

Running head: Longitudinal PA assessment

**Cross-Sectional and Longitudinal Comparisons of Self-Reported and Device-Assessed Physical Activity and Sedentary Behaviour**

\*Robin L. J. Lines<sup>1,2</sup>, Nikos Ntoumanis<sup>2,3</sup>, Cecilie Thøgersen-Ntoumani<sup>2,3</sup>, Joanne A. McVeigh<sup>2,4,5</sup>,  
Kagan J. Ducker<sup>1</sup>, David Fletcher<sup>6</sup>, and Daniel F. Gucciardi<sup>1,2</sup>

<sup>1</sup>*School of Physiotherapy and Exercise Science, Curtin University*

<sup>2</sup>*Physical Activity and Well-Being Research Group, Curtin University*

<sup>3</sup>*School of Psychology, Curtin University*

<sup>4</sup>*School of Occupational Therapy, Social Work, and Speech Pathology, Curtin University*

<sup>5</sup>*Movement Physiology Laboratory, School of Physiology, University of Witwatersrand*

<sup>6</sup>*School of Sport, Exercise and Health Sciences, Loughborough University*

Author Notes

\*Address correspondence to Robin Lines, School of Physiotherapy and Exercise Science, Curtin University, GPO Box U1987, Western Australia, 6845. Phone: +61 8 9266 3653. Email: [robin.lines@curtin.edu.au](mailto:robin.lines@curtin.edu.au)

Lines, R.L.J., Ntoumanis, N., Thøgersen-Ntoumani, C., McVeigh, J., Ducker, K.J., Fletcher, D., & Gucciardi, D.F. (in press). Cross-sectional and longitudinal comparisons of self-reported and device-assessed physical activity and sedentary behaviour. *Journal of Science and Medicine in Sport*. doi: 10.1016/j.jsams.2020.03.004

### Abstract

**Objectives:** To examine the longitudinal associations and differences between self-reported and device-assessed physical activity (PA) and sedentary behaviour (SB), using a multifaceted statistical approach.

**Design:** Longitudinal measurement burst. **Methods:** In total, 52 university students (78% female) aged 18 - 38 years (mean =  $21.94 \pm 4.57$  years) participated. The study consisted of three blocks of six days of measurement, during which participants wore an accelerometer on their wrist for the entire block, and self-reported their PA over the 6 days at the end of each block. **Results:** Meaningful latent differences between methods were observed for moderate PA and SB across all three assessment periods, such that participants underreported the time spent in each activity. Bland-Altman plots revealed a positive mean difference for vigorous PA, with over-reporting increasing as mean levels increased. Negative mean differences were observed for all other intensities. Underreporting of moderate PA increased as the mean level increased, whereas for light PA and SB, underreporting decreased at high levels. Repeated measures correlations revealed a meaningful association for vigorous PA only, suggesting that as self-reported minutes increase so too do device-measured minutes. **Conclusions:** We found evidence of cross-sectional and longitudinal differences and weak associations between self-reported and device-assessed PA and SB. Future work is needed to enhance the quality of self-reported methods to assess PA and SB (e.g., face and content validity), and consider improvements to the processing of device-based data.

**Keywords:** bias; exercise; sedentary behaviour; accelerometers; discrepancies

## Introduction

Due to its numerous physical (e.g., reduced risk of cardiovascular disease, obesity, type 2 diabetes, certain types of cancer <sup>1</sup>) and mental (e.g., lower levels of psychological distress <sup>2</sup>) health benefits, precise measurement of physical activity (PA) is essential to develop effective interventions aimed at maximising engagement in PA and minimising sedentary behaviour (SB). However, PA is challenging to assess, because it is a complex and multi-faceted behaviour <sup>3</sup>. Measurement of PA in the field is typically achieved via subjective (e.g., questionnaires) or device-based (e.g., accelerometers) methods. Subjective measurements of PA are the most commonly used method (e.g., self-reported surveys or diaries) <sup>4</sup>. These methods rely on participants recalling their activity levels over a designated period, usually up to the last seven days <sup>5</sup>. Self-reports have a number of benefits, including ease of use, low cost, and low burden to both researchers and participants, yet they are prone to measurement errors (e.g., social desirability <sup>6</sup>). In particular, precise recall of the duration, frequency, and intensity of PA can be cognitively challenging, especially for incidental rather than planned activities <sup>7</sup>.

Given the limitations associated with self-report measures, device-based measures of PA (e.g., accelerometers) are commonly used to maximise the accuracy of measurements <sup>6,8</sup>. Device-assessed PA circumvents many limitations associated with self-report measures <sup>7</sup> and, therefore, provides more accurate assessments of PA than self-reports alone <sup>6</sup>. However, device-based methods are often expensive, increase participant burden, and require expertise to process and analyse the data <sup>5</sup>. Additionally, some devices are limited in their ability to quantify activities, such as cycling, swimming, and weightlifting <sup>7,9</sup>, or are prone to measurement reactivity <sup>10</sup>. Therefore, accurate assessment of PA is challenging outside of a laboratory setting <sup>9</sup>.

The International Physical Activity Questionnaire (IPAQ <sup>11</sup>) is one of the most commonly used self-report measures of PA; consequently several studies have assessed its utility against device-based measures <sup>5,12</sup>. For example, in an assessment of the IPAQ across 12 countries, correlations with device-assessed PA ranged from 0.02 to 0.47 <sup>13</sup>. A later review of 23 studies indicated that correlations between the IPAQ and device-assessed total PA varied widely ( $r=0.09$  to  $0.39$ ) and were below an acceptable standard ( $r=0.5$ ) <sup>14</sup>. This problem is not confined to the IPAQ; a systematic

review of 57 studies reported correlations between PA questionnaires and accelerometers to range between  $r = 0.04$  and  $0.47$ <sup>15</sup>. Collectively, this body of evidence suggests that self-reports and device-based assessments do not seem to capture sufficiently similar levels of PA.

