Inter-instrument reliability of the Actigraph GT3X+ Ambulatory Activity Monitor during free-living conditions in adults

Manuscript Type: Original Research

Key Words:

Accelerometry, Physical Activity, Tri-axial, Derived Outputs.

Abstract: 199 words.

Manuscript Word Count: 3828

Date of Submission: 14th February 2013

Abstract

Background: Currently, no studies have investigated inter-instrument reliability of the ActiGraph (AG) GT3X+ in free-living conditions. **Methods:** Nineteen adults (11 males, 8 females; aged 36.8 ± 11.9 years) wore a pair of AG's (one on each hip), during all waking hours for one day. Raw outputs were generated for total counts, steps, wear time and mean counts per minute. Intensity outputs were derived for time (minutes) spent in <moderate, moderate, vigorous, very vigorous and moderate-to-vigorous physical activity (MVPA). Intraclass correlation (ICC), absolute percent difference (APD), coefficient of variation (CV), Bland-Altman plots and paired t-tests were used to evaluate reliability. **Results:** Inter-instrument reliability was high (CV<5%) for raw count and derived intensity outputs, except vigorous and very vigorous activity. ICC, CV and APD values for vigorous and very vigorous were .97, 12.28, 17.36% and .98, 18.15, 25.67% respectively. Amalgamating moderate, vigorous and very vigorous into a single MVPA category reduced the CV and APD values to 2.85 and 4.02%, and increased the ICC value to .99. No significant differences were found between contralateral units for any outputs (p>0.05). **Conclusion:** Reliability decreases beyond moderate intensities. MVPA displays superior inter-instrument reliability than individual intensity categories. Research question permitting, reporting time in MVPA may maximise reliability.

Introduction

Accelerometers are the most frequently used tool for the objective measurement of physical activity under free-living conditions in humans. There are now numerous research grade accelerometers available to the physical activity researcher.[1] The most widely used is the Actigraph range (Penacosa, FL), which has arguably become the standard issue unit within the field of physical activity epidemiology. The latest device in the Actigraph range, the tri-axial GT3X+ model, was launched in 2010. One of the advantages of the GT3X+ is the freedom to choose various pre-initialisation sampling rates (10 to 100 Hz), and the ability to define epoch length post-download. Compared to the earlier GT3X model (launched 2009), the GT3X+ has a larger memory capacity (256mb) and is water submersible.[2]

Recent data shows the GT3X model to exhibit acceptable intra and inter-instrument reliability (CV<5%) when tested on a vibration table at frequencies corresponding to typical ranges of human movement (1.1-10.2 Hz).[3] These are the first data to demonstrate that the GT3X is sufficiently reliable for research purposes. Mechanical devices can be used to test units across a wide range of accelerations and frequencies.[3,4] By maintaining tight experimental control, mechanical trials determine variance in device output solely attributable to the unit, negating variance components such as device placement, or gait characteristics.[5] Mechanical testing should therefore be the first step in determining device reliability. Researchers should then test ‘real world’ performance using participant mounted trials where the impact of device placement/tilt angle, gait characteristics, and non-controlled movements may be influential.

Following mechanical testing, Santos-Lozano et al. examined inter-instrument reliability of the GT3X in a participant mounted trial [N=1].[6] Batches of four GT3X were affixed to the left and right hip of a single male participant who completed 6 controlled activities (rest, sit-to stand, and treadmill walking/running at 4, 6, 8 and 10 $km.h^{-1}$). Inter-instrument reliability calculated within placement site was high (i.e. CV <10% and ICC >.90) for the vector magnitude $\left(\sqrt{X^2 + Y^2 + Z^2}\right)$ across all locomotion trials. Variance was greater in both horizontal axes (X and Z) than in the vertical axis (Y) across all activities. Although not a mechanically based experiment, the standardised activities were conducted in a controlled environment (i.e. not during free living) using a single participant. Further research under free-living conditions in multiple participants is required to evaluate the 'real world' performance of the Actigraph GT3X range.

To date no studies have examined the intra-or inter- instrument reliability of the Actigraph GT3X or GT3X+ under free-living conditions using multiple participants. Intra-instrument reliability is often assessed using mechanical methods and laboratory-based repeat standardised activity trials, but cannot be readily examined under free-living conditions.[6,7] The aim of the present study therefore was to determine the inter-instrument reliability of the Actigraph GT3X+ under free-living conditions.

