SPORT ENGLAND CONTRACT REFERENCE: SE730

The selection of a project level measure of physical activity

Final Report

December 2017

Karen Milton[1], Andrew Engeli[2], Nick Townsend[1], Emma Coombes[3], Andy Jones[3]

[1] Centre on Population Approaches for Non-Communicable Disease Prevention, University of Oxford

[2] Wavehill

[3] Norwich Medical School, University of East Anglia

Suggested citation: Milton K, Engeli A, Townsend T, Coombes E, Jones A (2017) The selection of a project level measure of physical activity. London: Sport England.

Introduction

The importance of physical activity promotion is increasingly being recognised, not only in the prevention and treatment of Non-Communicable Diseases (NCDs) but in its contribution to broader agendas including the United Nations Sustainable Development Goals (United Nations, 2015) and Sport England's key strategic outcomes for 2016-2021 (Sport England, 2016). To fully understand the impact of physical activity on these broad agendas, it is critical that robust measures of activity are in place both for population surveillance and programme evaluation.

The Government Strategy, 'Sporting Future' requires that part of Sport England's remit is to increase the number of people taking part in sport in England (Sport England, 2016). The Active Lives Survey is the surveillance system used by Sport England to track the number of adults (aged 16+) taking part in sport and physical activity at the population level. The survey has an annual sample of almost 200,000 individuals, and provides data on sport and physical activity by demographic group, where people live and activity type.

Sport England is keen to ensure that it has robust and consistent evaluation protocols for the projects which it funds. This research was commissioned with the aim of identifying a short physical activity tool which is appropriate for evaluating Sport England's projects. The selected tool will be used to measure the baseline activity levels of participants to determine the extent to which the Sport England funded projects are reaching their target audience as well as to determine the impact of the programmes on participants' physical activity behaviour.

Sport England's requirements for a project level tool are that it:

- measures participation in sport and physical activity;
- is capable of identifying people who do less than 30 minutes of physical activity per week;
- can assess physical activity against the Chief Medical Officer's (CMO) recommendation of 150 minutes of moderate to vigorous activity per week;
- is easy to administer in different modes;
- is easy for respondents to comprehend and complete;
- is suitable for all adults including those with impairments; and
- is capable of measuring changes in participation over time.

This research was conducted in two phases. Phase 1 involved identifying a short self-report physical activity tool which is fast and straightforward to complete and provides relatively strong agreement with Sport England's Active Lives Survey in terms of the classification of respondents as 'inactive', 'fairly active', and 'active'. Phase 2 involved testing the sensitivity of the tool(s) selected in phase 1 to detect changes in physical activity over time and thus provide an appropriate tool for programme evaluation. This report is divided into two parts to reflect these two separate but related phases of the project.

This research was approved by the Central University Research Ethics Committee (CUREC) at the University of Oxford.

Phase 1 Methods

Identification of potential tools

We undertook a scoping review of established physical activity questionnaires including the Baecke measurement of habitual physical activity (Baecke et al., 1982), both versions of the Seven Day Physical Activity Recall (PAR; Blair et al., 1985; Sallis et al., 1985), the Godin Leisure Time Exercise Questionnaire (Godin & Shepherd, 1985), the Global Physical Activity Questionnaire (GPAQ; Armstrong & Bull, 2006), the International Physical Activity Questionnaire – Short Form (IPAQ; Craig et al., 2003), the Minnesota Leisure Time Physical Activity Questionnaire (Talyor et al., 1978), and the Single-Item Measure (Milton et al., 2011). The results of this review are included in Appendix 1.

Only one tool was identified as potentially meeting Sport England's requirements – the IPAQ Short Form (Craig et al., 2003; see Appendix 2). The IPAQ is one of the most commonly used self-report physical activity tools. It is a seven-item instrument which collects data on moderate intensity activity, vigorous intensity activity, walking and sedentary behaviour in the past seven days. Whilst the IPAQ was originally developed for national and regional surveillance, it is frequently used for evaluation purposes.

The Single-Item Measure for physical activity (see Appendix 3) comprises one question designed to capture physical activity across the recreation and transport domains, and is easy to administer in the field (Milton et al., 2011). The Single-Item Measure was initially excluded from this research because it measures days of activity and not total minutes. However, at present this is the most commonly utilised tool among Sport England funded projects. The original Single-Item Measure asked respondents to report the number of days in the past week on which they have undertaken at least 30 minutes of at least moderate intensity physical activity. This tool was initially designed to assess whether respondents were meeting the physical activity recommendation of at least 30 minutes of at least moderate intensity physical activity on five or more days of the week (Department of Health, 2004). Since the Chief Medical Officers (CMO) physical activity recommendations changed in 2011 – to achieving 150 minutes per week (with no requirement on the number of days over which to spread this activity) – the Single-Item Measure has the potential to misclassify people against the current recommendation (Department of Health, 2011). In an attempt to overcome the risk of misclassification, a second question was added to the tool to assess whether people doing four or less days per week of physical activity as measured by the Single-Item Measure achieve 150 minutes or more per week. This new two-item tool is called the Single-Item-PLUS (or SI-PLUS) throughout this report (see Appendix 4). We analysed the data from the SI-PLUS in two ways; firstly using part 1 only (the original Single-Item Measure) and secondly using data from both part 1 and part 2 (the new additional component).

In addition we decided to create a Short Active Lives Survey using a sub-set of questions from the (longer) surveillance survey. The rationale for this was that such a tool would lead to data collection at the project level which is directly comparable to data collected in the Active Lives national surveillance system. The Active Lives Survey collects detailed data across five types of activities: a continuous walk lasting at least 10 minutes; gardening; a cycle ride; sport, fitness or recreation activity; and dance. For brevity, the Short Active Lives Survey does not include questions on gardening and questions related to dance are combined within the sport and fitness category. The Short Active Lives Survey assesses the number of days and the total time spent doing each activity in the past seven days. It also asks respondents to indicate whether the activity raised their breathing rate, in order to determine whether it was at least moderate intensity. The Short Active Lives Survey is included in Appendix 5.

Study design and sampling

Each survey was recreated in the Wavehill Qualtrics platform as separate modules. In order to ensure minimal participant burden, we selected a pairwise comparison design, in which respondents were assigned the module of interest (the Active Lives Survey) and randomly assigned one other tool (either the IPAQ, the Single-Item-PLUS, or the Short Active Lives Survey), as illustrated in Figure 1. All respondents completed the Active Lives Survey first, followed by the comparison tool.

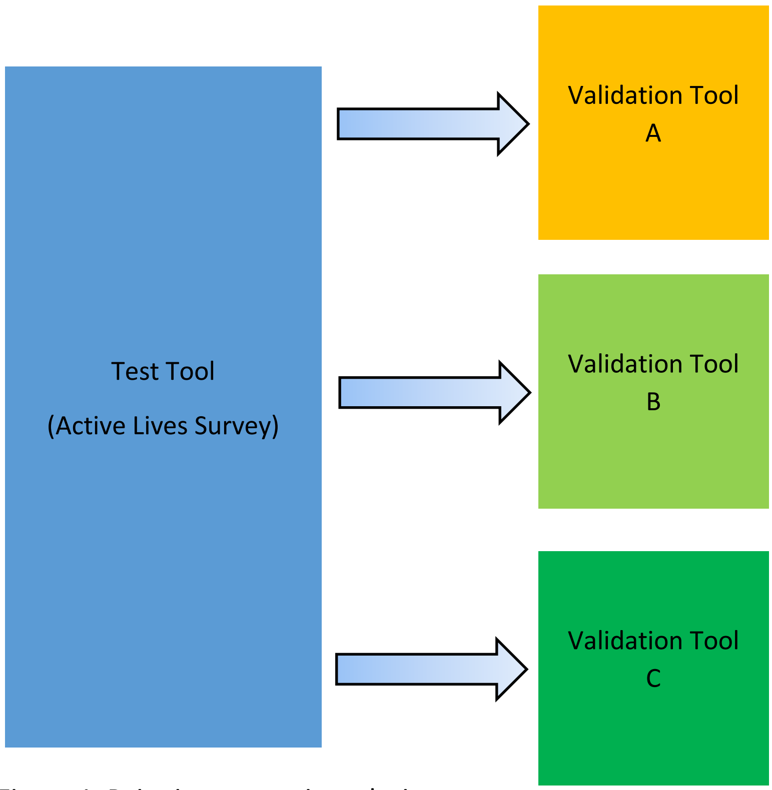


Figure 1. Pairwise comparison design

The survey link was circulated to a sample of approximately 25,000 people who had previously participated in the Active Lives Survey and had agreed to be re-contacted for follow-up research. We set, *a priori,* a target sample of 500 respondents per survey which was exceeded within several hours of the survey going live. A total of 2138 people took part in the survey. In total 681 respondents completed the Active Lives Survey and the IPAQ, 685 completed the Active Lives Survey and the Single-Item-PLUS, and 772 completed the Active Lives Survey and the short version of the same tool[1].

Data analysis

Total minutes of activity as measured by the IPAQ were calculated as follows: (days of moderate intensity activity x typical daily minutes of moderate intensity activity) + (days of vigorous intensity activity x typical daily minutes of vigorous intensity activity) + (days of walking x typical daily minutes of walking). For the Short Active Lives Survey, total minutes were calculated by summing all activities which were reported to raise the participants breathing rate, and were thus considered moderate intensity. As some of the comparison tools did not differentiate moderate and vigorous intensity activities, each minute of activity that was reported as being at least moderate intensity was counted as one minute of activity; no additional weighting was applied to minutes of vigorous intensity activity on any of the tools.

For the IPAQ and the Short Active Lives Survey, Pearson's bivariate correlation was used to measure the linear correlation of total minutes of physical activity with the Active Lives Survey. This analysis was not possible with the Single-Item-PLUS, which assesses days of activity and achievement of 150 minutes (for those reporting 4 days or less) using a binary (yes/no) response scale.

For the IPAQ and the Short Active Lives Survey, total minutes were used to classify respondents as 'inactive', 'fairly active', or 'active'. For the purposes of this study, those achieving less than 30 minutes per week are referred to as 'inactive', those achieving 30 or more minutes per week but less than 150 minutes are referred to as 'fairly active' and those achieving 150 minutes or more are referred to as 'active'. For the Single-Item Measure, those reporting zero days of activity were classified as 'inactive', those reporting between 1 and 4 days were considered 'fairly active', and those reporting 5 days or more were considered 'active'. For the SI-PLUS, those reporting zero days of physical activity, who also stated achieving less than 150 minutes per week were classified as 'inactive'. Those reporting 1-4 days per week, but less than 150 minutes were classified as 'fairly active', and those reporting 5 or more days of physical activity or 0-4 days but achievement of 150 minutes per week were classified as 'active'.

