Title page

Title: Comparability of children's sedentary time estimates derived from wrist worn GENEActiv and hip worn ActiGraph accelerometer thresholds

Lynne M. Boddy[1], Robert J. Noonan[1], Youngwon Kim[2], Alex V. Rowlands[3,4,5] Greg J. Welk[6] Zoe R. Knowles[1], Stuart J. Fairclough[7,8]

[1]Physical Activity Exchange, Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, UK

[2] MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus. Cambridge, UK

[3]Diabetes Research Centre, University of Leicester, Leicester General Hospital, Leicester, UK;

[4]NIHR Leicester Biomedical Research Centre, UK;

[5]Alliance for Research in Exercise, Nutrition and Activity (ARENA), Sansom Institute for Health Research, Division of Health Sciences, University of South Australia, Adelaide, Australia.

[6]Department of Kinesiology, College of Human Sciences, Iowa State University, Ames, IA

[7]Department of Sport and Physical Activity, Edge Hill University, Ormskirk, UK

[8]Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland

Corresponding author details: Dr Lynne M. Boddy
The Physical Activity Exchange,
Research Institute for Sport & Exercise Sciences,
Liverpool John Moores University,
62 Great Crosshall Street,
Liverpool,
L3 2AT
UK
Email: L.M.Boddy@ljmu.ac.uk
Tel: +44 (0)151 231 4275

Word count (excluding abstract and references): 2894

33 Title: Comparability of children's sedentary time estimates derived from wrist worn GENEActiv and
34 hip worn ActiGraph accelerometer thresholds
35

Abstract:

Objectives: to examine the comparability of children's free-living sedentary time (ST) derived from raw acceleration thresholds for wrist mounted GENEActiv accelerometer data, with ST estimated using the waist mounted ActiGraph 100 count·min^{-1} threshold.

Design: Secondary data analysis

Method: 108 10-11-year-old children (n=43 boys) from Liverpool, UK wore one ActiGraph GT3X+ and one GENEActiv accelerometer on their right hip and left wrist, respectively for seven days. Signal vector magnitude (SVM; *mg*) was calculated using the ENMO approach for GENEActiv data. ST was estimated from hip-worn ActiGraph data, applying the widely used 100 count·min^{-1} threshold. ROC analysis using 10-fold hold-out cross-validation was conducted to establish a wrist-worn GENEActiv threshold comparable to the hip ActiGraph 100 count·min^{-1} threshold. GENEActiv data were also classified using three empirical wrist thresholds and equivalence testing was completed.

Results: Analysis indicated that a GENEActiv SVM value of 51*mg* demonstrated fair to moderate agreement (Kappa: 0.32-0.41) with the 100 count·min^{-1} threshold. However, the generated and empirical thresholds for GENEActiv devices were not significantly equivalent to ActiGraph 100 count·min^{-1}. GENEActiv data classified using the 35.6 *mg* threshold intended for ActiGraph devices generated significantly equivalent ST estimates as the ActiGraph 100 count·min^{-1}.

Conclusions: The newly generated and empirical GENEActiv wrist thresholds do not provide equivalent estimates of ST to the ActiGraph 100 count·min^{-1} approach. More investigation is required to assess the validity of applying ActiGraph cutpoints to GENEActiv data. Future studies are needed to examine the backward compatibility of ST data and to produce a robust method of classifying SVM-derived ST.

Keywords: children, physical activity, inactivity, accelerometry, measurement

Introduction

Sedentary behaviour is increasingly viewed as an important health risk factor in children [1], and the detrimental effects of reallocating PA time to sedentary behaviours have been established [2]. Sedentary behaviour is defined as 'any waking behaviour characterized by an energy expenditure ≤1.5 METS while in a sitting, reclining or lying posture' [3] however for children the recommended upper boundary of energy expenditure is ≤ 2 METs or ≤1.5 child-METs [4]. It is common for researchers to assess sedentary time (ST) which is commonly defined as the time spent below the threshold of proprietary accelerometer counts representing light physical activity, rather than focussing on sedentary behaviour *per se*.

