

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

**Craig Donnachie, University of Glasgow
Paul Kelly, University of Edinburgh
Nanette Mutrie, University of Edinburgh
Kate Hunt, University of Stirling
Jason Gill, University of Glasgow**

Accepted author manuscript version reprinted, by permission, from *Journal for the Measurement of Physical Behaviour*, 2020 <https://doi.org/10.1123/jmpb.2019-0018>. © Human Kinetics, Inc.

1 **Abstract**

2 The capacity of physical activity (PA) measures to detect *changes* in PA within interventions is crucial.
3 This is the first study to examine responsiveness of *activPAL3™* and the International Physical Activity
4 Questionnaire (IPAQ; Short Form) in detecting PA change during a 12 week group-based, men-only
5 weight management programme - Football Fans in Training (FFIT). Participants wore an *activPAL3™* and
6 completed the IPAQ pre- and post-programme (n=30). Relationships between change scores were
7 assessed by Spearman’s correlations. Mean or median changes in PA were measured using paired
8 samples t-tests and Wilcoxon signed-rank tests. Responsiveness to change was assessed utilising
9 Standardised Response Mean (SRM). Both device-based and self-report measures demonstrated
10 significant changes pre-post intervention, although these changes were not significantly correlated. The
11 SRM values for changes in *activPAL3™* metrics were: 0.54 (MET-mins/day); 0.53 (step counts/day); and
12 0.44 (MVPA/day), indicating a small to medium responsiveness to change. SRM values for changes in
13 IPAQ scores were: 0.59 (total PA mins/day); 0.54 (total MET-mins/day); 0.59 (walking MET-mins/day);
14 0.38 (vigorous MET-mins/day); and 0.38 (moderate MET-mins/day), revealing a small to medium
15 responsiveness to change. These findings reveal that two commonly used device-based and self-report
16 measures demonstrated responsiveness to changes in PA. While inclusion of both device-based and
17 self-report measures is desirable within interventions it is not always feasible. The results from this
18 study support that self-reported measures can detect PA change within behavioural interventions,
19 although may have a tendency to overestimate changes, compared with device-based measures.

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

21

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

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

35 To understand whether strategies are effective in changing PA behaviour it is vital that appropriate
36 measurement methods are incorporated within evaluations of behavioural interventions which aim to
37 assess the extent of PA behaviour change. However, PA is a complex and multi-faceted behaviour often
38 characterised across several domains (i.e. leisure, travel, housework/gardening, and occupation),
39 dimensions and determinants/correlates (Kelly, Fitzsimons, & Baker, 2016). Consequently, assessment
40 of PA offers considerable methodological options and challenges (Warren et al., 2010). Subjective (i.e.
41 self-reported) PA measures are commonly employed in population and intervention studies as they are
42 easy to use and cost less than objective (i.e. device-based) assessment. Wearable device-based
43 technologies, such as accelerometers, have become increasingly popular in recent years as PA
44 assessment tools that are not prone to recall bias, more valid and reliable compared with self-report

45 instruments and are often more practical compared with alternative more robust measures such as
46 doubly labelled water (Silfee et al., 2018). Despite these benefits, there are limitations to relying solely
47 on device-based forms of PA assessment (Pedišić & Bauman, 2015), particularly as outcomes in
48 behavioural interventions. For instance, they are often unable to detect some forms of activity, and
49 hence may underestimate overall PA levels (Silfee et al., 2018). They may also inadvertently influence
50 PA when used as surveillance or measurement tools (e.g. measurement reactivity) and might enhance
51 burden on participants (Baumann et al., 2018). Moreover, with the rise in the use of wearable device-
52 based measures in recent years, there is substantial heterogeneity regarding the number of PA metrics
53 being reported, limiting comparability between studies (Silfee et al., 2018).

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

60 Despite a growing number of intervention studies incorporating device-based and/or self-report
61 measures of PA, there is a lack of research examining the (comparative) responsiveness of these
62 measures to detect PA behaviour change over time as distinct PA constructs. For instance, there are
63 only a small number of studies that have explicitly examined responsiveness to change of device-based
64 and/or self-report measures in adults and children (e.g. Lee, Clark, Winkler, Eakin, & Reeves, 2015;
65 Montoye, Pfeiffer, Sutton, & Trost, 2014; Swartz, Rote, Cho, Welch, & Strath, 2014). The term
66 responsiveness (or sensitivity) is typically defined as an indicator of an instrument's sensitivity to change
67 as well as being a gauge of the magnitude of intervention-related change over time (Beaton,

68 Bombardier, Katz, & Wright, 2001; Middel & van Sonderen, 2002). Although the validity and reliability
69 of device-based and self-report PA instruments are often examined comprehensively (e.g. Lee,
70 Macfarlane, Lam, & Stewart, 2011), responsiveness is comparatively under investigated, particularly
71 within the context of behavioural interventions.

72 In order to understand whether PA interventions are effective in changing behaviour, it is vital to
73 understand whether measures employed to evaluate changes in behaviour within intervention studies
74 are capable of detecting changes in PA. In this study we aim to examine and compare the
75 responsiveness of both device-based (*activPAL3™*) and self-report (International Physical Activity
76 Questionnaire; IPAQ, Short Form) PA measures to detect changes in PA behaviour, using data collected
77 before and after participation in the Football Fans in Training (FFIT) programme, a weight management
78 and healthy lifestyle intervention for men classified as overweight or obese ($BMI > 28 \text{ kg/m}^2$) and aged 35-
79 65 years (see Gray et al., 2013a; Hunt et al., 2014b; Wyke et al., 2015).