Kappa coefficients were calculated to determine agreement between each of the comparison tools (IPAQ, Single-Item Measure, SI-PLUS, and Short Active Lives Survey) on the classification of respondents achieving less than 30 minutes of at least moderate intensity physical activity per week and the classification of respondents achieving 150 minutes or more. Coefficient values of ≤0.2 were considered a weak correlation, 0.21–0.4 were considered fair, 0.41–0.6 were regarded as moderate, 0.61–0.8 were deemed strong and 0.81–1.0 very strong (Landis & Koch, 1977).

[1] The sample for the Short Active Lives Survey validation is larger than the other comparison tools due to a technical problem with the data collection, making it necessary to re-launch the survey with a new sample.

Phase 1 Results

A summary of the demographic characteristics of the final sample is shown in Table 1. There was a slight oversampling of females, with 48% of the sample being male. There was distribution across the whole age spectrum of the Active Lives Survey, although a disproportionate number of older participants, with over 55% of the sample being aged 56+ years. In comparison only 13% of the sample were aged 35 years or younger. Over half the sample were educated to degree level or higher and over 30% were retired.

Table 1. Demographic characteristics of the total sample and the sample for each comparison tool (n (%))

	IPAQ	**Single-Item Measure and SI-PLUS**	**Short Active Lives Survey**	**TOTAL**
Total sample	681 (31.9)	685 (32.0)	772 (36.1)	2138 (100)
Gender				
Male	313 (46.0)	305 (44.5)	401 (52.1)	1020 (47.7)
Female	368 (54.0)	380 (55.5)	369 (47.9)	1117 (52.2)
Other/Missing data	-	-	*2 (0.1)*	*2 (0.1)*
Age				
18-25	36 (5.3)	40 (5.8)	27 (3.5)	103 (4.8)
26-35	48 (7.0)	58 (8.5)	78 (10.1)	184 (8.6)
36-45	73 (10.7)	83 (12.1)	100 (13.0)	256 (12.0)
46-55	109 (16.0)	113 (16.5)	152 (19.7)	374 (17.5)
56-65	250 (36.7)	246 (35.9)	183 (23.7)	680 (31.8)
Over 65	165 (24.2)	145 (21.2)	231 (29.9)	540 (25.3)
Missing data	-	-	*1 (0.0)*	*1 (0.0)*
Self-reported BMI (weight/height)[2]				
Below 20	28 (4.1)	36 (5.3)	34 (4.4)	98 (4.6)
20-21.99	72 (10.6)	102 (14.9)	79 (10.2)	253 (11.8)
22-23.99	117 (17.2)	115 (16.8)	149 (19.3)	381(17.8)
24-25.99	138 (20.3)	116 (16.9)	154 (19.9)	408 (19.1)
26 and above	267 (39.2)	263 (38.4)	316 (40.9)	846 (39.6)
Missing data	*59 (8.7)*	*53 (7.7)*	*40 (5.2)*	*152 (7.1)*

[2] Includes 9 cases of pregnancy: 1 under 20, 1 between 20-22, 2 between 24-26, 5 over 26.

Table 1. Continued

Education				
Degree level or above	356 (52.3)	351 (51.2)	431 (55.8)	1138 (53.2)
A levels, NVQ level 3 and equivalents	81 (11.9)	94 (13.7)	89 (11.5)	264 (12.3)
Other Higher Education below degree level	93 (13.7)	85 (12.4)	126 (16.3)	304 (14.2)
GCSE/O level grade A*-C, NVQ level 2 and equivalents	90 (13.2)	98 (14.3)	78 (10.1)	266 (12.4)
Qualifications at level 1 and below	9 (1.3)	6 (0.9)	5 (0.6)	20 (0.9)
Another type of qualification	21 (3.1)	20 (2.9)	22 (2.8)	63 (2.9)
No qualifications	21 (3.1)	23 (3.4)	17 (2.2)	61 (2.9)
Prefer not to say	10 (1.5)	8 (1.2)	4 (0.5)	22 (1.0)
Occupational status				
Working full-time	259 (38.0)	274 (40.0)	286 (37.0)	819 (38.3)
Working part-time	116 (17.0)	114 (16.6)	119 (15.4)	349 (16.3)
Student, in full-time education studying for a recognised qualification	23 (3.4)	26 (3.8)	24 (3.1)	73 (3.4)
Student, in part-time education studying for a recognised qualification	1 (0.1)	3 (0.4)	2 (0.3)	6 (0.3)
Unemployed (long term), more than 12 months	6 (0.9)	5 (0.7)	7 (0.9)	18 (0.8)
Unemployed, less than 12 months	3 (0.4)	8 (1.2)	5 (0.6)	16 (0.7)
Not working, long term sick or disabled	19 (2.8)	11 (1.6)	9 (1.2)	39 (1.8)
Not working, looking after house/children	15 (2.2)	22 (3.2)	19 (2.5)	56 (2.6)
Not working, retired	212 (31.1)	204 (29.8)	271 (35.1)	687 (32.1)
Other	27 (4.0)	18 (2.6)	30 (3.9)	75 (3.5)

Table 2 shows the total mean minutes of moderate to vigorous physical activity as measured by the Active Lives Survey, the IPAQ and the Short Active Lives Survey. The standard deviation, standard error, confidence intervals and minimum and maximum scores are also presented.

Table 2. Summary of data on total minutes of physical activity on the Active Lives Survey, IPAQ, and the Short Active Lives Survey

	IPAQ[3]	**Short Active Lives Survey**[4]	**Active Lives Survey**[5]
Sample (N)	*680*	*769*	*2101*
Mean total minutes	745.23	426.33	497.82
Standard deviation	770.66	443.77	684.85
Standard error	29.55	16.00	15.92
Minimum	0	0	0
Maximum	4825	3780	4995
95% Confidence Interval	687.21 803.26	394.92 457.75	468.51 527.12

The scatterplots in Figures 2 and 3 show the total minutes of physical activity on the Active Lives Survey against the values on the IPAQ and the Short Active Lives Survey respectively. Whilst there is a positive correlation between the Active Lives Survey and each of the two comparison tools, the plots suggest that more individuals reported high levels of physical activity on the IPAQ in comparison to the Active Lives Survey, but that people were more likely to report high levels of activity on the full Active Lives Survey in comparison to the short version of the same tool.

[3] Excludes 1 case where activity levels exceeded 5000 total reported minutes
[4] Excludes 3 cases where activity levels exceeded 5000 total reported minutes
[5] Excludes 38 cases where activity levels exceeded 5000 total reported minutes