Accelerometers have been used for several years to quantify children's ST, but heterogeneous data processing and researcher decisions related to for example, device location, wear time criteria, and choice of thresholds, often mean that study methods lack consistency and comparability. The advent of newer accelerometer devices capable of raw acceleration data collection removes the reliance on proprietary counts and allows researchers more autonomy when examining data, whilst producing estimates of acceleration that in theory should be comparable between devices [5]. Therefore, devices that produce raw acceleration data for researchers to use, such as the GENEActiv and ActiGraph GT3X+ offer an opportunity to increase comparability between studies aiming to estimate ST using accelerometers.

Raw acceleration data from GENEActiv and ActiGraph accelerometers are increasingly being processed in the open source R package GGIR (http:/cran.r-project.org). GGIR auto-calibrates the data using local gravity as a reference [6], detects sustained abnormally high values and generates the average magnitude of dynamic acceleration (termed the Euclidean Norm Minus One (ENMO))[5, 7-9]. Recently, the ENMO metric has been used to estimate ST and physical activity in both children and adults[9-11], but significant differences have been reported for ST and PA estimated from counts and from raw acceleration signals. Authors have attributed these differences to the various intensity thresholds used to classify acceleration data across the reduction approaches and differences in wear-site [11], but they may also be due to the inherent differences between the proprietary counts and raw acceleration data. One recent study,

conducted in children, provided a method of calibrating raw acceleration data from wrist-worn monitors to counts based hip-warn physical activity estimates in an effort to harmonise data [9]. The study classified raw accelerations using a range of ENMO thresholds for wrist-worn monitors and aligned these to counts-based thresholds for hip-worn monitors, demonstrating that incremental thresholds enable simple group level comparisons to past estimates of physical activity derived from hip-worn accelerometer counts cutpoints. For traditional accelerometer counts-based protocols using hip-worn ActiGraphs, studies have widely adopted 100 vertical axis count·min^{-1} as the upper threshold for ST in children [12]. To date, the comparability of wrist-worn GENEActiv ENMO ST estimates to those generated using the ActiGraph 100 vertical axis count·min^{-1} method is unknown. Studies have utilised the ENMO regression equation published by Hildebrand et al. [8] which was generated using a laboratory protocol to classify ST, however, these thresholds have not been cross-validated for classifying ST or examined in comparison with other methods. More recent studies [13] utilised the Hildebrand et al. [8] laboratory protocol to general thresholds then examined the agreement between ST and activPAL (which was considered as a criterion reference standard measure) using free-living data. The thresholds demonstrated low specificity, overestimating sedentary time in comparison to the activPAL. The equivalence of wrist worn data classified using these approaches to the 100 count·min^{-1} standard is unknown. Therefore, researchers wishing to represent raw accelerations through ENMO cannot compare ST to previous counts-based research, and so a pragmatic solution to classifying ST is required.

The aims of this secondary data analysis were to examine the comparability of children's free-living ST derived using the ENMO metric for wrist mounted GENEActiv accelerometer data, with ST estimated using the waist mounted ActiGraph 100 count·min^{-1} threshold. This aim was addressed by examining, [1] if comparable ST estimates could be attained from wrist-mounted GENEActiv raw acceleration data anchored to the widely adopted 100 count·min^{-1} uniaxial hip-mounted ActiGraph ST threshold, and [2] the equivalence of ST estimates between the newly generated threshold, those published by Hildebrand et al. [8, 13] and the 100 count·min^{-1} uniaxial hip-mounted ActiGraph ST threshold.

METHODS

This is a secondary data analysis of data generated by a previous study [7]. After gaining University ethics approval, informed parental consent, and participant assent 108 10-11-year-old children (n=43 boys) were involved in this study. Data collection took place on school sites from January to May 2014. Stature and body mass were assessed to the nearest 0.1cm using a portable stadiometer (Leicester Height Measure, Seca, Birmingham, UK) and nearest 0.1kg (Seca, Birmingham, UK) respectively using standard techniques [14]. Body mass index (BMI), was calculated for each participant.

Sedentary time was assessed using two tri-axial accelerometers, one worn on the non-dominant wrist (GENEActiv; Activinsights, Cambs, UK) and one worn on the right hip (ActiGraph GT3X+; ActiGraph, Pensacola, FL). Both monitors were initialised using the same computer to record at a frequency of 100 Hz, and participants were asked to wear the monitors at all times for 7 consecutive days except when sleeping and engaging in water based activities (e.g., bathing, swimming).