Method

Participants

Eleven males (mean ± standard deviation, age 37 ± 12.1 years; height 1.79 ± 0.04 m; weight 80.44 ± 7.96 kg's; BMI 25.2 ± 2.3 kg.m^{-2}) and eight females (age 36.5 ± 12.3 years; height 1.67 ± 0.06 m; weight 63.69 ± 8.41 kg's; BMI 22.8 ± 2.6 kg.m^{-2}) were recruited from staff and students at the University of Worcester, UK. The experimental protocol received institutional ethics committee approval and written informed consent was obtained.

Instrumentation - Actigraph GT3X+ Accelerometer

The Actigraph GT3X+ is a tri-axial accelerometer that collects time varying acceleration in three orthogonal axes (X= mediolateral, Y= vertical and Z= anteroposterior) within the range of -6 to +6*g* [2]. The device can summarise data from all three axes using the vector magnitude $\left(\sqrt{X^2 + Y^2 + Z^2}\right)$. Sensed acceleration (as a voltage signal) is digitised by an analogue-to-digital converter; the signal is then rectified and subsequently integrated over a user defined epoch. Acceleration data is band pass filtered to a range of 0.05 to 2.0*g* within a frequency range of 0.25 to 2.5Hz.[2] All GT3X+ units were brand new prior to use in this study and therefore were factory calibrated according to manufacturer specifications.

Procedure

Participants wore single GT3X+ units over the right and left hip attached to an elasticised waist band for all waking hours during a 24 hour period. Participants were asked to wear the units for a minimum of 13 hours, on a day in which they would participate in ≥ 30 minutes of at least moderate intensity ambulatory physical activity; this was designed to ensure activity data was captured across the intensity spectrum. Twenty GT3X+ accelerometers were distributed to participants in pairs, the day immediately prior to the nominated data collection day, and collected the following day. Participants were provided with an information sheet and given a demonstration on correct unit placement; they were also instructed to wear the units during all waking hours, removing only for water-based activities (i.e. showering/bathing/swimming) and contact sports.

Data Treatment and statistical analyses

Raw proprietary activity count data were downloaded and converted into 5 second epoch data, using ActiLife v5.10.0. Non wear-time was defined as ≥60 minutes of consecutive zeros, and data were scanned for spurious values >1250 cts.5 sec^{-1} which have both been reported previously with accelerometer research.[8,9] Raw vector magnitude activity count data were reduced into minutes of <moderate, moderate, vigorous, very vigorous and moderate-to-vigorous (MVPA) intensity using published cut-points.[10] These cut-points were published as 60 second thresholds; they were therefore divided by 12 to create cut-points for

the 5 second data. Longer 60 second epochs could potentially mask inter-unit differences due to time smoothing[11]. Epoch adjusted intensity cut-points have been used in previous accelerometry measurement issue studies, including free-living studies.[11,12]. Original ‘as published’ and epoch adjusted cut-points are displayed in Table 1. Output variables were therefore total activity counts, mean activity counts (cts.min^{-1}), total step counts, and minutes of <moderate, moderate, vigorous, very vigorous and MVPA.

Output data were first imported into Microsoft Excel for the calculation of the coefficient of variation (CV) [SD/Mean*100] and absolute percent difference (APD) ((|LH-RH)/((RH+LH)/2))|)*100) between contralateral instruments. Output data were then imported into SPSS for Windows Version 19.0 (SPSS Inc., Chicago, IL) for further analysis. Intra-class correlation coefficients (ICC) were calculated for all right and left hip output variables. Paired samples t-tests were used to determine systematic bias between GT3X+ units for all output variables. To examine if inter-instrument concordance was dependent upon the accelerometer output Bland-Altman plots [13] and 95% limits of agreement were used. Data showed heteroscedasticity and is presented with the caveat that the limits of agreement may have a propensity to be too wide. The alpha level was set at *P*=0.05 for all tests.

Results

The average wear-time was 14.21 ± 1.69 hours, with wear-time ranging from 11.47 to 16.84 hours. Despite wear-time being less than the 13 hours wear-time requested, other research studies have established a 10 hour threshold as representative of a full days wear, and all data satisfied these criteria.[8]

Table 1

Descriptive data for right and left hip raw and derived output variables are shown in table 2. Paired samples t-tests indicated no significant differences (P>0.05) between right and left hip recordings for all of the raw counts and derived outputs detailed in table 2.