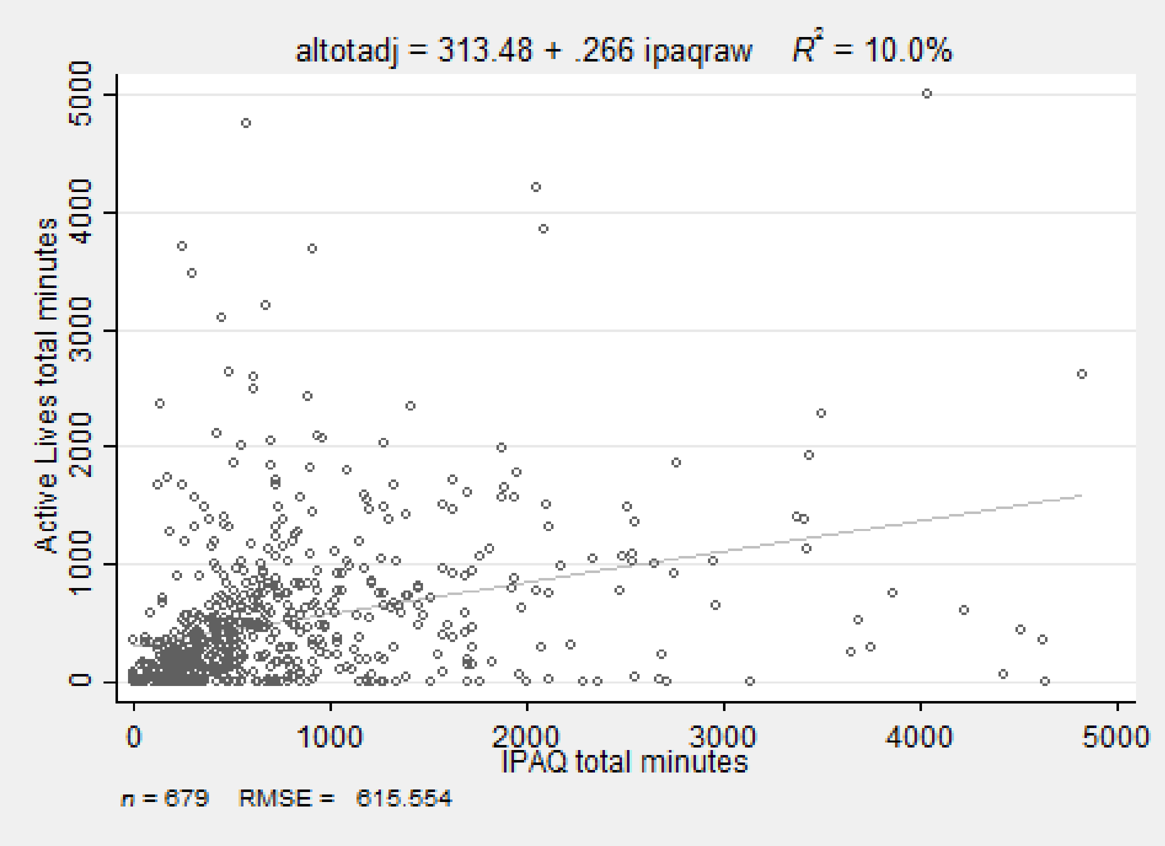


Figure 2. Scatter plot of the Active Lives Survey and the IPAQ

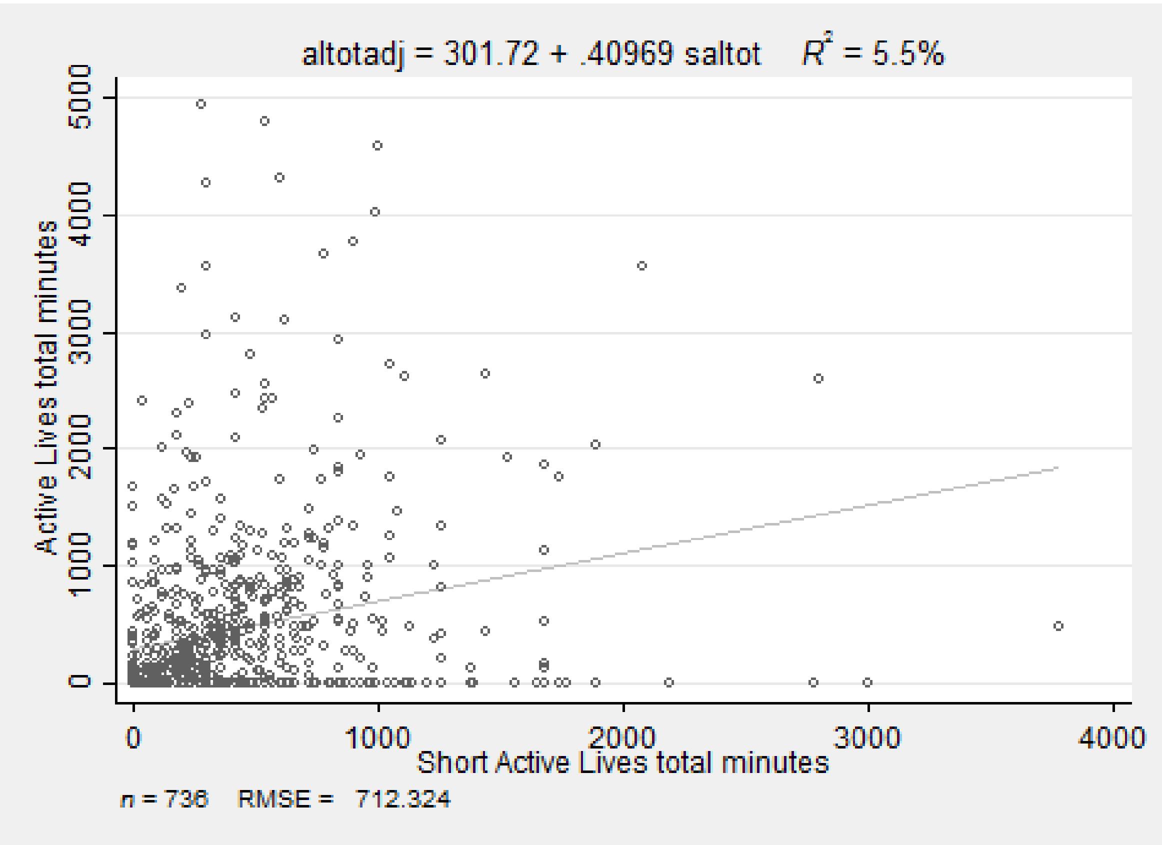


Figure 3. Scatter plot of the Active Lives Survey and the Short Active Lives Survey

Pearson’s bivariate correlation was used to measure the linear correlation between the Active Lives Survey and both the IPAQ and the Short Active Lives Survey. The Active Lives Survey demonstrated a correlation of r=0.35 with the IPAQ and r=0.23 for the Short Active Lives Survey. For both tools this level of correlation is considered ‘fair’.

Table 3 includes the classification of participants as ‘inactive’, ‘fairly active’, or ‘active’ on the Active Lives Survey and each of the comparison tools. Around 60% of the sample were categorised as ‘active’ on the Active Lives Survey, suggesting respondents were typical of participants taking part in the surveillance survey (although the data are not directly comparable as we did not apply additional weighting to vigorous intensity minutes). The distribution of the sample across the three activity categories varied by tool. Fewest people were classified as ‘inactive’ on the IPAQ, whilst the greatest proportion of the sample were classified as ‘inactive’ on the Single-Item Measure. Respondents were more likely to be classified as inactive on the full Active Lives Survey than on the short version of the same tool. The IPAQ categorised the greatest number of people as ‘active’ whereas the Active Lives Survey categorised the fewest people as active. In comparison to the Active Lives Survey, the distribution of the sample across the three physical activity categories was most balanced for the Single-Item Measure and least balanced for the IPAQ.

Table 3. Classification of participants as ‘inactive’, ‘fairly active’ and ‘active’ on each tool (n (%))

	IPAQ	**Single-Item Measure**	**SI-PLUS**	**Short Active Lives Survey**	**Active Lives Survey**
‘Inactive’	31 (4.6)	129 (18.8)	59 (8.6)	92 (11.9)	537 (25.1)
‘Fairly active’	50 (7.3)	100 (14.6)	91 (13.3)	110 (14.2)	269 (12.6)
‘Active’	600 (88.1)	456 (66.6)	535 (78.1)	570 (73.8)	1332 (62.3)
Total	681 (100)	685 (100)	685 (100)	772 (100)	2138 (100)

Weighted kappa was used to account for the distribution of the sample across activity categories and skew towards being classified as ‘active’. The kappa agreement between the Active Lives Survey and each of the comparison tools is shown in Table 4. Whilst none of the short tools demonstrated particularly strong kappa scores, the Single-Item-PLUS showed stronger agreement than the original Single-Item Measure, IPAQ and the Short Active Lives Survey in the classification of participants as ‘inactive’, ‘fairly active’ and ‘active’. A kappa value of 0.38 for the Single-Item-PLUS is considered fair. The agreement between the Active Lives Survey and the original Single-Item Measure (0.29), the IPAQ (0.26) and the Short Active Lives Survey (0.22) are also considered fair.

Table 4. Kappa agreement between Active Lives Survey and each comparison tool

	IPAQ	**Single-Item Measure**	**SI-PLUS**	**Short Active Lives Survey**
Agreement	78.56%	71.17%	79.93%	66.00%
Expected agreement	70.84%	59.12%	67.56%	56.13%
Kappa	0.2647	0.2948	0.3813	0.2249
Standard error	0.0261	0.026	0.0305	0.269
Z	10.12	11.33	12.51	8.37

Sensitivity and specificity analyses were used to explore the extent to which each tool allocated participants to the same physical activity category. In these analyses, the score on the Active Lives Survey was taken to be the 'true' measure. Sensitivity refers to the ability of the comparison tools to correctly identify those who are inactive and specificity refers to the ability of the comparison tools to correctly identify those who are active. The results for the IPAQ, the Single-Item Measure, the Single-Item-PLUS and the Short Active Lives Survey are shown in Tables 5, 6, 7 and 8 respectively.

Table 5. Sensitivity and specificity of the Active Lives Survey versus IPAQ (n (%))

		The Active Lives Survey			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
IPAQ	**'Inactive'**	24 (19.2)	6 (6.0)	1 (0.2)	31 (4.6)
	'Fairly active'	14 (11.2)	17 (17.0)	19 (4.2)	50 (7.3)
	'Active'	87 (69.6)	77 (77.0)	436 (95.6)	600 (88.1)
	Total	125 (100)	100 (100)	456 (100)	681 (100)

Table 6. Sensitivity and specificity of the Active Lives Survey versus the Single-Item Measure (n (%))

		The Active Lives Survey			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
Single-Item Measure	**'Inactive'**	56 (43.4)	17 (17.0)	12 (2.6)	85 (12.4)
	'Fairly active'	50 (38.8)	70 (70.0)	245 (52.7)	365 (53.3)
	'Active'	23 (17.8)	13 (13.0)	199 (42.8)	235 (34.3)
	Total	129 (100)	100 (100)	465 (100)	685 (100)

Table 7. Sensitivity and specificity of the Active Lives Survey versus the Single-Item-PLUS (n (%))

		The Active Lives Survey			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
Single-Item-PLUS	**'Inactive'**	42 (32.5)	13 (13.0)	4 (0.9)	59 (8.6)
	'Fairly active'	22 (17.1)	27 (27.0)	42 (9.2)	91 (13.3)
	'Active'	65 (50.4)	60 (60.0)	410 (90.0)	535 (78.1)
	Total	129 (100)	100 (100)	456 (100)	685 (100)

Table 8. Sensitivity and specificity of the Active Lives Survey versus the Short Active Lives Survey (n (%))

		The Active Lives Survey			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
Short Active Lives Survey	**'Inactive'**	65 (23.0)	4 (5.8)	23 (5.5)	92 (11.9)
	'Fairly active'	51 (18.0)	17 (24.6)	42 (10.0)	110 (14.2)
	'Active'	167 (59.0)	48 (69.6)	355 (84.5)	570 (73.8)
	Total	283 (100)	69 (100)	420 (100)	772 (100)

The IPAQ correctly identified 19% (24/125) of respondents who were categorised as 'inactive' on the Active Lives Survey (sensitivity) and 96% (436/456) of respondents who were classified as 'active' on the Active Lives Survey (specificity). Overall agreement between the Active Lives Survey and the IPAQ on these classifications was 79% (460/581). The Single-Item Measure correctly identified 43% (56/129) of respondents who were categorised as inactive on the Active Lives Survey and 43% (199/465) of those classified as active. Overall agreement between the Single-Item Measure and the Active Lives Survey was only 43%, however the Single-Item Measure was less likely to misclassify inactive or active respondents into the extreme opposing category (i.e. to classify the inactive as active and vice versa) compared to the SI-PLUS, IPAQ and the Short Active Lives Survey. The Single-Item-PLUS correctly identified 33% (42/129) of respondents who were categorised as 'inactive' on the Active Lives Survey and 90% (410/456) of respondents who were classified as 'active'. Overall agreement between the Active Lives Survey and the Single-Item-PLUS on these classifications was 77% (452/585). The Short Active Lives Survey correctly identified 23% (65/283) of respondents who were categorised as 'inactive' on the Active Lives Survey and 85% (355/420) of respondents who were classified as 'active'. Overall agreement between the Active Lives Survey and the Short Active Lives Survey on these classifications was 60% (420/703).

Time taken to complete the Active Lives Survey and each of the comparison tools was automatically recorded using question timers (click to click) and the summary statistics for each are shown in Table 9. It should be noted that respondents were able to suspend the survey and return to it later, thus for some respondents the actual time taken to complete the survey may be shorter than the recorded time. The Single-Item-PLUS was the quickest tool to complete, with a mean of just under 40 seconds. Completing the Single-Item Measure only took less time than the Single-Item PLUS although completion of the original tool was not timed. The Short Active Lives Survey took, on average, one minute to complete, and the IPAQ took just over two minutes to complete.

Table 9. Time spent completing each tool in seconds

	IPAQ	SI-PLUS	Short Active Lives Survey	Active Lives Survey
Mean (seconds)	133.71	38.30	59.47	830.16
Standard deviation	138.11	31.06	48.46	762.29
Standard error	5.2984	1.1866	1.6273	16.8940
Minimum	32.24	5.58	5.72	84.85
Maximum	2041.79[6]	464.45	689.19	9782.00[7]
95% Confidence Interval	123.31 144.12	35.97 40.63	56.28 62.68	797.03 863.29

[6] We excluded one case where the total time to complete was approximately 4 hours, which resulted from a paused survey.

[7] Does not include 103 cases that exceeded 10,000 seconds to complete.

Phase 1 Discussion

The aims of Phase 1 of this research were to identify a short physical activity tool which is fast and straightforward to complete and provides relatively strong agreement with the Active Lives Survey in terms of the classification of respondents as ‘inactive’, ‘fairly active’, and ‘active’. Whilst none of the short physical activity tools showed particularly strong agreement with the Active Lives Survey, each survey measures physical activity in a different way, which will influence both recall and how participants respond.