80 **Methods**

81 *Participants and intervention setting*

82 Football Fans in Training (FFIT) is a 12 week gender-sensitised, group programme delivered free of
83 charge by trained community coaches to men at Professional Football clubs in Scotland. The
84 development of the FFIT programme is detailed elsewhere (Gray et al., 2013a). In brief, FFIT was
85 designed in line with evidence of what was known to be effective for weight loss (National Institute for
86 Health and Clinical Excellence, 2006; Scottish Intercollegiate Guidelines Network, 2010) and to work
87 with rather than against prevailing notions of masculinity, appealing to men in: context (professional
88 football clubs); content (e.g. information around the science of weight management presented simply
89 and branded materials, such as club T-shirts); and style of delivery (e.g. coaches encourage peer-

90 support, participative learning and positive ‘banter’ to support discussion of more sensitive issues)
91 (Wyke et al., 2015).

92 Funding was secured to undertake an evaluation of FFIT (a randomised controlled trial (RCT)
93 incorporating an embedded process evaluation and cost-effectiveness); at that time funding was
94 available for three deliveries of the programme in 13 professional football clubs (i.e. the 12 clubs in the
95 top league in Scotland - then the Scottish Premier League (SPL) - and the most recently demoted club
96 who had taken part in pilot deliveries in the previous season) in August-December 2011, February-April
97 2012 and August-December 2012. Men taking part in the August-December deliveries in 2011 and 2012
98 were participants in the FFIT RCT as outlined elsewhere (Hunt et al., 2014b; Wyke et al., 2015). During
99 the baseline assessment period, the FFIT research team recruited adequate numbers of participants to
100 fill *all* places then available on the three deliveries of FFIT (funded by the Football Pools and the Scottish
101 Government). After recruitment of the intervention and control arms of the RCT had been achieved, the
102 remaining 306 men were offered a place on ‘non-trial’ deliveries of FFIT in February-April, 2012. The
103 RCT of FFIT demonstrated significant mean between-group difference in weight loss of 4.94kg (CI 3.95,
104 5.94, $p < 0.0001$) at 12 months after baseline (primary outcome), and in self-reported PA (International
105 Physical Activity Questionnaire, Short Form), and other secondary outcomes, all in favour of the
106 intervention group (Hunt et al., 2014a; Wyke et al., 2015). No device-based measures of PA were taken
107 in men participating in the RCT.

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

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

114 *Procedure*

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

121 Prior to attending pre-programme measurement sessions, men from four clubs (n=94) were sent a letter
122 outlining research for a sub-study on objective PA assessment and inviting them to take part. This
123 provided adequate time to decide if they were willing to take part in the sub-study before attending the
124 pre-programme stadium-based measurement sessions. At T0 (week 0 of the FFIT programme),
125 participants from these clubs were asked if they had received the information letter, given an additional
126 copy of the study information to read and asked if they would be willing to wear an *activPAL3™* (PAL
127 Technologies Ltd., Glasgow, Scotland, UK) device for seven consecutive days (i.e. providing six full days
128 of activity monitoring) so that it could be retrieved when they attended for their first programme
129 session (week 1 of the FFIT programme) the following week. They were also asked if they would be
130 willing to wear the *activPAL3™* again between week 11 and week 12 of the programme (T1). Those who
131 agreed to wear the device at week 11, provided data for a further seven days after the devices were
132 collected at week 12, the final week of the programme.

133 During the pre-programme measurement session, participants gave written informed consent after they
134 were fully briefed on the purpose of the *activPAL3™*, and given a demonstration on how to remove and

135 re-affix the device. The *activPAL3*[™] was placed inside a waterproof and protective nitrile sleeve,
136 wrapped in a single layer of Hypafix[®] water resistant adhesive. Next, the device was affixed directly to
137 the skin of the participants' right leg with one sheet of Hypafix[®] adhesive (10cm X 13cm) by the first
138 author, following standardised protocols to protect privacy; participants were told the device need only
139 be removed to prevent the device being immersed in water (i.e. during swimming or bathing) but could
140 be worn while showering and sleeping. The men were each given additional Hypafix[®] strips to re-apply
141 the device should it need to be removed for any reason throughout the week. When the monitors were
142 removed and retrieved by the first author at each of the four clubs the following week, participants
143 were asked to complete the IPAQ (short form) to obtain concurrent self-reported PA, recalled over the
144 previous week. The same procedures were repeated again for FFIT programme weeks 11-12. Full
145 ethical approval was granted by the University of Glasgow, College of Social Sciences Research Ethics
146 Committee (CSS201020106).

147 *Measures*

148 The device-based outcome measures in the current study were measured by the *activPAL3*[™] device and
149 included: number of steps taken per day; minutes of moderate to vigorous intensity PA; and energy
150 expenditure. The *activPAL3*[™] is a triaxial accelerometer/inclinometer which incorporates proprietary
151 technology (Intelligent Activity Classification[™]) to measure three types of free-living activity: time spent
152 sitting/lying; standing; and stepping. The *activPAL3*[™] quantifies the amount of steps performed, the
153 intensity of steps taken (cadence) and estimates of energy expenditure (Lord et al., 2011). It is a small
154 (35 X 53 X 7mm), lightweight device (15g), worn discreetly on the middle of the thigh between the hip
155 and the knee, above the quadriceps muscle. The device has a battery life of around nine days, and thus
156 can be worn continuously for 24 hour monitoring. The data are recorded in 15-second epochs and the
157 output downloaded onto a Personal Computer (PC) via a USB interface. Previous studies have shown

158 the activity and posture functions of the *activPAL*[™] to be valid compared to direct observation and
159 acceptable to participants in community-based research (e.g. Grant, Dall, Mitchell, & Granat, 2008).

