Responsiveness of device-based and self-report measures of physical activity to detect behaviour change in men taking part in the Football Fans in Training (FFIT) programme

Abstract

The capacity of physical activity (PA) measures to detect *changes* in PA within interventions is crucial. This is the first study to examine responsiveness of *activ*PAL3™ and the International Physical Activity Questionnaire (IPAQ; Short Form) in detecting PA change during a 12 week group-based, men-only weight management programme - Football Fans in Training (FFIT). Participants wore an *activ*PAL3™ and completed the IPAQ pre- and post-programme (n=30). Relationships between change scores were assessed by Spearman's correlations. Mean or median changes in PA were measured using paired samples t-tests and Wilcoxon signed-rank tests. Responsiveness to change was assessed utilising Standardised Response Mean (SRM). Both device-based and self-report measures demonstrated significant changes pre-post intervention, although these changes were not significantly correlated. The SRM values for changes in *activ*PAL3™ metrics were: 0.54 (MET-mins/day); 0.53 (step counts/day); and 0.44 (MVPA/day), indicating a small to medium responsiveness to change. SRM values for changes in IPAQ scores were: 0.59 (total PA mins/day); 0.54 (total MET-mins/day); 0.59 (walking MET-mins/day); 0.38 (vigorous MET-mins/day); and 0.38 (moderate MET-mins/day), revealing a small to medium responsiveness to change. These findings reveal that two commonly used device-based and self-report measures demonstrated responsiveness to changes in PA. While inclusion of both device-based and self-report measures is desirable within interventions it is not always feasible. The results from this study support that self-reported measures can detect PA change within behavioural interventions, although may have a tendency to overestimate changes, compared with device-based measures.

Keywords: physical activity measurement, accelerometer, questionnaire, sensitivity, intervention, adults

There is strong evidence that physical activity (PA) provides substantial health benefits (Warburton & Bredin, 2017). However, at least a third of adults around the world do not meet current recommendations for moderate to vigorous activity (Guthold, Stevens, Riley, & Bull, 2018).

Many strategies have been suggested to increase PA globally (WHO, 2018), although there is limited evidence of successful implementation. Kelly and Barker (2016) recently outlined six common errors repeatedly made by public health researchers/practitioners with regards to implementing scientific evidence when attempting to change health behaviours, including PA. We would like to propose another reason for this perceived failure in implementation: the difficulty in assessing which strategies work and which do not, and in those that work, the difficulty in assessing the extent to which they change behaviour. We suggest that, because measurement of PA behaviour can be challenging, it is often difficult to detect evidence of behaviour change. If measures of PA are used or interpreted incorrectly, interventions that appear to be ineffective may be incorrectly judged as successful (Type-I error) and those that are effective might be rejected (Type-II error).

To understand whether strategies are effective in changing PA behaviour it is vital that appropriate measurement methods are incorporated within evaluations of behavioural interventions which aim to assess the extent of PA behaviour change. However, PA is a complex and multi-faceted behaviour often characterised across several domains (i.e. leisure, travel, housework/gardening, and occupation), dimensions and determinants/correlates (Kelly, Fitzsimons, & Baker, 2016). Consequently, assessment of PA offers considerable methodological options and challenges (Warren et al., 2010). Subjective (i.e. self-reported) PA measures are commonly employed in population and intervention studies as they are easy to use and cost less than objective (i.e. device-based) assessment. Wearable device-based technologies, such as accelerometers, have become increasingly popular in recent years as PA assessment tools that are not prone to recall bias, more valid and reliable compared with self-report instruments and are often more practical compared with alternative more robust measures such as doubly labelled water (Silfee et al., 2018). Despite these benefits, there are limitations to relying solely on device-based forms of PA assessment (Pedišić & Bauman, 2015), particularly as outcomes in behavioural interventions. For instance, they are often unable to detect some forms of activity, and hence may underestimate overall PA levels (Silfee et al., 2018). They may also inadvertently influence PA when used as surveillance or measurement tools (e.g. measurement reactivity) and might enhance burden on participants (Baumann et al., 2018). Moreover, with the rise in the use of wearable device-based measures in recent years, there is substantial heterogeneity regarding the number of PA metrics being reported, limiting comparability between studies (Silfee et al., 2018).

Distinct forms of PA measurement can provide confusing or even contradictory findings (Thompson et al., 2009). Numerous studies have shown that correlations between device-based and self-report assessments of PA are low (Kowalski, Rhodes, Naylor, Tuokko, & MacDonald, 2012; Prince et al., 2008; Skender et al., 2016). It has been argued that although related, device-based and self-report measures assess distinct PA constructs and therefore not comparable (Fulton et al., 2016; Troiano, McClain, Brychta, & Chen, 2014).

Despite a growing number of intervention studies incorporating device-based and/or self-report measures of PA, there is a lack of research examining the (comparative) responsiveness of these measures to detect PA behaviour change over time as distinct PA constructs. For instance, there are only a small number of studies that have explicitly examined responsiveness to change of device-based and/or self-report measures in adults and children (e.g. Lee, Clark, Winkler, Eakin, & Reeves, 2015; Montoye, Pfeiffer, Suton, & Trost, 2014; Swartz, Rote, Cho, Welch, & Strath, 2014). The term responsiveness (or sensitivity) is typically defined as an indicator of an instrument's sensitivity to change as well as being a gauge of the magnitude of intervention-related change over time (Beaton, Bombardier, Katz, & Wright, 2001; Middel & van Sonderen, 2002). Although the validity and reliability of device-based and self-report PA instruments are often examined comprehensively (e.g. Lee, Macfarlane, Lam, & Stewart, 2011), responsiveness is comparatively under investigated, particularly within the context of behavioural interventions.

In order to understand whether PA interventions are effective in changing behaviour, it is vital to understand whether measures employed to evaluate changes in behaviour within intervention studies are capable of detecting changes in PA. In this study we aim to examine and compare the responsiveness of both device-based (*activ*PAL3™) and self-report (International Physical Activity Questionnaire; IPAQ, Short Form) PA measures to detect changes in PA behaviour, using data collected before and after participation in the Football Fans in Training (FFIT) programme, a weight management and healthy lifestyle intervention for men classified as overweight or obese (BMI>28kg/m^2) and aged 35-65 years (see Gray et al., 2013a; Hunt et al., 2014b; Wyke et al., 2015).

Methods

Participants and intervention setting

Football Fans in Training (FFIT) is a 12 week gender-sensitised, group programme delivered free of charge by trained community coaches to men at Professional Football clubs in Scotland. The development of the FFIT programme is detailed elsewhere (Gray et al., 2013a). In brief, FFIT was designed in line with evidence of what was known to be effective for weight loss (National Institute for Health and Clinical Excellence, 2006; Scottish Intercollegiate Guidelines Network, 2010) and to work with rather than against prevailing notions of masculinity, appealing to men in: context (professional football clubs); content (e.g. information around the science of weight management presented simply and branded materials, such as club T-shirts); and style of delivery (e.g. coaches encourage peersupport, participative learning and positive 'banter' to support discussion of more sensitive issues) (Wyke et al., 2015).