Table 2

ICCs and the range of APD and CVs for raw and derived output variables are presented in table 2. The ICC's for raw outputs of total activity counts, steps, valid wear time and mean counts per minute were .99, 1.00, .97 and .99 respectively. Further, ICC values for derived outputs of time spent in each category cut-point were generally high, ranging from .97 to .99.

Mean individual CV for raw outputs of total activity counts, steps, valid wear time and mean counts per minute were also low (2.19, 1.54, 0.99 and 2.80% respectively). The derived outputs show a pattern of increasing mean individual CV with increasing intensity (see table 2). In particular, the vigorous and very vigorous categories show high CVs (12.28 and 18.15% respectively). However, amalgamating the moderate, vigorous and very vigorous categories into a single 'MVPA' category reduces the CV to an improved mean CV of 2.85%, along with an improved range of 0.0-6.34%. The trend of increasing CV with increases in activity intensity is also seen in the range of values (see table 2).

The APD values for total activity counts, steps, valid wear time and mean counts per minute were 3.09, 2.18, 1.40 and 3.97% respectively. The mean individual APD for derived outputs followed a similar pattern to that of the mean individual CV (see table 2). When the moderate, vigorous and very vigorous categories were amalgamated into the single 'MVPA' category, the CV value improved, and this pattern was repeated with the APD values. The mean APD values for vigorous and very vigorous were elevated, at 17.36 and 25.67%

respectively, with the ranges also amplified. However, APD decreases to 4.02% with a range of 0.0-8.97% when collapsed into the MVPA category.

The Bland Altman plot for counts per minute is displayed in Figure 1. From Figure 1, these data suggest a trend for decreased inter-instrument agreement for individuals with greater activity levels (represented by cpm). The mean bias and 95% limits of agreement are shown in Table 3. When represented as a percent of both hips mean value, the limits of agreement become wider in each increasing intensity category. Collapsing moderate, vigorous and very vigorous categories into 'MVPA' reduces the mean bias and width of the limits of agreement.

Table 3

Figure 1

Discussion

The aim of the present study was to determine the inter-instrument reliability of the ActiGraph GT3X+. Inter-instrument reliability was evaluated by comparing the raw and derived outputs from concurrently worn AGs, positioned on the right and left hips during free-living conditions. The possibility of developing discrepancies between recordings when studying inter-instrument reliability using raw (continuous variables) and derived outputs (discrete variables) has been described elsewhere.[7] It is thought that derived outputs differ more between units than raw outputs due to their classification into discrete categories (<moderate, moderate, vigorous, very vigorous and MVPA).[7]

Despite the potential for derived outputs to differ between units following categorisation into discrete groups, the results of the present study indicate that both raw and derived outputs display generally high levels of inter-instrument reliability as indicated by ICC results (Table 2). Uniaxial accelerometers are reported to have ICCs ranging from .97-

.99 for raw outputs and .98-.99 for derived outputs.[7] The ICC results from this study show similarly high levels of reliability with values ranging from .97-1.00 for raw outputs and .97-.99 for derived outputs. However, the authors of the previous study did not report valid wear time, whereas our results include valid wear time and mean counts per minute as raw outputs. Despite valid wear time displaying the lowest ICC (.97), this is still considered to be indicative of a high level of reliability. The steps counted by the GT3X+ showed an absolute correlation between both hips. Considering the low CV and APD of the steps counted in-tandem with the correlation indicated by the ICC for this raw output, the GT3X+ can be considered as an extremely reliable step counter. In addition to the GT3X+ being a reliable step counter, previous research has demonstrated it to be a valid measure of steps in adults.[14]

In the present study reliability was evaluated using three statistics – ICC, CV and APD. The ICC, a measure of relative reliability, is widely reported in accelerometer reliability studies and therefore useful for comparison with previously published data. However, the ICC whilst easy to interpret (the closer to 1, the greater the reliability), gives no indication of the magnitude of agreement between GT3X units[15] and is influenced by the heterogeneity of the GT3X sample. Therefore absolute measures of reliability should also be reported,[16] such as the CV (and APD) which can be interpreted as follows; a CV of 2% (assuming normal distribution) means that 68% of differences between contralateral GT3X units lie within 2% of the mean of the two unit's output. Again the CV is widely used, but as it is calculated using the SD it may mask the total variance between units, which is arguably best reflected by the APD statistic.