A notable limitation of past research is the reliance on cross-sectional studies. In a recent review of self-report and device-based measures of PA, only 6 of 57 papers reviewed (10.5%) compared assessment methods over more than one time-point<sup>15</sup>. Focused specifically on comparing assessments of the IPAQ with objectively assessed PA, only one of 23 studies (4.3%) compared methods over multiple time-points<sup>14</sup>. This trend highlights the need for additional research comparing self-report and device-based measures over multiple time-points to account for the temporal dynamics of these assessments. An assumption of designs that involve repeated assessments of PA and SB for multiple individuals (e.g., randomised trials, longitudinal) is congruency in methods both between people and within individuals over time, yet our knowledge of this assumption remains underdeveloped. Additionally, the use of raw acceleration data in past research is limited<sup>16</sup>. Accelerometer data are often compressed into counts, which are generated when acceleration levels are above a chosen threshold for a defined time-frame (e.g., 1 to 60 seconds<sup>17</sup>). This analytical method results in smoothing of the data, hampering differentiation between activities in some devices (e.g., sedentary behaviours and low-intensity activities over a short duration<sup>16</sup>). Population-based studies (e.g., Whitehall II Study, and UK Biobank) have begun to gravitate towards the use of raw accelerometer signals, which can be obtained from a variety of accelerometer devices and processed and analysed with open-source software. Using high-resolution raw accelerometer data permits the identification of activity (intensity, duration and type) via algorithmic recognition of patterns in the raw data<sup>16</sup>. In addressing these issues, therefore, we provide a novel contribution by combining raw accelerometer data with contemporary analytical methods.

Analytically, there has been an overreliance on correlation coefficients within past research<sup>15</sup>. These coefficients provide an indication of the strength of the association, not the level of agreement between methods<sup>18</sup>. Simply put, a high correlation value does not mean a high level of agreement. Alternative statistical approaches, such as Bland-Altman plots, are able to reveal discrepancies between measures that may be overlooked when using correlation coefficients<sup>19</sup>. Also, as correlation

and regression methods rely on independence of observations, they are poorly suited to assessing repeated measures, where aggregation of observations violates the assumption of non-independence, leading to biased and spurious results<sup>20</sup>. Repeated measures correlation addresses this limitation by accounting for non-independence of observations and avoiding the requirement to aggregate data, thereby increasing statistical power<sup>20</sup>. An additional advantage of repeated measures correlation is that it provides a statistical index of the within-person association between methods that is common among a sample<sup>20</sup>; this consideration is absent from past research on the congruence between subjective and device-assessed PA.

Against this backdrop, the aim of the current study was to examine the associations and differences between self-reported and device-based assessments of PA and SB using an underutilised longitudinal design. We compared measures across three time-points to shed light on congruence effects over time. Addressing this objective is essential because knowledge of the congruence between methods is important for intervention development and assessment. Analytically, we use a novel, multidimensional perspective on the issue of methodological congruence, through the use of latent difference scores (absolute differences), Bland-Altman plots, and repeated measures correlations which are unused in this research domain.

### **Methods**

The lead author's University Human Research Ethics Committee approved this study. In total, 52 university students aged 18-38 years (mean =  $21.94 \pm 4.57$  years) participated. Participants were recruited via a University research participation pool in which students sign up to take part in studies in return for incentives (i.e., vouchers), and via invitations to participants from a previous study conducted by our team (blinded for peer review). The sample was 78.8% female (n=42), with a mean body mass index (BMI) of  $23.2 \pm 3.6$ ; 73% (n=38) spent time outside of university working in a paid job (mean =  $10.29 \pm 8.73$  hrs).

This study took place over a four-month period, which consisted of three blocks of six days of PA measurement, each separated by eight weeks. Participants attended a brief lab session at the beginning and end of each measurement block. In the initial session, participants were provided with an overview of the study; consenting participants completed a multi-section survey online via

Qualtrics (Qualtrics LLC, Utah, USA). At the start of each block, participants collected their accelerometers (GENEActive Original, Activinsights Ltd), and were instructed to wear them on their non-dominant wrist for the entire measurement block. One week later participants returned their accelerometers to the lab, self-reported their PA and SB for the past week, and received a reward (\$25 voucher). Of the 52 participants, 75.0% (n=39) completed all three blocks, 15.4% (n=8) completed two blocks, and 9.6% (n=5) completed a single block.

Self-reported PA was measured using the 7-item International Physical Activity Questionnaire – Short Form (IPAQ<sup>11</sup>). Six items target the frequency (i.e. days per week) and duration (i.e. hours and minutes) of three PA intensities (light, moderate, and vigorous). Each PA intensity was assessed with two items. The final item assessed sitting time (e.g., “how much time do you spend sitting on a week day?”). IPAQ data were processed using the standard scoring protocol ([www.ipaq.ki.se](http://www.ipaq.ki.se)) to calculate the total minutes, over the 6 day measurement period, for each PA intensity and SB.

Device-assessed PA and SB were collected using a GENEActive Original triaxial accelerometer (Activinsights Ltd, Kimbolton, Cambs, UK), worn on participants non-dominant wrist 24 hours a day. The GENEActive is compact, lightweight (16g), waterproof, and measures acceleration in three axes ranging between -8 g and +8 g. The accelerometer was set to a sampling rate of 60 Hz<sup>21</sup>. Data were downloaded using the GENEActiv software version 3.1, with raw .csv files converted into .bin files for data processing. Data were analysed using R package GGIR version 3.3.3 (<https://cran.r-project.org/web/packages/GGIR>). Raw data processing in GGIR facilitates data cleaning using autocalibration with local gravity as the reference<sup>22</sup>, as well as detection of non-wear time<sup>23</sup> and sustained abnormally high levels of acceleration, and supports extraction of defined levels of acceleration which are set to reflect intensity levels of PA. The summary measures reported here are time spent (in minutes) in sedentary (<50 mg), light (LPA; 50 – 100 mg), moderate (MPA; 100 – 400 mg), and vigorous (VPA; >400 mg) intensities<sup>21</sup>. Accelerometer data were confined to 6 full days and nights (4 weekdays and 2 weekend days), starting at waking time on the day after the devices

were collected. This decision was made so each data set represented the same 6 day window for analysis, as accelerometer return date sessions sometimes exceeded the 7 day measurement duration.