ActiGraph monitors were analysed using ActiLife v 6.11.4 software (ActiGraph, Pensacola, FL). Twenty minutes of consecutive zero counts (1 minute spike tolerance) defined non-wear time, and these periods were subtracted from daily wear time [15]. Sedentary time was coded as $\leq$100 count·min^{-1} [12]. Valid days were defined as $\geq$540 min for a weekday [16] and $\geq$480 min for weekend days [17]. For each participant the valid weekday and weekend day with the longest wear time were selected and retained for analysis. For participants with no valid weekend data, the valid weekday only with the longest wear time was included within analysis. After establishing daily wear time, data for the included days were converted to 1-s epoch csv output files for further analysis.

GENEActiv data were downloaded using GENEActiv v 2.2 software (Activinsights, Cambs, UK) and saved as binary files. These were then processed in R (http://cran.r-project.org) using the GGIR package (version 1.1-4). To correct for sensor calibration error autocalibration was completed [6]. GGIR processing produced files in csv format. Each csv file contained the ENMO-derived average magnitude of dynamic acceleration values expressed in average *mg* [18]. GENEActiv csv files corresponding to the selected ActiGraph weekday and/or weekend days were taken forward to the next stage of analysis.

ActiGraph and GENEActiv time stamped data were synched, resulting in one csv file for each participant containing date- and time-stamped ActiGraph and GENEActiv data in 1 s epochs. Non-wear times were removed from each merged file according to the ActiLife wear time details generated for each participant's ActiGraph data. For the ROC analysis each participant's ActiGraph and GENEActiv data were then summed into 1 min epochs to allow data scoring using the ActiGraph vertical axis 100 count·min^{-1} as the reference value for sedentary time [12]. These data were then stacked into one csv file to create a dataset including all participants (n = 108, 43 boys).

To establish GENEActiv classification criteria anchored to the ActiGraph 100 count·min^{-1} ST threshold, ROC analysis was performed on the whole sample, which represented 126,999 minutes of monitor wear time. Threshold values were cross-validated using 10-fold hold-out groups stratified by sex [19], whereby separate cross-validation analyses were conducted with a randomly selected hold-out group for each iteration (11 participants [6 girls and 5 boys] per analysis cycle) [20]. Therefore, each ROC analysis was completed with 97 participants with 11 excluded to enable cross-validation. For each hold-out group 2x2 contingency tables were used to check classification agreement based on the GENEActiv classifications generated from each cross-validation ROC analysis. Computed sensitivity and specificity, Cohen's kappa coefficients, and percentage agreement between classifications were assessed.

After generating the classification threshold, ST data were scored using 1 minute epochs. Data were classified for each participant using the newly generated GENEActiv threshold, ActiGraph 100 count·min^{-1}. Additionally GENEActiv ST was scored using the solved regression equation published by Hildebrand et al [8], where ST was defined as ≤1.5 child-METS [4], resulting in a threshold of 22.6 *mg*. GENEActiv ST was also scored using the 56.3 *mg* GENEActiv and 35.6 *mg* ActiGraph thresholds from the Hildebrand et al. 2016 study [13]. The ActiGraph threshold was included as theoretically using the raw data methods should allow the application of the threshold to the GENEActiv device. Pairwise equivalence testing was completed between all combinations of the thresholds. For this study a 95% equivalence test was performed to examine whether the 90% confidence intervals for mean ST for each classification method completely fell within the proposed equivalence zone (±10% of the mean of ST)

defined by the other classification method, representing statistically significant equivalence. Equivalence testing has been increasingly used in recent PA research where differences testing is not appropriate [11, 21-24]. Difference testing provides information on whether two methods are statistically different, where in this context it is more useful to know whether two methods are statistically equivalent at the group level, thus providing similar estimates. Analyses were conducted using IBM SPSS Statistics v.22 (IBM, Armonk, NY) and Microsoft Excel 2010 (Microsoft, Redmond, WA) and R for Windows (http://cran.r-project.org).