Despite the generally high inter-instrument reliability found in the GT3X+ as indicated by the ICC values (table 2), there was a trend for decreasing reliability between units with increases in activity intensity above the moderate level. This decreasing reliability between units is evidenced by increases in both CV and APD (see table 2) and the width of the limits of agreement when represented as a percent of the combined hip mean (see table 3). It should be noted that the equation used to calculate the APD in the present study differs slightly from previously used equations ([(|LH-RH)/RH)|)*100]).[7] However, since neither right nor left hip data is the ‘gold standard’, we calculated the APD as: ((|LH-RH)/((RH+LH)/2))|)*100). Though the difference in equations does not explain the trend of data in table 2, as using the previously used equation would have increased APD values further.

Santos-Lozano et al. found the GT3X accelerometer to display high inter-instrument reliability at frequencies between 2.1 and 4.1Hz during laboratory based experiments.[3] However, they found the CV to vary widely with changes in frequency of movement (1.1–10.2Hz), which supports our findings during free-living. Although Santos-Lozano et al. found that the CV was highest at an activity frequency of 1.1Hz, our findings highlight the <moderate activity category to demonstrate the lowest CV (1.48%) compared to all other derived intensity outputs. Additionally, their results show the CV increased at a frequency of 10.2Hz, which could be compared to the very vigorous category in this study, where the CV was highest (18.15%). Further, in their laboratory based experiment the ICC for total activity counts was high (.97), which was replicated in this study (.99). Overall, their findings suggest inter-instrument reliability to be compromised at the extremes of human movement during mechanical experiments. However, results from this study demonstrate high levels of inter-instrument reliability at the lower end of the PA spectrum, but increasing CV with increases in activity intensity during free-living conditions. The differences in findings between the present study and laboratory based experiments may be because our data were generated during free-living conditions which are difficult to reproduce in the laboratory. It is likely that the extreme accelerations which accelerometers undergo during laboratory based experiments (e.g. 10.2Hz) are not accelerations experienced during activities of daily living.

Previous research has found increased reliability at higher vertical accelerations with the Actical accelerometer, but no relationship between acceleration and reliability with the Actigraph 7164 accelerometer.[5] Our results show decreasing reliability with increases in activity intensity beyond the moderate level. The range of CV and APD values and width of the limits of agreement was higher in the vigorous and very vigorous activity categories compared to <moderate and moderate PA categories. In addition to decreased reliability at higher intensities, previous research has shown waist-worn GT3X+ accelerometers to significantly overestimate energy expenditure at higher intensities[14], suggesting both reliability and accuracy of GT3X+ data may be lower at higher intensities. However, the decreased reliability found in vigorous and very vigorous categories will likely only impact on PA monitoring within populations that are highly active.

An inter-instrument reliability study with uniaxial accelerometers found increased mean individual CV and APD values at the moderate intensity activity level during free-living.[7] The moderate intensity category had the least time recorded for their participants, presumably as they were regular runners and spent more time in the vigorous category. In contrast, participants in the present study were from a broader population, including a flamenco dancer and a squash player, who spent more time in the moderate category. Indeed table 2 shows participants in the present study spent little time completing activities of a vigorous or very vigorous intensity. It is possible that activities completed at high intensities elicit asymmetrical movements which could account for the decreases in reliability and further research should try to elucidate this. Additionally, the range of CV and APD values in the vigorous and very vigorous categories might be partially due to the wide range of activities undertaken by participants in the present study compared to values obtained in other research using homogeneous populations. If the latter is true, then these results may be more generalizable than previous findings. However an alternative explanation may be (as shown by McClain, Sisson and Tudor-Locke[7]) that lower reliability coincides with less time spent in a given category, such that small differences in output variables between units can be inflated when using the mean as a denominator to calculate both CV and APD. Moderate and vigorous categories are often amalgamated into a single category; moderate-vigorous PA (MVPA).[7,17] Similarly, higher intensity activities recorded using triaxial accelerometers are also frequently merged; typically moderate, vigorous and very vigorous categories are combined to produce the MVPA category.[18,19] As table 2 demonstrates, combining moderate, vigorous and very vigorous categories into MVPA reduces the CV and APD. It should be noted that the ICC (see table 2) for MVPA remains high as well, at .99. Therefore, grouping activities into a single MVPA group yields better inter-instrument reliability with data recorded using the GT3X+ accelerometer.

