Validity of the iPhone M7 Motion Coprocessor to Estimate Physical Activity during Structured and Free-Living Activities in Healthy Adults

Date of submission: 8th December 2020

Abstract

Modern smartphones such as the iPhone contain an integrated accelerometer which can be used to measure body movement and estimate the volume and intensity of physical activity.

Objectives: The primary objective was to assess the validity of the iPhone to measure step count and energy expenditure during laboratory-based physical activities. A further objective was to compare free-living estimates of physical activity between the iPhone and the Actigraph GT3X+ accelerometer.

Methods: Twenty healthy adults wore the iPhone 5S and GT3X+ in a waist-mounted pouch during bouts of treadmill walking, jogging, and other physical activities in the laboratory. Step counts were manually counted and energy expenditure was measured using indirect calorimetry. During two weeks of free-living, participants (n=17) continuously wore a GT3X+ attached to their waist and were provided with an iPhone 5S to use as they would their own phone.

Results: During treadmill walking, iPhone (703 ± 97 steps) and GT3X+ (675 ± 133 steps) provided accurate measurements of step count compared to the criterion method (700 ± 98 steps). Compared to indirect calorimetry (8 ± 3 kcal·min^{-1}), the iPhone (5 ± 1 kcal·min^{-1}) underestimated energy expenditure with poor agreement. During free-living, the iPhone (7990 ± 4673 steps·day^{-1}) recorded a significantly lower ($P < 0.05$) daily step count compared to the GT3X+ (9085 ± 4647 steps·day^{-1}).

Conclusions: The iPhone accurately estimated step count during controlled laboratory walking but records a significantly lower volume of physical activity compared to the GT3X+ during free living.

Keywords

Step Count; Energy Expenditure; Walking; Validation; Smartphone; Accelerometer; GT3X+

Introduction

Measuring levels of physical activity (PA) is becoming increasingly important given the well-defined relationships with health, disease, and mortality (Antero Kesaniemi (Chair) et al., 2001). Measuring PA, however, can be notoriously challenging in a large number of individuals. While indirect methods such as self-report questionnaires are easy to implement, they lack objectivity and accuracy (Prince et al., 2008; Dyrstad, Hansen, Holme, & Anderssen, 2014). On the other hand, accelerometers can be worn around the waist or the wrist to provide more objective and accurate estimates of the frequency, intensity, and duration of PA during periods of free-living (Aadland & Ylvisåker, 2015; Lee, Williams, Brown, & Laurson, 2014). Triaxial accelerometers such as the Actigraph GT3X+ (ActiGraph, FL, USA) are commonly used in research and large national PA surveys such as the 2013 – 2014 National Health and Nutrition Examination Survey. However, to gain meaningful data, participants are required to wear the accelerometer > 10 hours/day for at least four days/week, (Troiano et al., 2008) which some individuals find to be burdensome (O'Brien et al., 2017). Furthermore, accelerometers can be costly, require expertise in analysing the output data as well as lacking real-world transferability.

Nowadays, the majority of adults carry a smartphone that already contains the hardware that can measure locomotion, with 76% of the UK's population reporting to own a smartphone in 2018 (Taylor & Silver, 2019). For example, the iPhone 5S's M7 motion coprocessor (Apple Inc., Cupertino, CA, USA) collects sensor data from an integrated accelerometer which can estimate step count and PA. A range of downloadable applications (apps) can then integrate these data with the user's stature, body mass, and gender to generate estimates of energy expenditure (EE) using bespoke algorithms. One such app is 'ActivityTracker' (V2.6, Bits&Coffee, Romania) which provides instantaneous and cumulative measurements of step count and EE. Given the common usage of smartphones, this may reduce the participant burden and costliness associated with objective methods of PA monitoring as there is no additional wearable required. There is the added advantage that researchers would have continuous remote access to the measurements via data sharing platforms. However, the validity of the smartphone technology to measure parameters of PA needs to be further established. A recent study compared the iPhone 5S to manually counted steps (Major & Alford, 2016) and found good correlation between methods at fast walking speeds (4.68 and 6.48 $km \cdot h^{-1}$) but not at the slowest walking speed of 3.6 $km \cdot h^{-1}$. This study did not, however, explore the accuracy of iPhone mobile applications to estimate EE. Of further interest is the impact of user behaviour with mobile devices in a free-living environment and how this influences the accuracy of PA measurements.

To the authors' knowledge, no study has compared estimations of both step count and EE from the iPhone 5S with laboratory-based gold standards during a variety of different physical activities. Therefore, the primary purpose of this study was to assess the validity of the iPhone M7 motion coprocessor, for estimating step count and subsequently estimating EE (with the use of the 'ActivityTracker' app) during treadmill walking, jogging, running, stationary cycling, and an aerobics session. A further aim was to compare measurements of step count between the iPhone and GT3X+ during a two-week period of free-living.

