

Reliability and validity testing of a single-item physical activity measure

K Milton,¹ F C Bull,^{1,2} A Bauman³

¹British Heart Foundation National Centre for Physical Activity and Health, School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

²School of Population Health, University of Western Australia, Perth, Western Australia, Australia

³Centre for Physical Activity and Health, School of Public Health, University of Sydney, Sydney, New South Wales, Australia

Correspondence to

Karen Milton, British Heart Foundation National Centre for Physical Activity and Health, School of Sport, Exercise and Health Sciences, Loughborough University, Leicestershire LE11 3TU, UK; k.milton@lboro.ac.uk

Accepted 3 December 2009

ABSTRACT

Objective To develop and test a new single-item physical activity screening tool, suitable for assessing respondents' eligibility for behaviour change interventions.

Design Two single-item assessment tools were developed, one using a "past week" recall period, the other using a "past month" recall period. A quota sampling system was used to recruit 480 adults from across England, Scotland and Wales. Half the sample completed the past-week question and half completed the past-month version. Test-retest reliability was assessed over a 2- to 5-day period. Concurrent validity was assessed using the Global Physical Activity Questionnaire and the UK Active People Survey. All surveys were completed via telephone interviews.

Results Both versions of the single-item instrument demonstrated strong reproducibility ($r=0.72-0.82$), using Spearman's rank correlation coefficients. The past-week recall question showed strong agreement in the classification of respondents meeting the current physical activity recommendation ($\kappa=0.63$, 95% CI 0.54 to 0.72). Concurrent validity over the past week compared to the Global Physical Activity Questionnaire was modest ($r=0.53$) and slightly weaker for the past month compared to the Active People Survey ($r=0.33-0.48$).

Conclusion Both versions of the new single-item measure performed as well as other short physical activity tools in terms of reliability and concurrent validity. Criterion validity testing of the single-item measure is recommended to establish its ability to assess objectively measured physical activity levels. In addition, further research to assess the responsiveness of the single-item measure in detecting changes in physical activity will inform its usefulness in programme evaluation.

The measurement of physical activity is central to our understanding of physical activity prevalence, as well as the effectiveness of interventions to change physical activity levels. Identifying an appropriate measurement tool is a key priority for policy makers interested in population surveillance and for practitioners interested in programme evaluation and research.¹

Self-report measures of physical activity, such as questionnaires, remain popular because of their ability to capture data from a large number of people at low cost.²

Questionnaires vary greatly in terms of the detail of information that they collect and the way in which physical activity is classified. The choice of questionnaire depends on the primary research questions as well as on available time

and expertise.³ Because of increasing interest in the collection of physical activity data across a broad range of settings (eg, walking schemes, active travel programmes, recreational activities, sports participation), there is a need to establish standardised brief measures, which are feasible for completion when time and resources are limited.

In an attempt to address this need, a number of very short physical activity measures have been developed, some comprising just one single question.⁴⁻¹⁷ Short physical activity tools have predominantly been used for screening purposes to identify whether respondents achieved a specified level of activity to determine their appropriateness for entry into an intervention.

In addition to reducing the respondent burden associated with longer questionnaires, a suitable single-item measure could be collected in comparable ways by different sectors across physical activity programmes. Furthermore, a brief single-item measure could more easily be incorporated into existing routine data collection systems where there is currently an absence of data collection on any physical activity behaviour.

A search for single-item tools was conducted and 14 validated (short and single-item) tools were identified (see table 1). These instruments assessed physical activity in different ways. One approach identified sedentary individuals by asking whether respondents participated in regular physical activity using a binary (yes/no) response scale.^{8,9} An alternative approach identified current physical activity levels by asking respondents to consider whether they were more or less active than their peers.^{11,13} Neither of these types of measures allow the quantification of the amount of physical activity undertaken in terms of frequency, duration or meeting a "health-enhancing threshold".

Another approach used in short measures asks respondents about the frequency of activity that was sufficient to make them sweat.^{6,10,12} However, questions that aimed to capture more details are typically longer, requiring respondents to read various descriptions of physical activity (including concepts of multiple domains and/or intensities) before categorising themselves on a scale.^{5,7} Although described as "single-item", the length and complexity of these instruments may in fact make them more burdensome and time consuming for respondents compared with other longer instruments.