However, whilst forming the MVPA group may 'solve' the decreased reliability of these accelerometers at extreme accelerations, it does not aid researchers or clinicians in measuring and differentiating between moderate, vigorous and very vigorous physical activities. Clear physical activity recommendations exist for adults, in which intensities are clearly distinguished *'do 150 minutes of moderate-intensity aerobic PA throughout the week or do at least 75 minutes of vigorous-intensity aerobic PA throughout the week'*. The ability of researchers and clinicians to reliably measure and distinguish between moderate, vigorous and very vigorous PA is therefore very important.

Despite the trend of the data shown in table 2, the ActiGraph GT3X+ accelerometer displayed generally high inter-instrument reliability and paired samples t-tests revealed differences between right and left hip activity recordings were not statistically significant ($P>0.05$) for any of the variables. However, the present study was limited insofar as participants only wore the accelerometers for one day, and it is not known if longer periods of wear time would result in greater differences between concurrently worn units.

It should be noted that despite participants being asked to wear the accelerometers for a minimum of 13 hours, not all participants met this initial criteria, but they did exceed previous guidelines of 10 hours wear-time.[8] Future researchers should be cognisant of the potential for protocol non-compliance when giving instructions to participants regarding preferred lengths of accelerometer wear-time specific to their research question.

There is currently only one set of triaxial cut-points for use with GT3X VM3 data.[10] Whilst the cut-points are available to differentiate between intensities at the middle to upper end of the PA spectrum (i.e. moderate, vigorous and very vigorous), further cut-points need to be developed to to enable researchers to utilise triaxial GT3X data when addressing research questions concerned with the lower end of the PA spectrum. In comparison, the ActiGraph default cut-points for uniaxial data have three categories for activities at the lower end of the PA spectrum (sedentary, light and lifestyle),[20] whilst VM3 data only has the <moderate category.[10] Unless activities beyond the moderate level are required specifically for the research question, moderate activity and beyond is often amalgamated into the MVPA category, which leaves only two intensity categories (<moderate and MVPA) when using VM3 data.

Further research should consider evaluating the inter-instrument reliability of the GT3X+ over periods longer than one day., **Conclusion**

In summary, this is the first study to examine the inter-instrument reliability of the Actigraph GT3X+ accelerometer during free-living conditions. The GT3X+ displayed high levels of inter-instrument reliability for raw outputs including total activity counts, steps, valid wear time and mean counts per minute. Additionally, the derived classifications of time spent in <moderate and moderate exhibited high inter-instrument reliability. Reliability was lower in the vigorous and very vigorous categories, but collapsing the data from moderate, vigorous and very vigorous into MVPA resulted in improved reliability. This study found no

differences (P>0.05) between recordings from accelerometers worn on the right and left hip, for either raw or derived outputs. Unless a research question requires lucidity between moderate, vigorous and very vigorous categories, it is recommended that MVPA be reported to enhance data reliability. Further GT3X VM3 cut-points need to be developed for the classification of activity behaviours at the lower end of the PA spectrum. Future research with a longer data-collection period is justified to clarify our findings of increasing between unit variance with increases in activity intensity during free-living conditions.

Acknowledgments

No acknowledgements to be made.

Funding Source

No funding source.

References

1. Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. *Archives of Disease in Childhood.* 2008;93(7):614-619.
2. John D, Freedson P. ActiGraph and Actical physical activity monitors: a peek under the hood. *Medicine and Science in Sports and Exercise.* 2012;44(1 Suppl 1):S86-89.
3. Santos-Lozano A, Marin PJ, Torres-Luque G, Ruiz JR, Lucia A, Garatachea N. Technical variability of the GT3X accelerometer. *Med Eng Phys.* 2012;34(6):787-790.
4. Routen AC, Upton D, Edwards MG, Peters DM. Intra- and inter-instrument reliability of the actiwatch 4 accelerometer in a mechanical laboratory setting. *Journal of Human Kinetics.* 2012;31:17-24.
5. Esliger DW, Tremblay MS. Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and Science in Sports and Exercise.* 2006;38(12):2173-2181.
6. Santos-Lozano A, Torres-Luque G, Marin PJ, Ruiz JR, Lucia A, Garatachea N. Intermonitor Variability of GT3X Accelerometer. *Int J Sports Med.* 2012.
7. McClain JJ, Sisson SB, Tudor-Locke C. Actigraph accelerometer interinstrument reliability during free-living in adults. *Medicine and Science in Sports and Exercise.* 2007;39(9):1509-1514.