Methods

Study Design

The current study consisted of two distinct phases. The first phase comprised a single experimental trial conducted in the laboratory to determine the validity of iPhone estimates of step count and EE in comparison to gold standard measures (step count = manually counted, EE = indirect calorimetry. The second phase was a two-week observational period during which free-living PA data was concurrently monitored using the iPhone and the GT3X+ accelerometer. The study was approved by the School of Science and Sport Ethics Committee at the University of the West of Scotland. Written informed consent was obtained from each participant.

Phase 1: Laboratory

Participants

Twenty healthy adults, twelve females and eight males (mean ± SD: age 28 ± 5 years, stature 168 ± 8 cm and body mass 72.0 ± 12.7 kg), volunteered to take part in the current study. The health status of the participants was established by the completion of a self-declared medical questionnaire which excluded participants with a history of cardiorespiratory or neurological disease.

Procedures

Upon arrival at the laboratory, participants were briefed on the protocol before anthropometric variables were measured using conventional techniques. A Polar H7 monitor (Polar Electro Oy, Kempele, Finland) was attached to the participant's chest to continuously monitor heart rate. A triaxial GT3X+ accelerometer (Actigraph, FL, USA) was attached to the right hip of participants in a pouch that also held an iPhone 5S. Breath-by-breath pulmonary gas exchange was measured continuously throughout the experiment using a metabolic cart (Ultima CPX, MedGraphics, MN, USA). The metabolic cart was calibrated as per the manufacturer's guidelines. EE ($kcal \cdot min^{-1}$) was calculated by indirect calorimetry using the Weir equation ($\dot{V}O_2 \times 3.941$) + ($\dot{V}CO_2 \times 1.1$) (Weir, 1949) - and the mean value from the final 2 min of each bout of exercise or activity was used in later analyses.

The first component required participants to walk and jog on a treadmill at light, moderate and vigorous intensities based on their heart rate reserve for a total of 15 min (5 min at each intensity). Resting heart rate was recorded and age-predicted maximal heart rate was calculated as (208 – (0.7 x AGE)) (Tanaka, Monahan, & Seals, 2001). The Karvonen formula (% target intensity (max heart rate – resting heart rate) + resting heart rate) (Ewing, Wilmore, Blair, Haskell, & Kraemer, 1998) was used to determine the treadmill speed associated with each target intensity range. The treadmill speed began at 3 $km \cdot h^{-1}$ and was increased until the participant's heart rate was in the desired range: Light (30-39% heart rate reserve); Moderate (40-59% heart rate reserve); and Vigorous (60-89% heart rate reserve) (ACSM, 2017). The second component consisted of cycling on a stationary ergometer at a fixed power of 50 W for 5 min. The last component required the participant to complete a 5 min aerobics session by following a YouTube video. Throughout all activities, measurements of accelerometry, heart rate, iPhone step count, and EE from the 'ActivityTracker' app were continuously recorded. The treadmill component was also video recorded in order to manually count steps, video footage was reviewed and manual steps were counted for each intensity with the use of a hand tally counter. Two members of the research team separately watched and counted each video 3 times and recorded the values. These values were then compared and when discrepancies were noted, the researchers reanalysed the videos until agreement was reached.

In order to assess the validity of the iPhone during the laboratory trials, iPhone estimates of step count and EE were compared to the gold standards (manually counted and indirect calorimetry, respectively). For all activities, step count is reported as the number of measured steps for that component.

Component 1: Treadmill walking and jogging

Participants were instructed to step onto the motorised treadmill (PPS Med, Woodway, Waukesha, WI) on which the incline was increased to 1% to mimic the metabolic cost of outdoor walking (Jones & Doust, 1996). After finding the speed which elicited the desired heart rate range, participants were asked to straddle the treadmill in order to record iPhone step count and EE from the 'ActivityTracker' app before recommencing walking. This was repeated for light (5.4 ± 1.0 km·h^{-1}; 9 ± 2 RPE), moderate (6.5 ± 0.9 km·h^{-1}; 11 ± 2 RPE) and vigorous (8.0 ± 1.1 km·h^{-1}; 13 ± 2 RPE) intensities. Between each stage, participants were again asked to straddle the treadmill and to stand motionless so the iPhone step count and EE could be recorded from the 'ActivityTracker' app.

Components 2 and 3: Cycling and aerobics

Participants carried out 5 min of cycling on an electronically-braked ergometer (Lode Excalibur Sport; Lode Medical Technology, Groningen, The Netherlands) with a constant external power output of 50 W. Participants were instructed to cycle at a comfortable cadence as they would on a leisurely cycle. Following this, participants followed a 5 min segment of a YouTube aerobics-style cardiovascular workout video (https://www.youtube.com/watch?v=istOU9nxhm8). The iPhone step count and EE from the 'ActivityTracker' app were recorded at the beginning and end of each activity.

iPhone 5S

The iPhone application 'ActivityTracker' was downloaded onto the iPhone 5S from the Apple store. This app was selected as it provided a live reading of daily total steps and estimates of EE. The 'ActivityTracker' app is reported by the developer to collect step count directly from the iPhones' Health Kit and uses a bespoke algorithm based on step count, gender, stature and body mass to estimate EE.