160 Prior to assessment, each of the *activPAL3*[™] monitors was fully charged and initialised to record six
161 consecutive days. On retrieval of the monitors, data were uploaded to a PC using *activPAL*[™] proprietary
162 software (PALtechnologies v5.9.1.1). Microsoft Office Excel was used for subsequent data processing
163 and management. Custom software (HSC analysis software v2.19, Philippa Dall and Malcolm Granat,
164 Glasgow Caledonian University) was used to extract information on individual participants' PA intensity
165 using the *activPAL3*[™] time-stamped 'event' data files. Based on the conclusions of a systematic review
166 (Tudor-Locke & Rowe, 2012), in the current study, time spent stepping at a cadence of at least 100
167 steps/minute was deemed indicative of moderate intensity PA. Daily energy expenditure is classified by
168 the *activPAL3*[™] software as metabolic equivalent (MET-hours) and expressed in this study as MET-
169 minutes per day. Best practice guidelines for accelerometer use in PA measurement suggest that for
170 adults a minimum of 3-5 days of monitoring is necessary to quantify free-living PA (Troost, McIver, &
171 Pate, 2005; Ward, Evenson, Vaughn, Rodgers, & Troiano, 2005). Data files were inspected visually and
172 individual cases excluded if less than three days of wear time were evident. Days with <500 steps were
173 removed, consistent with previous studies which have incorporated similar cut-offs to classify non-wear
174 days (Edwardson et al., 2017). Wake/sleep times were included as recorded by the *activPAL3*[™],
175 enabling capture of daily 24 hour activity (i.e. midnight to midnight). Self-report logs/diaries (e.g. to
176 record sleep, wake or removal time) were not incorporated in the current study to reduce overall
177 participant burden.

178 Self-reported PA outcomes were measured using the International Physical Activity Questionnaire
179 (IPAQ, Short Form) (Craig et al., 2003), which included total PA minutes and total work done in PA per
180 week (MET-minutes). The IPAQ is a well-established measure designed principally as a gold standard for

181 population surveillance of PA among adults (18-65 years) (Craig et al., 2003). According to the IPAQ
182 scoring guidelines, data are reported as total metabolic equivalent of task (in MET-minutes) per week.
183 Calculation of the total score involves summation of the duration (i.e. minutes) and frequency (i.e. days)
184 of walking, moderate-intensity and vigorous-intensity activities, recalled over the past seven days. MET-
185 minute scores are quantified by multiplying the MET score of an activity by the minutes performed. All
186 reported activity (i.e. walking, moderate and vigorous activity) exceeding 'three hours' (or 180 minutes)
187 were truncated to allow a maximum of 21 hours of activity per week for each category to minimise over
188 reporting, consistent with the IPAQ scoring protocol ([https://sites.google.com/site/theipaq/scoring-
189 protocol](https://sites.google.com/site/theipaq/scoring-protocol)). Changes in IPAQ and *activPAL3™* metrics are expressed in total minutes 'per day' to enable
190 comparison between both measures. All IPAQ total scores ('per week') were divided by 7 for daily PA
191 estimates.

192 *Statistical analysis*

193 Descriptive statistics are presented as means (standard deviation, SD), medians (interquartile range,
194 IQR) and percentages (number). Exploratory analysis revealed that the majority of device-based PA
195 metrics were approximately normally distributed, whereas the self-reported data were positively
196 skewed. Paired samples t-tests or Wilcoxon signed-rank tests (for data that violated assumptions of
197 normality) were used to examine differences pre- and post-intervention. Spearman's rank-order
198 correlation coefficients (ρ) were used to assess relationships between change scores for device-based
199 (*activPAL3™*) and self-report (IPAQ) instruments, interpreted as weak (<0.3), low (0.30–0.49), moderate
200 (0.50–0.69), strong (0.70–0.89) or very strong (≥ 0.90).

201 Responsiveness to change in device-based and self-report PA scores between T0 and T1 was assessed
202 using the Standardised Response Mean (SRM) or Cohen's *d*_z (Cohen, 1977; Dankel & Loenneke, 2018;
203 Lakens, 2013; Liang, Fossel, & Larson, 1990), a type of effect size that has been used in previous studies

204 to assess responsiveness to change of device-based and self-report PA instruments in adults and
205 children (e.g. Almeida et al., 2017; Clevenger et al., 2018; Swartz et al., 2014). SRM is calculated for each
206 measure by dividing the absolute mean change score by the standard deviation of differences between
207 the paired measurements and can be interpreted in line with Cohen's *d* as trivial, small, moderate or
208 large (<0.20, ≥0.20 to <0.50, ≥0.50 to <0.80, and ≥0.80, respectively) (Cohen, 1977; Husted, Cook,
209 Farewell, & Gladman, 2000; Stratford & Riddle, 2005). In addition to SRM values, non-parametric effect
210 size (ES) values were calculated for each of the device-based and self-reported PA metrics using
211 Wilcoxon's statistic and related Z-score divided by the square root of *n* (z/\sqrt{n}), interpreted as small (*r* =
212 <0.3), medium (*r* = ≥0.3 to <0.5) and large (*r* = ≥0.5) (Cohen, 1977; Field, 2009).