A common feature in this set of short instruments is the attempt to collect physical activity data across multiple domains. Some measures include all domains, whereas others limit the



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://bjism.bmj.com/info/unlocked.dtl>

Table 1 Summary details of short and single-item physical activity measures

Reference	Summary details of item(s)	Test–retest reliability	Criterion or convergent validity
Eaton <i>et al</i> ⁴	“In the past month, how often on the average do you do continuous vigorous exercise for 20 min or more? (This includes activities such as brisk walking, jogging, swimming, bicycling, or aerobic exercise.) Would you say it was <once a week, 1 or 2 times a week, 3 or more times a week, or not at all?”	–	Maximal oxygen consumption r=0.13–0.19 Blood pressure, BMI and HDL cholesterol r=0.09–0.14
Iwai <i>et al</i> ¹⁵	Three questions assess: Average weekly time spent engaging in sports or physical exercise Average time per day spent walking indoors or outside Frequency of engagement in sports or physical exercise. Recall time frame is “over the past year or two”	kappa=0.25– 0.56	Interview data r=0.43–0.60
Jackson <i>et al</i> ⁶	PA5 scale Single-response option chosen from five descriptors of physical activity level based on the stages of change model. ¹⁸ Responses are based on usual behaviour and the length of time that respondents have maintained their current level of physical activity behaviour PA8 scale Single-response option chosen from eight descriptors of physical activity level based on the stages of change model. ¹⁸ Responses are based on usual behaviour and the length of time that respondents have maintained their current level of physical activity behaviour	r=0.81 r=0.84	Fitness test r=0.57 Fitness test r=0.53
Kohl <i>et al</i> ⁶	“How many times a week do you engage in vigorous physical activity long enough to work up a sweat?”	–	Maximal treadmill test r=0.51 Physical Activity Questionnaire r=0.66
Li <i>et al</i> ⁷	Single-response rating of usual physical activity level using a 5-point Likert scale ranging from highly active to inactive. Descriptions of highly active, relatively active and inactive are provided. Responses are based on “usual physical activity”	r=0.88	
Macera <i>et al</i> ⁸	“In the past month, other than your regular job, did you do any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?”	–	More detailed set of questions 5% of patients were misclassified (as sedentary) using the single-question
Marshall <i>et al</i> ^{16*}	Two questions A frequency and duration of vigorous intensity physical activity Frequency and duration of moderate intensity physical activities (including walking). Time frame for both is “usual week”	–	CSA monitor kappa=0.14–0.40
Rose <i>et al</i> ¹⁷	“As a rule, do you do at least half an hour of moderate or vigorous exercise (such as walking or a sport) on five or more days of the week?”	–	NZPAQ-LF kappa=0.56
Schechtman <i>et al</i> ⁹	“Do you currently participate in any regular activity or programme (either on your own or in a formal class) designed to improve or maintain your physical fitness?”	–	Significantly associated with BMI and oxygen capacity and with HDL cholesterol in women
Siconolfi <i>et al</i> ^{10†}	“At least once a week, do you engage in any regular activity similar to brisk walking, jogging, bicycling, etc, long enough to work up a sweat?” Then, if the answer is yes, “How many times per week?”	–	Energy expenditure measured by the Paffenbarger physical activity index questionnaire r=0.57 Maximal oxygen uptake r=0.46
Smith <i>et al</i> ^{14*}	Two question assessment Frequency and duration of vigorous intensity physical activity Frequency and duration of moderate intensity physical activity or walking. Time frame for both is “usual week” Three question assessment Frequency and duration of vigorous intensity physical activity Frequency and duration of moderate intensity physical activity Frequency and duration of walking. Time frame is “usual week”	r=0.61 kappa=58.0 r=0.63 kappa=55.6	Active Australia questionnaire r=0.54, kappa=46.7 CSA accelerometer r=0.39, kappa=18.2 Active Australia questionnaire r=0.43, kappa=38.7 CSA accelerometer r=0.31, kappa=24.3
Sternfeld <i>et al</i> ¹¹	Single question asking women to rate their level of physical activity as “much less”, “less”, “the same as”, “more” or “much more” than other women their age	–	Significantly associated with BMI and self-reported health
Washburn <i>et al</i> ^{12†}	“At least once a week, do you engage in any regular activity long enough to work up a sweat (eg, brisk walking, jogging, bicycling, construction work, etc)? If yes, how many hours per week do you engage in these activities?”	–	Energy expenditure measured by the Paffenbarger Physical Activity Index Questionnaire r=0.39
Weiss <i>et al</i> ¹³	Four questions designed to capture how much hard physical work is required in respondents job or usual daily activity and whether respondents are “more” or “less active” than their peers. Responses are based on “usual physical activity”	–	Energy expenditure calculated from a more detailed set of questions r=0.14–0.41

*Studies involved testing of the same physical activity tool.