8. Masse LC, Fuemmeler BF, Anderson CB, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S544-554.
9. Esliger DW, Copeland JL, Barnes JD, Tremblay MS. Standardising and optimizing the use of accelerometer data for free-living physical activity monitoring. *Journal of Physical Activity and Health.* 2005;3:366-383.
10. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport.* 2011;14(5):411-416.
11. Routen AC, Upton D, Edwards MG, Peters DM. Discrepancies in accelerometer-measured physical activity in children due to cut-point non-equivalence and placement site. *Journal of Sports Sciences.* 2012;30(12):1303-1310.
12. Edwardson CL, Gorely T. Epoch length and it's effect on physical activity intensity. *Medicine and Science in Sports and Exercise.* 2010;42(5):928-934.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;8;1(8476):307-310.
14. McMinn D, Acharya R, Rowe DA, Gray SR, Allan JL. Measuring Activity Energy Expenditure: Accuracy of the GT3X+ and Actiheart Monitors. *International Journal of Exercise Science.* 2013;6(3):5.
15. Metcalf BS, Curnow JS, Evans C, Voss LD, Wilkin TJ. Technical reliability of the CSA activity monitor: The EarlyBird Study. *Medicine and science in sports and exercise.* 2002;34(9):1533-1537.
16. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine.* 1998;26(4):217-238.
17. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. *Journal of aging and physical activity.* 2009;17(1):17-30.
18. Chen C, Jerome GJ, Laferriere D, Young DR, Vollmer WM. Procedures used to standardize data collected by RT3 triaxial accelerometers in a large-scale weight-loss trial. *Journal of physical activity & health.* 2009;6(3):354-359.
19. Vanhelst J, Beghin L, Duhamel A, Bergman P, Sjostrom M, Gottrand F. Comparison of uniaxial and triaxial accelerometry in the assessment of physical activity among adolescents under free-living conditions: the HELENA study. *BMC medical research methodology.* 2012;12:26.
20. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998;30(5):777-781.

Tables

Table 1: Epoch adjusted and original unadjusted Actigraph GT3X intensity cut-points.

Cut-Points	Duration	<Moderate (Cts)	Moderate (Cts)	Vigorous (Cts)	V. Vigorous (Cts)
Sasaki et al. (2011)	60 secs	0-2689	2690-6166	6167-9642	≥9643
	5 secs*	0-223	224-514	515-804	≥805

*Epoch adjusted (60secs/12). Cts: accelerometer counts.

Table 2: Means ± standard deviations GT3X+ output variables.

	RH (mean ± SD)	LH (mean ± SD)	ICC	Mean APD % (Range)	Mean CV % (Range)
Raw Outputs					
Total Act Cts	803557.4 ± 260927.27	794191.53 ± 258068.02	.99	3.09 (0.19-11.10)	2.19 (0.13-7.85)
Steps	11700.00 ± 3924.64	11726.95 ± 3974.49	1.00	2.18 (0.18-5.65)	1.54 (0.13-3.99)
Valid Wear Time	852.27 ± 101.17	853.02 ± 103.33	.97	1.40 (0.0-6.82)	0.99 (0.0-4.82)
$Cts.min^{-1}$	947.37 ± 289.99	933.69 ± 283.63	.99	3.97 (0.19-11.60)	2.80 (0.13-8.20)
Derived Outputs					
<Mod (mins)	732.30 ± 96.67	735.43 ± 98.77	.97	2.09 (0.01-8.97)	1.48 (0.01-6.34)
Mod (mins)	89.45 ± 33.54	87.86 ± 34.46	.99	5.13 (0.28-12.52)	3.63 (0.19-8.85)
Vig (mins)	19.25 ± 12.19	18.54 ± 11.83	.97	17.36 (0.92-61.54)	12.28 (0.65-43.51)
V. Vig (mins)	11.07 ± 11.63	11.03 ± 10.69	.98	25.67 (1.76-76.92)	18.15 (1.25-54.39)
MVPA (mins)	119.78 ± 41.33	117.43 ± 41.36	.99	4.05 (0.0-8.97)	2.85 (0.0-6.34)

Right Hip (RH), Left Hip (LH), Intra Class Correlation (ICC), Absolute Percent Difference (APD), Coefficient of Variation (CV), Activity Counts (Act Cts), Counts per Minute ($Cts.min^{-1}$), < Moderate (<Mod), Moderate (Mod), Vigorous (Vig), Very Vigorous (V. Vig), and Moderate-Vigorous Physical Activity (MVPA).