Results

Mean anthropometric data, weekend and weekday accelerometer wear times and the number of days included within analysis for boys and girls are displayed in Table 1.

[TABLE 1 ABOUT HERE]

The ROC curve for the whole cohort (N = 108) indicated that a GENEActiv threshold of 51 *mg* (sensitivity = 81.2%, specificity = 57.4%, AUC 0.760, 95% CI = 0.758, 0.763) provided the most accurate classification of ST. The ROC generated cutpoints, sensitivity and specificity, agreement, and Kappa values for each hold-out analysis for ST can be viewed in supplementary material A. The hold-out analysis found that the ST ENMO threshold performed significantly better than random classification, with agreement ranging from 64.7-69.7% and Kappa values ranging from 0.32-0.41 (fair to moderate agreement [25]). The mean GENEActiv ST cutpoint generated was 51 *mg*, corresponding with the whole group threshold, therefore 51 *mg* was used for subsequent equivalence analysis.

Figure 1 displays the results of the equivalence testing using ActiGraph count·min^{-1} as the reference threshold. Mean time spent in ST for each classification is displayed in supplementary file B. None of the 90% CIs for the newly generated GENEActiv 51*mg* (630.6-666.7 min), Hildebrand 2014 22.6 *mg*

(323.2-362 min) or Hildebrand 2016 GENEActiv 56.3*mg* (673.5-711.1 min) were completely included within the zone of equivalence for the ActiGraph 100 count·min^{-1} (443.2-541.6 min), suggesting no statistically significant equivalence between the cut-points compared and the ActiGraph 100 count·min^{-1}, on average. The Hildebrand ActiGraph 2016 35.6*mg* threshold, applied to GENEActiv data yielded 90% CIs (492.9-527.5) that fell within the zone of equivalence, so is considered statistically equivalent to the GENEActiv, on average. The newly generated GENEActiv threshold ST estimates were, on average, significantly equivalent to the 2016 Hildebrand GENEActiv threshold, with the 90% CIs for the Hildebrand 2016 GENEActiv threshold of 56.3*mg* falling within the zone of equivalence for the threshold generated by our study (689.4-695.1 min, zone of equivalence 583.8-713.5 min). No other combinations exhibited statistically significant equivalence.

[FIGURE 1 ABOUT HERE]

Discussion

The aims of this secondary data analysis were to examine the comparability of children's free-living sedentary time (ST) derived from raw acceleration thresholds for wrist mounted GENEActiv accelerometer data, with ST estimated using the waist mounted ActiGraph 100 count·min^{-1} threshold. A GENEActiv wrist ST threshold of 51 *mg* was generated which demonstrated fair to moderate agreement between the cross-validation and whole samples. The fact that the free-living data reflected a typical range of sedentary activities undertaken by children gave it a high degree of ecological validity. Irrespective of this, ST estimated using the 51 *mg* was not equivalent to the ActiGraph 100 count·min^{-1} threshold and therefore is not an acceptable value to use to generate ST estimates from GENEActiv wrist accelerations that are compatible with estimates from waist-worn ActiGraphs. However, when applied to the GENEActiv data, the Hildebrand 35.6 *mg* ActiGraph wrist acceleration threshold produced significantly equivalent estimates of ST as the waist ActiGraph 100 count·min^{-1} suggesting that this threshold could potentially be applied to GENEActiv data to provide comparable estimates of ST. Whether this provides an accurate estimate of ST when compared to criterion reference methods

such as activPAL warrants further investigation, however this was not the purpose of the analysis conducted.