Funding was secured to undertake an evaluation of FFIT (a randomised controlled trial (RCT) incorporating an embedded process evaluation and cost-effectiveness); at that time funding was available for three deliveries of the programme in 13 professional football clubs (i.e. the 12 clubs in the top league in Scotland - then the Scottish Premier League (SPL) - and the most recently demoted club who had taken part in pilot deliveries in the previous season) in August-December 2011, February-April 2012 and August-December 2012. Men taking part in the August-December deliveries in 2011 and 2012 were participants in the FFIT RCT as outlined elsewhere (Hunt et al., 2014b; Wyke et al., 2015). During the baseline assessment period, the FFIT research team recruited adequate numbers of participants to fill *all* places then available on the three deliveries of FFIT (funded by the Football Pools and the Scottish Government). After recruitment of the intervention and control arms of the RCT had been achieved, the remaining 306 men were offered a place on 'non-trial' deliveries of FFIT in February-April, 2012. The RCT of FFIT demonstrated significant mean between-group difference in weight loss of 4.94kg (CI 3.95, 5.94, $p<0.0001$) at 12 months after baseline (primary outcome), and in self-reported PA (International Physical Activity Questionnaire, Short Form), and other secondary outcomes, all in favour of the intervention group (Hunt et al., 2014a; Wyke et al., 2015). No device-based measures of PA were taken in men participating in the RCT.

The February 2012 delivery of FFIT provided an opportunity to examine factors not feasible to investigate within the FFIT RCT. This included the incorporation of measures of PA to assess pre- and post-programme activity levels, and changes in PA assessed both subjectively and objectively. All participants in the current study were sampled from men who took part in the 'non-trial' deliveries of

FFIT at 12 clubs, between February-April 2012 (Donnachie, Wyke, & Hunt, 2018; Donnachie, Wyke, Mutrie, & Hunt, 2017).

Procedure

Data collection occurred between January 2012 and May 2012. Of the 306 men offered places on the February 2012 deliveries of FFIT, 203 men attended the pre-programme measurement sessions at each professional football club stadium and undertook a battery of objective physical (e.g. anthropometric measurements and blood pressure) and subjective (e.g. PA and diet) assessments pre-programme (T0) and post-programme (T1, 12 week follow-up). All of the assessments were performed by fieldwork staff trained to standard protocols concordant with the FFIT RCT (Hunt et al., 2014a; Hunt et al., 2014b).

Prior to attending pre-programme measurement sessions, men from four clubs (n=94) were sent a letter outlining research for a sub-study on objective PA assessment and inviting them to take part. This provided adequate time to decide if they were willing to take part in the sub-study before attending the pre-programme stadium-based measurement sessions. At T0 (week 0 of the FFIT programme), participants from these clubs were asked if they had received the information letter, given an additional copy of the study information to read and asked if they would be willing to wear an *activ*PAL3™ (PAL Technologies Ltd., Glasgow, Scotland, UK) device for seven consecutive days (i.e. providing six full days of activity monitoring) so that it could be retrieved when they attended for their first programme session (week 1 of the FFIT programme) the following week. They were also asked if they would be willing to wear the *activ*PAL3™ again between week 11 and week 12 of the programme (T1). Those who agreed to wear the device at week 11, provided data for a further seven days after the devices were collected at week 12, the final week of the programme.

During the pre-programme measurement session, participants gave written informed consent after they were fully briefed on the purpose of the *activ*PAL3™, and given a demonstration on how to remove and re-affix the device. The *activ*PAL3™ was placed inside a waterproof and protective nitrile sleeve, wrapped in a single layer of Hypafix® water resistant adhesive. Next, the device was affixed directly to the skin of the participants' right leg with one sheet of Hypafix® adhesive (10cm X 13cm) by the first author, following standardised protocols to protect privacy; participants were told the device need only be removed to prevent the device being immersed in water (i.e. during swimming or bathing) but could be worn while showering and sleeping. The men were each given additional Hypafix® strips to re-apply the device should it need to be removed for any reason throughout the week. When the monitors were removed and retrieved by the first author at each of the four clubs the following week, participants were asked to complete the IPAQ (short form) to obtain concurrent self-reported PA, recalled over the previous week. The same procedures were repeated again for FFIT programme weeks 11-12. Full ethical approval was granted by the University of Glasgow, College of Social Sciences Research Ethics Committee (CSS201020106).

Measures

The device-based outcome measures in the current study were measured by the *activ*PAL3™ device and included: number of steps taken per day; minutes of moderate to vigorous intensity PA; and energy expenditure. The *activ*PAL3™ is a triaxial accelerometer/inclinometer which incorporates proprietary technology (Intelligent Activity Classification™) to measure three types of free-living activity: time spent sitting/lying; standing; and stepping. The *activ*PAL3™ quantifies the amount of steps performed, the intensity of steps taken (cadence) and estimates of energy expenditure (Lord et al., 2011). It is a small (35 X 53 X 7mm), lightweight device (15g), worn discreetly on the middle of the thigh between the hip and the knee, above the quadriceps muscle. The device has a battery life of around nine days, and thus can be worn continuously for 24 hour monitoring. The data are recorded in 15-second epochs and the output downloaded onto a Personal Computer (PC) via a USB interface. Previous studies have shown the activity and posture functions of the *activ*PAL™ to be valid compared to direct observation and acceptable to participants in community-based research (e.g. Grant, Dall, Mitchell, & Granat, 2008).

Prior to assessment, each of the *activ*PAL3™ monitors was fully charged and initialised to record six consecutive days. On retrieval of the monitors, data were uploaded to a PC using *activ*PAL™ proprietary software (PALtechnologies v5.9.1.1). Microsoft Office Excel was used for subsequent data processing and management. Custom software (HSC analysis software v2.19, Philippa Dall and Malcolm Granat, Glasgow Caledonian University) was used to extract information on individual participants' PA intensity using the *activ*PAL3™ time-stamped 'event' data files. Based on the conclusions of a systematic review (Tudor-Locke & Rowe, 2012), in the current study, time spent stepping at a cadence of at least 100 steps/minute was deemed indicative of moderate intensity PA. Daily energy expenditure is classified by the *activ*PAL3™ software as metabolic equivalent (MET-hours) and expressed in this study as MET-minutes per day. Best practice guidelines for accelerometer use in PA measurement suggest that for adults a minimum of 3-5 days of monitoring is necessary to quantify free-living PA (Trost, McIver, & Pate, 2005; Ward, Evenson, Vaughn, Rodgers, & Troiano, 2005). Data files were inspected visually and individual cases excluded if less than three days of wear time were evident. Days with <500 steps were removed, consistent with previous studies which have incorporated similar cut-offs to classify non-wear days (Edwardson et al., 2017). Wake/sleep times were included as recorded by the *activ*PAL3™, enabling capture of daily 24 hour activity (i.e. midnight to midnight). Self-report logs/diaries (e.g. to record sleep, wake or removal time) were not incorporated in the current study to reduce overall participant burden.