We carried out three main analyses to examine the dis/agreement between the two measures of PA and SB. First, we examined latent differences in total time for each intensity within each time block, using a latent growth modelling framework<sup>24</sup> available in *Mplus* 8<sup>25</sup>. Device measured PA intensities were used as the intercept in these models. Second, we used the Bland-Altman method<sup>18</sup> to illustrate and assess the agreement between the two methods, and allow for the identification of systematic differences between methods. Finally, we used repeated measures correlations to assess the within-person associations between the two assessment methods, using the R package *rmcorr*<sup>20</sup>. Latent growth models and repeated measures correlation permit insight into the longitudinal dynamics *across each* measurement block, whereas Bland-Altman plots provide an understanding of statistic associations *within each* assessment block.

## Results

Descriptive statistics for total time spent in each activity across each testing period for all methods are presented in Table 1. Latent differences in time spent in each activity are presented in Table 2. We observed meaningful differences across all three blocks for MPA and SB, such that participants' under-reported time spent in each activity relative to the data obtained from their device. Participants' over-reported time spent in VPA, but only at block two. All other differences between measures were small and incompatible with a meaningful effect.

Results from the Bland-Altman analyses are presented in Table 3 and the supplementary material (see Figures S1-S4). There was a positive mean difference in VPA across each time point ( $M_{\text{block1}} = 65.9$ ;  $M_{\text{block2}} = 33.5$ ;  $M_{\text{block3}} = 62.4$ ), indicating an over-reporting of VPA. Plots indicated smaller differences between methods at lower levels of VPA, though as the mean level increases so does over-reporting of VPA. There were negative mean differences between methods for LPA ( $M_{\text{block1}} = -473.2$ ;  $M_{\text{block2}} = -384.9$ ;  $M_{\text{block3}} = -334.0$ ), MPA ( $M_{\text{block1}} = -540.5$ ;  $M_{\text{block2}} = -555.0$ ;  $M_{\text{block3}} = -468.5$ ), and SB ( $M_{\text{block1}} = -2035.5$ ;  $M_{\text{block2}} = -1729.7$ ;  $M_{\text{block3}} = -1789.6$ ). At higher mean levels of MPA, under-reporting also increases. With regard to LPA and SB, at higher levels of both intensities under-reporting reduces, yet there is over-reporting at the highest levels of LPA.

Visual depictions of repeated measures correlations to assess the within-person associations between the two methods are presented in the supplementary material (see Figures S5-S8). For VPA, self-reported minutes were significantly positively correlated with device-based time ( $r_{(78)} = 0.25$ , 95% CI = [0.02 – 0.44],  $p = 0.028$ ). This finding suggests that as levels of self-reported VPA increase so do levels of accelerometer measured PA. The correlations for MPA ( $r_{(78)} = 0.04$ , 95% CI = [-0.19 – 0.26],  $p = 0.751$ ), LPA ( $r_{(71)} = 0.14$ , 95% CI = [-0.10 – 0.36],  $p = 0.253$ ), and SB ( $r_{(68)} = -0.05$ , 95% CI = [-0.29 – 0.19],  $p = 0.671$ ) were small and incompatible with a meaningful effect.

### Discussion

The aim of this study was to assess the dis/agreement between self-reported and device-assessed PA and SB across three time points, using a multi-faceted statistical approach. In so doing, we offered methodological (i.e., longitudinal and use of raw accelerometer data) and analytical (i.e., multiple approaches) extensions to past work on the associations between these methods of PA and SB assessment. Overall, we found evidence of cross-sectional and longitudinal mean differences, and mainly weak associations between self-reported and device-based measures of PA and SB.

The latent difference scores indicated that participants largely underreported PA intensities and SB across all three time points, with meaningful differences observed for MPA and SB. Bland-Altman plots were generally consistent with these findings, indicating the presence of incongruence for MPA, LPA, and SB. Differences between assessment methods were most pronounced for SB. As sedentary behaviours are dispersed throughout the entire day, rather than planned bouts that may be memorable to participants (e.g., going for a run or swim), it may be difficult to recall all instances of SB accurately<sup>8</sup>. Consistent with past work<sup>5,8,12</sup>, it appears that the use of a single item focused on sitting time is a poor representation for the assessment of sedentary behaviours that are much broader in scope (e.g., lying, passive standing<sup>26</sup>). Composite measures encompassing different domains of SB are better than a single item-assessment<sup>27,28</sup>. Therefore, although the IPAQ was never truly designed to be a measure of SB, due to its popularity, there may be a need to revise the existing single SB item in favour of a composite assessment.

Previous findings regarding the consensus between self-reports and device-based assessments of MPA have been mixed. In a narrative review of reviews including 63 papers, of which 17



compared self-reported and device-assessed PA, self-reports of MPA were underestimated in approximately 23% of comparisons and overestimated in roughly 77%<sup>29</sup>. Latent difference scores and Bland-Altman plots indicated substantial underestimation differences between self-reported and device-assessed MPA in the current study. One explanation for these findings relates to the cognitive demands required to recall accurately distinct intensities of PA. For example, VPA is typically planned and structured (e.g., running, sports training), whereas MPA is more incidental and can be accumulated throughout the day with minimal structure and planning (e.g., running to catch a bus), making it harder to recall with precision. Methodological factors might also play a role in these discrepancies. The MPA item within the IPAQ, for example, asks participants to recall events that last for at least 10 minutes, which make them 'breathe somewhat harder than normal'. What is normal for one person may differ to others (e.g., aerobic fitness) and may introduce bias in individual perceptions of intensity. The operationalisation of MPA within the IPAQ therefore requires enhancements to capture the essence of this PA intensity (e.g., suite of activities relevant to different fitness cohorts).

