;40:525-37.

42. Aggio D, Fairclough SJ, Knowles Z, Graves L. Validity and reliability of a modified english version of the physical activity questionnaire for adolescents. Arch Public Health. 2016;74(1):3.

43. Bassett DR, Jr., Rowlands AV, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc. 2012;44:S32-8.

44. Kim S, Kim H. A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting. 2016;32:669-79.

45. Nusser SM, Beyler NK, Welk GJ, Carriquiry AL, Fuller WA, King BM. Modeling errors in physical activity recall data. J Phys Act Health. 2012;9:S56-67.

46. Beyler N, Beyler A. Adjusting for measurement error and nonresponse in physical activity surveys: a simulation study. Journal of Official Statistics. 2017.33:533-550.

47. Saint-Maurice PF, Welk GJ. Web-Based Assessments of physical activity in youth: considerations for design and scale calibration. J Med Internet Res. 2014;16(12):e269.

48. Brazendale K, Beets MW, Weaver RG, Pate RR, Turner-McGrievy GM, Kaczynski AT, et al. Understanding differences between summer vs. school obesogenic behaviors of children: the structured days hypothesis. Int J Behav Nutr Phys Act. 2017;14(1):100.

49. Friedman WJ. The development of temporal metamemory. Child Dev. 2007;78:1472-91.

50. Dumuid D, Olds TS, Lewis LK, Maher C. Does home equipment contribute to socioeconomic gradients in Australian children’s physical activity, sedentary time and screen time? BMC Public Health. 2016;16(1).

51. Baranowski T. Validity and reliability of self-report measures of physical activity: an information-processing perspective. Res Q Ex Sport. 1988;59:314-27.

52. Bai Y, Welk GJ, Nam YH, Lee JA, Lee J-M, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc. 2016;48:151-8.

53. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. Med Sci Sports Exerc. 2010;42:2134-40.

54. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Sci Rep. 2018;8(1):7961.

55. van Kuppevelt D, Heywood J, Hamer M, Sabia S, Fitzsimons E, van Hees V. Segmenting accelerometer data from daily life with unsupervised machine learning. PLOS ONE. 2019;14(1):e0208692.

56. Nilsson A, Ekelund U, Yngve A, Sjostrom M. Assessing physical activity among children with accelerometers using different time sampling intervals and placements. Ped Exerc Sci. 2002;14:87-96.

Tables

Table 1. Time segments used in the YAP calibration (adapted from (23)).

Question/Segment	Date	Individualised time	Start time*	End time*
1. Before travel to school	Every day	Yes	60 min before start time for travel to school	Start time for travel to school
2. Travel to school	Every day	Yes	30 min before start time for school	Start time for school
3. Play/Break time	When provided	Yes	Determined by school schedule	Determined by school schedule
4. Physical Education	When provided	Yes	Determined by school schedule	Determined by school schedule
5. Lunch	When provided	Yes	Determined by school schedule	Determined by school schedule
6. Travel from school	Every day	Yes	End time for school	30 min after end time for school
7. After-school	Every day	Yes	End time for travel from school	6:00 PM
8. Evening	Every day	No	6:00 PM	10:00 PM
9. Saturday	Saturday	No	7:00 AM	10:00 PM
10. Sunday	Sunday	No	7:00 AM	10:00 PM

* Individualised school Start and End times were obtained from individual schools (e.g., start at 9:00am, end at 3:30pm).

Table 2. Descriptive statistics of complete, calibration, and cross-validation samples. (Mean (SD) unless stated).

	All	Calibration sample	Cross-validation sample
	331	202	129
x			
oys (%)	51.4	59.4	38.8
irls (%)	48.6	40.6	61.2
nicity			
hite British (%)	93.7	94.6	92.2
ther (%)	6.3	5.4	7.8
e (years)	12.3 (2.1)	12.3 (2.1)	12.2 (2.3)
ight (cm)	154.9 (13.7)	155.0 (14.1)	154.6 (13.2)
ight (kg)	49.6 (15.4)	50.0 (15.3)	49.0 (15.6)
II (kg•m^2)	20.3 (4.3)	20.4 (4.4)	20.1 (4.3)
ight status			
% Normal weight	71.6	71.3	72.1
% Overweight/Obese	28.4	28.7	27.9
ist Circumference (cm)	71.0 (10.3)	71.4 (10.3)	70.4 (10.2)
S (IMD score)	19.4 (13.9)	21.9 (15.5)	15.7 (10.3)
A wear time (days)	5.8 (1.2)	5.8 (1.2)	5.9 (1.3)
A total wear time in×day^{-1})	1014.8 (114.7)	1021.9 (115.2)	1004.5 (113.5)

Table 3. Group-level agreement between SWA and US YAP predicted MVPA and SB (n=9 schools).