Self-reported PA outcomes were measured using the International Physical Activity Questionnaire (IPAQ, Short Form) (Craig et al., 2003), which included total PA minutes and total work done in PA per week (MET-minutes). The IPAQ is a well-established measure designed principally as a gold standard for population surveillance of PA among adults (18-65 years) (Craig et al., 2003). According to the IPAQ scoring guidelines, data are reported as total metabolic equivalent of task (in MET-minutes) per week. Calculation of the total score involves summation of the duration (i.e. minutes) and frequency (i.e. days) of walking, moderate-intensity and vigorous-intensity activities, recalled over the past seven days. MET-minute scores are quantified by multiplying the MET score of an activity by the minutes performed. All reported activity (i.e. walking, moderate and vigorous activity) exceeding 'three hours' (or 180 minutes) were truncated to allow a maximum of 21 hours of activity per week for each category to minimise over reporting, consistent with the IPAQ scoring protocol (https://sites.google.com/site/theipaq/scoring-protocol). Changes in IPAQ and *activ*PAL3™metrics are expressed in total minutes 'per day' to enable comparison between both measures. All IPAQ total scores ('per week') were divided by 7 for daily PA estimates.

Statistical analysis

Descriptive statistics are presented as means (standard deviation, SD), medians (interquartile range, IQR) and percentages (number). Exploratory analysis revealed that the majority of device-based PA metrics were approximately normally distributed, whereas the self-reported data were positively skewed. Paired samples t-tests or Wilcoxon signed-rank tests (for data that violated assumptions of normality) were used to examine differences pre- and post-intervention. Spearman's rank-order correlation coefficients (rho) were used to assess relationships between change scores for device-based (*activ*PAL3™) and self-report (IPAQ) instruments, interpreted as weak (<0.3), low (0.30–0.49), moderate (0.50–0.69), strong (0.70–0.89) or very strong (≥0.90).

Responsiveness to change in device-based and self-report PA scores between T0 and T1 was assessed using the Standardised Response Mean (SRM) or Cohen's *dz* (Cohen, 1977; Dankel & Loenneke, 2018; Lakens, 2013; Liang, Fossel, & Larson, 1990), a type of effect size that has been used in previous studies to assess responsiveness to change of device-based and self-report PA instruments in adults and children (e.g. Almeida et al., 2017; Clevenger et al., 2018; Swartz et al., 2014). SRM is calculated for each measure by dividing the absolute mean change score by the standard deviation of differences between the paired measurements and can be interpreted in line with Cohen's *d* as trivial, small, moderate or large (<0.20, ≥0.20 to <0.50, ≥0.50 to <0.80, and ≥0.80, respectively) (Cohen, 1977; Husted, Cook, Farewell, & Gladman, 2000; Stratford & Riddle, 2005). In addition to SRM values, non-parametric effect size (ES) values were calculated for each of the device-based and self-reported PA metrics using Wilcoxon's statistic and related *Z*-score divided by the square root of *n* (z/√n), interpreted as small (*r* = <0.3), medium (*r* = ≥0.3 to <0.5) and large (*r* = ≥0.5) (Cohen, 1977; Field, 2009).

The Guyatt Responsiveness Index (GRI) is an alternative measure of responsiveness based on the variability of changes among stable participants (Guyatt, Walter, & Norman, 1987; Husted et al., 2000). When utilising this responsiveness statistic, participants are preferably assessed on several occasions to ascertain the level of variability across a 'stable' time period, ideally before taking part in an intervention, to detect minimally clinically important change exceeding any spurious changes in measurement which may occur over time (Guyatt et al., 1987). However, where only two observations are available (i.e. baseline and post-intervention), the GRI is calculated as the mean score of participants identified as improved, divided by the standard deviation of the change in participants identified as stable or showing no improvement pre- to post-intervention. In this analysis, the mean change of participants identified as having increased PA between T0 and T1, as indicated by each of the device-based and self-report PA metrics, were incorporated as the numerator, whereas the standard deviation of the change in participants identified as unchanged or having decreased PA were included as the denominator. Similarly to SRM, GRI values of 0.20, 0.50 and 0.80 or greater have been used to delineate low, moderate and high responsiveness, respectively (Husted et al., 2000). However, it is important to note that the GRI method is anticipated to yield higher coefficients than the SRM or other ES values as a consequence of the removal of mean change values of unchanged participants (Stucki, Liang, Fossel, & Katz, 1995). Statistical analyses were undertaken using IBM SPSS 21.0 (Armonk, NY, USA). All tests were two-tailed with an alpha p-value of $p<0.05$ to assess statistical significance.

Results

Demographic characteristics

Descriptive characteristics of participants pre-programme are displayed in Table 1. The mean age of participants was 45.9 years (*SD* = 9.8). Mean body weight was 111.8kg (*SD* = 14.3), mean BMI was 35.9 kg/m^2 (*SD* = 5.3) and mean waist circumference was 118.5cm (*SD* = 11.1), thus comparable with clinical characteristics of men taking part in other research deliveries of FFIT (Gray et al., 2013b; Hunt et al., 2014a). Participants in this study were from across the socioeconomic spectrum, consistent with previous research demonstrating that FFIT attracted men from a range of socioeconomic backgrounds (Hunt et al., 2014b).

Changes in device-based and self-reported physical activity

Changes in device-based and self-reported PA between T0 and T1 are presented in Table 2. Data are presented for n=30 participants with concurrent PA data (i.e. device-based and self-report assessments) at both time points. Paired samples t-tests confirmed significant increases in *activ*PAL3™-assessed number of average daily ‘steps’ from 8315.5 (*SD* = 3063.3) at T0 to 9834.4 (*SD* = 3855.9) at T1 with an increase of 1518.8 steps, ($t(29) = -2.9$, $p = 0.007$), time spent stepping at least at a moderate cadence increased from 28.3 (*SD* = 18.8) minutes/day at T0 to 37.8 (*SD* = 27.3) at T1 with an increase of 9.5 minutes/day ($t(29) = -2.4$, $p = 0.022$) and increased daily MET-minutes from 2040 (*SD* = 78) at T0 to 2076 (*SD* = 90) at T1 with an increase of 36 MET-minutes/day, ($t(29) = -2.9$, $p = 0.006$). Wilcoxon signed rank tests showed significant increases in self-reported PA (IPAQ) at T1 from T0 for total PA minutes (Z = -

3.56, p = <0.001, median difference = 83 minutes/day), total MET-minutes (Z = -3.59, p = <0.001, median difference = 341 MET-minutes/day), walking MET-minutes (Z = -2.86, p = 0.004, median difference = 154 MET-minutes/day), moderate MET-minutes (Z = -2.53, p = 0.011, median difference = 56 MET-minutes/day) and vigorous MET-minutes (Z = -2.58, p = 0.010, median difference = 80 MET-minutes/day).

The SRM and non-parametric effect size (ES) values for changes in device-based and self-reported PA between T0 and T1 are also displayed in Table 2. The SRM values for device-based (*activ*PAL3™) time spent active stepping at least at a moderate cadence, average steps per day and daily MET-minutes were d = 0.44, d = 0.53 and d = 0.54, respectively, demonstrating a small to moderate responsiveness to change. The SRM values for total self-reported PA minutes/day, MET-minutes/day, walking MET-minutes/day, moderate MET-minutes/day and vigorous MET-minutes/day were, d = 0.59, d = 0.54, d = 0.59, d = 0.38 and d = 0.38, respectively, revealing a small to moderate responsiveness to change between T0 and T1. The non-parametric ES values for changes in time spent active at least at a moderate stepping cadence, average steps per day and daily MET-minutes were r = 0.35, r = 0.47 and r = 0.48, respectively, thus indicating a moderate effect size. The non-parametric ES values for changes in self-reported (IPAQ) total minutes/day, MET-minutes/day, walking MET-minutes/day, moderate MET-minutes/day and vigorous MET-minutes/day were r = 0.65, r = 0.66, r = 0.52, r = 0.46 and r = 0.47 respectively, indicating a moderate to large effect size.