We observed the smallest mean differences and largest cross-sectional and longitudinal associations between methods for the assessment of VPA. Nevertheless, these associations were small-to-moderate in magnitude, and differences between methods indicated a degree of overestimation of VPA via self-reports, a pattern also seen in previous research<sup>6,12,29</sup>. These differences may be explained by the inability of wrist worn accelerometers to detect certain activities that would be classified as vigorous intensity (e.g., weightlifting, cycling)<sup>8,9</sup>. It is also possible that vigorous activities are easier for people to remember as these types of activities are planned, structured, and distinct from other everyday activities<sup>8</sup>. Relatedly, the over-reporting of VPA and under-reporting of MPA may occur because we have no inbuilt schema to differentiate between PA intensities, such that individuals overestimate moderate activities and report them as vigorous in nature<sup>12</sup>. Such an interpretation again reinforces the need to consider notions of content and face validity for high-quality operationalisations of MPA and VPA via self-reports, and their integration as a composite measure of MVPA<sup>8,30</sup>.

Strengths of the current study include temporal alignment between measurement epochs, which seems logical but has been overlooked in past research<sup>7,12</sup>, the use of raw accelerometer data in

analyses, assessments taken over multiple time-points, and a multi-dimensional statistical approach. Nevertheless, some limitations should be considered when interpreting findings. First, wrist worn accelerometer assessment is incapable of detecting certain activities (e.g., cycling) and therefore potentially underestimates PA. Second, the PA intensity thresholds selected for device-assessed data processing influence the time spent in each intensity, noting that we chose thresholds found to be reliable in previous work for an adult population<sup>21</sup>. Third, although we captured daily assessments of device-assessed PA, we aggregated each 6-day period to align with the temporal focus of self-reported PA; future research may benefit from assessments of congruence between daily assessments of both methods. Fourth, given the relatively small sample, our design and test combination was likely sufficiently powered to detect moderate-to-large effects. The current findings may prove informative for simulation work to guide future research, particularly for examinations of potential moderators (e.g., gender) of the association between methods, which would require larger samples. Finally, the nature of our sample (e.g., young, predominantly female, healthy university students) needs to be taken into consideration when assessing the extent to which these findings generalise to other populations. It may therefore be particularly important to examine this question in physically inactive and overweight/obese adults.

### **Conclusion**

The current findings indicate there are meaningful differences between self-reported and device-assessed PA and SB, when examined cross-sectionally and longitudinally. The findings regarding longitudinal effects have important implications for future work focused on the temporal dynamics of naturally occurring (e.g., observational panel study) or experimentally manipulated (e.g., RCTs) PA and SB. From a methodological standpoint, the current findings underscore the importance of additional work geared towards improving operationalisations of PA and SB via self-report tools in ways that maximise content and face validity.

### **Practical Implications**

- People appear to recall vigorous physical activities most accurately.
- Strengths and limitations of assessment methods of PA and SB should be considered when selecting a measurement approach for one's purposes.

- Face and content validity are essential considerations for enhancing existing self-report tools of PA and SB or developing new measures.

## References

1. Rhodes RE, Janssen I, Bredin SSD, Warburton DER, Bauman A. Physical activity: Health impact, prevalence, correlates and interventions. *Psychology & Health*. 2017;32(8):942-975.
2. Gucciardi DF, Law KH, Guerrero MD, et al. Longitudinal relations between psychological distress and moderate-to-vigorous physical activity: A latent change score approach. *Psychology of Sport and Exercise*. in press.
3. Kelly P, Fitzsimons C, Baker G. Should we reframe how we think about physical activity and sedentary behaviour measurement? Validity and reliability reconsidered. *International Journal of Behavioral Nutrition and Physical Activity*. 2016;13:32.
4. Nascimento-Ferreira MV, De Moraes ACF, Toazza Oliveira PV, et al. Assessment of physical activity intensity and duration in the paediatric population: evidence to support an a priori hypothesis and sample size in the agreement between subjective and objective methods. *Obesity Reviews*. 2018;19(6):810-824.
5. Cleland C, Ferguson S, Ellis G, Hunter RF. Validity of the International Physical Activity Questionnaire (IPAQ) for assessing moderate-to-vigorous physical activity and sedentary behaviour of older adults in the United Kingdom. *BMC Medical Research Methodology*. 2018;18(1):176.
6. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*. 2008;5:56.
7. Koolhaas CM, van Rooij FJ, Cepeda M, Tiemeier H, Franco OH, Schoufour JD. Physical activity derived from questionnaires and wrist-worn accelerometers: comparability and the role of demographic, lifestyle, and health factors among a population-based sample of older adults. *Clinical Epidemiology*. 2018;10:1-16.
8. Raask T, Maestu J, Latt E, et al. Comparison of IPAQ-SF and Two Other Physical Activity Questionnaires with Accelerometer in Adolescent Boys. *PLoS One*. 2017;12(1):e0169527.