213 The Guyatt Responsiveness Index (GRI) is an alternative measure of responsiveness based on the
214 variability of changes among stable participants (Guyatt, Walter, & Norman, 1987; Husted et al., 2000).
215 When utilising this responsiveness statistic, participants are preferably assessed on several occasions to
216 ascertain the level of variability across a 'stable' time period, ideally before taking part in an
217 intervention, to detect minimally clinically important change exceeding any spurious changes in
218 measurement which may occur over time (Guyatt et al., 1987). However, where only two observations
219 are available (i.e. baseline and post-intervention), the GRI is calculated as the mean score of participants
220 identified as improved, divided by the standard deviation of the change in participants identified as
221 stable or showing no improvement pre- to post-intervention. In this analysis, the mean change of
222 participants identified as having increased PA between T0 and T1, as indicated by each of the device-
223 based and self-report PA metrics, were incorporated as the numerator, whereas the standard deviation
224 of the change in participants identified as unchanged or having decreased PA were included as the
225 denominator. Similarly to SRM, GRI values of 0.20, 0.50 and 0.80 or greater have been used to delineate
226 low, moderate and high responsiveness, respectively (Husted et al., 2000). However, it is important to
227 note that the GRI method is anticipated to yield higher coefficients than the SRM or other ES values as a

228 consequence of the removal of mean change values of unchanged participants (Stucki, Liang, Fossel, &
229 Katz, 1995). Statistical analyses were undertaken using IBM SPSS 21.0 (Armonk, NY, USA). All tests were
230 two-tailed with an alpha p-value of $p < 0.05$ to assess statistical significance.

231 **Results**

232 *Demographic characteristics*

233 Descriptive characteristics of participants pre-programme are displayed in Table 1. The mean age of
234 participants was 45.9 years ($SD = 9.8$). Mean body weight was 111.8kg ($SD = 14.3$), mean BMI was 35.9
235 kg/m^2 ($SD = 5.3$) and mean waist circumference was 118.5cm ($SD = 11.1$), thus comparable with clinical
236 characteristics of men taking part in other research deliveries of FFIT (Gray et al., 2013b; Hunt et al.,
237 2014a). Participants in this study were from across the socioeconomic spectrum, consistent with
238 previous research demonstrating that FFIT attracted men from a range of socioeconomic backgrounds
239 (Hunt et al., 2014b).

240 *Changes in device-based and self-reported physical activity*

241 Changes in device-based and self-reported PA between T0 and T1 are presented in Table 2. Data are
242 presented for $n=30$ participants with concurrent PA data (i.e. device-based and self-report assessments)
243 at both time points. Paired samples t-tests confirmed significant increases in *activPAL3*TM-assessed
244 number of average daily 'steps' from 8315.5 ($SD = 3063.3$) at T0 to 9834.4 ($SD = 3855.9$) at T1 with an
245 increase of 1518.8 steps, ($t(29) = -2.9, p = 0.007$), time spent stepping at least at a moderate cadence
246 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
247 minutes/day ($t(29) = -2.4, p = 0.022$) and increased daily MET-minutes from 2040 ($SD = 78$) at T0 to 2076
248 ($SD = 90$) at T1 with an increase of 36 MET-minutes/day, ($t(29) = -2.9, p = 0.006$). Wilcoxon signed rank
249 tests showed significant increases in self-reported PA (IPAQ) at T1 from T0 for total PA minutes ($Z = -$

250 3.56, $p = <0.001$, median difference = 83 minutes/day), total MET-minutes ($Z = -3.59$, $p = <0.001$, median
251 difference = 341 MET-minutes/day), walking MET-minutes ($Z = -2.86$, $p = 0.004$, median difference = 154
252 MET-minutes/day), moderate MET-minutes ($Z = -2.53$, $p = 0.011$, median difference = 56 MET-
253 minutes/day) and vigorous MET-minutes ($Z = -2.58$, $p = 0.010$, median difference = 80 MET-
254 minutes/day).

255 The SRM and non-parametric effect size (ES) values for changes in device-based and self-reported PA
256 between T0 and T1 are also displayed in Table 2. The SRM values for device-based (*activPAL3™*) time
257 spent active stepping at least at a moderate cadence, average steps per day and daily MET-minutes
258 were $d = 0.44$, $d = 0.53$ and $d = 0.54$, respectively, demonstrating a small to moderate responsiveness to
259 change. The SRM values for total self-reported PA minutes/day, MET-minutes/day, walking MET-
260 minutes/day, moderate MET-minutes/day and vigorous MET-minutes/day were, $d = 0.59$, $d = 0.54$, $d =$
261 0.59 , $d = 0.38$ and $d = 0.38$, respectively, revealing a small to moderate responsiveness to change
262 between T0 and T1. The non-parametric ES values for changes in time spent active at least at a
263 moderate stepping cadence, average steps per day and daily MET-minutes were $r = 0.35$, $r = 0.47$ and $r =$
264 0.48 , respectively, thus indicating a moderate effect size. The non-parametric ES values for changes in
265 self-reported (IPAQ) total minutes/day, MET-minutes/day, walking MET-minutes/day, moderate MET-
266 minutes/day and vigorous MET-minutes/day were $r = 0.65$, $r = 0.66$, $r = 0.52$, $r = 0.46$ and $r = 0.47$
267 respectively, indicating a moderate to large effect size.

268 The GRI values for changes in device-based and self-reported PA between T0 and T1 are depicted in
269 Table 3. The GRI values for device-assessed daily MET-minutes, average steps per day and time spent
270 active stepping at least at a moderate cadence (GRI = 2.24, GRI = 2.36, and GRI = 4.21, respectively)
271 showed a large responsiveness to change. Self-reported total PA minutes/day and total MET-
272 minutes/day were GRI = 0.66 and GRI = 0.64, respectively, demonstrating a moderate responsiveness to

273 change as a consequence of higher variability in changed total self-reported PA among participants.
274 However, IPAQ sub-domains of walking, moderate and vigorous MET-minutes/day were GRI = 1.49, GRI
275 = 1.05 and GRI = 11.83, respectively, revealing a large responsiveness to change between T0 and T1.