†Studies involved testing very similar physical activity tools using different response scale. BMI, body mass index; HDL, high-density lipoprotein; CSA, Computer Science Applications; NZPAQ-LF, New Zealand Physical Activity Questionnaire-Long Form.

focus to certain types of activity. Another variant is the time frame used. Many instruments in table 1 asked about “usual physical activity”; however, because of within-person variation in physical activity behaviour, it is often desirable to quantify activity based on a specific time frame—for example, the past week or the past month.

Although few of the short instruments have been assessed for reliability, most have been assessed for validity. As shown in table 1, criterion validation has used a number of approaches including fitness tests, motion sensors and convergent validity against previously validated self-report tools. The results appear mixed, although these short and single-item measures have generally demonstrated moderate correlations compared with fitness and longer questionnaires, but like many self-report tools, weaker correlations with accelerometers.

In the UK, like elsewhere, the current physical activity recommendations state that adults should undertake at least 30 min of moderate-intensity activity on five or more days of the week. Recent national policy has highlighted the Department of Health’s intention to increase the number of people meeting the national physical activity guidelines, and a large number of initiatives are underway.¹⁹ To assess whether programmes are reaching their target groups, many diverse government agencies and non-governmental organisations are interested in the baseline physical activity levels of participants. Currently, no suitable instrument exists that allows for classification against the current national recommendation of “5 × 30 min moderate intensity activity” using a clearly specified recall period.

The aim of this study was to develop and test a single-item measure to assess physical activity that was suitable for use across a range of intervention and policy settings. The primary research questions were to determine test–retest reliability and concurrent validity of the new single-item measure.

METHODS

The single-item measure was developed after a review of available literature (see table 1) and in consultation with key agencies in the UK interested in addressing the challenges of assessing physical activity in a comparable way as part of their ongoing programme delivery and evaluation. The collaborative group prioritised the need for a short (one to two items) physical activity measure with application across different sectors.

Two versions of the question were developed to meet the contextual requirements of the stakeholders; one version used a “past week” recall period, consistent with other instruments such as the International and Global Physical Activity Questionnaires (IPAQ and GPAQ)^{20 21}; the other version used a “past month” recall period, which was compatible with the current national surveillance tool, the Health Survey for England.²²

Cognitive testing was undertaken to explore comprehension and understanding and to refine the final wording of the single-item measure. Telephone interviews were conducted with a sample of 92 respondents across England, Scotland and Wales. After resolving issues highlighted by the results from the cognitive testing, the final wording of the single question was:

“In the past week/past month, 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”.

An open-response scale was used, with valid responses ranging from 0 to 7 days for the past-week version and from 0 to 31 days for the past-month version.

A quota sampling system was used to recruit 480 adults from across England, Scotland and Wales. A data set of 3000 names, addresses and telephone numbers representing the socioeconomic range across the UK was obtained from ACORN, a geo-demographic tool used to categorise the UK population. The quota sample was balanced on age, sex and socioeconomic groups. A second, additional, sample of 18- to 24-year olds was needed to fill the sample quota because telephone numbers supplied came from landline installations and many younger people utilise mobile phones. Testing was undertaken with a sample of Welsh-speaking respondents, sampled from the 100 postcodes where Welsh was most likely to be spoken at home according to the 2001 census.²³

All testing was undertaken via telephone interviews undertaken by a market research company. Half of each country sample completed the past-week recall question and half completed the past-month recall version. Test–retest reliability assessed the consistency with which respondents answered the single-item measure when administered on two separate occasions, between 2 and 5 days apart. If no repeat contact was made within 5 days, the respondent’s data were not used.

Concurrent validity testing was undertaken with the same sample of respondents who completed the reliability testing. Concurrent validity involved testing whether responses to the single-item measure correlate with responses to a previously validated physical activity measure that is matched on recall period and domains of activity. GPAQ was selected as the comparator for the past-week version because it covers the same domains and asks about days of activity as opposed to total number of times or the number of minutes spent being physically active. GPAQ has demonstrated strong reliability (0.67–0.81) and moderate to strong validity against IPAQ.²⁰ Although validation of GPAQ against objective measures produced weaker results, the magnitude was similar to that observed for other self-report measures.²⁴

The Active People Survey (APS) was selected as the comparison measure for the past-month recall version.²⁵ Previous studies have shown that items from APS have strong reliability.²⁵ Moreover, APS was preferred over other surveys because it is telephone administered and therefore more appropriate for use within the study protocols. APS also included a question assessing the number of “unique” days of activity, which is closely conceptually aligned to the single-item measure.