Accelerometers

Before each trial, a triaxial GT3X+ accelerometer (Actigraph, FL, USA) was initialised to record data at a sampling frequency of 30 Hz in three axes: vertical, mediolateral and anteroposterior, using ActiLife software (V6.13.3 Lite Edition, Actigraph, FL, USA). The Actigraph GT3X+ accelerometer was selected for use in the current study as it has been previously shown to accurately assess step count when worn on the waist in laboratory studies (Mcminn, Acharya, Rowe, Gray, & Allan, 2013; Tudor-Locke, Barreira, & Schuna, 2015). These accelerometers also have high inter-instrument reliability during activities of daily living (Ozemek, Kirschner, Wilkerson, Byun, & Kaminsky, 2014) and under free-living conditions (Jarrett, Fitzgerald, & Routen, 2015; Aadland & Ylvisåker, 2015). Furthermore, the GT3X+ is the most commonly used accelerometer by researchers in laboratory and free-living settings (Wijndaele et al., 2015; Migueles et al., 2017; Reid et al., 2017). The accelerometer was worn on the anterior-superior iliac spine of the right hip in a neoprene pouch. When downloading the GT3X+ data using the ActiLife software, the manufacturer's default filter and ActiGraph's proprietary algorithm for step-count were used. EE was estimated using the Freedson VM3 equation (Sasaki, John, & Freedson, 2011) (0.001064 x VM + 0.087512 (BM) – 5.500229), where VM is vector magnitude and BM is body mass.

Data Analysis

For manually counted steps, a step was recorded each time the participant's foot touched the treadmill. Reproducibility was evaluated using the concordance correlation coefficient of Lin (CCC) (Lin, 1989) with the thresholds: almost perfect > 0.90; substantial > 0.8 – 0.9; moderate 0.65 – 0.8; poor < 0.65. Bland and Altman (1986) analysis was used to express agreement between methods of measuring step count and EE. The 95% limits of agreement (LOA) were calculated as mean bias ± (1.96 × standard deviation). Log-transformation of EE data was attempted as the difference between measurement methods increased as EE increased. However, this did not reduce the linear change of the data, so the original, non-log scaled data were used. The mean percentage error (MPE) was computed as (steps detected – observed steps (manually counted))/ observed steps (manually counted) × 100, for step count and (estimated EE – measured EE (indirect calorimetry))/ measured EE (indirect calorimetry) × 100, for EE. The mean absolute percentage error (MAPE) was also computed using the same formulas, with the exception that negative values were converted to positive values. Calculating both MPE and MAPE allows for a true representation of the direction and magnitude of difference between methods to be established (Le Masurier, Lee, & Tudor-Locke, 2004). A one-way analysis of variance (ANOVA) was used to assess differences between measurement methods (iPhone, GT3X+, and criterion methods). Statistical significance was set at $P < 0.05$. All statistical procedures were carried out using Jamovi project (2018; version 0.9.5.12; retrieved from https://www.jamovi.org, open source).

Phase 2: Free-Living

Participants

Twenty adults volunteered to take part in the second phase of the study. Two participants withdrew (reasons undisclosed), one participant did not have sufficient wear time of the GT3X+ accelerometer, and the iPhone application malfunctioned for another participant. Therefore, sixteen participants, ten female and six males (mean ± SD: 42 ± 17 years old), completed phase 2 of the study.

Experimental Design and Procedures

Participants were monitored for a total period of 14 days. Participants were given an iPhone 5S and were asked to carry the iPhone with them as they would their own mobile phone. Step count data from the iPhone were automatically uploaded to a bespoke online digital platform (Lenus, StormID, Edinburgh, UK) which enabled continuous data exchange between the user and the researcher. The user experience of the Lenus health platform was evaluated as a separate component of this study and will be reported elsewhere. Participants were also given a GT3X+ accelerometer which was attached to an elastic waistband. Participants were instructed to wear the GT3X+ all day, every day on the right anterior-superior iliac spine, removing only for sleep and showering/swimming. The accelerometers (Actigraph, FL, USA) were initialised to record data at a sampling frequency of 30 Hz in three axes of motion and data was downloaded as previously described.

Data Analysis

Daily steps from the GT3X+ were only included in the analysis if wear time was ≥ 10 hours per day (Van Dyck et al., 2015). Non-wear time was defined as ≥ 60 min of consecutive zeros (Van Dyck et al., 2015). Days with <1000 steps were excluded from further analysis (Barreira et al., 2013). Reproducibility was evaluated using the concordance correlation coefficient of Lin (CCC) (Lin, 1989) with the thresholds: almost perfect > 0.90; substantial > 0.8 – 0.9; moderate 0.65 – 0.8; poor < 0.65. Bland and Altman analysis (Martin Bland & Altman, 1986) was used to assess agreement between step count estimates from the iPhone and the GT3X+ as previously described. Paired T-tests were used to determine whether there was a difference in step count between measurement methods (Jamovi project 2018; version 0.9.5.12; retrieved from https://www.jamovi.org, open source).