276 *Comparison between device-based and self-reported physical activity*

277 The Spearman's correlation coefficients between device-based and self-reported activity scores at both
278 T0 and T1 are displayed in Table 4. Generally, the highest correlations among device-based and self-
279 report PA measures were observed at T0. The correlation coefficients between *activPAL3™*-assessed PA
280 (number of steps, time spent stepping at least at a moderate intensity and total MET-minutes) and one
281 of the five IPAQ metrics (walking MET-minutes), were positive but low ($\rho = 0.42$, $\rho = 0.49$ and $\rho =$
282 0.37 , respectively). The correlations between each of the *activPAL3™* metrics and IPAQ assessed total
283 PA minutes, total MET-minutes, moderate MET-minutes and vigorous MET-minutes/day were all non-
284 significant. All of the correlation coefficients between device-based and self-report PA measures at T1
285 were not statistically significant, ranging from low to weak ($\rho = 0.36$ to -0.11).

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

289 **Discussion**

290 The capacity of self-report and device-based PA instruments to detect change in PA within intervention
291 settings is crucial to determining which interventions work. To our knowledge, this is the only study that
292 has compared responsiveness of both IPAQ (i.e. self-report) and *activPAL3™* (i.e. device-based)
293 measures across a number of outcome metrics to assess changes in PA over time within the context of a
294 behavioural intervention. This is also the first study to examine changes in both device-based and self-

295 reported PA within research deliveries of the FFIT programme, and extends previous research
296 demonstrating significant increases in self-reported PA (i.e. IPAQ, Short Form) during pilot (Gray et al.,
297 2013b) and full trial (Hunt et al., 2014b) deliveries of FFIT.

298 In this study, changes in all device-based and self-reported PA metrics were statistically significant.
299 According to device-based assessment (*activPAL3™*) taking part in the 12 week FFIT programme resulted
300 in an increase in an average of 1519 steps; 9.5 minutes spent stepping at least a moderate stepping
301 intensity (i.e. ≥ 100 steps/minute); and an extra 36 MET-minutes per day. According to the self-reported
302 PA measure (IPAQ), taking part in FFIT increased total PA by 83 minutes and 341 MET-minutes per day.
303 IPAQ sub-domains of walking, moderate and vigorous intensity activity also showed an increase of 154,
304 56, and 80 MET-minutes per day, respectively. The most salient finding from the current study is that
305 we observed comparable responsiveness to change for both device-based and self-report instruments.
306 SRM values for *activPAL3™* were greatest when measuring change in average MET-minutes (0.54) and
307 steps per day (0.53), whereas IPAQ SRM values were highest when assessing change in total PA MET-
308 minutes (0.54), walking MET-minutes (0.59) and total PA minutes (0.59), classified as moderate (i.e. SRM
309 values ≥ 0.50). The SRM values for IPAQ sub-domains of moderate and vigorous PA intensity (both SRMs
310 0.38) and *activPAL3™* assessed MVPA (time spent active at least a moderate intensity) (0.44), are
311 considered small (i.e. SRM values < 0.50). Similar trends were found for both non-parametric ES and GRI
312 responsiveness values. Thus, our findings indicate that despite uncorrelated changes, both instruments
313 were able to detect a comparable magnitude of change in PA. In contrast with previous research (Lee et
314 al, 2015), self-reported PA demonstrated slightly greater responsiveness compared with device-based
315 measures for total PA within the context of a 12 week, men-only behavioural intervention. However,
316 total self-reported PA (i.e. IPAQ total PA minutes and total MET-minutes) demonstrated lower GRI
317 values (GRI = 0.66 and GRI = 0.64, respectively) compared to each of the three device-assessed PA
318 metrics: *activPAL3™* assessed MET-minutes (GRI = 2.24); number of steps (GRI = 2.36); and time

319 stepping at a moderate cadence (GRI = 4.21), thus demonstrating lower variability in changed device-
320 based PA pre- to post-intervention. The findings suggest that both IPAQ and *activPAL3™* measures
321 should be responsive to change when evaluating PA change in future intervention settings.
322 Nonetheless, due to the substantial differences in self-reported PA scores compared with device-based
323 assessment, caution is warranted when interpreting intervention change based solely on self-reported
324 PA, as may overestimate change consistent with other research (Winkler et al., 2013).

325 There are a limited number of studies that have investigated responsiveness to change of both device-
326 based and self-report PA measures in adults. A recent study investigated responsiveness to change of
327 self-reported (Baecke Habitual Physical Activity Questionnaire) and device-based (ActiGraph GT3X-BT)
328 PA measures in patients with chronic low back pain receiving physical therapy (Morelhão et al., 2018).
329 The authors concluded that none of the PA measures were able to detect changes in PA over time,
330 according to SRM values (<0.20). Similarly, Almeida et al (2017) examined the responsiveness of self-
331 report (Community Health Activities Model Program for Older Adults Questionnaire) and two distinct
332 device-based (Actigraph GT3X; Sensewear Armband) measures in detecting changes in PA in older adults
333 with osteoarthritis during a rehabilitation programme following knee replacement surgery. The findings
334 revealed that each PA measure exhibited low responsiveness to change (i.e. in light, moderate and
335 vigorous intensity PA) as indicated by SRM values (<0.30). Nicaise and colleagues examined the
336 sensitivity of the IPAQ (Long Form) for detecting changes in PA compared with device-based (Actigraph
337 7164) assessment among Spanish-speaking Latina women during a 12 week pedometer-based
338 intervention (Nicaise, Crespo, & Marshall, 2014). In this study, both IPAQ ($r = 0.27$) and device-based (r
339 $= 0.40$) measures detected intervention-related changes in moderate intensity PA, indicating a small and
340 moderate effect size of change. Consistent with our study findings, changes in self-report and device-
341 based PA metrics were not correlated at 12 weeks.