The GRI values for changes in device-based and self-reported PA between T0 and T1 are depicted in Table 3. The GRI values for device-assessed daily MET-minutes, average steps per day and time spent active stepping at least at a moderate cadence (GRI = 2.24, GRI = 2.36, and GRI = 4.21, respectively) showed a large responsiveness to change. Self-reported total PA minutes/day and total MET-minutes/day were GRI = 0.66 and GRI = 0.64, respectively, demonstrating a moderate responsiveness to change as a consequence of higher variability in changed total self-reported PA among participants. However, IPAQ sub-domains of walking, moderate and vigorous MET-minutes/day were GRI = 1.49, GRI = 1.05 and GRI = 11.83, respectively, revealing a large responsiveness to change between T0 and T1.

Comparison between device-based and self-reported physical activity

The Spearman's correlation coefficients between device-based and self-reported activity scores at both T0 and T1 are displayed in Table 4. Generally, the highest correlations among device-based and self-report PA measures were observed at T0. The correlation coefficients between *activ*PAL3™-assessed PA (number of steps, time spent stepping at least at a moderate intensity and total MET-minutes) and one of the five IPAQ metrics (walking MET-minutes), were positive but low (*rho* = 0.42, *rho* = 0.49 and *rho* = 0.37, respectively). The correlations between each of the *activ*PAL3™ metrics and IPAQ-assessed total PA minutes, total MET-minutes, moderate MET-minutes and vigorous MET-minutes/day were all non-significant. All of the correlation coefficients between device-based and self-report PA measures at T1 were not statistically significant, ranging from low to weak (*rho* = 0.36 to -0.11).

The Spearman's correlation coefficients between the change scores for device-based and self-reported PA measures are displayed in Table 5 and ranged from -0.30 to 0.35; none of these were statistically significant.

Discussion

The capacity of self-report and device-based PA instruments to detect change in PA within intervention settings is crucial to determining which interventions work. To our knowledge, this is the only study that has compared responsiveness of both IPAQ (i.e. self-report) and *activ*PAL3™ (i.e. device-based) measures across a number of outcome metrics to assess changes in PA over time within the context of a behavioural intervention. This is also the first study to examine changes in both device-based and self-reported PA within research deliveries of the FFIT programme, and extends previous research demonstrating significant increases in self-reported PA (i.e. IPAQ, Short Form) during pilot (Gray et al., 2013b) and full trial (Hunt et al., 2014b) deliveries of FFIT.

In this study, changes in all device-based and self-reported PA metrics were statistically significant. According to device-based assessment (*activ*PAL3™) taking part in the 12 week FFIT programme resulted in an increase in an average of 1519 steps; 9.5 minutes spent stepping at least a moderate stepping intensity (i.e. ≥100 steps/minute); and an extra 36 MET-minutes per day. According to the self-reported PA measure (IPAQ), taking part in FFIT increased total PA by 83 minutes and 341 MET-minutes per day. IPAQ sub-domains of walking, moderate and vigorous intensity activity also showed an increase of 154, 56, and 80 MET-minutes per day, respectively. The most salient finding from the current study is that we observed comparable responsiveness to change for both device-based and self-report instruments. SRM values for *activ*PAL3™ were greatest when measuring change in average MET-minutes (0.54) and steps per day (0.53), whereas IPAQ SRM values were highest when assessing change in total PA MET-minutes (0.54), walking MET-minutes (0.59) and total PA minutes (0.59), classified as moderate (i.e. SRM values ≥0.50). The SRM values for IPAQ sub-domains of moderate and vigorous PA intensity (both SRMs 0.38) and *activ*PAL3™-assessed MVPA (time spent active at least a moderate intensity) (0.44), are considered small (i.e. SRM values <0.50). Similar trends were found for both non-parametric ES and GRI responsiveness values. Thus, our findings indicate that despite uncorrelated changes, both instruments were able to detect a comparable magnitude of change in PA. In contrast with previous research (Lee et al, 2015), self-reported PA demonstrated slightly greater responsiveness compared with device-based measures for total PA within the context of a 12 week, men-only behavioural intervention. However, total self-reported PA (i.e. IPAQ total PA minutes and total MET-minutes) demonstrated lower GRI values (GRI = 0.66 and GRI = 0.64, respectively) compared to each of the three device-assessed PA metrics: *activ*PAL3™-assessed MET-minutes (GRI = 2.24); number of steps (GRI = 2.36); and time stepping at a moderate cadence (GRI = 4.21), thus demonstrating lower variability in changed device-based PA pre- to post-intervention. The findings suggest that both IPAQ and *activ*PAL3™ measures should be responsive to change when evaluating PA change in future intervention settings. Nonetheless, due to the substantial differences in self-reported PA scores compared with device-based assessment, caution is warranted when interpreting intervention change based solely on self-reported PA, as may overestimate change consistent with other research (Winkler et al., 2013).

There are a limited number of studies that have investigated responsiveness to change of both device-based and self-report PA measures in adults. A recent study investigated responsiveness to change of self-reported (Baecke Habitual Physical Activity Questionnaire) and device-based (ActiGraph GT3X-BT) PA measures in patients with chronic low back pain receiving physical therapy (Morelhão et al., 2018). The authors concluded that none of the PA measures were able to detect changes in PA over time, according to SRM values (<0.20). Similarly, Almeida et al (2017) examined the responsiveness of self-report (Community Health Activities Model Program for Older Adults Questionnaire) and two distinct device-based (Actigraph GT3X; Sensewear Armband) measures in detecting changes in PA in older adults with osteoarthritis during a rehabilitation programme following knee replacement surgery. The findings revealed that each PA measure exhibited low responsiveness to change (i.e. in light, moderate and vigorous intensity PA) as indicated by SRM values (<0.30). Nicaise and colleagues examined the sensitivity of the IPAQ (Long Form) for detecting changes in PA compared with device-based (Actigraph 7164) assessment among Spanish-speaking Latina women during a 12 week pedometer-based intervention (Nicaise, Crespo, & Marshall, 2014). In this study, both IPAQ (r = 0.27) and device-based (r = 0.40) measures detected intervention-related changes in moderate intensity PA, indicating a small and moderate effect size of change. Consistent with our study findings, changes in self-report and device-based PA metrics were not correlated at 12 weeks.

Lee and colleagues reported significant changes in total PA minutes/week for two distinct self-report (Active Australia Survey; United States National Health Interview Survey) and device-based PA measures (Actigraph GT1M), longitudinally within the context of a weight loss intervention that were small in magnitude, although device-based PA was classed as slightly more responsive (Lee et al., 2015). Research conducted by the same group of authors investigated responsiveness to changes in PA using three unique self-report instruments (Community Health Activities Model Program for Older Adults Questionnaire; Active Australia Survey; United States National Health Interview Survey) in adults following a four month behavioural intervention, demonstrating a small responsiveness to change (Reeves, Marshall, Owen, Winkler, & Eakin, 2010). The findings observed in the current study are similar to other research comparing responsiveness of device-based activity measures in adult populations (Swartz et al., 2014; van Nassau, Chau, Lakerveld, Bauman, & van der Ploeg, 2015). For instance, Swartz et al (2014) examined responsiveness to change in two different device-based PA measures (Actigraph GT3X; *activ*PAL™) in sedentary adults during a behavioural intervention to reduce sitting time. They observed comparable SRM (0.44) values post-intervention for changes in *activ*PAL™ assessed PA (average daily steps) indicating a small responsiveness to change.