The tools all had benefits and drawbacks. Whilst the IPAQ demonstrated a stronger correlation with the Active Lives Survey than the correlation observed for the Short Active Lives Survey, the IPAQ demonstrated the greatest level of over-reporting, with almost 90% of the sample being classified as ‘active’ in comparison to 62% classified as ‘active’ on the Active Lives Survey. Whilst the IPAQ correctly identifying 93% of people who were classified as active on the Active Lives Survey, the IPAQ correctly identified just 19% of respondents who were categorised as inactive on the Active Lives Survey. The closest match to the Active Lives Survey in terms of the distribution of the sample across physical activity categories was the Single-Item Measure. The Single-Item Measure also demonstrated the lowest risk of extreme misclassification i.e. classifying someone identified as ‘inactive’ on the Active Lives Survey, as ‘active’ and vice versa. The Single-Item-PLUS demonstrated the strongest kappa agreement, suggesting it is the closest match to the Active Lives Survey in the classification of respondents as ‘inactive’, ‘fairly active’ or ‘active’. The Short Active Lives Survey demonstrated the lowest kappa agreement, predominantly due to the relatively large proportion (over 20%) of the sample who were classified as ‘inactive’ on the full Active Lives Survey and ‘active’ on the short version of the same tool. This is an unexpected finding, given greater reporting would be expected on the more detailed tool.

Our recommendation to Sport England was to take both the Single-Item-PLUS and the Short Active Lives Survey forward to Phase 2 of the research, which focused on assessing the sensitivity of the tools to detect changes in objectively measured physical activity over time. We decided not to proceed with further testing of the IPAQ in Phase 2, due to the over-reporting demonstrated on the IPAQ and also the longer duration required to complete the IPAQ in comparison to the other short tools. We recommended taking forward the Single-Item-PLUS as it demonstrated the strongest kappa agreement with the Active Lives Survey. We also recommended taking a second tool forward in case the Single-Item-PLUS proved insensitive to detect changes in physical activity over time, being a days-based and not a minutes-based tool. The second tool which we recommended taking forward was the Short Active Lives Survey. Whilst it showed the poorest kappa score, this appears to be due to the large proportion of the sample who were classified as ‘inactive’ on the full survey and ‘active’ on the short tool. The Short Active Lives Survey generates data which is directly comparable to data collected in the national surveillance system, it takes only slightly longer that the Single-Item-PLUS to complete, and is more likely to be sensitive to detecting changes over time due to assessing both days and minutes of physical activity. For the Short Active Lives Survey to be taken forward to Phase 2, the questions required reformatting for administration in a non-web-based format. The reformatted version is included in Appendix 6.

Phase 2 Methods

Tool selection

Following a period of reflection on the results of Phase 1, it was agreed that the Single-Item-PLUS was not providing Sport England with a sufficiently robust assessment of whether respondents were achieving 30 minutes or more of physical activity per week. Whilst the Single-Item-PLUS can identify those who do not achieve 30 minutes of physical activity on a single day, we cannot be confident that these people do not achieve 30 minutes of activity across a week. As having the capability of identifying people who do less than 30 minutes of physical activity per week was a key criteria for the selection of a tool, we decided not to proceed with the Single-Item-PLUS for further testing.

Following further discussion on the best way forward, the research team decided to develop a new minutes-based version of the Single-Item Measure. This new tool is called SIMBA (Single-Item Minutes Based Assessment). The SIMBA uses very similar wording to the original Single-Item Measure but asks respondents to categorise their minutes of activity in the past week, as opposed to the number of days on which they achieved 30 minutes or more (see Appendix 7). It was felt this minutes based tool was more likely to accurately capture those who do less than 30 minutes of physical activity per week and was also more likely to detect increases in physical activity, given it can capture increases in duration and/or bouts of activity as opposed to only detecting an increase in the number of days on which thirty minutes of physical activity is undertaken.

To test the ability of the self-report tools to detect changes in physical activity over time, an objective measure of physical activity was used to provide a 'true' measure of physical activity change. Objective physical activity data were collected using Actigraph GT1M devices. The Actigraph GT1M is a uni-axial accelerometer which uses a piezoelectric lever to detect acceleration in the vertical plane. Flexion of this lever, caused by movement, generates a signal proportional to the amount of acceleration, which is summed over a defined period of time known as an epoch. For this project, the Actigraph GT1M devices were initialised to collect data in 10 second epochs.

Study design and sampling

Data were collected by Active Norfolk, via recruitment of participants in their Fun & Fit programme. Fun & Fit is targeted at people who are inactive or do very little physical activity and wish to incorporate more physical activity into their lives. The programme offers 8-10 week courses in a wide variety of sports and activities including running, swimming, yoga and walking football. All sessions last for one hour and are free of charge to attend. Participants were eligible to participate in the research if they were new to the Fun & Fit project, irrespective of their baseline level of physical activity. Participants were provided with information about the project and those agreeing to participate completed a written consent form.

Data collection

At baseline, all participants completed a short questionnaire including contact details and demographic characteristics (age, gender, ethnicity), and were instructed on how to wear the Actigraph accelerometer (these devices are attached to an elastic belt which is worn around the waist, in the right mid-axillary line and level with the iliac crest). Participants began wearing the accelerometer immediately to gain familiarity with the device whilst in the presence of the instructor and to try to normalise the wearing of the device in an attempt to reduce reactivity. The participants were instructed to wear the accelerometer during waking hours, except when in water, until the next class which took place exactly one week later. At the Fun & Fit class the following week, participants returned the Actigraph device and completed a short questionnaire including the Short Active Lives Survey and the SIMBA. Six to eight weeks later, the data collection was repeated; participants wore an Actigraph for a seven day period and at the Fun & Fit class the following week, completed the Short Active Lives Survey and the SIMBA.

Data analysis

The accelerometer data were downloaded using the ActiLife software (ActiGraph corporation, FL, USA). The data were cleaned for periods when the monitor was not worn, by excluding periods when the accelerometer count was zero for 60 minutes or more (allowing for up to two minutes of non-zero counts per hour) according to the protocol of Troiano et al., 2008. The remaining data points were classified into four intensity categories based on recorded counts per minute (CPM): sedentary (≤499 CPM), light (500–2019 CPM), moderate (2020–5999 CPM) or vigorous intensity activity (≥6000 CPM), in accordance with recommended cut-points for adults (Troiano et al, 2008). These data were used to calculate total accelerometer wear time per day, mean daily counts per minute, and total daily time in moderate to vigorous physical activity.

As the Fun & Fit classes took place at a range of different times, not all participants started and stopped wearing the accelerometer at the same time of day. For this reason we decided to exclude the first and last partial day of accelerometer data for each participant. Based on the remaining data, participants were included in the study if they had an accelerometer wear time of at least eight hours per day on at least five days. No requirement was set around the number of weekday or weekend days of data as there was no clear association between weekday vs weekend day and mean minutes of moderate to vigorous physical activity.

We approached the accelerometry analysis in two ways. Firstly we included all activity which was classified as being of moderate to vigorous intensity, regardless of bout duration. Secondly we included moderate to vigorous minutes only if they occurred in a bout of at least 10 minutes (could be longer), allowing up to 20% of the bout time to drop below the moderate intensity threshold (2020 counts per minute). We included the bout approach in an attempt to exclude incidental short bouts which may be overlooked when considering responses to self-report tools.

Pearson's bivariate correlation was used to assess associations between self-reported time spent in moderate to vigorous physical activity on the Short Active Lives Survey and the objective measure of physical activity. This analysis was not possible for the SIMBA, which collects categorical as opposed to continuous data. Concord correlation coefficient of agreement was used to determine how far the observed data deviate from the line of perfect agreement, the Bradley-Blackwood F statistic was used to test for significant concordance, and Bland and Altman 95% Limits of Agreement were used to calculate the level of agreement between the self-report and objective data.

The accelerometer data were used to classify participants as 'inactive', 'fairly active' or 'active', using the same classifications as used in Phase 1 (inactive: less than 30 minutes per week; fairly active: between 30 and 149 minutes per week; and active: 150 minutes per week or more). Percent agreement and κ statistic were used to determine the level of agreement between each of the self-report tools (the Short Active Lives Survey and the SIMBA) and accelerometry on these classifications. Sensitivity and specificity analyses were used to explore the extent to which each tool allocated participants to the same physical activity category as the objective measure. In these analyses, the score on the Actigraph was taken to be the 'true' measure.

Phase 2 Results

A total of 87 participants agreed to take part in the study, of which 61% were female. There was distribution across the age spectrum of the Active Lives Survey. Of the 78 participants who provided a date of birth, the mean age was 49 years with a minimum of 22 and a maximum of 75. Ninety two percent of the sample reported to be white. The gender, age, and ethnicity distribution of the participants is shown in Table 10.

Table 10. Demographic characteristics of the sample, n (%)

Total sample	87 (100)
Gender	
Female	53 (61)
Male	30 (34)
Missing	*4 (5)*
Age	
18-25	1 (1)
26-35	17 (20)
36-45	16 (18)
46-55	12 (14)
56-65	18 (21)
Over 65	14 (16)
Missing data	*9 (10)*
Ethnicity	
White	80 (92)
Mixed	1 (1)
Rather not say	1 (1)
Missing	*5 (6)*

Of the total sample of 87 participants, one did not return their accelerometer and one had an accelerometer that failed to record. A further six had accelerometers that stopped recording prematurely. The reason for this is not known. Of the remaining 79 participants, 57 (72%) met the wear-time criteria of at least eight hours per day on at least five days, and were included in the analysis. Of the final sample of 57 who met the accelerometer wear time criteria, 52 completed the Short Active Lives Survey, but only 41 completed the SIMBA[8].

The mean time in moderate to vigorous physical activity captured via accelerometry and the Short Active Lives Survey, when including all minutes of activity, is shown in Table 11. These values were calculated twice; firstly based on the full sample of 52 and again after removal of one outlier who reported doing 4600 minutes (almost 77 hours per week, or 11 hours per day) of moderate to vigorous physical activity on the Short Active Lives Survey.

[8] Data collection commenced prior to final decision making about the second tool to be included, thus the first 11 participants completed the short Active Lives Survey only and not the SIMBA

Table 11. Minimum, Maximum and Mean time in moderate to vigorous physical activity captured via accelerometry and the Short Active Lives Survey, including all minutes of activity

	N=52		**N=51***	
	Mean (sd)	**Min - Max**	**Mean (sd)**	**Min - Max**
Accelerometry	357.8 (157.7)	86.3 - 917.0	357.8 (157.7)	86.3 - 917.0
Short Active Lives Survey	420.9 (699.3)	0 - 4500	340.9 (399.4)	0 - 1920

*One outlier was removed from the sample

Whilst the mean minutes in moderate to vigorous physical activity were relatively similar between the accelerometer and the Short Active Lives Survey following removal of the outlier, Pearson's correlation between the two tools was 0.318, which is considered fair. Using concord correlation coefficient of agreement to determine how far the observed data deviate from the line of perfect agreement (rho_c = 1), there was rho_c of 0.243, showing weak agreement. The Bradley-Blackwood F statistic of 17.96 (p<0.001) suggests no significant concordance and the Bland and Altman 95% Limits of Agreement were large. The association between the two tools, when including all minutes of activity, is shown in Figure 4 (correlation) and Figure 5 (Bland-Altman 95% Limits of Agreement) and a summary of the statistics is shown in Table 12.