Results

Phase 1: Laboratory

Treadmill Walking, Jogging and Running

Step count data from the iPhone, GT3X+, and criterion method during the treadmill trial are presented in Table 1. The agreement in measurements of step count between iPhone and manually counted was almost perfect (CCC = 0.993; 95% CI 0.988 to 0.996 steps) throughout the treadmill trial with a mean difference of 3 steps (95% LOA -19 to 25 steps) (Fig. 1.a) and a MAPE of 1.1%. When comparing the intensities separately, there was almost perfect agreement between the iPhone and criterion methods with a MAPE of < 2 % at each intensity (Table 2).

The GT3X+ and manually counted agreement for the measurement of step count was moderate (CCC = 0.76; 95% CI 0.64 to 0.84 steps) throughout the treadmill trial with a mean difference of -25 steps (95% LOA -179 to 129 steps) (Fig. 1.b) and a MAPE of 4.6%. When comparing the intensities separately there was poor agreement at light and moderate intensities (MAPE > 5%) but substantial agreement during vigorous intensity (MAPE < 2%, Table 2).

Estimates of EE from the iPhone, GT3X+, and criterion method (indirect calorimetry) during the treadmill trial can be viewed in Table 3. The agreement in measurements of EE between iPhone and indirect calorimetry was poor (CCC = 0.48; 95% CI 0.36 to 0.58 kcal·min^{-1}) throughout the treadmill trial with a mean difference of -1.9 kcal·min^{-1} (95% LOA -5.6 to 1.8 kcal·min^{-1}) (Fig. 3) and a MAPE of 23.7%. When comparing the intensities separately there was moderate agreement at the light intensity, while the iPhone estimates of EE were significantly lower than indirect calorimetry with poor agreement at moderate and vigorous intensities (Table 4).

The agreement in measurements of EE between GT3X+ and indirect calorimetry was substantial (CCC = 0.85; 95% CI 0.77 to 0.91 kcal·min^{-1}) throughout the treadmill trial with a mean difference of 0.5 kcal·min^{-1} (95% LOA -2.5 to 3.6 kcal·min^{-1}) (Fig. 4) and a MAPE of 18.3%. When comparing the intensities separately there was moderate agreement at light, substantial at moderate, and poor agreement at vigorous intensity. The CCC, mean bias, 95% LOA, p-value, MPE and MAPE data presented in table 4.

Cycling and Aerobics

In comparison to indirect calorimetry (5.3 ± 0.9 kcal·min^{-1}), both the iPhone (3.3 ± 2.1 kcal·min^{-1}) and the GT3X+ (0.4 ± 0.8 kcal·min^{-1}) significantly underestimated EE during stationary cycling trial ($P < 0.001$). There was poor agreement between the criterion method and estimates of EE with the iPhone (CCC = 0.20; 95% CI 0.01 to 0.38 kcal·min^{-1}) and the GT3X+ (CCC = 0.01; 95% CI 0.02 to 0.03 kcal·min^{-1}). The mean bias between iPhone and indirect calorimetry was -2.0 kcal·min^{-1} (95% LOA -5.5 to 1.6 kcal·min^{-1}) whereas between GT3X+ and indirect calorimetry the mean bias was -4.8 kcal·min^{-1} (95% LOA -7.2 to -2.5 kcal·min^{-1}).

The iPhone (3.8 ± 0.8 kcal·min^{-1}) significantly underestimated EE during the aerobics activity in comparison to the criterion method (7.5 ± 1.6 kcal·min^{-1}). There was no difference between EE estimated by the GT3X+ (8.4 ± 1.3 kcal·min^{-1}) and the criterion method ($P = 0.120$). There was poor agreement between the criterion method and estimates of EE with both the iPhone (CCC = 0.10; 95%

CI 0.03 to 0.17 $kcal \cdot min^{-1}$) and the GT3X+ (CCC = 0.62; 95% CI 0.33 to 0.80 $kcal \cdot min^{-1}$). The mean bias between the iPhone and criterion method was -3.8 $kcal \cdot min^{-1}$ (95% LOA -6.0 to -1.5 $kcal \cdot min^{-1}$), whereas between GT3X+ and criterion method the mean bias was 0.9 $kcal \cdot min^{-1}$ (95% LOA -1.1 to 2.9 $kcal \cdot min^{-1}$).

Phase 2: Free-Living

The average daily wear-time of the GT3X+ was 731 ± 89 $minutes \cdot day^{-1}$. There was substantial agreement in the measurement of step count between the iPhone (7990 ± 4673 $steps \cdot day^{-1}$) and GT3X+ (9085 ± 4647 $steps \cdot day^{-1}$) (CCC = 0.894; 95% CI 0.86 to 0.92 $steps \cdot day^{-1}$), with a mean difference of -1095 $steps \cdot day^{-1}$ (95% LOA -4780 to 2591 $steps \cdot day^{-1}$) (Fig. 3). Daily step count measured by the iPhone was significantly lower than the GT3X+ ($P < 0.001$). The Bland and Altman plot (fig. 3) shows a large spread, with 10 data points above or below the 95% limits of agreement which have a range of ~7000 steps.