As noted by Lee et al (2015), the majority of studies have focused on assessing the validity of PA measures to examine behaviour change within interventions over time, usually relying on correlations between changes in self-report and device-based measures (e.g. Hoos, Espinoza, Marshall, & Arredondo, 2012; Nicaise et al., 2014; Sloane, Snyder, Demark-Wahnefried, Lobach, & Kraus, 2009). However, research findings have indicated greater disagreement between device-based and self-reported PA at increased activity levels (e.g. Slootmaker, Schuit, Chinapaw, Seidell, & van Mechelen, 2009), hence agreement between instruments may be attenuated by intervention effects (Lee et al., 2015). Winkler et al (2013) observed that agreement between self-report and device-based measures deteriorated as levels of PA increased during a behavioural PA intervention, particularly among intervention group participants compared to controls. The authors reported that intervention effects were greater when PA was assessed by self-report compared with device-based measures, despite both instruments yielding statistically significant differences. They suggest PA interventions might appear more effective when relying exclusively on self-report.

Device-based measures of PA are often heralded as the 'gold standard' for PA behaviour as they demonstrate somewhat stronger agreement with doubly labelled water (a precise measure of total energy expenditure) in comparison to self-report measures (Kelly et al., 2016). Device-based measures of PA quantify acceleration and movement, whereas self-reported methods provide an understanding of the purpose, domain and context of PA behaviour (Troiano, Gabriel, Welk, Owen, & Sternfeld, 2012). Both forms of PA assessment have distinct limitations and are susceptible to different forms of measurement error. For instance, self-report PA assessment is more prone to social desirability bias, poor recall, or misreading of questionnaires. Specifically, the IPAQ Short Form has been shown to overestimate PA by approximately 84 percent compared to objective assessments (Lee et al., 2011). Also, it is possible that some participants could have responded more favourably when completing self-reported PA assessments post-intervention as they may not have wanted to appear less physically active (Adams et al., 2005). Additionally, lifestyle interventions incorporating behaviour change techniques, such as self-monitoring of PA and goal setting, like the FFIT programme, may enhance participants' awareness of PA, hence potentially influencing PA reporting (Winkler et al., 2013).

In contrast, device-based measures may fail to accurately recognise certain forms of activity (e.g. swimming or resistance training) and therefore underestimate overall intervention effects. However, during the FFIT programme, participants were encouraged to increase their activity levels predominantly by increasing steps during the graduated walking component of the programme. It is therefore unlikely that many of the participants in this study would have been performing other forms of activity during the intervention, such as swimming or strength-based exercises, thus the magnitude of error was likely small. Devices can be lost or malfunction, and only some participants may be willing to wear them. Further, data processing requires subjective decisions about thresholds and cut-offs that are much debated (Wijndaele et al., 2015). It has been advocated that due to the complexity in measuring PA, no single methodology is able to sufficiently capture all PA domains and subcomponents (Warren et al., 2010). We do not argue one or other method should be used; combining different methods of PA assessment may provide a more comprehensive reflection of individuals' amount of activity and its context, offering greater insights regarding evidence of behaviour change and efficacy of behavioural interventions targeting this complex behaviour. However, this study incorporated the IPAQ, Short Form which does not measure contextual information in the same way as the IPAQ, Long Form (e.g. leisure, transportation, housework/gardening, and occupation-related activity). Future studies investigating responsiveness of the IPAQ (Long Form) to changes in PA compared with device-based measures would be advantageous.

Previously noted low correlations between device-measured and self-reported PA have been used to criticise self-report measures. The results presented here suggest that if ability to detect *change* in PA behaviour is considered, self-reported PA can provide comparable sensitivity compared with device-based assessment. Importantly, it is frequently noted that the IPAQ should not be used to detect intervention effects as it was not designed for this purpose. The reality is that, due to its ubiquity, ease of use, and the lack of a viable alternative, it often is used. The results from this study based on a population of adult Scottish men add to arguments that the IPAQ can detect PA behaviour change. Nevertheless, it is important to note that higher responsiveness for self-reported PA may have occurred as a consequence of over reporting.

Strengths and limitations

This study has a number of strengths. First, the incorporation of device-based and self-report measures of PA enabled examination of the responsiveness of each measure to behaviour change longitudinally within the context of a 12 week intervention. The *activ*PAL™ monitor has been shown to be accurate in assessing step count in adults at normal walking speeds (Grant et al., 2008). The device is also able to assess MVPA utilising a threshold of cadence generally indicative of a moderate intensity (i.e. 100 steps/minute). Additionally, inclusion of distinct indices of effect size and responsiveness is a further strength of this study as these calculations are simple to perform and easy to interpret, providing valuable information on the magnitude of behaviour change, thus enabling comparison of intervention efficacy across the field.

However, the study has some limitations. The assessment of responsiveness using SRM values is dependent to some degree on the extent to which data are normally distributed, although this is almost never the case with PA data (Lee et al., 2015). Also, the lack of a control condition restricted our use of alternative responsiveness methods used in other comparable studies. For example, the responsiveness statistic (Husted et al., 2000) also enables a comparison of the mean change in intervention scores compared to a control group condition. Hence, future research with a control condition would be advantageous. Future assessment regarding the degree of responsiveness to PA behaviour change (detected via self-report and device-based measures) compared to an established criterion (direct observation) would also be enlightening. Moreover, comparison of self-report and device-based methods in assessing responsiveness to long term behaviour change (beyond 12 weeks) would be of considerable value in understanding maintenance of PA behaviour change post-intervention. Another limitation of this study is the relatively small sample size which may have increased the chance of a type-II error, although a strength of using SRM as a measure of responsiveness is that it is independent of sample size (Prous, Salvanés, & Ortells, 2008). Additionally, the high degree of attrition may also indicate some bias towards participants who were successful in changing behaviour as indicated by both self-report and device-based PA methods. Lastly, the findings are specific to adult men participating in a weight management and healthy lifestyle programme in Scotland (UK) and generalisability to wider population groups may be limited.

Conclusion

In this study, two commonly used device-based (*activ*PAL3™) and self-report (i.e. IPAQ, Short Form) PA measures were found to be responsive to behaviour change in men following participation in a 12 week weight loss and healthy living programme (FFIT), although there were non-significant correlations between these change scores. The magnitude of responsiveness to change was marginally higher for self-reported PA according to SRM values. While inclusion of both device-based and self-report measures is desirable, it is not always feasible, hence these findings provide support for the utility of self-reported PA instruments within the context of behavioural interventions promoting increased PA, although they may overestimate PA changes, relative to device-based measures.