9. Gorzelitz J, Peppard PE, Malecki K, Gennuso K, Nieto FJ, Cadmus-Bertram L. Predictors of discordance in self-report versus device-measured physical activity measurement. *Annals of Epidemiology*. 2018;28(7):427-431.
10. Baumann S, Gross S, Voigt L, et al. Pitfalls in accelerometer-based measurement of physical activity: The presence of reactivity in an adult population. *Scandinavian Journal of Medicine & Science in Sports*. 2018;28(3):1056-1063.
11. Booth M. Assessment of physical activity: an international perspective. *Research Quarterly for Exercise and Sport*. 2000;71 Suppl 2:114-120.
12. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. *Medicine & Science in Sports & Exercise*. 2014;46(1):99-106.
13. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*. 2003;35(8):1381-1395.
14. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*. 2011;8:115.
15. Skender S, Ose J, Chang-Claude J, et al. Accelerometry and physical activity questionnaires - a systematic review. *BMC Public Health*. 2016;16:515.
16. Innerd P, Harrison R, Coulson M. Using open source accelerometer analysis to assess physical activity and sedentary behaviour in overweight and obese adults. *BMC Public Health*. 2018;18(1):543.
17. Rowlands AV, Stiles VH. Accelerometer counts and raw acceleration output in relation to mechanical loading. *Journal of Biomechanics*. 2012;45(3):448-454.
18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies*. 2010;47(8):931-936.
19. Schmidt ME, Steindorf K. Statistical methods for the validation of questionnaires. *Methods of Information in Medicine*. 2006;45:409-413.

20. Bakdash JZ, Marusich LR. Repeated Measures Correlation. *Frontiers in Psychology*. 2018;8:456.
21. Hildebrand M, VT VANH, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and Science in Sports and Exercise*. 2014;46(9):1816-1824.
22. 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.
23. 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.
24. Voelkle MC. Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Science*. 2007;49:375-414.
25. Muthen LK, Muthen BO. *Mplus user's guide*. 8th ed. Los Angeles, CA: Muthen & Muthen; 1998-2017.
26. Tremblay MS, Aubert S, Barnes JD, et al. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. *International Journal of Behavioral Nutrition and Physical Activity*. 2017;14(1):75.
27. Dall PM, Coulter EH, Fitzsimons CF, Skelton DA, Chastin S, Seniors USPT. Taxonomy of Self-reported Sedentary behaviour Tools (TASST) framework for development, comparison and evaluation of self-report tools: content analysis and systematic review. *BMJ Open*. 2017;7(4):e013844.
28. Healy GN, Clark BK, Winkler EA, Gardiner PA, Brown WJ, Matthews CE. Measurement of adults' sedentary time in population-based studies. *American Journal of Preventive Medicine*. 2011;41(2):216-227.
29. Dowd KP, Szeklicki R, Minetto MA, et al. A systematic literature review of reviews on techniques for physical activity measurement in adults: a DEDIPAC study. *International Journal of Behavioral Nutrition and Physical Activity*. 2018;15(1):15.

30. World Health Organisation. Global recommendations on physical activity for health. 2010.

Table 1. Descriptive statistics of total minutes of self-reported and device measured physical activity across assessment periods.

Variables	Mean $\pm$ Standard Deviation		
	Time point 1	Time point 2	Time point 3
<b>IPAQ</b>			
Vigorous (total minutes)	101.67 $\pm$ 138.66	65.63 $\pm$ 79.53	96.43 $\pm$ 132.69
Moderate (total minutes)	68.33 $\pm$ 94.09	82.02 $\pm$ 120.44	124.53 $\pm$ 224.69
Light (total minutes)	235.98 $\pm$ 289.85	313.98 $\pm$ 371.23	353.25 $\pm$ 380.41
Sedentary (total minutes)	2507.73 $\pm$ 1399.21	2627.14 $\pm$ 1219.72	2558.92 $\pm$ 1017.69
<b>Accelerometer</b>			
Vigorous (total minutes)	37.02 $\pm$ 37.82	33.50 $\pm$ 31.29	36.54 $\pm$ 37.60
Moderate (total minutes)	606.23 $\pm$ 224.54	633.30 $\pm$ 251.67	596.89 $\pm$ 254.84
Light (total minutes)	714.47 $\pm$ 180.60	717.76 $\pm$ 188.20	694.46 $\pm$ 179.19
Sedentary (total minutes)	4510.70 $\pm$ 637.32	4368.59 $\pm$ 613.78	4348.73 $\pm$ 517.49

Note. Total minutes over the 6 day measurement period.



Table 2. Latent difference scores in total time spent in physical activity between self-reported and device-based methods.

	Time point 1			Time point 2			Time point 3		
	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>	$\beta$	95% CI	<i>p</i>
Vigorous	130.15	17.38, 242.92	0.248	125.90	76.23, 175.56	0.011	-164.23	-295.01, -33.45	0.209
Moderate	-487.45	-629.12, -345.78	0.001	-610.06	-784.17, -435.95	0.000	-1107.00	-1404.89, -809.11	0.000
Light	-410.54	-801.95, -19.13	0.294	-290.40	-738.59, 157.78	0.517	-472.90	-1002.74, 56.93	0.372
Sedentary	-3809.00	-4026.49, -3591.51	0.000	-3378.92	-3619.62, -3138.21	0.000	-4049.58	-4210.06, -3889.11	0.000

Table 3. Bland-Altman mean differences and limits of agreement.

	Time point 1			Time point 2			Time point 3		
	Mean Bias	Lower LOA	Upper LOA	Mean Bias	Lower LOA	Upper LOA	Mean Bias	Lower LOA	Upper LOA
Vigorous	65.94	-183.10	314.98	33.53	-93.71	160.76	62.40	-179.97	304.77
Moderate	-540.54	-981.03	-100.04	-555.05	-1030.76	-79.34	-468.47	-1004.78	67.83
Light	-473.19	-1090.41	144.04	-384.91	-1125.10	355.29	-333.96	-1106.40	438.49
Sedentary	-2035.47	-5102.30	1031.37	-1729.66	-4290.48	831.16	-1789.64	-4018.94	439.66

Note. LOA = Limits of Agreement.