To assess validity against GPAQ, the number of reported days of walking and/or cycling (item 6a), moderate-intensity activity (item 10a) and vigorous-intensity activity (item 8a) were summed to provide a total score of the number of days of physical activity. Work-related activity (items 2 and 5) was not included as this domain is excluded from the single-item measure.

Concurrent validity using APS was undertaken using two output variables. First, responses on the number of days of moderate- and vigorous-intensity physical activity (including walking, cycling, sport and recreation) were summed to provide a total score of “days of physical activity” (questions 3, 7 and 11). Only moderate-intensity walking recorded as “brisk” or “fast” pace was included. All other activities were treated as at least moderate intensity if the respondent reported that the

activity was sufficient to raise their breathing rate. Second, the single-item measure was tested against responses to one item (question 15) that asked respondents to report the number of separate days on which they had done at least one activity for at least 30 min. This question aimed to capture unique days of physical activity.

Given these data were not normally distributed, Spearman's rank correlation coefficients were used to assess reliability. Kappa and weighted kappa coefficients were calculated to determine agreement between physical activity prevalence at time 1 and time 2 for the past week and past month, respectively. Spearman's rank order correlation was used to assess concurrent validity using data from surveys completed at time 1. For the purposes of this study, 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.²⁶

RESULTS

Testing was completed with a sample of 480 respondents, with half the sample completing each version of the single-item measure. To achieve the quota of 480 completed interviews at time 2 required 522 respondents at time 1, which represents a dropout of 8%. A summary of the demographic characteristics of the final sample is shown in table 2. Country samples were similar in terms of sex and age but varied by socioeconomic group.

Test–retest reliability

For the past-week version, the most frequent response was 0 days (23% at time 1; 26% at time 2) and the least frequent response was 6 days (3% at time 1; 5% at time 2). The proportion of respondents classified as meeting the current physical activity recommendation was 28 and 26% at time 1 and time 2, respectively, demonstrating strong agreement (kappa=0.63, 95% confidence interval 0.54 to 0.72).

Table 2 Demographic characteristics of the sample

	Total sample (n)	Past week sample (n)	Past month sample (n)
Total sample	480	240	240
Country			
England	160	80	80
Scotland	120	60	60
Wales-English	80	40	40
Wales-Welsh	120	60	60
Sex			
Male	239	119	120
Female	241	121	120
Age			
18–24	96	49	47
25–34	98	49	49
35–44	93	46	47
45–54	95	47	48
55–64	98	49	49
SES			
Intermediate	79	39	40
Managerial and professional	123	61	62
Never worked/long-term unemployed	21	10	11
Not classified	179	87	92
Routine and manual	78	43	35

SES, socioeconomic status.

Table 3 reports the correlation between time 1 and time 2 by subgroups. Overall, there was a strong positive correlation ($r=0.72$) between responses to the past-week recall version of the single-item; however, this was influenced by the weaker correlation of 0.37 in the Wales-Welsh subgroup. Recalculating test–retest reliability excluding the Wales-Welsh sample produced an overall pooled test–retest coefficient of 0.86.

Responses to the past-month recall version also showed 0 days of activity to be the most frequently reported at time 1 (22%) and time 2 (18%).

Responses were grouped into four prevalence categories: 0 days of activity, 1–11 days, 12–19 days and 20+ days. Weighted kappa demonstrates strong agreement between the categorical variables (kappa=0.76, confidence interval 0.69 to 0.82). A pooled correlation coefficient of 0.82 indicates strong reproducibility of the data. Results by subgroups are shown in table 3, which shows mostly “good” to “very good” repeatability coefficients.

Concurrent validity

Concurrent validity results are shown in table 4. Concurrent validity of the past-week measure was assessed against GPAQ and results showed a modest correlation of 0.53. For the past month, comparing responses to the APS produced a coefficient of 0.48 for the APS total physical activity score and 0.33 for AP question 15, demonstrating a fair to moderate correlation. Analyses of the concurrent validity data collected at time 2 (data not shown) produced very similar correlations for the GPAQ ($r=0.52$) but slightly stronger correlations for both the APS total physical activity score ($r=0.59$) and question 15 ($r=0.42$).