References

Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology, 161*(4), 389-398. doi: 10.1093/aje/kwi054

Almeida, G. J., Terhorst, L., Irrgang, J., J, , Fitzgerald, G. K., Jakicic, J. M., & Piva, S. R. (2017). Responsiveness of physical activity measures following exercise programs after total knee arthroplasty. *Journal of Exercise, Sports & Orthopedics, 4*(3), 1-8. doi: http://dx.doi.org/10.15226/2374-6904/4/3/00164

Baker, G., Mutrie, N., & Lowry, R. (2008). Using pedometers as motivational tools: Are goals set in steps more effective than goals set in minutes for increasing walking? *International Journal of Health Promotion and Education, 46*(1), 21-26. doi: 10.1080/14635240.2008.10708123

Baumann, S., Gross, S., Voigt, L., Ullrich, A., Weymar, F., Schwaneberg, T., . . . Ulbricht, S. (2018). Pitfalls in accelerometer-based measurement of physical activity: The presence of reactivity in an adult population. *Scandinavian journal of medicine & science in sports, 28*(3), 1056-1063. doi: 10.1111/sms.12977

Beaton, D. E., Bombardier, C., Katz, J. N., & Wright, J. G. (2001). A taxonomy for responsiveness. *Journal of Clinical Epidemiology, 54*(12), 1204-1217. doi: http://dx.doi.org/10.1016/S0895-4356(01)00407-3

Clevenger, K. A., Moore, R. W., Suton, D., Montoye, A. H. K., Trost, S. G., & Pfeiffer, K. A. (2018). Accelerometer responsiveness to change between structured and unstructured physical activity in children and adolescents. *Measurement in Physical Education and Exercise Science, 22*(3), 224-230. doi: 10.1080/1091367X.2017.1419956

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Oja, P. (2003). International physical activity questionnaire: 12-Country reliability and validity. *Medicine & Science in Sports & Exercise, 35*(8), 1381-1395. doi: 10.1249/01.MSS.0000078924.61453.FB

Dankel, S. J., & Loenneke, J. P. (2018). Effect sizes for paired data should use the change score variability rather than the pre-test variability. *Journal of Strength and Conditioning Research, 00*(00), 1-6. doi: 10.1519/jsc.0000000000002946

Donnachie, C., Wyke, S., & Hunt, K. (2018). Men’s reactions to receiving objective feedback on their weight, BMI and other health risk indicators. *BMC Public Health, 18*(1), 291. doi: 10.1186/s12889-018-5179-1

Donnachie, C., Wyke, S., Mutrie, N., & Hunt, K. (2017). ‘It’s like a personal motivator that you carried around wi’ you’: utilising self-determination theory to understand men’s experiences of using pedometers to increase physical activity in a weight management programme. *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 61. doi: 10.1186/s12966-017-0505-z

Edwardson, C. L., Winkler, E. A. H., Bodicoat, D. H., Yates, T., Davies, M. J., Dunstan, D. W., & Healy, G. N. (2017). Considerations when using the activPAL monitor in field-based research with adult populations. *Journal of Sport and Health Science, 6*(2), 162-178. doi: https://doi.org/10.1016/j.jshs.2016.02.002

Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3rd ed.). London, England: SAGE.

Fitzsimons, C., Baker, G., Gray, S., Nimmo, M., & Mutrie, N. (2012). Does physical activity counselling enhance the effects of a pedometer-based intervention over the long-term: 12-Month findings from the Walking for Wellbeing in the west study. *BMC Public Health, 12*(1), 206. doi: https://doi.org/10.1186/1471-2458-12-206

Fulton, J. E., Carlson, S. A., Ainsworth, B. E., Berrigan, D., Carlson, C., Dorn, J. M., . . . Wendel, A. (2016). Strategic priorities for physical activity surveillance in the United States. *Medicine & Science in Sports & Exercise, 48*(10), 2057-2069. doi: 10.1249/mss.0000000000000989

Grant, P. M., Dall, P. M., Mitchell, S. L., & Granat, M. H. (2008). Activity-monitor accuracy in measuring step number and cadence in community-dwelling older adults. *Journal of Aging and Physical Activity, 16*(2), 201-214.

Gray, C., Hunt, K., Mutrie, N., Anderson, A., Leishman, J., Dalgarno, L., & Wyke, S. (2013a). Football Fans in Training: the development and optimization of an intervention delivered through professional sports clubs to help men lose weight, become more active and adopt healthier eating habits. *BMC Public Health, 13*(1), 232.

Gray, C., Hunt, K., Mutrie, N., Anderson, A., Treweek, S., & Wyke, S. (2013b). Weight management for overweight and obese men delivered through professional football clubs: a pilot randomized trial. *International Journal of Behavioral Nutrition and Physical Activity, 10*(1), 121. doi: https://doi.org/10.1186/1479-5868-10-121

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1·9 million participants. *The Lancet Global Health, 6*(10), e1077-e1086. doi: https://doi.org/10.1016/S2214-109X(18)30357-7

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases, 40*(2), 171-178. doi: https://doi.org/10.1016/0021-9681(87)90069-5

Hoos, T., Espinoza, N., Marshall, S., & Arredondo, E. M. (2012). Validity of the global physical activity questionnaire (GPAQ) in adult Latinas. *Journal of physical activity & health, 9*(5), 698-705.

Hunt, K., Gray, C., Maclean, A., Smillie, S., Bunn, C., & Wyke, S. (2014a). Do weight management programmes delivered at professional football clubs attract and engage high risk men? A mixed-methods study. *BMC Public Health, 14*(1), 50. doi: https://doi.org/10.1186/1471-2458-14-50

Hunt, K., McCann, C., Gray, C. M., Mutrie, N., & Wyke, S. (2013). “You've got to walk before you run”: Positive evaluations of a walking program as part of a gender-sensitized, weight-management program delivered to men through professional football clubs. *Health Psychology, 32*(1), 57-65. doi: 10.1037/a0029537

Hunt, K., Wyke, S., Gray, C. M., Anderson, A. S., Brady, A., Bunn, C., . . . Treweek, S. (2014b). A gender-sensitised weight loss and healthy living programme for overweight and obese men delivered by Scottish Premier League football clubs (FFIT): A pragmatic randomised controlled trial. *The Lancet, 383*(9924), 1211-1221. doi: https://doi.org/10.1016/S0140-6736(13)62420-4

Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology, 53*(5), 459-468. doi: https://doi.org/10.1016/S0895-4356(99)00206-1

Kelly, M. P., & Barker, M. (2016). Why is changing health-related behaviour so difficult? *Public Health, 136*, 109-116. doi: http://dx.doi.org/10.1016/j.puhe.2016.03.030

Kelly, P., Fitzsimons, C., & Baker, G. (2016). 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, 13*(1), 1-10. doi: 10.1186/s12966-016-0351-4

Kowalski, K., Rhodes, R., Naylor, P. J., Tuokko, H., & MacDonald, S. (2012). Direct and indirect measurement of physical activity in older adults: A systematic review of the literature. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 1-21. doi: 10.1186/1479-5868-9-148

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863). doi: 10.3389/fpsyg.2013.00863

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *The Lancet, 380*(9838), 219-229. doi: https://doi.org/10.1016/S0140-6736(12)61031-9

Lee, P., Macfarlane, D., Lam, T., & Stewart, S. (2011). Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 8*(1), 115. doi: https://doi.org/10.1186/1479-5868-8-115

Lee, W. Y. H., Clark, B. K., Winkler, E., Eakin, E. G., & Reeves, M. M. (2015). Responsiveness to change of self-report and device-based physical activity measures in the Living Well With Diabetes Trial. *Journal of Physical Activity and Health, 12*(8), 1082-1087. doi: 10.1123/jpah.2013-0265

Liang, M. H., Fossel, A. H., & Larson, M. G. (1990). Comparisons of five health status instruments for orthopedic evaluation. *Medical Care, 28*(7), 632-642.

