PEDIATRICOBESITY
LETTERTO THE EDITOR doi:10.1111/ijpo.256

Deriving and evaluating new accelerometer cut-points in young children – a comment on Johansson *et al*. (2014)

We read Johansson *et al*.'s (1) article with great interest, as validation of accelerometers for children under 3 years of age is urgently needed. However, several methodological issues warrant serious consideration. Among the remarks listed on Table 1, three main issues deserve highlighting:

1. Only a few inappropriate tests were used for 'agreement' between cut-points and Children's Activity Rating Scale (2); agreement of categorical classification was not tested (e.g. Cohen's kappa (4)) and Spearman's rank measures strength of association not agreement between two scale measures (in contrast to Lin's concordance coefficient (3)). Because strong correlations do not warrant strong agreement (3), conclusions regarding agreement should not be drawn;
2. Not providing sensitivity/specificity and area under the curve values for sedentary behaviour, light- and high-intensity physical activity in the cross-validation denies the reader a chance to evaluate the validity and reliability of derived cut-points outside of the calibration sample, and subsequently their utility in other free-living populations;
3. Although differences in epoch durations (among others) may hinder direct comparisons, whether or not the derived cut-points show advantages/disadvantages over already published cut-points (e.g. through sensitivity/specificity) is neither discussed nor demonstrated.

Clear and detailed information on the protocols/procedures used together with thorough evaluations of new cut-points/equations (including how these improve on existing ones) is imperative and should be encouraged to move the paediatric physical activity measurement field forward (5). This will avoid introducing confusion not only in the measurement literature (5), but also into research investigating the influences on young children's physical activity/sedentary behaviour, their relationship with health/disease and the effectiveness of interventions.

S. Costa[1], N. Cameron[2] and S. A. Clemes[2]
[1]Institute of Child Health, University College London, London, UK; [2]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

Table 1 Description of main methodological and inconsistency issues in Johansson *et al*.'s study

Description of calibration procedures	What activities were observed in the obstacle course and during free play? Were parents/caregivers and multiple/individual children involved in the indoor and outdoor sessions? Such information is needed to evaluate representativeness and suitability to assess the type and intensities of daily activities in the target population (5).
Use of CARS observation system	Were entire 5-s intervals classified as one only CARS score? If so, what were the decision rules for classification (e.g. if 3 s were spent sedentary and 2 s were in moderate-to-vigorous PA, how would that 5-s interval be classified)?
	Procedures are stated as similar to previous studies, but considering CARS level two as 'low-intensity PA' is different to such studies and the original CARS article (2).
Statistical analyses – categorical classification	AUC, sensitivity and specificity at calibration are only presented for SB and high-intensity PA. No similar information is given for light-intensity PA, thus the accuracy and reliability of the thresholds to assess light-intensity PA cannot be assessed.
	AUC, sensitivity and specificity was not assessed for any of the thresholds in the cross-validation sample, thus the accuracy and reliability of the thresholds in a sample other than that used to calibrate the accelerometer cannot be evaluated.
	Formal tests of agreement between CARS and threshold classification of epochs as SB, light- and high-intensity PA was not assessed in the calibration or the cross-validation sample – tests like Cohen's *kappa* (3) are required in such studies to assess strength of agreement between the criterion and the derived classification systems.