**Supplementary Materials**

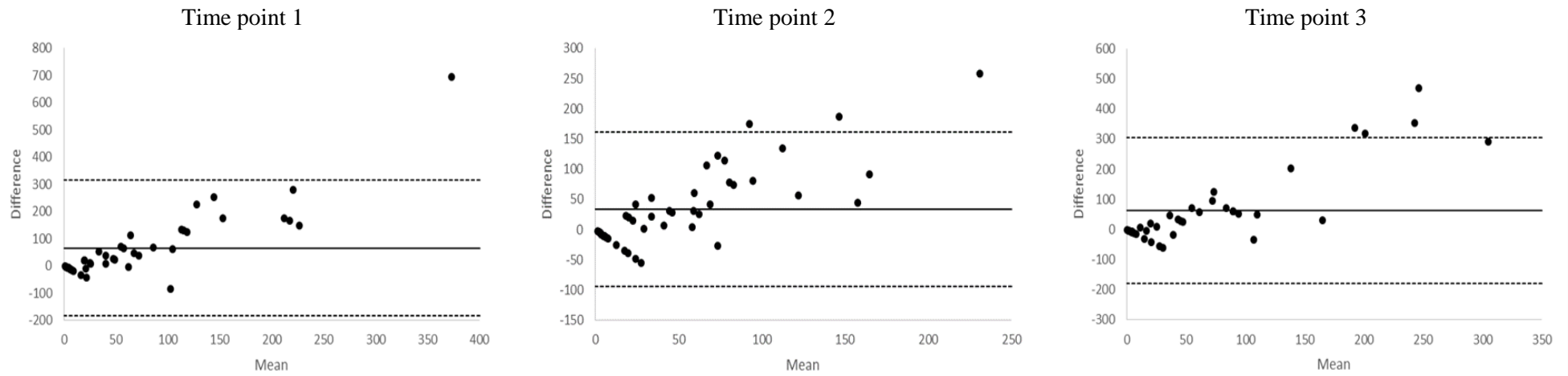


Figure S1. Bland-Altman Plots for Vigorous Physical Activity.

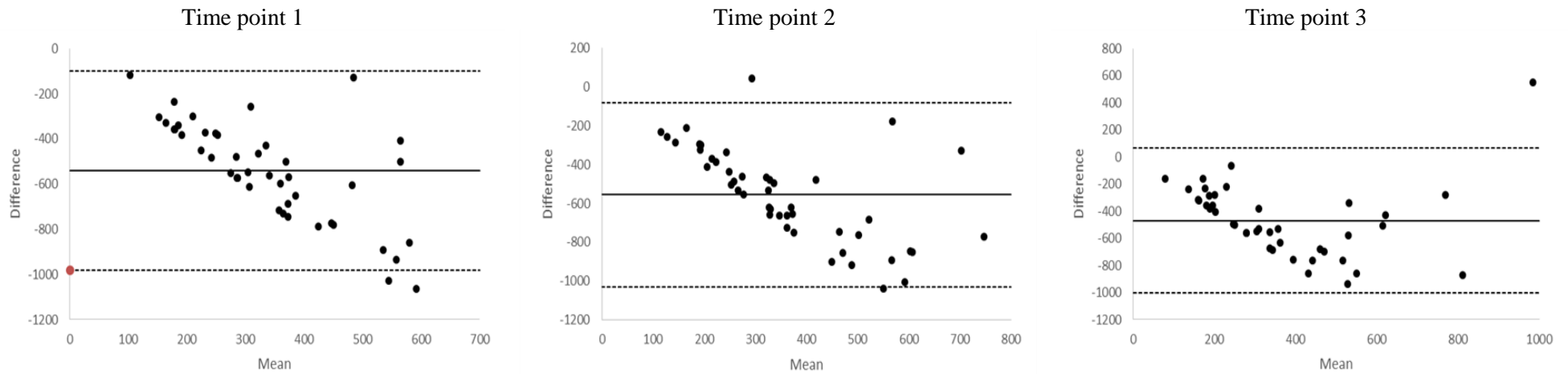


Figure S2. Bland-Altman Plots for Moderate Physical Activity.

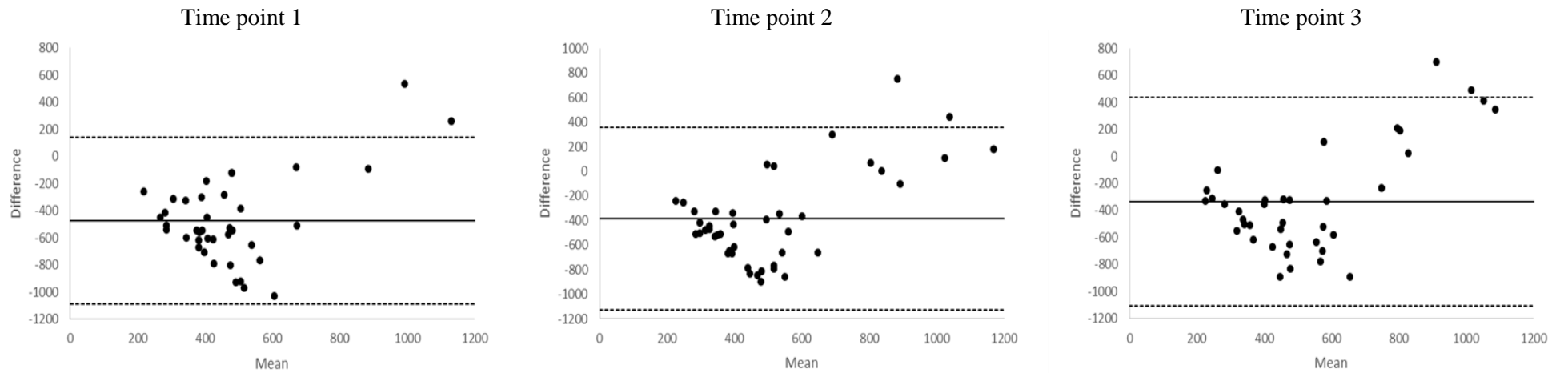


Figure S3. Bland-Altman Plots for Light Physical Activity.