Lord, S., Chastin, S. F. M., McInnes, L., Little, L., Briggs, P., & Rochester, L. (2011). Exploring patterns of daily physical and sedentary behaviour in community-dwelling older adults. *Age and Ageing, 40*(2), 205-210. doi: 10.1093/ageing/afq166

Middel, B., & van Sonderen, E. (2002). Statistical significant change versus relevant or important change in (quasi) experimental design: Some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care, 2*, e15.

Middel, B., & van Sonderen, E. (2008). Erratum. *International Journal of Integrated Care, 8*, e72.

Montoye, A. H., Pfeiffer, K. A., Suton, D., & Trost, S. G. (2014). Evaluating the responsiveness of accelerometry to detect change in physical activity. *Measurement in Physical Education & Exercise Science, 18*(4), 273-285. doi: 10.1080/1091367X.2014.942454

Morelhão, P. K., Franco, M. R., Oliveira, C. B., Hisamatsu, T. M., Ferreira, P. H., Costa, L. O. P., . . . Pinto, R. Z. (2018). Physical activity and disability measures in chronic non-specific low back pain: A study of responsiveness. *Clinical Rehabilitation*, 0269215518787015. doi: 10.1177/0269215518787015

National Institute for Health and Clinical Excellence. (2006). Obesity: The prevention, identification, assessment and management of overweight and obesity in adults and children. London: NICE.

Nicaise, V., Crespo, N. C., & Marshall, S. (2014). Agreement between the IPAQ and accelerometer for detecting intervention-related changes in physical activity in a sample of Latina women. *Journal of Physical Activity and Health, 11*(4), 846-852. doi: 10.1123/jpah.2011-0412

Pedišić, Ž., & Bauman, A. (2015). Accelerometer-based measures in physical activity surveillance: Current practices and issues. *British Journal of Sports Medicine, 49*(4), 219. doi: 10.1136/bjsports-2013-093407

Prince, S., Adamo, K., Hamel, M., Hardt, J., 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*(1), 56. doi: https://doi.org/10.1186/1479-5868-5-56

Prous, M. J., Salvanés, F. R., & Ortells, L. C. (2008). Responsiveness of outcome measures. *Reumatología Clínica (English Edition), 4*(6), 240-247. doi: https://doi.org/10.1016/S2173-5743(08)70197-7

Reeves, M. M., Marshall, A. L., Owen, N., Winkler, E. A. H., & Eakin, E. G. (2010). Measuring physical activity change in broad-reach intervention trials. *Journal of Physical Activity and Health, 7*(2), 194-202. doi: 10.1123/jpah.7.2.194

Scottish Intercollegiate Guidelines Network. (2010). Management of obesity: a national clinical guideline. Edinburgh: Scottish Intercollegiate Guidelines Network.

Silfee, V. J., Haughton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C. N., Sreedhara, M., . . . Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive Medicine Reports, 11*, 74-80. doi: https://doi.org/10.1016/j.pmedr.2018.05.003

Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., . . . Ulrich, C. M. (2016). Accelerometry and physical activity questionnaires - A systematic review. *BMC Public Health, 16*(1), 1-10. doi: 10.1186/s12889-016-3172-0

Sloane, R., Snyder, D. C., Demark-Wahnefried, W., Lobach, D., & Kraus, W. E. (2009). Comparing the 7-day physical activity recall with a triaxial accelerometer for measuring time in exercise. *Medicine & Science in Sports & Exercise, 41*(6), 1334-1340. doi: 10.1249/MSS.0b013e3181984fa8

Slootmaker, S. M., Schuit, A. J., Chinapaw, M. J. M., Seidell, J. C., & van Mechelen, W. (2009). Disagreement in physical activity assessed by accelerometer and self-report in subgroups of age, gender, education and weight status. *The International Journal of Behavioral Nutrition and Physical Activity, 6*, 17-17. doi: 10.1186/1479-5868-6-17

Stratford, P. W., & Riddle, D. L. (2005). Assessing sensitivity to change: Choosing the appropriate change coefficient. *Health and Quality of Life Outcomes, 3*, 23-23. doi: 10.1186/1477-7525-3-23

Stucki, G., Liang, M. H., Fossel, A. H., & Katz, J. N. (1995). Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of Clinical Epidemiology, 48*(11), 1369-1378. doi: https://doi.org/10.1016/0895-4356(95)00054-2

Swartz, A. M., Rote, A. E., Cho, Y. I., Welch, W. A., & Strath, S. J. (2014). Responsiveness of motion sensors to detect change in sedentary and physical activity behaviour. *British Journal of Sports Medicine, 48*(13), 1043-1047. doi: 10.1136/bjsports-2014-093520

Thompson, D., Batterham, A. M., Markovitch, D., Dixon, N. C., Lund, A. J. S., & Walhin, J.-P. (2009). Confusion and conflict in assessing the physical activity status of middle-aged men. *PLoS One, 4*(2), e4337. doi: 10.1371/journal.pone.0004337

Troiano, R. P., Gabriel, K. K. P., Welk, G. J., Owen, N., & Sternfeld, B. (2012). Reported physical activity and sedentary behavior: Why do you ask? *Journal of Physical Activity and Health, 9*(s1), S68-S75. doi: doi:10.1123/jpah.9.s1.s68

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine, 48*(13), 1019-1023. doi: 10.1136/bjsports-2014-093546

Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in field-based research. *Medicine & Science in Sports & Exercise, 37*(11), S531-S543. doi: 10.1249/01.mss.0000185657.86065.98

Tudor-Locke, C., & Rowe, D., A. (2012). Using cadence to study free-living ambulatory behaviour. *Sports Medicine, 42*(5), 381-398. doi: 10.2165/11599170-000000000-00000

van Nassau, F., Chau, J. Y., Lakerveld, J., Bauman, A. E., & van der Ploeg, H. P. (2015). Validity and responsiveness of four measures of occupational sitting and standing. *International Journal of Behavioral Nutrition and Physical Activity, 12*(1), 144. doi: 10.1186/s12966-015-0306-1

Warburton, D. E. R., & Bredin, S. S. D. (2017). Health benefits of physical activity: A systematic review of current systematic reviews. *Current Opinion in Cardiology, 32*(5), 541-556. doi: 10.1097/hco.0000000000000437

Ward, D. S., Evenson, K. R., Vaughn, A., Rodgers, A. B., & Troiano, R. P. (2005). Accelerometer use in physical activity: Best practices and research recommendations. *Medicine & Science in Sports & Exercise, 37*(11), S582-S588. doi: 10.1249/01.mss.0000185292.71933.91

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity – a review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation, 17*(2), 127-139. doi: 10.1097/HJR.0b013e32832ed875

WHO. (2018). *Global action plan on physical activity 2018–2030: More active people for a healthier world.* World Health Organization. Geneva, Switzerland.