Table 1 Continued

Statistical analyses – continuous time estimates	Spearman's rank correlation was used to assess agreement between CARS observed and threshold estimates of SB, light- and high-intensity PA; however, (i) Spearman's rank correlation is a measure of 'strength of association' and not 'agreement' between two variables (as assessed by Lin's concordance coefficient (5)); and (ii) it is a test for non-normally distributed data, but there is no indication of tests or other procedures made to assess normality of data. Threshold estimated time spent in SB, light- and high-intensity PA in the calibration sample is not presented, and agreement with CARS observed time is not evaluated (even if using Spearman's rank correlation) – this denies the possibility of assessing any differences in biases and correlations (if using Spearman's rank correlation) with CARS in the calibration vs. the cross-validation sample.
Use of ROC analysis	Although the authors state that ROC analysis was used '*to maximize sensitivity and specificity*', one could argue that the results do not reflect this – the information on Table 3 indicates that only sensitivity (100%) was maximized for the sedentary behaviour threshold whereas specificity (92.3%) was maximized for the high-intensity PA threshold. Such imbalance between sensitivity and specificity will unavoidably result in highly biased estimates, particularly for less prevalent observations (such as high-intensity PA).
Inconsistencies between Methods and Discussion	The use of '*both structured activities and free play were included when deriving the intensity thresholds*' is presented as a strength of the study; however, this is not the case for the SB threshold derived from '*watching cartoon and drawing*' (according to the Methods section and legend of Table 3) and may confuse the reader with regards to the calibration procedures. The methods section states that only CARS score 1 was considered as SB and a score of 2 was considered light-intensity PA (in line with the presented '*CARS 2 + 3*' as light-intensity PA in Table 4); however, in the Discussion section, the authors contradict this by stating that both CARS scores 1 and 2 were considered SB in line with a previous study. In addition, the Discussion states that '*standing still*' was considered as SB but there is no reference in the Methods to standing still as an activity used to derive the SB threshold.
Discussion of results	The magnitude of bias between the CARS observed and threshold estimated time in SB, light- and high-intensity PA is neither presented in the Results nor considered in the Discussion section. The observed +8.5/+10.2% (*y*-axis/vector magnitude) bias in estimates of SB, −16.3/−24.5% for light-intensity PA and striking +78.5/+118.2% bias for high-intensity PA time (calculated from Table 4) warrant serious consideration when recommending the use of the cut-points in future studies, and arguably jeopardize the conclusion that the thresholds '*appear valid to categorize SB and PA intensity categories in children 2 years of age*'. The statement that '*no time spent in lower levels of PA will be classified as high intensity*' in the Discussion is contradictory to the study results – a 92.3% specificity for the high-intensity threshold (Table 3) means that 7.7% of SB or light-intensity PA will have erroneously been classified as high-intensity PA. In the Discussion, the authors state that '*there is a possibility that the thresholds developed in this study will overestimate time spent sedentary and underestimate time spent in higher intensities*'; the overestimation of high-intensity time seen in the cross-validation sample (Table 4) goes against this assumption – however, this is neither noted nor discussed by the authors. The Discussion statement that '*Observational criterion methods, such as CARS, might not be optimal in order to capture the intermittent activity pattern of small children*' is contradictory to the supportive statements and references given both in the Methods and further in the Discussion for the use of observational methods as a good criterion measure in young children. This makes it unclear whether the authors consider the CARS as a suitable criterion measure or not, and why an alternative method was not used if CARS might be considered unsuitable. The cut-points for vector magnitude show consistently higher biases than those for the vertical axis in relation to CARS assessed time in SB, light- and high-intensity PA (Table 4); however, this is neither mentioned nor discussed anywhere in the manuscript, and there is no advice from the authors regarding which cut-points (vector magnitude or vertical axis) seem to perform better and should be prioritized for use in future studies. Comparison with previously published cut-points is limited to the acceleration cut-points itself, monitor placement and epoch durations – all important to highlight but not very informative for future research. Although differences in epoch durations, placement sites and criterion measures may hinder direct comparisons, there is no consideration or discussion regarding the advantages/disadvantages of the derived cut-points over previously published cut-points (e.g. improved or poorer agreement, AUC or sensitivity/specificity values) to help researchers in choosing the most appropriate accelerometer and cut-points to use in future studies with such young children.

AUC, area under the curve; CARS, Children's Activity Rating Scale; PA, physical activity; ROC, receiver operating characteristics; SB, sedentary behaviour.

References

1. Johansson E, Ekelund U, Nero H, Marcus C, Hagstromer M. Calibration and cross-validation of a wrist-worn Actigraph in young preschoolers. *Pediatr Obes* 2014; doi: 10.1111/j.2047-6310.2013.00213.x; Epub ahead of print.

2. Puhl J, Greaves K, Hoyt M, Baranowski T. Children's Activity Rating Scale (CARS): description and calibration. *Res Q Exerc Sport* 1990; 61: 26–36.

3. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989; 45: 255–268.

4. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.

5. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. *Med Sci Sports Exerc* 2005; 37: S501–S511.