Field-based approaches to generating acceptable ST thresholds may be desirable because of their greater ecological validity, and because they may reduce the risk of misclassification associated with laboratory-derived thresholds being used in the field [26]. However, our findings suggest that the current thresholds used to classify ST using ENMO do not produce comparable estimates to those reported when using the standard 100 count·min^{-1} approach. The challenges of estimating ST from wrist accelerometry are becoming more established [27]. Accelerometers are predominantly designed to measure movement rather than postural allocations. Accelerations from hip- and wrist-worn accelerometers are highly correlated in children during ST and physical activities of moderate through to vigorous intensities [28]. However, correlations are weaker during stationary light intensity physical activity which can involve a combination of sitting and standing activities, as well as transitions between the two [28]. Sitting and standing often encompass a combination of sedentary time and time in light intensity physical activity, whereby a high degree of hip and wrist acceleration decoupling occurs. For example, an individual may be sitting but gesturing with their hands, or standing and throwing a ball, both of which involve movements that a hip monitor may not detect but that could be detected by a wrist mounted device. This lack of consistency between hip and wrist accelerations during some sedentary and light intensity activities provides some explanation of the moderate levels of agreement observed in the cross-validation analyses, and the lack of equivalence with the hip 100 count·min^{-1} threshold in particular.

The accuracy of classifying ST is not explored in this study, we simply looked at the comparability of the GENEActiv thresholds to the standard ActiGraph vertical axis 100 count·min^{-1} threshold. Whether the standard approach provides a more or less accurate estimate of ST is not examined and warrants further evaluation. To examine the accuracy of ST thresholds within a field-based protocol, a criterion measure, such as an inclinometer is needed. Theoretically this would increase participant burden through

the need to wear two devices, increase the cost of undertaking the research and data would still not allow for cross-comparisons between previous counts based studies. An alternative approach, that negates the need for additional devices, is to use accelerometers to examine assumed postural changes relative to arm elevation and wrist orientation (i.e., the Sedentary Sphere approach [27]). Recent evidence suggests that the Sedentary Sphere method provides comparable estimates of ST in adults when compared to the activPAL [29], however, this method has not been validated in children, and so further work is required to examine its utility of this method in this population.

The Hildebrand 22.6 *mg* ST threshold is based on GENEActiv wrist ENMO values, but was generated using VO_2 data rather than ActiGraph counts as in the current study. This may explain why the thresholds were not equivalent. In addition, the laboratory protocol used by Hildebrand et al.[8, 13] only included lying watching TV and sitting using a computer as sedentary activities [8]. Whilst such activities are common among children they do not reflect the wide range of free-living sedentary behaviours that the children involved in this study were likely to have engaged in. Further, the Hildebrand et al. (2014) 22.6 *mg* estimated ST threshold was calculated from a regression equation anchored to energy expenditure. As sedentary behaviours are characterised by posture and low energy expenditure, determining sedentary time using energy expenditure alone without posture classification may be a less accurate approach than using criterion measures such as inclinometers or direct observation [27]. Hildebrand et al.'s 2016 GENEActiv 56.5*mg* wrist threshold was similar to our 51*mg* threshold, though the former demonstrated low specificity, overestimating sedentary time in comparison to the activPAL when examining free-living data, which may be due to the limited number of sedentary stations included in the original laboratory protocol.

There are a number of limitations to this study. Our study was conducted in one geographical area of the UK and as such the results may not be representative of other populations. To classify GENEActiv data against the 100 count·min^{-1} criterion, we used a 1-minute epoch setting. Though this would likely result in the inability to detect movement at higher intensities, as sedentary behaviour is characterised

by a lack of movement the 1-minute epoch setting would have less impact upon the ST estimates generated. We did not use a criterion reference standard device such as activPAL within this study. This was by design, as the primary aim was to examine the comparability of simple accelerometer estimates rather than investigate the accuracy of the measurement of ST. Future studies should aim to utilise the activPAL and other reference methods to develop and validate ST thresholds for use in children.

Conclusions

Despite displaying fair to moderate agreement, the generated GENEActiv ST threshold does not provide an equivalent estimate of ST to the hip mounted ActiGraph 100 count·min^{-1} approach. Furthermore, ST data generated using Hildebrand thresholds were not equivalent to the 100 count·min^{-1} method. Future studies are needed to examine the backwards compatibility of ST data and to produce a robust method of classifying ENMO-derived ST.

Practical implications

- Estimates of children's sedentary time generated from GENEActiv wrist ENMO and ActiGraph 100 count·min^{-1} are not comparable.
- Researchers should not compare data generated using the two different methods.
- Future studies are required to provide methods of data harmonization and to establish valid and reliable sedentary time thresholds for children.