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 & Science in Sports & Exercise, 47*(10), 2129-2139. doi: 10.1249/MSS.0000000000000661

Winkler, E., Waters, L., Eakin, E., Fjeldsoe, B., Owen, N., & Reeves, M. (2013). Is measurement error altered by participation in a physical activity intervention? *Medicine & Science in Sports & Exercise, 45*(5), 1004-1011. doi: 10.1249/MSS.0b013e31827ccf7d

Wyke, S., Hunt, K., Gray, C., Fenwick, E., Bunn, C., Donnan, P., . . . Trew. (2015). Football Fans in Training (FFIT): A randomised controlled trial of a gender-sensitised weight loss and healthy living programme for men – end of study report. *Public Health Research, 3*(2). doi: 10.3310/phr03020

Table 1 Pre-programme characteristics of study participants (n=30)

Physical measures	**Mean (SD)**
Age (years)	45.9 (9.8)
Weight (kg)	111.8 (14.3)
BMI (kg/m^2)	35.9 (5.3)
Waist (cm)	118.5 (11.1)
BP Systolic (mmHG)	139.8 (15.4)
BP Diastolic (mmHG)	86.8 (7.7)
Socioeconomic status[a]	**% (n)**
1 (most deprived)	16.7 (5)
2	23.3 (7)
3	23.3 (7)
4	10 (3)
5 (least deprived)	26.7 (8)
Marital Status	**% (n)**
Single	3. (1)
Married	66.7 (20)
Separated	6.7 (2)
Living with someone	20 (6)
Divorced	3.3 (1)
Widowed	0 (0)

[a] Estimated using the Scottish Index of Multiple Deprivation based on home postcode (http://www.gov.scot/Topics/Statistics/SIMD).

Table 2 Device-based (activPAL3™) and self-report (IPAQ) physical activity measurements at T0 and T1 and changes between T0 and T1 (n=30)

	T0 Mean (SD)	Median (IQR)	T1 Mean (SD)	Median (IQR)	Change Mean (SD)	Change Median	Wilcoxon (*z*)	*p*	SRM	ES
activPAL Number of steps (steps/day)	8315.5 (3063.3)	8167.2 (5874.7-9741.4)	9834.4 (3855.9)	9016.5 (6772.2-11667.5)	1518.8 (2891.1)	848.3	-2.58	0.007[a]	0.53	0.47
activPAL Time stepping at a moderate cadence (min/day)	28.3 (18.8)	21.63 (12.7-43.5)	37.8 (27.3)	32.3 (15.6-46.6)	9.5 (21.6)	10.7	-1.90	0.022[a]	0.44	0.35
activPAL MET-minutes (min/day)	2040 (78)	2031 (1976-2084)	2076 (90)	2056 (2025-2147)	36 (72)	24.6	-2.61	0.006[a]	0.54	0.48
IPAQ Total PA minutes (min/day)	70.2 (78.4)	32.1 (23.8-112)	137.4 (86.3)	119.3 (62.7-181.1)	67.2 (113)	83	-3.56	<0.001[b]	0.59	0.65
IPAQ Total MET-minutes (min/day)	304.5 (339.8)	166.5 (99-414.6)	622.1 (457.7)	507.9 (275.5-880.5)	317.7 (588.7)	341.4	-3.59	<0.001[b]	0.54	0.66
IPAQ Walking MET-minutes (min/day)	134.7 (160.5)	75.4 (23.6-176.8)	254.4 (178.7)	229.9 (99-396)	119.7 (203.9)	154.4	-2.86	0.004[b]	0.59	0.52
IPAQ Moderate MET-minutes (min/day)	65.4 (117.5)	0 (0-68.6)	115 (138.2)	55.7 (12.9-205.7)	49.5 (129.1)	55.7	-2.53	0.011[b]	0.38	0.46

IPAQ Vigorous MET-minutes (min/day)	104.3 (170.2)	60 (0-137.1)	252.8 (348.1)	140 (0-291.4)	148.5 (388.6)	80	-2.58	0.010[b]	0.38	0.47

[a]Paired-samples t-test, [b]Wilcoxon signed-rank test, *p < 0.05. **p < 0.01. ES, effect size (non-parametric); SRM, standardised response mean.

Table 3 Responsiveness to change scores for participants demonstrating increased physical activity, no change or decreased physical activity according to device-based (activPAL3™) and self-report (IPAQ) measurements between T0 and T1 (n=30)

	% Increased PA (n)[a]	% Decreased PA (n)	% No change (n)	T0 and T1 mean change[b]	SD	GRI
activPAL Number of steps (steps/day)	70 (21)	30 (9)	0 (0)	2913.1	1235.10	2.36
activPAL Time stepping at a moderate cadence (min/day)	56.7 (17)	43.3 (13)	0 (0)	23.5	5.58	4.21
activPAL MET-minutes (min/day)	73.3 (22)	26.7 (8)	0 (0)	72.7	32.4	2.24
IPAQ Total PA minutes (min/day)	86.7 (26)	10% (3)	3.3% (1)	95.5	145.18	0.66
IPAQ Total MET-minutes (min/day)	86.7 (26)	13.3% (4)	0 (0)	457.1	717.16	0.64
IPAQ Walking MET-minutes (min/day)	73.3 (22)	16.7% (5)	10% (3)	204.7	137.67	1.49
IPAQ Moderate MET-minutes (min/day)	60 (18)	17% (5)	23% (7)	119.0	113.83	1.05
IPAQ Vigorous MET-minutes (min/day)	60 (18)	10% (3)	30% (9)	321.71	27.19	11.83

[a]Percentage and number of participants demonstrating increased PA according to each measure, [b]Mean change in PA score for participants identified as having increased PA from baseline to follow-up, [c]Standard deviation of change in PA score of participants indicating no change or identified as having decreased PA from baseline to follow-up; GRI, Guyatt responsiveness index.

Table 4 Correlation coefficients between device-based (activPAL3™) and self-report (IPAQ) physical activity measurements

T0 correlations (n=30)

	IPAQ				
activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	Walking MET-minutes (min/day)	Moderate MET-minutes (min/day)	Vigorous MET-minutes (min/day)
Number of steps (steps/day)	0.34	0.29	0.42*	0.07	0.14
Time stepping at a moderate cadence (min/day)	0.26	0.18	0.49**	-0.15	-0.00
MET-minutes (min/day)	0.31	0.27	0.37*	0.08	0.14

T1 correlations (n=30)

	IPAQ				
activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	Walking MET-minutes (min/day)	Moderate MET-minutes (min/day)	Vigorous MET-minutes (min/day)
Number of steps (steps/day)	0.36	0.30	0.34	0.14	0.15
Time stepping at a moderate cadence (min/day)	0.15	0.05	0.28	-0.10	-0.11
MET-minutes (min/day)	0.28	0.23	0.31	0.13	0.08

Spearman's rank-order correlations, *$p < 0.05$. **$p < 0.01$.

Table 5 Correlation coefficients between device-based (activPAL3™) and self-report (IPAQ) physical activity changes

T0-T1 correlations (n=30)

	IPAQ				
activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	Walking MET-minutes (min/day)	Moderate MET-minutes (min/day)	Vigorous MET-minutes (min/day)
Number of steps (steps/day)	0.16	0.12	0.35	-0.20	-0.22
Time stepping at a moderate cadence (min/day)	0.07	0.04	0.18	-0.30	-0.19
MET-minutes (min/day)	0.13	0.10	0.31	-0.25	-0.21

Spearman's rank-order correlations, *p < 0.05. **p < 0.01.