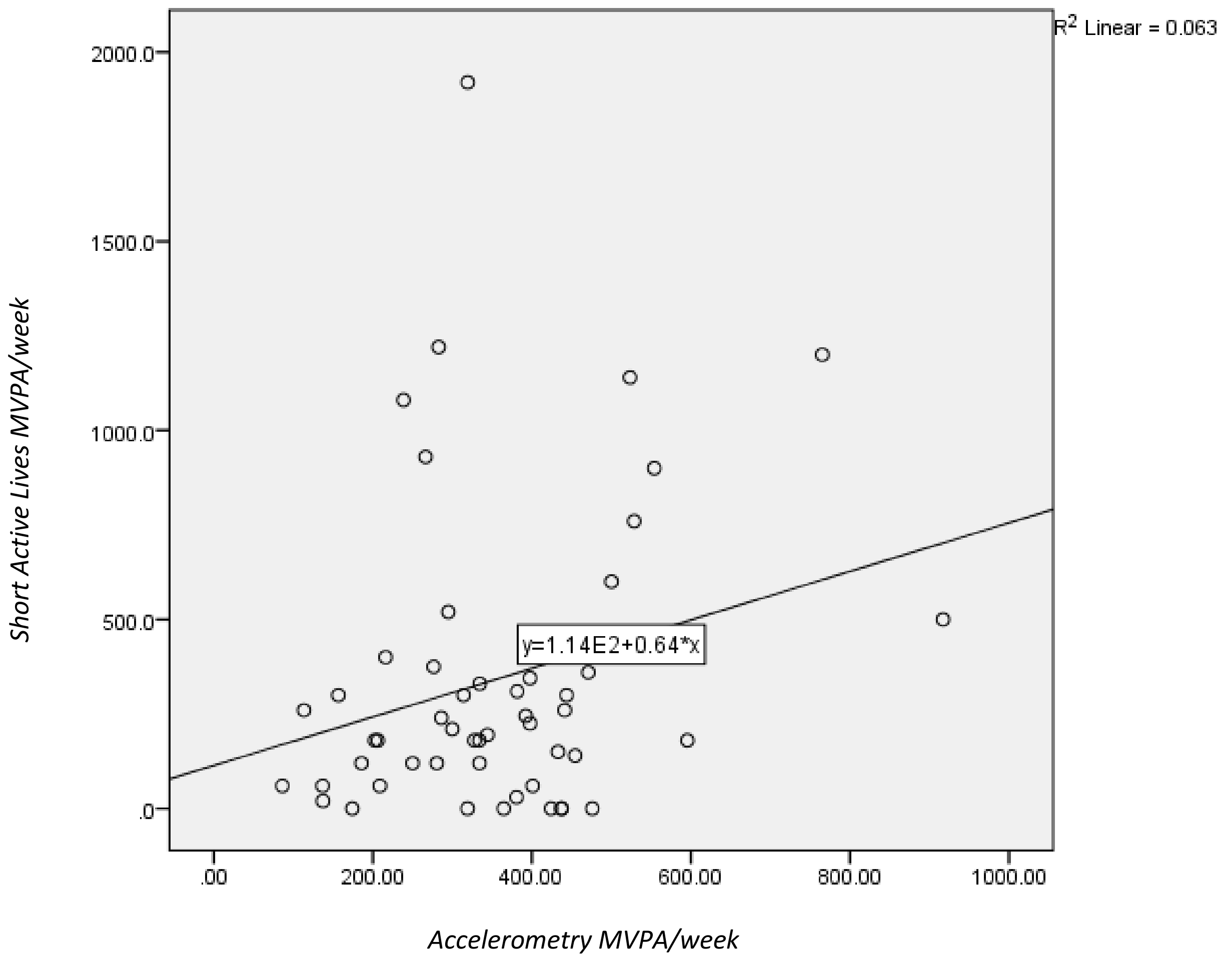


Figure 4. Scatter plot of the Short Active Lives Survey and accelerometry, including all minutes of activity, n=51

Figure 5. Limits of Agreement between the minutes per week of moderate to vigorous physical activity measured by the Short Active Lives Survey and accelerometry, including all minutes of activity, n=51

Table 12. Summary of the correlation and agreement statistics for the Short Active Lives Survey with accelerometry, n=51

Rho_c	0.243	95% CI (0.040,0.446)
Pearson's r	0.318	P=0.028
Bradley-Blackwood F	17.96	P<0.001
Bland Altman 95% Limits of Agreement	-594.3	676.8

These analyses were repeated including only minutes of at least moderate intensity activity that were undertaken in bouts of 10 minutes or more. The mean time in moderate to vigorous physical activity captured via accelerometry and the Short Active Lives Survey, when including minutes accumulated in at least 10 minute bouts, is shown in Table 13. These values were calculated twice; firstly based on the full sample of 52 participants and again after removal of one outlier who reported doing 4600 minutes (almost 77 hours per week, or 11 hours per day) of moderate to vigorous physical activity on the Short Active Lives Survey.

Table 13. Minimum, Maximum and Mean time in moderate to vigorous physical activity captured via accelerometry and the Short Active Lives Survey, including only activity accumulated in at least 10 minute bouts

	N=52		N=51*	
	Mean (sd)	**Min - Max**	**Mean (sd)**	**Min - Max**
Accelerometry	151.5 (126.6)	0 – 532.0	150.3 (127.5)	0 – 532.0
Short Active Lives Survey	422.8 (422.8)	0 – 4500	340.9 (399.4)	0 – 1920

*One outlier was removed from the sample

When including only bouts of activity undertaken in bouts of 10 minutes or more as measured by accelerometry, the agreement in mean minutes of activity between the accelerometer and the Active Lives Survey was less similar; the Pearson's correlation between the two tools was 0.164 which is considered weak. Using concord correlation coefficient of agreement to determine how far the observed data deviate from the line of perfect agreement (rho_c = 1), there was rho_c of 0.078, also showing a weak agreement. The Bradley-Blackwood F statistic of 67.10 ($p<0.001$) suggests no significant concordance, and the Bland and Altman 95% Limits of Agreement were large. The association between the two tools, when including minutes of activity undertaken in bouts of at least 10 minutes, is shown in Figure 6 (correlation) and Figure 7 (Bland-Altman 95% Limits of Agreement) and a summary of the statistics is shown in Table 14.

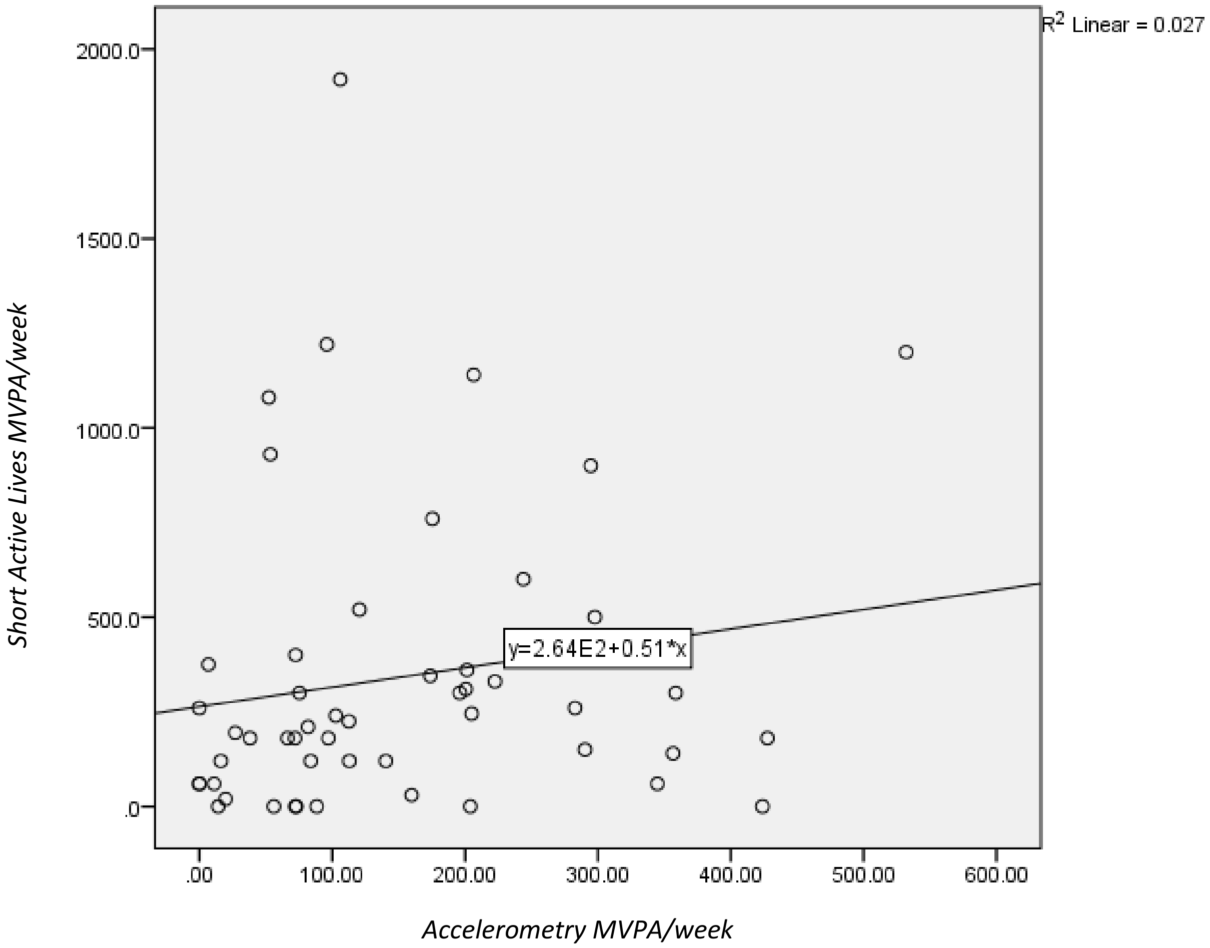
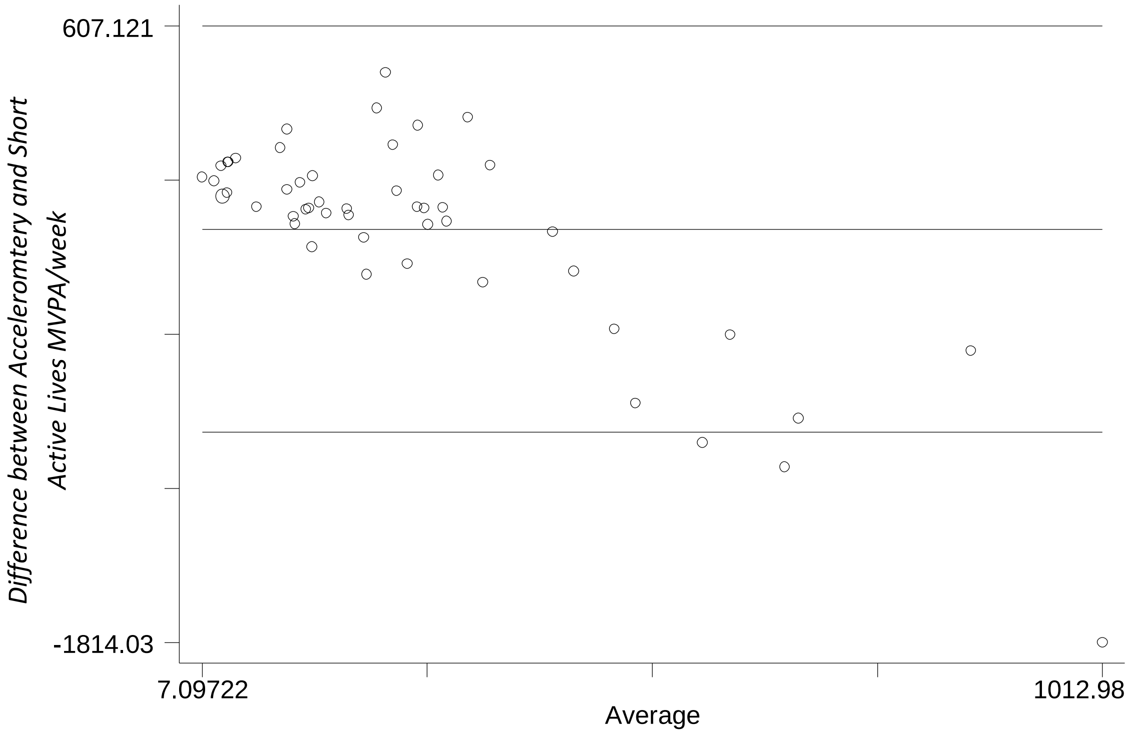