References

1. Cliff DP, Hesketh KD, Vella SA, et al. Objectively measured sedentary behaviour and health and development in children and adolescents: systematic review and meta-analysis. *Obesity Reviews.* 2016; 17(4):330-344.
2. Fairclough SJ, Dumuid D, Taylor S, 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.
3. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary behavior network (SBRN)- Terminology Consensus Project process and outcome. *International Journal of Behavioural Nutrition and Physical Activity.* 2017; 14(75).
4. Saint-Maurice PF, Kim Y, Welk GJ, Gaesser GA. Kids are not little adults: what MET threshold captures sedentary behavior in children? *European Journal of Applied Physiology.* 2016; 116(1):29-38.
5. Rowlands AV, Yates T, Davies M, Khunti K, Edwardson CL. Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Med Sci Sports Exerc.* 2016.
6. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *Journal of applied physiology.* 2014; 117(7):738-744.
7. Fairclough SJ, Noonan R, Rowlands AV, V VANH, Knowles Z, Boddy LM. Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Med Sci Sports Exerc.* 2016; 48(2):245-253.
8. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Med Sci Sports Exerc.* 2014; 46(9):1816-1824.
9. Rowlands AV, Cliff DP, Fairclough SJ, et al. Moving Forward with Backward Compatibility: Translating Wrist Accelerometer Data. *Med Sci Sports Exerc.* 2016; 48(11):2142-2149.
10. 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(21):2067-2072.
11. Kim Y, Hibbing P, Saint-Maurice PF, et al. Surveillance of Youth Physical Activity and Sedentary Behavior With Wrist Accelerometry. *Am J Prev Med.* 2017; 52(6):872-879.
12. Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc.* 2011; 43(7):1360-1368.
13. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. *Scand J Med Sci Sports.* 2016.
14. Lohman T, Roche AF, Martorell R. *Anthropometric standardization reference manual*, Champaign, Illinois, Human Kinetics; 1988.
15. Catellier DJ, Hannan PJ, Murray DM, et al. Inputation of missing data when measuring activity by accelerometry. *Med Sci Sports Exerc.* 2005; 37(Suppl 11):S555-S562.
16. Rich C, Geraci M, Griffiths L, Sera F, Dezateux C, Cortina-Borja M. Quality control methods in accelerometer data processing: defining minimum wear time. *PloS one.* 2013; 8(6):e67206.
17. Rowlands AV, Pilgrim EL, Eston RG. Patterns of habitual activity across weekdays and weekend days in 9-11-year-old children. *Prev Med.* 2008; 46(4):317-324.
18. van Hees VT, Gorzelniak L, Dean Leon EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One.* 2013; 8(4):e61691.
19. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Med Sci Sports Exerc.* 2012; 44(1 Suppl 1):S39-49.
20. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. *Med Sci Sports Exerc.* 2012; 44(1 Suppl 1):S61-67.

21. Kim Y, Welk GJ. Criterion Validity of Competing Accelerometry-Based Activity Monitoring Devices. *Med Sci Sports Exerc.* 2015; 47(11):2456-2463.
22. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc.* 2014; 46(9):1840-1848.
23. Welk GJ, Kim Y, Stanfill B, et al. Validity of 24-h physical activity recall: physical activity measurement survey. *Med Sci Sports Exerc.* 2014; 46(10):2014-2024.
24. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. *Medicine and Science in Sports and Exercise.* 2017; Ahead of Print.
25. Altman DG. *Practical Statistics for Medical Research.*, London, Chapman and Hall; 1991.
26. Freedson PS, John D. Comment on "Estimating Activity and Sedentary Behavior from an Accelerometer on the Hip and Wrist". *Medicine & Science in Sports & Exercise.* 2013; 45(5):962-963 910.1249/MSS.1240b1013e31827f31024d.
27. Rowlands AV, Olds TS, Hillsdon M, et al. Assessing sedentary behavior with the GENEActiv: introducing the sedentary sphere. *Med Sci Sports Exerc.* 2014; 46(6):1235-1247.
28. Rowlands AV, Rennie K, Kozarski R, et al. Children's physical activity assessed with wrist- and hip-worn accelerometers. *Med Sci Sports Exerc.* 2014; 46(12):2308-2316.
29. Pavey TG, Gomersall SR, Clark BK, Brown WJ. The validity of the GENEActiv wrist-worn accelerometer for measuring adult sedentary time in free living. *Journal of science and medicine in sport / Sports Medicine Australia.* 2015:http://dx.doi.org/10.1016/j.jsams.2015.1004.1007.