	YAP-predicted estimates (min·week^{-1})	SWA estimates (min·week^{-1})	Bias (min·week^{-1})	MAPE (%)
In-school MVPA	241.1 (61.4)	273.3 (81.3)	-32.1 (116.8)	39.7
Out-of-school MVPA	235.0 (42.6)	341.8 (90.9)	-106.7 (60.1)	29.3
Weekend MVPA	175.4 (49.6)	249.1 (74.0)	-73.7 (36.5)	26.4
Out-of-school SB	1496.0 (297.8)	1050.4 (365.8)	445.7 (106.4)	51.0

Table 4. Regression coefficients (SE) for in-school, out-of-school, weekend, and sedentary behaviour YAP segments

Model	In-school (n=200)	Out-of-school (n=196)	Weekend (n=187)	Sedentary behaviour (n=196)
Intercept (Primary stage)	**45.51 (9.14)**	**12.75 (5.01)**	**15.17 (5.02)**	**34.92 (12.71)**
Intercept (Secondary stage)	**13.72 (5.58)**	2.61 (3.41)	**8.44 (3.18)**	**48.95 (8.91)**
Sex	**-11.34 (2.47)**	-1.61 (1.95)	**-4.33 (1.56)**	1.61 (2.99)
YAP x Primary level	1.78 (2.33)	2.01 (2.06)	1.13 (1.22)	-0.84 (5.22)
YAP x Secondary level	**7.31 (1.66)**	**4.44 (1.18)**	1.44 (0.90)	**5.65 (2.54)**

Note. Bold type indicates significance ($p<.05$)

Table 5. Cross-validation sample group-level estimates of MVPA and SB, bias, and MAPE.

Segment	YAP-predicted estimates (min·week^{-1})	SWA estimates (min·week^{-1})	Bias (min·week^{-1})	MAPE (%)	Equivalence zone
In-school MVPA	258.7 (74.9)	241.4 (97.6)	17.2 (34.4)	17.3	20%
Out-of-school MVPA	319.2 (67.0)	287.6 (93.2)	31.6 (28.3)	14.0	20%
Weekend MVPA	220.9 (83.6)	225.8 (96.8)	-4.9 (13.2)	3.6	15%
Out-of-school SB	1180.3 (408.6)	1071.1 (416.1)	109.2 (20.5)	11.8	15%

Table 6. Group-level estimates of MVPA and ST, relative bias, and MAPE for primary and secondary stages.

Segment	YAP-predicted estimates (min·week^{-1})	SWA estimates (min·week^{-1})	Bias (min·week^{-1})	MAPE (%)
Primary stage				
In-school MVPA	353.4 (19.1)	340.5 (24.8)	12.9 (10.6)	4.2
Out-of-school MVPA	379.1 (3.9)	408.4 (70.3)	-29.3 (69.5)	13.4
Weekend MVPA	300.1 (10.7)	317.9 (19.2)	-17.8 (10.7)	5.5
Out-of-school SB	722.1 (5.5)	674.9 (92.8)	47.2 (91.5)	12.5
Secondary stage				
In-school MVPA	230.4 (36.0)	219.5 (68.0)	10.9 (33.1)	15.6
Out-of-school MVPA	308.3 (42.2)	288.5 (69.4)	19.8 (36.4)	13.4
Weekend MVPA	192.0 (28.0)	196.4 (50.7)	-4.5 (28.8)	10.5
Out-of-school SB	1469.1 (89.2)	1350.7 (86.2)	118.4 (50.0)	8.0

Table 7. Group-level estimates of MVPA and ST, relative bias, and MAPE for boys and girls.

Segment	YAP-predicted estimates (min·week^{-1})	SWA estimates (min·week^{-1})	Bias (min·week^{-1})	MAPE (%)
Boys				
In-school MVPA	342.1 (63.1)	328.7 (66.2)	13.4 (36.9)	10.3
Out-of-school MVPA	366.7 (36.3)	369.2 (66.8)	-2.5 (51.6)	11.5
Weekend MVPA	282.4 (56.6)	262.7 (69.0)	19.7 (54.2)	17.5
Out-of-school SB	1101.0 (426.9)	1011.3 (404.8)	89.7 (78.2)	11.8
Girls				
In-school MVPA	242.4 (67.2)	228.7 (81.6)	13.7 (25.4)	11.7
Out-of-school MVPA	328.2 (51.4)	332.2 (109.8)	-3.9 (71.1)	16.0
Weekend MVPA	211.7 (56.1)	247.8 (86.9)	-36.1 (36.1)	15.5
Out-of-school SB	1093.8 (377.5)	1032.6 (348.6)	61.2 (84.1)	9.7