Table 3: Mean Bias and Limits of Agreement (LOA). Values in brackets represent the bias or LOA as a percentage of the combined mean of both hips.

Intensity Category	Mean of two hips	Bias	95% Limits of agreement
Less than Moderate	733.9	-3.12 (0.43)	-26.0 (3.54) to 80.1 (10.91)
Moderate	88.7	1.59 (1.79)	-5.4 (6.08) to 16.8 (18.94)
Vigorous	18.9	0.71 (3.76)	-3.0 (15.87) to 9.1 (48.14)
Very Vigorous	11.1	0.04 (0.36)	-2.2 (19.81) to 6.9 (62.16)
MVPA	118.6	2.34 (1.97)	-6.1 (5.14) to 18.7 (15.76)

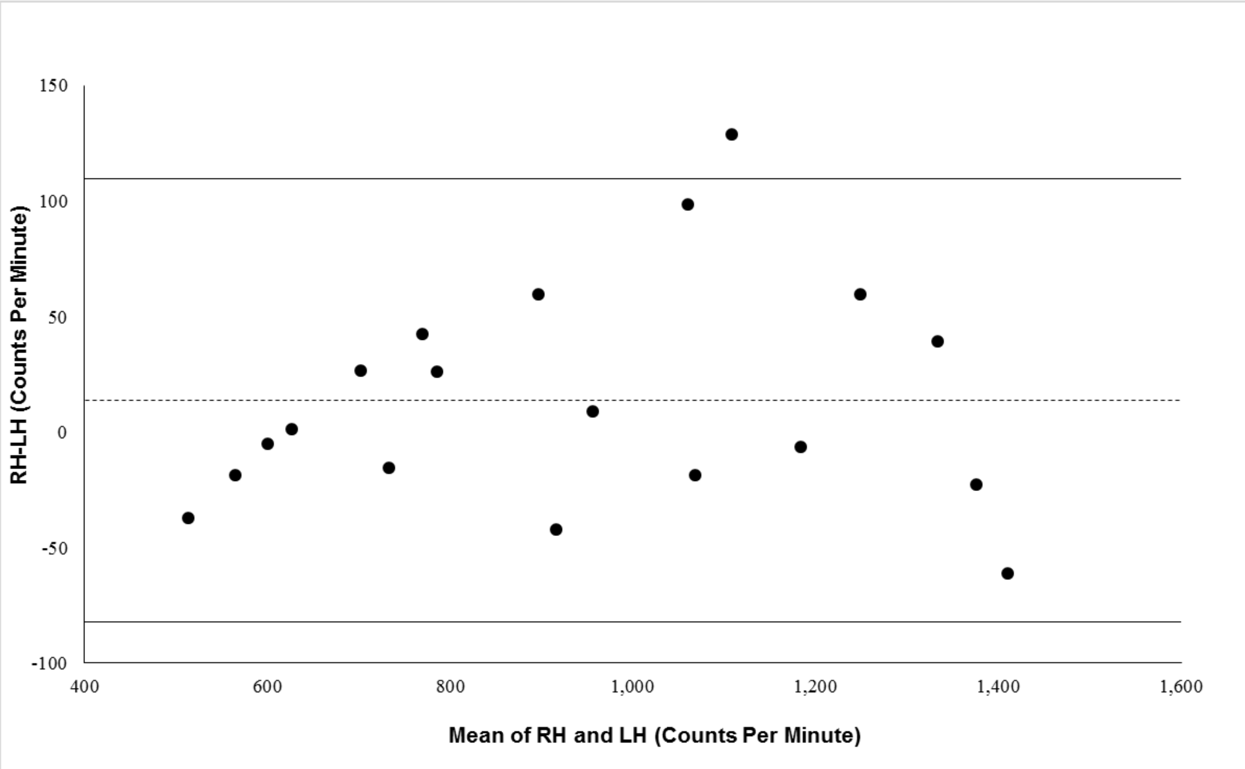

Figure 1: Bland-Altman plot showing mean difference (13.86 ± 48.87 cpm) and 95% limits of agreement (109.46, -82.10 cpm) between counts per minute from right hip and left hip located units.