Figure 6 . Scatter plot of the Short Active Lives Survey and accelerometry, including only minutes of activity undertaken in bouts of at least 10 minutes, n=51



Average of Acceleromtery and Short Active Lives MVPA/week

Figure 7. Limits of Agreement between the minutes per week of moderate to vigorous physical activity measured by the Short Active Lives Survey and accelerometry, including only minutes of activity undertaken in bouts of at least 10 minutes, n=51

Table 14. Summary of the correlation and agreement statistics for the Short Active Lives Survey with accelerometry, including only minutes of activity undertaken in bouts of at least 10 minutes, n=51

Rho_c	0.078	95% CI (-0.054, 0.211)
Pearson's r	0.164	P=0.25
Bradley-Blackwood F	67.10	P<0.001
Bland Altman 95% Limits of Agreement	-972.3	591.2

Sport England had a specific interest in finding a tool which is capable of identifying people who do less than 30 minutes of physical activity per week and those who meet the CMO recommendation of 150 minutes of moderate to vigorous activity per week. As such, data from both the accelerometer and the Short Active Lives Survey were used to categorise participants as 'inactive', 'fairly active' or 'active'. For the accelerometry data, this was calculated by multiplying mean daily minutes of moderate to vigorous physical activity by seven. The distribution of the data for each tool across the three physical activity categories, when including all minutes of activity and only activity accumulated in at least 10 minute bouts, is shown in Table 15. When including all minutes of accelerometry data, zero participants were classified as 'inactive' and over 90% were classified as 'active'. When including accelerometry data undertaken in bouts of 10 or more minutes, nine people were classified as inactive, which is identical to the number classified as inactive on the Short Active Lives Survey. Around 40% were classified as active based on accelerometry data accumulated in 10 minute bouts, which is less than half the number classified as active when including all minutes of accelerometry derived data, and around a third less than were classified as active on the Short Active Lives Survey.

Table 15. Classification of participants as 'inactive', 'fairly active' and 'active' from accelerometry and the Short Active Lives Survey, n = 51 (n (%))

	'inactive' (<30 mins)	**'fairly active' (30-149 mins)**	**'active' (150+ mins)**
Accelerometry – all minutes	0 (0)	4 (8)	47 (92)
Accelerometry – bouts	9 (18)	21 (41)	21 (41)
Short Active Lives Survey	9 (18)	10 (20)	32 (63)

Sensitivity and specificity analyses were used to explore the extent to which each tool allocated participants to the same physical activity category. In these analyses, the score on the accelerometer was again taken to be the 'true' measure, to allow for a common comparison against each of the self-report tools. Sensitivity refers to the ability of the comparison tools to correctly identify those who are inactive and specificity refers to the ability of the comparison tools to correctly identify those who are active.

The results for the Short Active Lives Survey are shown in Tables 16 and 17. Table 16 includes all objectively measured physical activity and Table 17 includes activity accumulated in bouts of 10 minutes or more. When including all minutes of physical activity (Table 16), the Short Active Lives Survey correctly identified 66% of people who were classified as active via accelerometry and 50% of those who were classified as fairly active by accelerometry, although the sample size in the fairly active category according to the accelerometry data was very small, and no participants were classified as inactive according to accelerometry. Overall agreement between the two tools was 65%, although Kohen's Kappa showed a low measure of agreement between the two tools (0.132), with no significant difference from zero (p=0.075).

When including physical activity undertaken in 10 minute bouts or more (Table 17), the Short Active Lives Survey correctly identified 71% of the 21 people who were classified as active via accelerometry but only 14% of the 21 people classified as fairly active by accelerometry. It correctly identified just two of the nine participants identified as inactive by accelerometry. Overall agreement between the two tools was 39% and Kohen's Kappa showed a low measure of agreement between the two tools (0.035), with no significant difference from zero (p=0.711).

Table 16. Sensitivity and specificity of the Short Active Lives Survey versus accelerometry, including all objectively measured moderate to vigorous physical activity (MVPA), n=51 (n (%))

		Accelerometry – all MVPA			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
Short Active	**'Inactive'**	0 (0)	1 (25)	8 (17)	9 (18)
	'Fairly active'	0 (0)	2 (50)	8 (17)	10 (20)
	'Active'	0 (0)	1 (25)	31 (66)	32 (63)
	Total	0 (0)	4 (100)	47 (100)	51 (100)

Table 17. Sensitivity and specificity of the Short Active Lives Survey versus accelerometry, including only objectively measured moderate to vigorous physical activity (MVPA) undertaken in 10 minute bouts or more, n=51 (n (%))

<table>
<tr><th colspan="2"></th><th colspan="4">Accelerometry – bouts</th></tr>
<tr><th colspan="2"></th><th>'Inactive'</th><th>'Fairly active'</th><th>'Active'</th><th>Total</th></tr>
<tr><td rowspan="4">Short Active</td><td>'Inactive'</td><td>2 (4)</td><td>4 (8)</td><td>3 (6)</td><td>9 (18)</td></tr>
<tr><td>'Fairly active'</td><td>4 (8)</td><td>3 (6)</td><td>3 (6)</td><td>10 (20)</td></tr>
<tr><td>'Active'</td><td>3 (6)</td><td>14 (27)</td><td>15 (29)</td><td>32 (63)</td></tr>
<tr><td>Total</td><td>9 (18)</td><td>21 (41)</td><td>21 (41)</td><td>51 (100)</td></tr>
</table>

The SIMBA uses six categories of classification for respondents' physical activity levels. To aid comparability against the Short Active Lives Survey, the middle four categories of the SIMBA were collapsed, to simulate the three category system. The distribution of the participants across the three physical activity categories according to accelerometry and the SIMBA are shown in Table 18. Whereas no participants were classified as inactive according to accelerometery when including all minutes of physical activity, nine people were classified as inactive when applying the bout criteria. Ninety percent of the sample were classified as active based on all minutes of objectively measured criteria, but when including only those activities undertaken in bouts of 10 minutes or more this fell to 39%. The SIMBA under-estimated the number of people achieving recommended activity levels.

Table 18. Classification of participants as 'inactive', 'fairly active' and 'active' from accelerometry and the SIMBA, n = 41 (n (%))

	<30 mins	**30-149 mins**	**150+ mins**
Accelerometry – all minutes	0 (0)	4 (10)	37 (90)
Accelerometry - bouts	9 (22)	16 (39)	16 (39)
SIMBA	5 (12)	27 (66)	9 (22)

The results of the sensitivity and specificity analysis of the SIMBA are shown in Tables 19 and 20. When including all minutes of objectively measured physical activity (Table 19), there was 100% agreement between the participants classified as fairly active on the SIMBA and accelerometer but a very small proportion of the sample fell into this category. Approximately 90% of the sample were classified as active by accelerometry. Only 24% of these participants were correctly categorised as active according to the SIMBA. Sixty two percent were incorrectly classified as fairly active and 14% were misclassified as inactive on the self-report tool. Overall there was a 32% agreement between the two tools, and Kohen's Kappa showed a low measure of agreement (-0.02) with no significant difference to zero (p=0.506). When including activity accumulated in 10 minute bouts or more (Table 20), there was 31% agreement between the participants classified as active on the SIMBA and accelerometer and 56% agreement of the participants categorised as fairly active by accelerometer, but no agreement in the inactive category. Whereas the accelerometer data classified 9 people as inactive, none of these were classified as inactive on the SIMBA. Overall there was a 34% agreement between the two tools, and Kohen's Kappa showed a low measure of agreement (-0.044) with no significant difference to zero (p=0.669).

Table 19. Sensitivity and specificity of the SIMBA versus accelerometry, n=41 (n (%))

		Accelerometry – all MVPA			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
SIMBA	**'Inactive'**	0 (0)	0 (0)	5 (14)	5 (12)
	'Fairly active'	0 (0)	4 (100)	23 (62)	27 (66)
	'Active'	0 (0)	0 (0)	9 (24)	9 (22)
	Total	0 (0)	4 (100)	37 (100)	41 (100)

Table 20. Sensitivity and specificity of the SIMBA versus accelerometry, including only objectively measured physical activity undertaken in 10 minute bouts or more, n=41 (n (%))

		Accelerometry – bouts			
		'Inactive'	**'Fairly active'**	**'Active'**	**Total**
SIMBA	**'Inactive'**	0 (0)	4 (10)	1 (2)	5 (12)
	'Fairly active'	8 (20)	9 (22)	10 (24)	27 (67)
	'Active'	1 (2)	3 (7)	5 (12)	9 (22)
	Total	9 (22)	16 (39)	16 (39)	41 (100)

A key objective of this research was to identify a tool which is capable of detecting changes in physical activity over time, and can therefore provide a suitable before and after measure for the evaluation of Sport England funded projects. Whilst 57 participants met the accelerometer wear time criteria at baseline, only 22 (39%) of these participants also returned usable accelerometer data at follow-up. These 22 participants all had data for the Short Active Lives Survey but only 15 participants had data from the SIMBA. This analysis focused on participants that moved from one activity category ('inactive', 'fairly active', 'active') to another according to accelerometry and aimed to assess whether the self-report tools also captured this change. The highest activity level is referred to as '3' in this analysis and the lowest activity level is referred to as '1'.