342 Lee and colleagues reported significant changes in total PA minutes/week for two distinct self-report
343 (Active Australia Survey; United States National Health Interview Survey) and device-based PA measures
344 (Actigraph GT1M), longitudinally within the context of a weight loss intervention that were small in
345 magnitude, although device-based PA was classed as slightly more responsive (Lee et al., 2015).

346 Research conducted by the same group of authors investigated responsiveness to changes in PA using
347 three unique self-report instruments (Community Health Activities Model Program for Older Adults
348 Questionnaire; Active Australia Survey; United States National Health Interview Survey) in adults
349 following a four month behavioural intervention, demonstrating a small responsiveness to change
350 (Reeves, Marshall, Owen, Winkler, & Eakin, 2010). The findings observed in the current study are similar
351 to other research comparing responsiveness of device-based activity measures in adult populations
352 (Swartz et al., 2014; van Nassau, Chau, Lakerveld, Bauman, & van der Ploeg, 2015). For instance, Swartz
353 et al (2014) examined responsiveness to change in two different device-based PA measures (Actigraph
354 GT3X; *activPAL™*) in sedentary adults during a behavioural intervention to reduce sitting time. They
355 observed comparable SRM (0.44) values post-intervention for changes in *activPAL™* assessed PA
356 (average daily steps) indicating a small responsiveness to change.

357 As noted by Lee et al (2015), the majority of studies have focused on assessing the validity of PA
358 measures to examine behaviour change within interventions over time, usually relying on correlations
359 between changes in self-report and device-based measures (e.g. Hoos, Espinoza, Marshall, & Arredondo,
360 2012; Nicaise et al., 2014; Sloane, Snyder, Demark-Wahnefried, Lobach, & Kraus, 2009). However,
361 research findings have indicated greater disagreement between device-based and self-reported PA at
362 increased activity levels (e.g. Slootmaker, Schuit, Chinapaw, Seidell, & van Mechelen, 2009), hence
363 agreement between instruments may be attenuated by intervention effects (Lee et al., 2015). Winkler
364 et al (2013) observed that agreement between self-report and device-based measures deteriorated as
365 levels of PA increased during a behavioural PA intervention, particularly among intervention group

366 participants compared to controls. The authors reported that intervention effects were greater when
367 PA was assessed by self-report compared with device-based measures, despite both instruments
368 yielding statistically significant differences. They suggest PA interventions might appear more effective
369 when relying exclusively on self-report.

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

384 In contrast, device-based measures may fail to accurately recognise certain forms of activity (e.g.
385 swimming or resistance training) and therefore underestimate overall intervention effects. However,
386 during the FFIT programme, participants were encouraged to increase their activity levels predominantly
387 by increasing steps during the graduated walking component of the programme. It is therefore unlikely
388 that many of the participants in this study would have been performing other forms of activity during

389 the intervention, such as swimming or strength-based exercises, thus the magnitude of error was likely
390 small. Devices can be lost or malfunction, and only some participants may be willing to wear them.
391 Further, data processing requires subjective decisions about thresholds and cut-offs that are much
392 debated (Wijndaele et al., 2015). It has been advocated that due to the complexity in measuring PA, no
393 single methodology is able to sufficiently capture all PA domains and subcomponents (Warren et al.,
394 2010). We do not argue one or other method should be used; combining different methods of PA
395 assessment may provide a more comprehensive reflection of individuals' amount of activity and its
396 context, offering greater insights regarding evidence of behaviour change and efficacy of behavioural
397 interventions targeting this complex behaviour. However, this study incorporated the IPAQ, Short Form
398 which does not measure contextual information in the same way as the IPAQ, Long Form (e.g. leisure,
399 transportation, housework/gardening, and occupation-related activity). Future studies investigating
400 responsiveness of the IPAQ (Long Form) to changes in PA compared with device-based measures would
401 be advantageous.

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

411 *Strengths and limitations*

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

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

436 self-report and device-based PA methods. Lastly, the findings are specific to adult men participating in a
437 weight management and healthy lifestyle programme in Scotland (UK) and generalisability to wider
438 population groups may be limited.

439 **Conclusion**

440 In this study, two commonly used device-based (*activPAL3™*) and self-report (i.e. IPAQ, Short Form) PA
441 measures were found to be responsive to behaviour change in men following participation in a 12 week
442 weight loss and healthy living programme (FFIT), although there were non-significant correlations
443 between these change scores. The magnitude of responsiveness to change was marginally higher for
444 self-reported PA according to SRM values. While inclusion of both device-based and self-report
445 measures is desirable, it is not always feasible, hence these findings provide support for the utility of
446 self-reported PA instruments within the context of behavioural interventions promoting increased PA,
447 although they may overestimate PA changes, relative to device-based measures.