366 **Table legend**

367 Table 1. Mean (SD) anthropometric, wear time and number of days included within analysis for boys

368 and girls

369

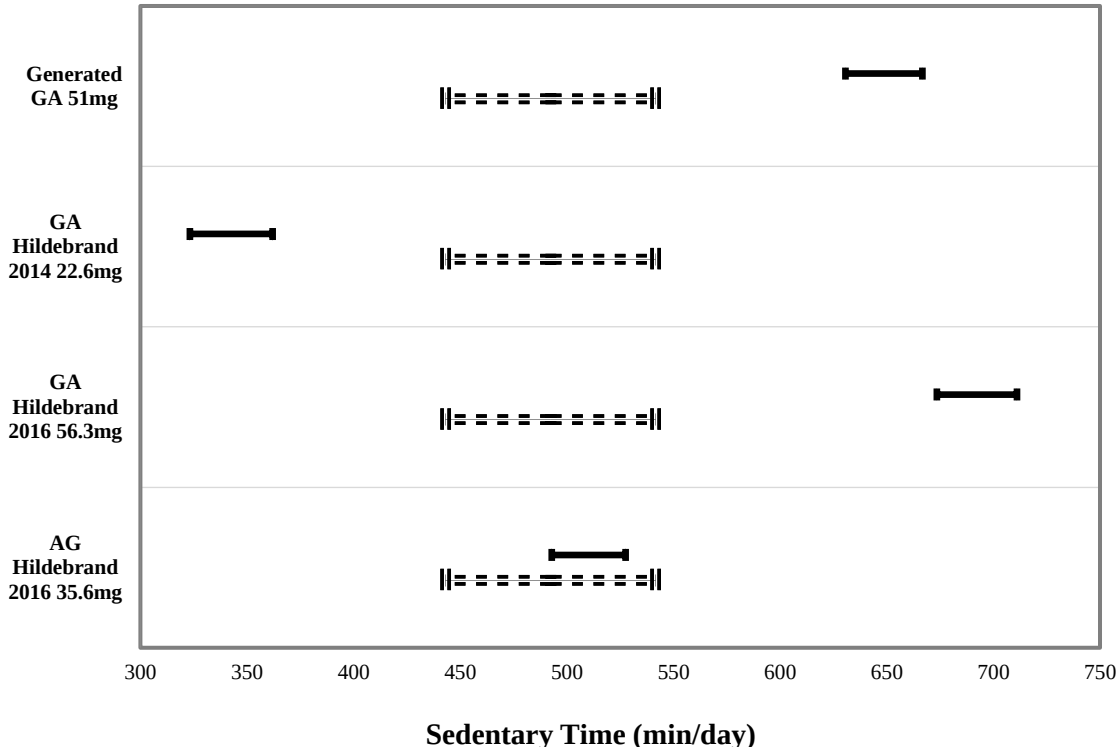
	Boys N = 43		Girls N = 65	
	Mean or Frequency	SD	Mean or Frequency	SD
Age (y)	10.03	0.35	10.04	0.31
Height (cm)	139.49	7.89	137.97	7.37
Body mass (kg)	35.64	8.24	34.23	8.60
BMI ($kg \cdot m^2$)	18.15	3.00	17.78	3.18
ActiGraph weekday wear ($min \cdot day^{-1}$)	739.88	115.55	738.75	100.35
ActiGraph weekend day wear ($min \cdot day^{-1}$)	631.83	110.82	661.50	108.28
ActiGraph valid weekdays included	41	N/A	64	N/A
ActiGraph valid weekend days included	30	N/A	46	N/A
Total valid included days	71	N/A	110	N/A

370

371

372 **Figure legend**

373 Figure 1. ActiGraph 100 count·min^{-1} zone of equivalence (dotted lines) and 90% confidence intervals
374 for the GENEActiv sedentary time data



375