Discussion

The primary purpose of this study was to assess the validity of the iPhone 5S for estimating step count and EE during laboratory-based physical activities. We compared step counts from the iPhone 5S and a research-grade accelerometer (GT3X+) to the criterion method. Both devices were found to provide valid estimates of step count during walking and jogging on a treadmill. The GT3X+ was also found to provide accurate estimates of EE during treadmill walking and jogging but the iPhone significantly underestimated EE compared to indirect calorimetry. A further objective was to compare estimates of step count between the iPhone 5S and GT3X+ during a two-week period of free-living. We found that the iPhone recorded significantly fewer daily steps compared to the GT3X+, suggesting the iPhone may not be a suitable method of estimating daily physical activity.

In the treadmill component of the laboratory trial, the iPhone provided near perfect estimates of step count compared to manually counted, whereas the GT3X+ provided moderate estimates of step count compared to the criterion method. Both the iPhone and GT3X+ were most accurate at the vigorous (8.0 ± 1.1 km·h^{-1}) intensity and least accurate at the moderate intensity (6.5 ± 0.9 km·h^{-1}). This suggests that the relationship between the accuracy of the GT3X+/iPhone and speed is not linear, as previously reported (Lee et al., 2014; Major & Alford, 2016). However, the difference in accuracy of the iPhone between intensities was very minimal, with MAPEs ranging from 0.6% to 1.5%, whereas the GT3X+ ranged from 1.2% to 6.9%. Contrastingly, the iPhone was least accurate at estimating EE at the vigorous intensity and performed best at light intensity (5.4 ± 1.0 km·h^{-1}), when compared to the criterion method of indirect calorimetry.

The GT3X+ on average overestimated EE at all speeds and was least accurate at the light intensity, while performing best at moderate intensity when compared to indirect calorimetry. Previous studies comparing the GT3X+ to indirect calorimetry have also found the device to overestimate EE at speeds comparable to those in the current study but to underestimate at faster running speeds, higher intensity activities, and at much slower walking speeds (2.6 km·h^{-1}) (Gastin, Cayzer, Dwyer, & Robertson, 2018; Mcminn, Acharya, Rowe, Gray, & Allan, 2013). Despite this apparent systematic bias, novel EE equations that are gender specific or that incorporate other metrics into the prediction equations have previously improved the accuracy of the GT3X+ in comparison to those available in the Actilife software (Santos-Lozano et al., 2013; Howe, Moir, & Easton, 2017).

During stationary cycling, the iPhone and GT3X+ significantly underestimated EE and had poor agreement with indirect calorimetry. The likely reason for the consistent underestimation of EE during stationary cycling is due to the stable position of the trunk where both the iPhone and GT3X+ were located. The adoption of a lower-limb accelerometer placement has previously been shown to improve the accuracy of pedal-revolution count during cycling when compared to waist-placement (Gatti, Stratford, Brenneman, & Maly, 2016). During aerobics, the iPhone significantly overestimated steps compared to the GT3X+ and underestimated EE compared to indirect calorimetry, with poor agreement for both comparisons. There was poor agreement between the GT3X+ estimations of EE compared to indirect calorimetry, however methods were not significantly different. The poor agreement between the iPhone and GT3X+ compared to indirect calorimetry during the aerobics trial suggests that both methods are unsuitable for monitoring EE during exercise that is not steady-state. The iPhone's overestimation of steps compared to the GT3X+ during aerobics suggests that it may not be suitable for monitoring exercise that requires non-uniform movement patterns.

In the free-living component of the study, the agreement in daily step count between the iPhone and GT3X+ devices was substantial although the iPhone, recorded significantly fewer steps (1095 steps·day^{-1}). It is not possible to ascertain the precise reason for this discord although user behaviour with the iPhone devices seems a likely explanation. While participants wore the GT3X+ attached to their waist, they were instructed to carry the iPhone as they would their own personal phone to ensure an ecologically valid measurement method. Depending on the individual, the iPhone may have been regularly left on a surface or carried in a bag. Participants may have been less likely to carry the iPhone on their person as it was additional to their own phone. Unfortunately, there was no way to monitor "wear-time" of the iPhone so this hypothesis remains speculative.

Further research on the validity, reliability, and sensitivity of smartphones to measure PA is clearly warranted. Developing a convenient and widely-used method for monitoring free-living PA would facilitate greater understanding of population PA levels and enable data sharing with health care professionals using a digital health platform. However, the daily step count metric does not enable a nuanced interpretation of PA as it lacks information on context, duration, frequency and intensity of the activity. Intensity in particular is important as the UK's National Health Service's current PA guidelines recommend that adults should undertake 150 min of moderate intensity exercise or 75 min of vigorous activity per week (UK Chief Medical Officer's Guidlines, 2011). The development of smartphone-specific algorithms to estimate the intensity of PA from the inbuilt accelerometer would generate a substantially more informative data set. The interpretation of the data may be further enhanced if other metrics such as Global Positioning System (GPS) data, (Gordon, Bruce, & Benson, 2016) or heart rate data can be combined with measurements of acceleration.