The change in physical activity captured via each tool is shown in Table 21. From the accelerometer data including all activity, 20 out of 22 participants (91%) demonstrated no change in their physical activity level and 2 participants demonstrated a decrease. When including only objectively measured activity accumulated in bouts of 10 minutes or more, 14 out of 22 participants demonstrated no change in activity level, six demonstrated a decrease, and 2 demonstrated an increase. On the Short Active Lives Survey, 17 participants (77%) reported no change in activity, two reported a decrease in activity and three reported an increase. On the SIMBA, 11 participants (73%) reported no change in activity and 4 reported an increase. Sensitivity and specificity analyses were used to explore the extent to which the tools agreed on the change in categories. In these analyses, Kohen's Kappa consistently showed a low measure of agreement, with no significant difference to zero. However given the small number of participants changing activity category over the intervention period, the sample size is really too small to draw any meaningful conclusions on the ability of the tools to detect change.

Table 21. Changes in activity category between 1) ‘inactive’, 2) ‘fairly active’, and 3) ‘active’ according to accelerometry, the Short Active Lives Survey and the SIMBA (n (%))

	Decrease			**No change**	**Increase**		
	3 to 1	**3 to 2**	**2 to 1**		**1 to 2**	**1 to 3**	**2 to 3**
Accelerometer - all (n=22)		2 (9)		20 (91)			
Accelerometer - bouts (n=22)		3 (14)	3 (14)	14 (64)			2 (9)
SAL (n=22)		1 (5)	1 (5)	17 (77)		1 (5)	2 (9)
SIMBA (n=15)				11 (73)		2 (13)	2 (13)

Phase 2 Discussion

The aim of Phase 2 of this research was to test the sensitivity of the self-report tools selected in Phase 1 to detect changes in physical activity over time and thus provide an appropriate tool for programme evaluation. In order to obtain a ‘true’ measure of change in behaviour against which to validate the self-report tools, objective data were collected via Actigraph accelerometers. The accelerometer data show that the sample were unusually active, with between 39% and 90% of participants achieving the physical activity recommendation of 150 minutes or more of moderate to vigorous physical activity per week depending on whether the analysis included all objectively measured physical activity or only activity undertaken in bouts of 10 minutes or more. To provide some context, when accelerometry was included in the Health Survey for England the results indicated that only 6% of men and 4% of women were meeting recommended physical activity levels (The Information Centre for Health and Social Care, 2009). The prevalence of the population meeting recommended physical activity levels was particularly high in the current research, especially given we excluded the days on which we knew the participants attended the Fun & Fit programme.

This study found quite different results depending on whether we considered all objectively measured moderate to vigorous physical activity in the analysis or only activity accumulated in bouts of ten minutes or more. When including all objectively measured minutes, we found similar mean minutes of activity between the objective data and the Short Active Lives Survey. The correlation between Short Active Lives Survey and accelerometry was similar to that reported for other short physical activity instruments (Milton et al., 2011). The correlation between the Short Active Lives Survey and accelerometry when only including objective data accumulated in 10 minute bouts or more was weaker. The Short Active Lives Survey also showed better agreement on the categorisation of participants as ‘inactive’, ‘fairly active’ and ‘active’ when including all objectively measured activity (65% agreement) in comparison to including only the bout data (39% agreement). The level of agreement observed when including all objectively measured activity is similar to, and slightly higher than that observed for the Single-Item Measure which has been the commonly utilised tool to evaluate Sport England funded projects in recent years (Milton et al., 2013).

The SIMBA showed overall agreement of 32% in comparison with all objectively measured physical activity and 34% with objectively measured activity accumulated in bouts of at least ten minutes. A problem with the analysis of all objectively measured data was that none of the sample were classified as inactive. Overall, however the SIMBA performed less well than the Short Active Lives Survey in terms of the classification of participants as ‘inactive’, ‘fairly active’ and ‘active’, predominantly due to under-reporting on the self-report tool. This is an uncommon finding, as due to social desirability people usually claim to be more active than they really are. A potential explanation for the under-reporting in the current study might come from anecdotal evidence from the data collection team which suggested that participants were reluctant to admit that an activity caused them to raise their breathing rate, as they wanted to be perceived as fit. Thus this activity, despite being classified as moderate intensity via accelerometry, would not be captured via the self-report tool. If this was the case, we might expect this to lead to under-reporting on both tools, although it was much more pronounced on the SIMBA.

Almost all studies undertaken to validate self-report physical activity instruments against accelerometry have collected data at a single-time point only. Thus there is virtually no evidence on the ability of these self-report tools to detect changes in physical activity, despite many tools being used for this purpose. To our knowledge, only one previous study has attempted to measure change in physical activity behaviour over time and validate the sensitivity of a self-report tool (GPAQ) to detect this change (Cleland et al., 2014). The authors found the GPAQ to have a moderate level of agreement with accelerometry in terms of mean change in physical activity. Given the novelty of this part of the research, it is not surprising that we encountered several challenges. Firstly, the sample were relatively active at baseline, meaning many of them already fell into the highest physical activity category. Secondly there was a large drop-out of participants from baseline to follow-up, meaning we had a sample of only 22 participants with usable data at both time points. According to the accelerometry data, the majority of participants did not change their physical activity category over the course of the project. The self-report tools also showed no change in activity for the majority of the sample, thus agreeing with the objective data. Whilst this shows agreement in detecting no change in activity it does not provide evidence of the tools' ability to detect a real change in physical activity over time.

Conclusions

Based on the results of both Phase 1 and Phase 2 of this research, we suggest Sport England utilise the Short Active Lives Survey for the evaluation of their projects. In Phase 1 of the research the Short Active Lives Survey demonstrated a fair correlation with the full Active Lives Survey. In terms of the classification of participants as ‘inactive’, ‘fairly active’ and ‘active’, the Short Active Lives Survey demonstrated reasonable agreement with the full version of the tool in our sensitivity and specificity analysis, despite demonstrating the lowest kappa score. In Phase 2 of the research, the Short Active Lives Survey recorded similar mean minutes to the objective measure when including all minutes of activity in the analysis, and demonstrated a correlation with the objective data which is similar to that observed for other self-report tools (Milton et al., 2011). It also demonstrated a comparable level of agreement in the classification of respondents as ‘inactive’, ‘fairly active’ and ‘active’. The Short Active Lives Survey has the capability of identifying those who do less than 30 minutes of physical activity per week and those who achieve 150 minutes per week or more – two variables which are critical to evaluating success against Sport England’s key performance indicators. In addition, the Short Active Lives Survey takes just one minute to complete and has the advantage of generating data which is directly comparable to Sport England’s national surveillance system.

Whilst the Short Active Lives Survey under-estimated activity in comparison to all objectively measured minutes and slightly over-estimated activity when only ten minute bouts were included, the likelihood is that the participants’ ‘true’ physical activity levels lie somewhere between these objectively measured levels. The cut-point approach to accelerometry analysis applies the same definition of moderate intensity to all participants regardless of age, fitness, or other characteristics. Thus some participants may be credited with minutes of activity which were not sufficient to raise their breathing rate, and thus were not technically moderate intensity for that individual. Therefore the inclusion of all objectively measured minutes may over-estimate ‘true’ physical activity levels. Applying the ten minute bout approach also has limitations. There is insufficient evidence that undertaking ten minute or longer bouts of activity is necessary to additionally benefit health. Whilst the bout restriction does allow us to remove incidental activity which is unlikely to be recalled on a self-report tool, it is possible that it also removes activity which was moderate intensity and should therefore count towards an individual’s overall minutes. Given that the Short Active Lives Survey estimated participants’ activity levels to lie somewhere in between the two objectively measured estimates, it likely provides data which are relatively reflective of participants’ true physical activity levels.

This project was unable to deliver on its objective of identifying a short self-report tool which was found to be valid in detecting changes in physical activity over time. It may be that either of the two tools we looked at in Phase 2 are capable of doing this, but the small sample size and the lack of change in activity prevented validation. We would be happy to discuss potential opportunities to conduct further research with Sport England to fulfil this objective.

Acknowledgements

We would like to thank Sam Watts, Lisa Ward, Rachel Cooke and the team at Active Norfolk for their support with the data collection for Phase 2 of this research.

References

Armstrong, T., & Bull, F. (2006). Development of the World Health Organization Global Physical Activity Questionnaire (GPAQ). *Journal of Public Health*, *14*, 66–70.

Baecke, J., Burema, J., & Frijters, J. (1982). A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The American Journal of Clinical Nutrition*, *36*(5), 936–942.

Blair, S., Haskell, W., Ho, P., Paffenbarger, R. J., Vranizan, K., Farquhar, J., & Wood, P. (1985). Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. *American Journal of Epidemiology*, *122*(5), 794–804.

Cleland, C., Hunter, R., Kee, F., Cupples, M., Sallis, J., & Tully, M. (2014). Validity of the Global Physical Activity Questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health*, *14*, 1255.

Craig, C., Marshall, A., Sjöström, M., Bauman, A., Booth, M., Ainsworth, B., … Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Medicine and Science in Sports and Exercise*, *35*(8), 1381–1395.

Department of Health. (2004). *At least five a week*. London.

Department of Health. (2011). *Start active, stay active*. London.

Godin, G., & Shepherd, R. (1985). Godin Leisure-Time Exercise Questionnaire. *Canadian Journal of Applied Sports Science*, *10*, 141–146.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Milton, K., Bull, F., & Bauman, A. (2011). Reliability and validity testing of a single-item physical activity measure. *British Journal of Sports Medicine*, *45*, 203–208.

Milton, K., Clemes, S., & Bull, F. (2013). Can a single question provide an accurate measure of physical activity? *British Journal of Sports Medicine*, *47*(1), 44–8.

Sallis, J., Haskell, W., Wood, P., Fortmann, S., Rogers, T., Blair, S., & Paffenbarger, R. J. (1985). Physical activity assessment methodology in the Five-City Project. *American Journal of Epidemiology*, *121*, 91–106.

Sport England. (2016). *Sport England: Towards an active nation. Strategy 2016–2021*. London.

Talyor, H., Jacobs, D., Schunker, B., Knudsen, J., Leon, A., & Debacker, G. (1978). Questionnaire for the assessment of leisure time physical activities. *Journal of Chronic Disease*, *31*(12), 741–755.

The Information Centre for Health and Social Care. (2009). *Health Survey for England 2008. Volume 1. Physical activity and fitness*. Leeds, UK.

Troiano, R., Berrigan, D., Dodd, K., Mâsse, L., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise*, *40*, 181–188.

United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development , 2015*. New York.