Figures

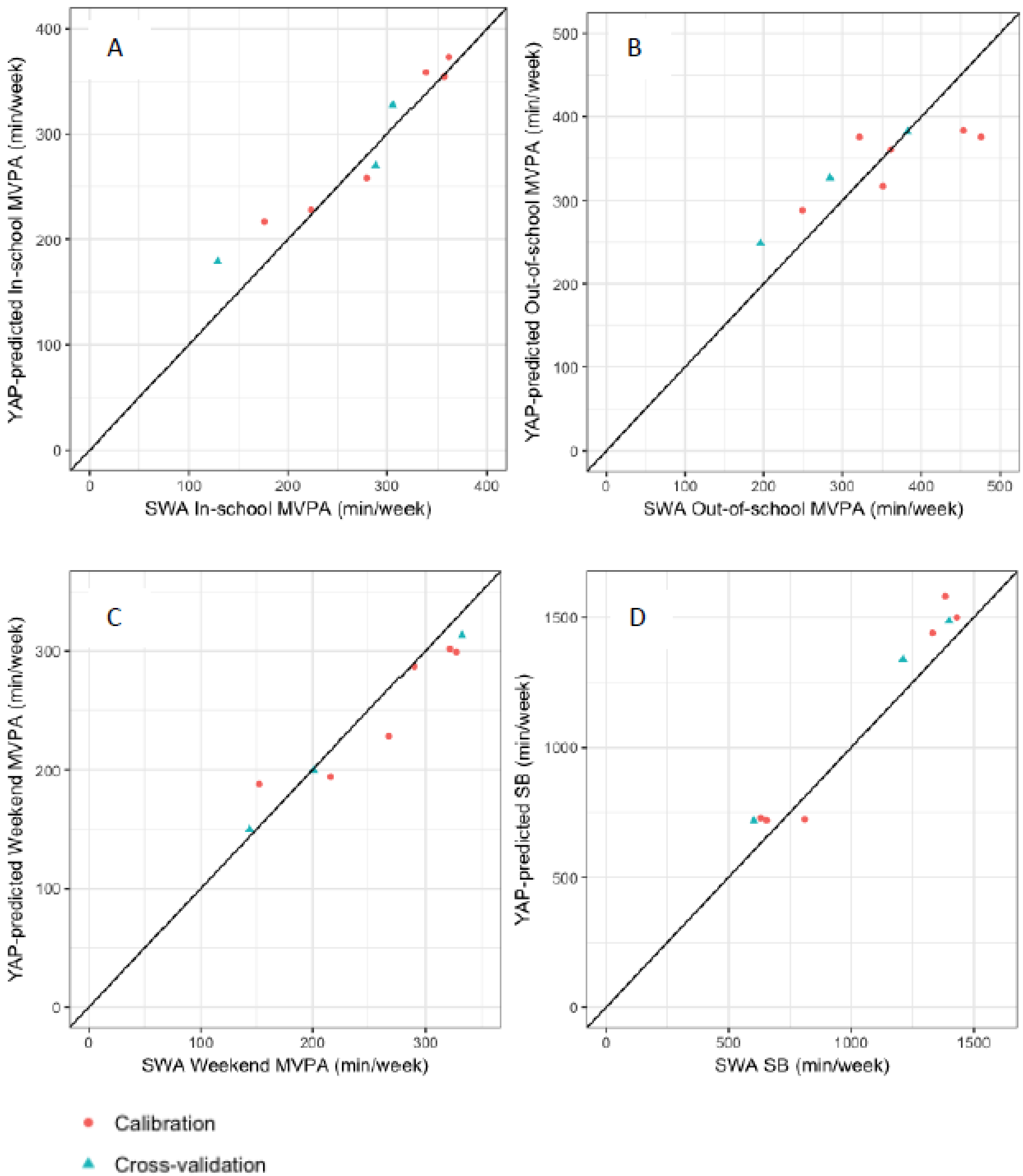

Figure 1

Group-level agreement between SWA estimates of MVPA and SB and predicted MVPA and SB minutes using the English YAP prediction algorithms. LEGEND: Panel A: in-school MVPA; Panel b: out-of-school MVPA; Panel C: weekend MVPA; Panel D: out-of-school SB. Each data point represents one school; red dots: the calibration sample schools used for generating the English YAP equations, blue triangles: the cross-validation sample schools used to assess prediction accuracy.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile7.Grouplevelcvpercent.pdf
- Additionalfile6.Individuallevelcvminpercent.pdf
- Additionalfile3.ScoringconversionofrawYAPvalues.pdf
- Additionalfile5.EnglishYAPmodels.pdf
- Additionalfile1.AmendmentstooriginalYAPwording.pdf
- Additionalfile2.YAP.pdf
- Additionalfile4.IndividuallevelresultsUSYAPvsSWA.pdf