There are a few limitations of the current study. Firstly, the iPhone 5S model was used which utilises the M7 motion coprocessor technology whereas the most recent generation of iPhones (iPhone 11, 11 Pro and SE) have M13 co-processors. It is unclear how advancements in the coprocessor technology would influence the measurement of step count and EE. Secondly, estimations of EE from the iPhone were generated by the 'ActivityTracker' app using an algorithm based on acceleration, gender, body mass and stature. The algorithm itself is unbeknown to the researchers and is likely to be different to estimates of EE from other apps. Thirdly, the treadmill speeds which corresponded with light, moderate and vigorous intensities ranged from 5.4 ± 1.0 to 8 ± 1.1 km·h^{-1} based on measurements of heart rate reserve. Future studies should incorporate slower and faster speeds in more homogenous groups of participants.

In conclusion, the iPhone 5S is a suitable method of measuring step count but not EE during walking and jogging. In the free-living phase of the study, the iPhone significantly underestimated daily step count compared to an accelerometer worn continuously around the waist. This is likely because the phone was not carried on the person as frequently as the accelerometer. Further optimisation of the prediction algorithms in the mobile apps to incorporate measurements of heart rate and/or GPS data may enhance iPhone estimates of EE (Howe et al., 2017) and provide a more accurate and informative data set on PA and sedentary behaviour patterns. Finally, when using smartphones such as the iPhone 5S to measure step-count, users should be cognisant that there may be a significant underestimation of daily steps.

Acknowledgements

Dr David Muggeridge is supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB). The support of StormID for the access and use of the Lenus Health Digital Health Platform is gratefully acknowledged.

Funding Source

This study was supported by a research grant from the Digital Health and Care Institute.

Conflicts of Interest

The authors have no conflicts of interest to report.

Authors

All co-authors made a substantial contribution to the concept of the work, or acquisition, analysis, or interpretation of the data. All authors helped to draft and revised the article and approve the version to be submitted. CE is the guarantor for the work.

References

Aadland, E., & Ylvisåker, E. (2015). Reliability of the Actigraph GT3X + Accelerometer in Adults under Free-Living Conditions. *PLoS ONE*, *10*(8), 1–10. https://doi.org/10.1371/journal.pone.0134606

ACSM. (2017). *ACSM's Guidelines for Exercise Testing and Prescription (10th edition)*. *Wolters Kluwer* (Vol. 10). https://doi.org/10.1007/s13398-014-0173-7.2

Antero Kesaniemi (Chair), Y., Danforth, E., Jensen, M. D., Kopelman, P. G., Lefebvre, P., & Reeder, B. A. (2001). Dose-response issues concerning physical activity and health: an evidence-based symposium. *Medicine and Science in Sports and Exercise*, *33*(Supplement), S351–S358. https://doi.org/10.1097/00005768-200106001-00003

Barreira, T. V., Tudor-Locke, C., Champagne, C. M., Broyles, S. T., Johnson, W. D., & Katzmarzyk, P. T. (2013). Comparison of GT3X Accelerometer and YAMAX Pedometer Steps/Day in a Free-Living Sample of Overweight and Obese Adults. *Journal of Physical Activity and Health*, *10*, 263–270.

Dyrstad, S. M., Hansen, B. H., Holme, I. M., & Anderssen, S. A. (2014). Comparison of self-reported versus accelerometer-measured physical activity. *Medicine and Science in Sports and Exercise*, *46*(1), 99–106. https://doi.org/10.1249/MSS.0b013e3182a0595f

Ewing, C., Wilmore, J., Blair, S., Haskell, W., & Kraemer, W. (1998). ACSM Position Stand: The Recommended Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory and Muscular Fitness, and Flexibility in Healthy Adults The position stand replaces the 1990 ACSM position paper of similar title (3). S. *Med. Sci. Sports Exerc*, *30*(6), 975–991. Retrieved from http://ovidsp.tx.ovid.com.libproxy.csun.edu/sp-3.15.1b/ovidweb.cgi

Gastin, P. B., Cayzer, C., Dwyer, D., & Robertson, S. (2018). Validity of the ActiGraph GT3X+ and BodyMedia SenseWear Armband to estimate energy expenditure during physical activity and sport. *Journal of Science and Medicine in Sport*, *21*(3), 291–295.

https://doi.org/10.1016/j.jsams.2017.07.022

Gatti, A. A., Stratford, P. W., Brenneman, E. C., & Maly, M. R. (2016). GT3X+ accelerometer placement affects the reliability of step-counts measured during running and pedal-revolution counts measured during bicycling. *Journal of Sports Sciences*. https://doi.org/10.1080/02640414.2015.1096018

Gordon, B. A., Bruce, L., & Benson, A. C. (2016). Physical activity intensity can be accurately monitored by smartphone global positioning system 'app.' *European Journal of Sport Science*, *16*(5), 624–631. https://doi.org/10.1080/17461391.2015.1105299