448

449

450

451

452

453

454

455

- 457 Adams, S. A., Matthews, C. E., Ebbeling, C. B., Moore, C. G., Cunningham, J. E., Fulton, J., & Hebert, J. R.
 458 (2005). The effect of social desirability and social approval on self-reports of physical activity.
 459 *American Journal of Epidemiology*, 161(4), 389-398. doi: 10.1093/aje/kwi054
- 460 Almeida, G. J., Terhorst, L., Irrgang, J., J., Fitzgerald, G. K., Jakicic, J. M., & Piva, S. R. (2017).
 461 Responsiveness of physical activity measures following exercise programs after total knee
 462 arthroplasty. *Journal of Exercise, Sports & Orthopedics*, 4(3), 1-8. doi:
 463 <http://dx.doi.org/10.15226/2374-6904/4/3/00164>
- 464 Baker, G., Mutrie, N., & Lowry, R. (2008). Using pedometers as motivational tools: Are goals set in steps
 465 more effective than goals set in minutes for increasing walking? *International Journal of Health
 466 Promotion and Education*, 46(1), 21-26. doi: 10.1080/14635240.2008.10708123
- 467 Baumann, S., Gross, S., Voigt, L., Ullrich, A., Weymar, F., Schwaneberg, T., . . . Ulbricht, S. (2018). Pitfalls
 468 in accelerometer-based measurement of physical activity: The presence of reactivity in an adult
 469 population. *Scandinavian journal of medicine & science in sports*, 28(3), 1056-1063. doi:
 470 10.1111/sms.12977
- 471 Beaton, D. E., Bombardier, C., Katz, J. N., & Wright, J. G. (2001). A taxonomy for responsiveness. *Journal
 472 of Clinical Epidemiology*, 54(12), 1204-1217. doi: [http://dx.doi.org/10.1016/S0895-
 473 4356\(01\)00407-3](http://dx.doi.org/10.1016/S0895-4356(01)00407-3)
- 474 Clevenger, K. A., Moore, R. W., Sutton, D., Montoye, A. H. K., Trost, S. G., & Pfeiffer, K. A. (2018).
 475 Accelerometer responsiveness to change between structured and unstructured physical activity
 476 in children and adolescents. *Measurement in Physical Education and Exercise Science*, 22(3),
 477 224-230. doi: 10.1080/1091367X.2017.1419956
- 478 Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.
- 479 Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Oja, P. (2003).
 480 International physical activity questionnaire: 12-Country reliability and validity. *Medicine &
 481 Science in Sports & Exercise*, 35(8), 1381-1395. doi: 10.1249/01.MSS.0000078924.61453.FB
- 482 Dankel, S. J., & Loenneke, J. P. (2018). Effect sizes for paired data should use the change score variability
 483 rather than the pre-test variability. *Journal of Strength and Conditioning Research*, 00(00), 1-6.
 484 doi: 10.1519/jsc.0000000000002946
- 485 Donnachie, C., Wyke, S., & Hunt, K. (2018). Men's reactions to receiving objective feedback on their
 486 weight, BMI and other health risk indicators. *BMC Public Health*, 18(1), 291. doi:
 487 10.1186/s12889-018-5179-1
- 488 Donnachie, C., Wyke, S., Mutrie, N., & Hunt, K. (2017). 'It's like a personal motivator that you carried
 489 around wi' you': utilising self-determination theory to understand men's experiences of using
 490 pedometers to increase physical activity in a weight management programme. *International
 491 Journal of Behavioral Nutrition and Physical Activity*, 14(1), 61. doi: 10.1186/s12966-017-0505-z
- 492 Edwardson, C. L., Winkler, E. A. H., Bodicoat, D. H., Yates, T., Davies, M. J., Dunstan, D. W., & Healy, G. N.
 493 (2017). Considerations when using the activPAL monitor in field-based research with adult
 494 populations. *Journal of Sport and Health Science*, 6(2), 162-178. doi:
 495 <https://doi.org/10.1016/j.jshs.2016.02.002>
- 496 Field, A. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)* (3rd ed.). London,
 497 England: SAGE.
- 498 Fitzsimons, C., Baker, G., Gray, S., Nimmo, M., & Mutrie, N. (2012). Does physical activity counselling
 499 enhance the effects of a pedometer-based intervention over the long-term: 12-Month findings
 500 from the Walking for Wellbeing in the west study. *BMC Public Health*, 12(1), 206. doi:
 501 <https://doi.org/10.1186/1471-2458-12-206>

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

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

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

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

516 Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical
517 activity from 2001 to 2016: A pooled analysis of 358 population-based surveys with 1.9 million
518 participants. *The Lancet Global Health*, 6(10), e1077-e1086. doi: [https://doi.org/10.1016/S2214-](https://doi.org/10.1016/S2214-109X(18)30357-7)
519 [109X\(18\)30357-7](https://doi.org/10.1016/S2214-109X(18)30357-7)

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

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

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

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

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

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

539 Kelly, M. P., & Barker, M. (2016). Why is changing health-related behaviour so difficult? *Public Health*,
540 136, 109-116. doi: <http://dx.doi.org/10.1016/j.puhe.2016.03.030>

541 Kelly, P., Fitzsimons, C., & Baker, G. (2016). Should we reframe how we think about physical activity and
542 sedentary behaviour measurement? Validity and reliability reconsidered. *International Journal*
543 *of Behavioral Nutrition and Physical Activity*, 13(1), 1-10. doi: 10.1186/s12966-016-0351-4

544 Kowalski, K., Rhodes, R., Naylor, P. J., Tuokko, H., & MacDonald, S. (2012). Direct and indirect
545 measurement of physical activity in older adults: A systematic review of the literature.
546 *International Journal of Behavioral Nutrition and Physical Activity*, 9(1), 1-21. doi: 10.1186/1479-
547 5868-9-148

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

550 Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Effect of physical
551 inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and
552 life expectancy. *The Lancet*, *380*(9838), 219-229. doi: [https://doi.org/10.1016/S0140-](https://doi.org/10.1016/S0140-6736(12)61031-9)
553 [6736\(12\)61031-9](https://doi.org/10.1016/S0140-6736(12)61031-9)