Appendix 1 – Appraisal of previously validated self-report tools for utility for Sport England

Instrument	Description				Utility for Sport England	Final decision
	Items	Recall period	Domains assessed	Intensities assessed		
Baecke measurement of habitual physical activity (Baecke et al., 1982)	22	Typical behaviour	Work Sport Leisure	No breakdown by intensity, just a broad question on what sport people play most frequently - low intensity/ medium intensity/ high intensity	• Not able to identify people who do less than 30 minutes per week • Not capable of assessing physical activity level against the CMO recommendation of 150 minutes per week • The tool is relatively long • Categorical response options will present challenges for detecting changes in physical activity behaviour over time	Exclude
Seven-Day Physical Activity Recall (Blair et al., 1985)	14	Last 7 days – captures weekday and weekend day separately	Work Leisure Sleep	Moderate Hard Very hard	• Designed to be interview administered • The tool is relatively long • Breakdown of weekday and weekend day is more detail than required by Sport England • Sport England are not interested in two of the three domains assessed by the tool (work and sleep)	Exclude

Godin Leisure Time Exercise Questionnaire (Godin & Shepherd, 1985)	4	Typical 7-Day period	Exercise	Mild Moderate Strenuous	• Tool captures days of at least 15 minutes, not total time • Not able to identify people who do less than 30 minutes per week • Not capable of assessing physical activity level against the CMO recommendation of 150 minutes per week • Unable to detect increases in bout length	Exclude
Global Physical Activity Questionnaire (Armstrong & Bull, 2006)	16	Usual behaviour	Work Travel Recreation	Moderate Vigorous	• Includes work as well as recreational sport and physical activity • The tool is relatively long • 'Usual behaviour' may pose challenges for detecting recent changes in physical activity	Exclude
International Physical Activity Questionnaire (Craig et al., 2003)	7	Past 7 days	All activity combined, including at work, at home, and during sport and recreation	Sedentary Walking Moderate Vigorous	• Relatively short tool • Capable of identifying people who are inactive • Capable of detecting whether people meet the CMO physical activity recommendation of 150 minutes	Include
Minnesota Leisure Time Physical Activity Questionnaire (Talyor et al., 1978)	63	Past year recall	Walking Sport Recreation	Light Medium Heavy	• The tool is relatively long • Designed to be interview administered • Not feasible for completion in the field	Exclude

Seven-Day Physical Activity Recall (Sallis et al., 1985)	Includes a matrix for respondents to complete	Past 7 days	Sleep Work Leisure	Moderate Hard Very hard	• Designed to be interview administered • Respondents are required to complete a complex matrix of activities undertaken in the morning, afternoon and evening of each day in the past week. • Not feasible for completion in the field	Exclude
Single-Item Measure (Milton et al., 2011)	1	Past 7 days	All activity for sport, recreation and travel combined. Work is excluded.	Activities that raise breathing rate (i.e. moderate intensity)	• Assesses days of 30 minutes or more, not total time • Capable of identifying people who are inactive • Will detect when people increase the days on which they are active, but not their bout duration • Will not assess physical activity against the current CMO recommendation of 150 minutes	Include – with a second question to assess minutes

Appendix 2 – The International Physical Activity Questionnaire (IPAQ)

We are interested in finding out about the kinds of physical activities that people do as part of their everyday lives. The questions will ask you about the time you spent being physically active in the **last 7 days**. Please answer each question even if you do not consider yourself to be an active person. Please think about the activities you do at work, as part of your house and yard work, to get from place to place, and in your spare time for recreation, exercise or sport.

Think about all the **vigorous** activities that you did in the **last 7 days**. **Vigorous** physical activities refer to activities that take hard physical effort and make you breathe much harder than normal. Think *only* about those physical activities that you did for at least 10 minutes at a time.

1. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?

_____ **days per week**
No vigorous physical activities ***Skip to question 3***

2. How much time did you usually spend doing **vigorous** physical activities on one of those days?

_____ **hours per day**
_____ **minutes per day**
Don't know/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities refer to activities that take moderate physical effort and make you breathe somewhat harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.

3. During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.

_____ **days per week**
No moderate physical activities ***Skip to question 5***

4. How much time did you usually spend doing **moderate** physical activities on one of those days?

_____ **hours per day**
_____ **minutes per day**
Don't know/Not sure

Think about the time you spent **walking** in the **last 7 days**. This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.

5. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time?

_____ days per week
No walking ***Skip to question 7***

6. How much time did you usually spend **walking** on one of those days?

_____ hours per day
_____ minutes per day
Don't know/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

7. During the **last 7 days**, how much time did you spend **sitting** on a **week day**?

_____ hours per day
_____ minutes per day
Don't know/Not sure

Guidance on the analysis of data from the International Physical Activity Questionnaire

There are two approaches to scoring the IPAQ, one which involves the calculation of MET-minutes per week (an estimate of total energy expenditure) and an alternative approach which uses a combination of days and METS to categorise participants as low, moderate or highly active. The MET approach is the most straightforward and involves the following calculations:

Walking MET-minutes/week = 3.3 * walking minutes * walking days

Moderate MET-minutes/week = 4.0 * moderate-intensity activity minutes * moderate days

Vigorous MET-minutes/week = 8.0 * vigorous-intensity activity minutes * vigorous-intensity days

Total physical activity MET-minutes/week = sum of Walking + Moderate + Vigorous MET-minutes/week scores

Appendix 3 – The Single-Item Measure

In the past week, on how many days have you done a total of 30 minutes or more of physical activity, which was enough to raise your breathing rate?

This may include sport, exercise, and brisk walking or cycling for recreation or to get to and from places, but should not include housework or physical activity that may be part of your job.

Please circle the relevant number:

0 1 2 3 4 5 6 7

Guidance on the analysis of data from the Single-Item Measure

The Single-Item Measure uses raw scores, with no computation required.

Appendix 4 – The Single-Item-PLUS

In the past week, on how many days have you done a total of 30 minutes or more of physical activity, which was enough to raise your breathing rate?

This may include sport, exercise, and brisk walking or cycling for recreation or to get to and from places, but should not include housework or physical activity that may be part of your job.

Please circle the relevant number:

0 1 2 3 4 5 6 7

If four days of less, have you been physically active for at least two and a half hours (150 minutes) over the course of the past week?

Please circle the relevant answer:

Yes No

Guidance on the analysis of data from the Single-Item-PLUS

Respondents indicating 5 or more days on part one of the Single-Item-PLUS are considered to be meeting recommended physical activity levels (according to the old guidelines) and are therefore considered to be sufficiently active. Respondents who report 4 days or less on part one of the tool are asked to complete part two. Those who answer 'yes' to part two are considered to be sufficiently active, whilst those who respond 'no' to part 2 are considered insufficiently active.

Appendix 5 – The Short Active Lives Survey

1) In the past 7 days, have you done any of these activities? *Please tick the relevant boxes:*
   - ☐ A continuous walk lasting at least 10 minutes
   - ☐ A cycle ride
   - ☐ A sport, fitness activity (such as gym or fitness classes), or dance
   - ☐ None of these

Ask if ticked yes to walking at Q1:

2) In the past 7 days, on how many days did you do a walk lasting at least ten minutes?

Days:

3) How much time did you usually spend walking on each day that you did the activity?

Hours:	Minutes:	per day

4) Was the effort you put into walking usually enough to raise your breathing rate? *Please circle:*

   Yes / No

Ask if ticked yes to a cycle ride at Q1:

5) In the past 7 days, on how many days did you do a cycle ride?

Days:

6) How much time did you usually spend cycling on each day that you did the activity?

Hours:	Minutes:	per day

7) Was the effort you put into cycling usually enough to raise your breathing rate? *Please circle:*

   Yes / No

Ask if ticked yes to sport or fitness activity at Q1:

8) In the past 7 days, on how many days did you do a sport, fitness activity (such as gym or fitness classes), or dance?

Days:

9) How much time did you usually spend doing sport, fitness activities, or dance on each day that you did the activity?

Hours:	Minutes:	per day

10) Was the effort you put into doing sport, fitness activities, or dance usually enough to raise your breathing rate? *Please circle:*

Yes / No

Appendix 6 – Reformatted version of the Short Active Lives Survey for non-web-based administration

1) In the past 7 days, have you done a continuous walk lasting at least 10 minutes? *Please circle:*

   Yes/ No

   If yes:

   a) In the past 7 days, on how many days did you do a walk lasting at least ten minutes? *Please circle:*

      0 1 2 3 4 5 6 7

   b) How much time did you usually spend walking on each day that you did the activity?

      _____ hours and _____ minutes per day

   c) Was the effort you put into walking usually enough to raise your breathing rate? *Please circle*

      Yes No

2) In the past 7 days, have you done a cycle ride? *Please circle:*

   Yes/ No

   If yes:

   a) In the past 7 days, on how many days did you do a cycle ride? *Please circle*

      0 1 2 3 4 5 6 7

   b) How much time did you usually spend cycling on each day that you did the activity?

      _____ hours and _____ minutes per day

   c) Was the effort you put into cycling usually enough to raise your breathing rate? *Please circle*

      Yes No

3) In the past 7 days, have you done sport, fitness activity (such as gym or fitness classes), or dance? *Please circle:*

   Yes/ No

If yes:

a) In the past 7 days, on how many days did you do a sport, fitness activity (such as gym or fitness classes), or dance? *Please circle*

   0 1 2 3 4 5 6 7

b) How much time did you usually spend doing sport, fitness activities, or dance on each day that you did the activity?

   _____ hours and _____ minutes per day

c) Was the effort you put into doing sport, fitness activities, or dance usually enough to raise your breathing rate? *Please circle*

   Yes No

Guidance on the analysis of data from the Short Active Lives Survey

For each activity that respondents indicate was sufficient to raise their breathing rate, the total number of days are multiplied by the usual minutes spent undertaking the activity to give a measure of total minutes over the 7 day period for that activity. The respondent's final score is calculated by summing all activities which were sufficient to raise breathing rate, as follows:

(Days of walking * usual minutes of walking IF sufficient to raise breathing rate) + (Days of cycling * usual munities of cycling IF sufficient to raise breathing rate) + (days of sport, fitness or dance * usual minutes of sport, fitness or dance IF sufficient to raise breathing rate).

Any activities which were not identified as sufficient to increase breathing rate by the respondent are excluded from the calculation.

Appendix 7 – The Single-Item Minutes Based Assessment (SIMBA)

In the past week, how many minutes of physical activity have you done in total, which was enough to raise your breathing rate? Include sport, fitness and recreation activities, and brisk walking or cycling for any purpose, but do not include physical activity that is part of your work. *Please tick the relevant box:*

- ☐ 0 – 29
- ☐ 30 – 59
- ☐ 60 – 89
- ☐ 90 – 119
- ☐ 120 – 149
- ☐ 150+

Guidance on the analysis of data from the Single-Item Minutes Based Assessment

The Single-Item Minutes Based Assessment (SIMBA) uses raw scores, with no computation required.