Howe, C. C. F., Moir, H. J., & Easton, C. (2017). Classification of Physical Activity Cut-Points and the Estimation of Energy Expenditure During Walking Using the GT3X+ Accelerometer in Overweight and Obese Adults. *Measurement in Physical Education and Exercise Science*, *21*(3), 127–133. https://doi.org/10.1080/1091367X.2016.1271801

Jarrett, H., Fitzgerald, L., & Routen, A. (2015). Inter-instrument reliability of the Actigraph GT3X+ Ambulatory Activity Monitor during free-living conditions in adults Manuscript. *Critical Studies on Security*, *2*(2), 210–222.

Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, *14*(4), 321–327. https://doi.org/10.1080/02640419608727717

Le Masurier, G. C., Lee, S. M., & Tudor-Locke, C. (2004). Motion Sensor Accuracy under Controlled and Free-Living Conditions. *Medicine and Science in Sports and Exercise*, *36*(5), 905–910. https://doi.org/10.1249/01.MSS.0000126777.50188.73

Lee, J. A., Williams, S. M., Brown, D. D., & Laurson, K. R. (2014). Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. *Journal of Sports Sciences*, *33*(10), 991–1000. https://doi.org/10.1080/02640414.2014.981848

Lin, L. I. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, *45*(1), 255–268. https://doi.org/10.2307/2532051

Major, M. J., & Alford, M. (2016). Validity of the iPhone M7 motion co-processor as a pedometer for able-bodied ambulation. *Journal of Sports Sciences*, *34*(23), 2160–2164. https://doi.org/10.1080/02640414.2016.1189086

Martin Bland, J., & Altman, D. (1986). Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *The Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Mcminn, D., Acharya, R., Rowe, D. A., Gray, S. R., & Allan, J. L. (2013). Measuring Activity Energy Expenditure: Accuracy of the GT3X+ and Actiheart Monitors. *International Journal of Exercise Science*, *6*(3), 217–229.

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Delisle Nyström, C., Mora-Gonzalez, J., Löf, M., … Ortega, F. B. (2017). Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Medicine*, *47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

O'Brien, W. J., Shultz, S. P., Firestone, R. T., George, L., Breier, B. H., & Kruger, R. (2017). Exploring the challenges in obtaining physical activity data from women using hip-worn accelerometers. *European Journal of Sport Science*, *17*(7), 922–930. https://doi.org/10.1080/17461391.2017.1323952

Ozemek, C., Kirschner, M. M., Wilkerson, B. S., Byun, W., & Kaminsky, L. A. (2014). Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. *Physiological Measurement*, *35*(2), 129–138. https://doi.org/10.1088/0967-3334/35/2/129

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). 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*, *5*(56). https://doi.org/10.1186/1479-5868-5-56

Reid, R. E. R., Insogna, J. A., Carver, T. E., Comptour, A. M., Bewski, N. A., Sciortino, C., & Andersen, R. E. (2017). Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. *Journal of Science and Medicine in Sport*, *20*(6), 578–582. https://doi.org/10.1016/j.jsams.2016.10.015

Santos-Lozano, A., Santín-Medeiros, F., Cardon, G., Torres-Luque, G., Bailón, R., Bergmeir, C., … Garatachea, N. (2013). Actigraph GT3X: Validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine*, *34*(11), 975–982. https://doi.org/10.1055/s-0033-1337945

Sasaki, J. E., John, D., & Freedson, P. S. (2011). Validation and comparison of ActiGraph activity monitors. *Journal of Science and Medicine in Sport*, *14*(5), 411–416. https://doi.org/10.1016/j.jsams.2011.04.003

Tanaka, H., Monahan, K. D., & Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, *37*(1), 153–156. https://doi.org/10.1016/S0735-1097(00)01054-8

Taylor, K., & Silver, L. (2019). *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally*. Retrieved from http://www.pewglobal.org/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & Mcdowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*(1), 181–188. https://doi.org/10.1249/mss.0b013e31815a51b3

Tudor-Locke, C., Barreira, T. V., & Schuna, J. M. (2015). Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine and Science in Sports and Exercise*, *47*(4), 839–842. https://doi.org/10.1249/MSS.0000000000000476

UK Chief Medical Officer's Guidlines. (2011). *Start Active, Stay Active: A report on physical activity from the four home countries' Chief Medical Officers*. *Department of Health, Physical Activity, Health Improvement and Protection*. https://doi.org/10.1134/S0021364011030027

Van Dyck, D., Cerin, E., De Bourdeaudhuij, I., Hinckson, E., Reis, R. S., Davey, R., … Sallis, J. F. (2015). International study of objectively measured physical activity and sedentary time with body mass index and obesity: IPEN adult study. *International Journal of Obesity*. https://doi.org/10.1038/ijo.2014.115

Weir, J. B. de V. (1949). NEW METHODS FOR CALCULATING METABOLIC RATE WITH SPECIAL REFERENCE TO PROTEIN METABOLISM. *Journal of Physiology*, *109*(5), 1–9.