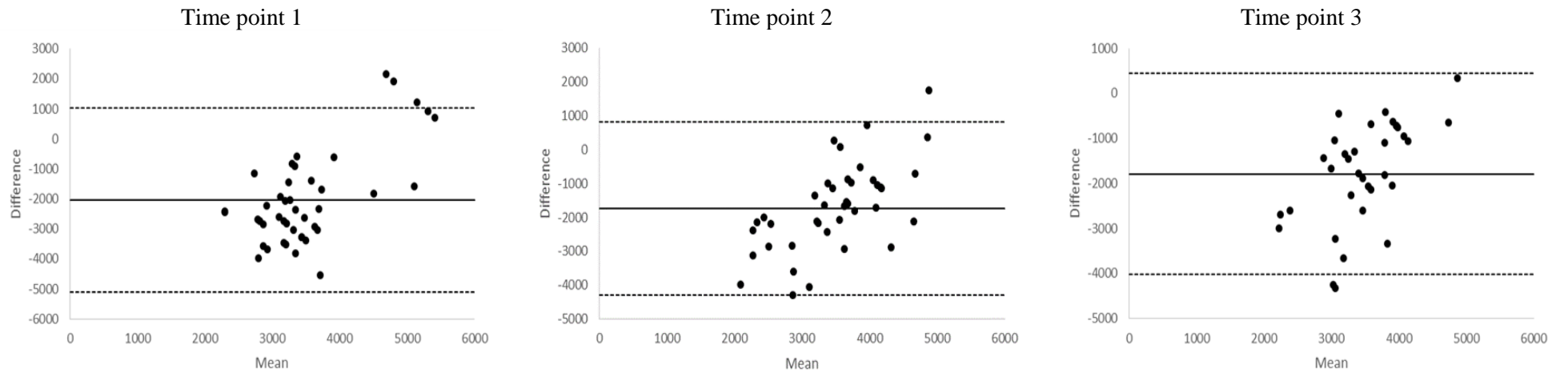


Figure S4. Bland-Altman Plots for Sedentary Activity.

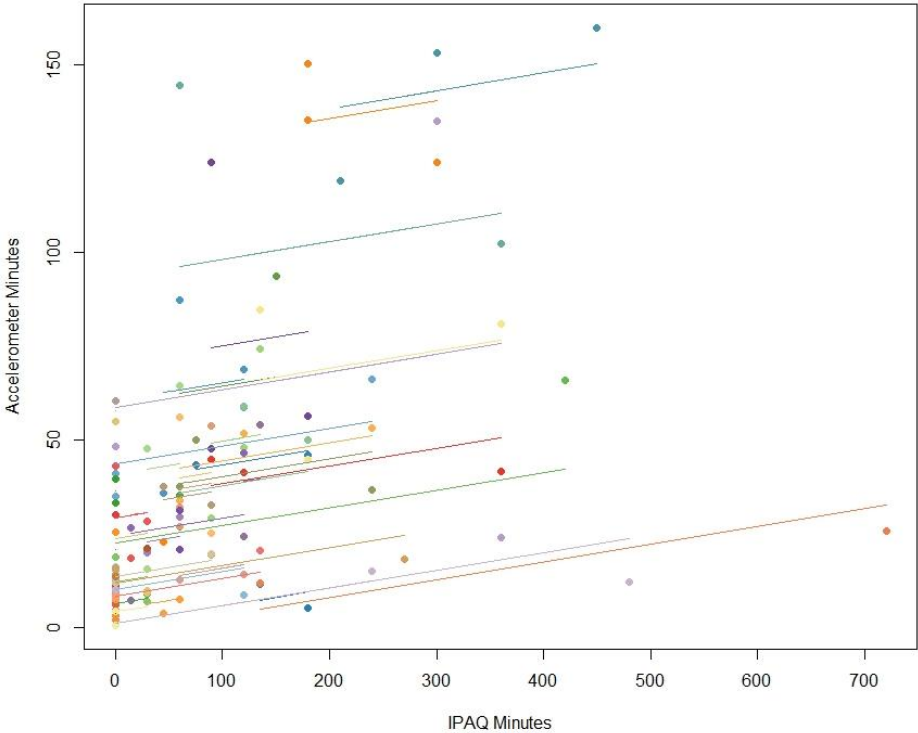


Figure S5. Repeated Measures Correlation Plot for Vigorous Physical Activity.