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

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

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

562 Lord, S., Chastin, S. F. M., McInnes, L., Little, L., Briggs, P., & Rochester, L. (2011). Exploring patterns of
563 daily physical and sedentary behaviour in community-dwelling older adults. *Age and Ageing*,
564 *40*(2), 205-210. doi: [10.1093/ageing/afq166](https://doi.org/10.1093/ageing/afq166)

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

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

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

573 Morelhão, P. K., Franco, M. R., Oliveira, C. B., Hisamatsu, T. M., Ferreira, P. H., Costa, L. O. P., . . . Pinto,
574 R. Z. (2018). Physical activity and disability measures in chronic non-specific low back pain: A
575 study of responsiveness. *Clinical Rehabilitation*, *0269215518787015*. doi:
576 [10.1177/0269215518787015](https://doi.org/10.1177/0269215518787015)

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

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

582 Pedišić, Ž., & Bauman, A. (2015). Accelerometer-based measures in physical activity surveillance:
583 Current practices and issues. *British Journal of Sports Medicine*, *49*(4), 219. doi:
584 [10.1136/bjsports-2013-093407](https://doi.org/10.1136/bjsports-2013-093407)

585 Prince, S., Adamo, K., Hamel, M., Hardt, J., Gorber, S., & Tremblay, M. (2008). A comparison of direct
586 versus self-report measures for assessing physical activity in adults: A systematic review.
587 *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 56. doi:
588 <https://doi.org/10.1186/1479-5868-5-56>

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

591 Reeves, M. M., Marshall, A. L., Owen, N., Winkler, E. A. H., & Eakin, E. G. (2010). Measuring physical
592 activity change in broad-reach intervention trials. *Journal of Physical Activity and Health*, *7*(2),
593 194-202. doi: [10.1123/jpah.7.2.194](https://doi.org/10.1123/jpah.7.2.194)

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

596 Silfee, V. J., Houghton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C. N., Sreedhara, M., . . .
597 Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle

598 interventions among adults: A systematic review. *Preventive Medicine Reports*, 11, 74-80. doi:
599 <https://doi.org/10.1016/j.pmedr.2018.05.003>

600 Skender, S., Ose, J., Chang-Claude, J., Paskow, M., Brühmann, B., Siegel, E. M., . . . Ulrich, C. M. (2016).
601 Accelerometry and physical activity questionnaires - A systematic review. *BMC Public Health*,
602 16(1), 1-10. doi: 10.1186/s12889-016-3172-0

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

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

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

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

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

618 Thompson, D., Batterham, A. M., Markovitch, D., Dixon, N. C., Lund, A. J. S., & Walhin, J.-P. (2009).
619 Confusion and conflict in assessing the physical activity status of middle-aged men. *PLoS One*,
620 4(2), e4337. doi: 10.1371/journal.pone.0004337

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

624 Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for
625 physical activity research. *British Journal of Sports Medicine*, 48(13), 1019-1023. doi:
626 10.1136/bjsports-2014-093546

627 Trost, S. G., McIver, K. L., & Pate, R. R. (2005). Conducting accelerometer-based activity assessments in
628 field-based research. *Medicine & Science in Sports & Exercise*, 37(11), S531-S543. doi:
629 10.1249/01.mss.0000185657.86065.98

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

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

635 Warburton, D. E. R., & Bredin, S. S. D. (2017). Health benefits of physical activity: A systematic review of
636 current systematic reviews. *Current Opinion in Cardiology*, 32(5), 541-556. doi:
637 10.1097/hco.0000000000000437

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

641 Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of
642 physical activity – a review of methodologies with reference to epidemiological research: A
643 report of the exercise physiology section of the European Association of Cardiovascular
644 Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation*,
645 17(2), 127-139. doi: 10.1097/HJR.0b013e32832ed875

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

648 Wijndaele, K., Westgate, K., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F. M., . . . Healy, G. N.
649 (2015). Utilization and harmonization of adult accelerometry data: Review and expert
650 consensus. *Medicine & Science in Sports & Exercise*, 47(10), 2129-2139. doi:
651 10.1249/MSS.0000000000000661

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

655 Wyke, S., Hunt, K., Gray, C., Fenwick, E., Bunn, C., Donnan, P., . . . Trew. (2015). Football Fans in Training
656 (FFIT): A randomised controlled trial of a gender-sensitised weight loss and healthy living
657 programme for men – end of study report. *Public Health Research*, 3(2). doi: 10.3310/phr03020
658

659

660

661

662

663

664

665

666

667

668

669

670

671 **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)

672 ^a Estimated using the Scottish Index of Multiple Deprivation based on home postcode
673 (<http://www.gov.scot/Topics/Statistics/SIMD>).

674 **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	Wilcoxo n (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
---	------------------	-----------------	------------------	------------------	------------------	----	-------	--------------------	------	------

^aPaired-samples t-test, ^bWilcoxon signed-rank test, *p < 0.05. **p < 0.01. ES, effect size (non-parametric); SRM, standardised response mean.

675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704

705 **Table 3 Responsiveness to change scores for participants demonstrating increased physical activity, no change or decreased physical activity**
 706 **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 ^c	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

707 ^aPercentage and number of participants demonstrating increased PA according to each measure, ^bMean change in PA score for participants
 708 identified as having increased PA from baseline to follow-up, ^cStandard deviation of change in PA score of participants indicating no change or
 709 identified as having decreased PA from baseline to follow-up; GRI, Guyatt responsiveness index.

710

711

712

713

714

715

716

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

T0 correlations (n=30)

activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	IPAQ		
			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)

activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	IPAQ		
			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

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

719

720

721

722

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

T0-T1 correlations (n=30)

activPAL	Total PA minutes (min/day)	Total MET-minutes (min/day)	IPAQ		
			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

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