Wijndaele, K., Westgate, K., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F. M., … Healy, G. N. (2015). Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus. *Medicine and Science in Sports and Exercise*, *47*(10), 2129–2139. https://doi.org/10.1249/MSS.0000000000000661

Figure Legends

Figure 1. Bland and Altman plots of iPhone and GT3X+ steps versus manually counted steps during a laboratory-based treadmill trial at light, moderate and vigorous intensities with mean bias (solid line) and 95% limits of agreement (dashed lines): (a) iPhone; (b) GT3X+.

Figure 2. Bland and Altman plots of iPhone and GT3X+ estimated energy expenditure (EE) versus measured EE (indirect calorimetry) during a laboratory-based treadmill trial including light, moderate and vigorous intensities with mean bias (solid line) and 95% limits of agreement (dashed lines): (a) iPhone; (b) GT3X+.

Figure 3. Bland and Altman plot of estimates of daily step count from the iPhone and GT3X+ during a free-living period. Mean bias (solid line) and 95% limits of agreement (dashed lines).

Table 1. Mean ± standard deviations estimates of step count from the iPhone, GT3X+, and criterion measure (manually counted steps) during treadmill walking and jogging at light, moderate and vigorous intensities.

Activity	**iPhone (Steps)**	**GT3X+ (Steps)**	**Manually Counted (Steps)**
Full Treadmill Trial	703 ± 97	675 ± 133	700 ± 98
Light	613 ± 46	579 ± 91	607 ± 47
Moderate	701 ± 89	659 ± 147	700 ± 88
Vigorous	796 ± 40	789 ± 43	794 ± 42

Table 2. Comparison of iPhone and GT3X+ estimates of step count with the criterion method (manually counted steps) during treadmill walking and jogging at light, moderate and vigorous intensities.

Intensity	Mean Bias (Steps)	95% LOA (Steps)	Lin's CCC (95% CI)	P-Value	MPE	MAPE
iPhone vs manually counted						
Light	6	-7 to 20	0.98 (0.95 to 0.99)	0.651	1.1 %	1.2 %
Moderate	1	-32 to 34	0.98 (0.96 to 0.99)	0.925	0.2 %	1.5 %
Vigorous	2	-9 to 13	0.99 (0.97 to 1.0)	0.895	0.3 %	0.6 %
GT3X+ vs manually counted						
Light	-28	-184 to 129	0.36 (0.03 to 0.62)	0.549	-4.6 %	5.5 %
Moderate	-41	-250 to 169	0.57 (0.29 to 0.77)	0.417	-6.0 %	6.9 %
Vigorous	-5	-47 to 36	0.87 (0.70 to 0.95)	0.693	-0.6 %	1.2 %

MPE = Mean percentage error
MAPE = Mean absolute percentage error

Table 3. Mean ± standard deviation estimates of energy expenditure from the iPhone, GT3X+ and criterion method (indirect calorimetry) during treadmill walking and jogging at light, moderate and vigorous intensities.

Activity	iPhone ($kcal \cdot min^{-1}$)	GT3X+ ($kcal \cdot min^{-1}$)	Indirect Calorimetry ($kcal \cdot min^{-1}$)
Full Treadmill Trial	5.6 ± 1.4	8.0 ± 3.2	7.5 ± 2.9
Light	4.8 ± 1.0	5.4 ± 2.2	4.9 ± 1.0
Moderate	5.5 ± 1.5	7.3 ± 2.4	7.2 ± 2.1
Vigorous	6.4 ± 1.4	11.2 ± 1.9	10.3 ± 2.2

Table 4. Comparison of energy expenditure estimates from the iPhone and GT3X+ with the criterion method (indirect calorimetry) during treadmill walking and jogging at light, moderate and vigorous intensities.

Intensity	Mean Bias (kcal·min^{-1})	95% LOA (kcal·min^{-1})	Lin's CCC (95% CI)	P-Value	MPE	MAPE
iPhone vs indirect calorimetry						
Light	-0.1	-1.4 to 1.1	0.80 (0.56 to 0.92)	0.922	-1.5 %	11.2 %
Moderate	-1.6	-3.7 to 0.4	0.58 (0.36 to 0.74)	0.034*	-21.3 %	22.1 %
Vigorous	-3.9	-6.7 to -1.1	0.20 (0.06 to 0.33)	< 0.001*	-37.8 %	37.8 %
GT3X+ vs indirect calorimetry						
Light	0.5	-2.1 to 3.2	0.67 (0.51 to 0.78)	0.611	6.5 %	23.7 %
Moderate	0.2	-2.5 to 2.8	0.82 (0.60 to 0.92)	0.961	2.3 %	14.0 %
Vigorous	0.9	-2.9 to 4.7	0.49 (0.11 to 0.75)	0.283	11.1 %	18.1 %

* denotes significance between measurement methods.
MPE = Mean percentage error
MAPE = Mean absolute percentage error