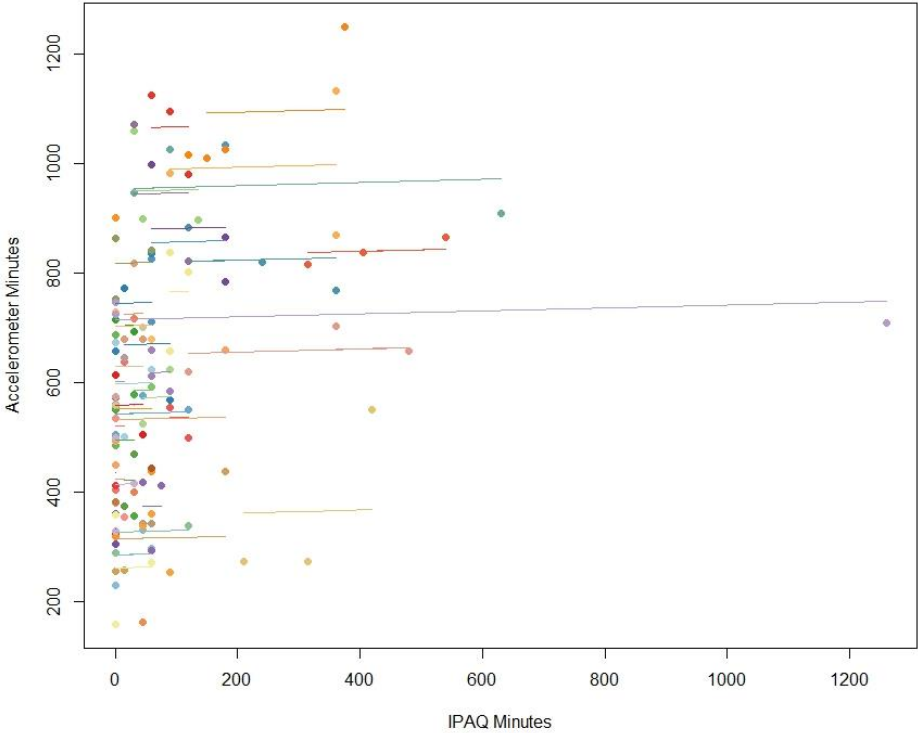


Figure S6. Repeated Measures Correlation Plot for Moderate Physical Activity.

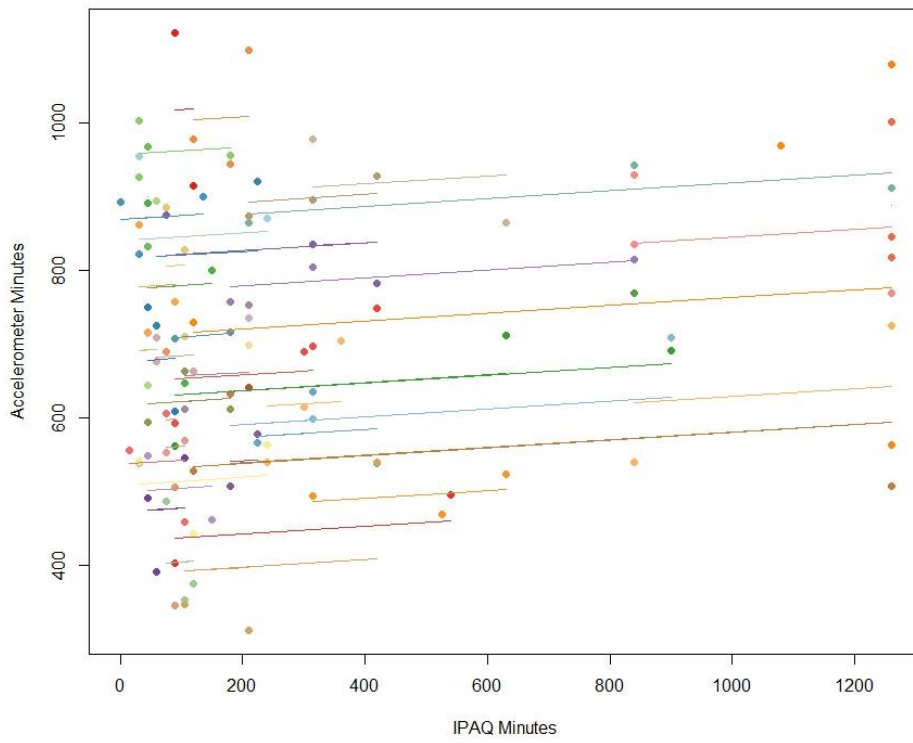


Figure S7. Repeated Measures Correlation Plot for Light Physical Activity.

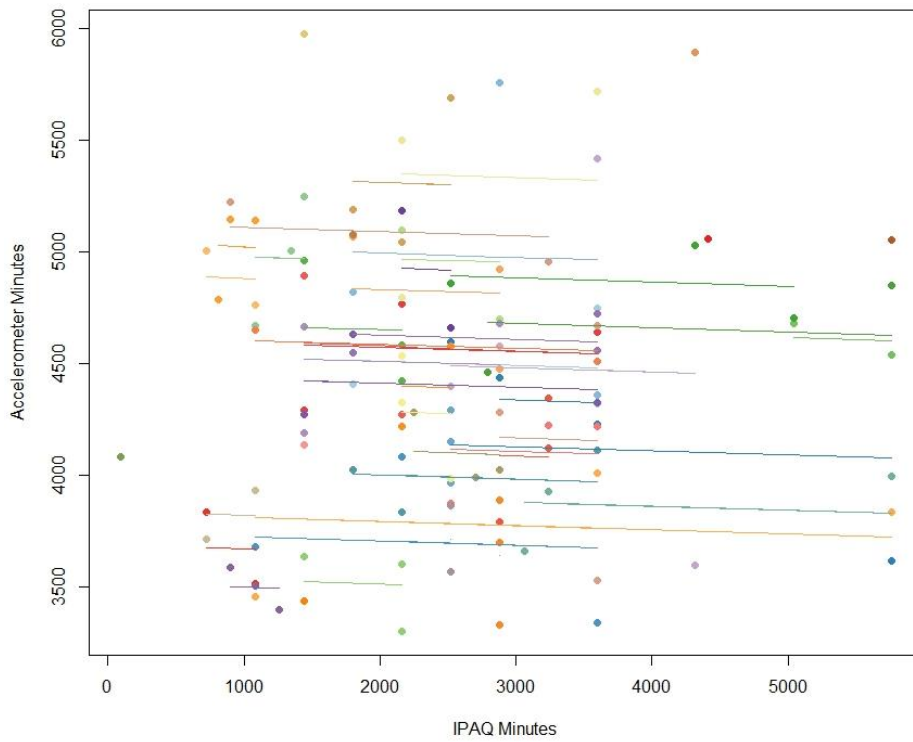


Figure S8. Repeated Measures Correlation Plot for Sedentary Activity.