DISCUSSION

The measurement of physical activity is important for screening for broadly classifying populations as active, for national surveillance and for large-scale programme evaluation.¹ This

Table 3 Test–retest reliability of the single-item by country, sex, age and socioeconomic status

	Past week (r)	Past month (r)
Total sample	0.72	0.82
Country		
England	0.87	0.80
Scotland	0.88	0.93
Wales-English	0.73	0.79
Wales-Welsh	0.37	0.71
Sex		
Male	0.67	0.85
Female	0.77	0.79
Age		
18–24	0.88	0.91
25–34	0.73	0.77
35–44	0.57	0.79
45–54	0.72	0.83
55–64	0.70	0.79
SES		
Intermediate	0.85	0.71
Managerial and professional	0.68	0.86
Never worked/long-term unemployed	0.80	0.85
Not classified	0.67	0.80
Routine and manual	0.72	0.82

SES, socioeconomic status.

Table 4 Concurrent validity of the single-item against GPAQ and APS

	Past week (r)		Past month (r)
	GPAQ	AP total	AP Q15
Total sample	0.53	0.48	0.33
Country			
England	0.55	0.59	0.43
Scotland	0.60	0.46	0.25
Wales-English	0.37	0.38	0.25
Wales-Welsh	0.45	0.40	0.22
Sex			
Male	0.53	0.47	0.30
Female	0.52	0.49	0.33
Age			
18–24	0.62	0.46	0.56
25–34	0.61	0.49	0.26
35–44	0.41	0.40	0.23
45–54	0.65	0.41	0.29
55–64	0.46	0.54	0.26
SES			
Intermediate	0.37	0.30	0.39
Managerial and professional	0.52	0.67	0.21
Never worked/long-term unemployed	0.74	0.59	0.41
Not classified	0.57	0.45	0.33
Routine and manual	0.48	0.43	0.45

APS, Active People Survey; GPAQ, Global Physical Activity Questionnaire; SES, socioeconomic status.

study developed and tested a single-item physical activity measure for use in these diverse settings.

The results of the test–retest reliability revealed correlation coefficients of 0.86 (excluding the Wales-Welsh sample) and 0.82, for the past-week and past-month versions, respectively, indicating very strong repeatability of the single-item measure. Although few studies have assessed the reliability of “single-item” tools, studies of other short surveys report correlations in the range of between 0.61 and 0.88.^{7 14} This new single-item measure therefore demonstrates repeatability comparable to other short measures.

The frequency results from the past-month recall version generally showed higher prevalence of response values which are multiples of four. This suggests that when respondents are asked about the past-month of activity, they may in fact recall a 1-week period and multiply by four. This response pattern, referred to as “estimation”, has been reported in previous research.²⁷ Our results provide further evidence that this approach may have been used, as the maximum number of reported days was 28 despite a month typically consisting of 30 or 31 days.

One unexpected result was the notably weaker reliability of the single-item measure among the Wales-Welsh sample. This was in spite of careful translation and cultural adaptation processes that included back-translation. This result was observed only for the past-week version and not the past-month and is difficult to explain. However, the interview protocols scheduled interviews using the past-week format first, and once completed, the past-month interviews were undertaken. Thus, it is possible that the telephone interviewers were initially not as familiar with the Welsh script, and a learning effect helped to eliminate this problem for the past-month testing. Further testing may be required on the use of the single-item measure in the Welsh language.

We tested concurrent validity of the past-week single-item using GPAQ, which has established measurement properties

and is structured to capture physical activity over the same domains and time frame as the single item.²⁴ The results showed a moderate positive correlation ($r=0.53$). Assessing concurrent validity against established physical activity tools presents several challenges because of variation in the range of activities that are included or excluded from the questions, differences in the way physical activity is quantified and the potential for correlated measurement error of the comparison tools. Overall, it seems that the concurrent validity of the past-week single-item measure was within the range of previous studies, which have assessed the validity of short or single-item measures against longer self-report tools.^{7 13 14 17}

For the past month, validity was tested against the APS. Two approaches were taken and showed different results. Responses to the single-item measure compared with the APS total activity score demonstrated a moderate correlation ($r=0.48$), yet the results using question 15 showed a weaker correlation ($r=0.33$). This was unexpected as question 15 aimed to capture unique days of activity and was conceptually the most closely aligned to the single-item measure. One possible explanation may be because of differences in the domains of physical activity included in each question. The single-item measure asks respondents to consider walking and cycling undertaken for recreation and for travel purposes, whereas APS asks about recreational walking and cycling only. True differences in transport-related activity may therefore explain some of the variation in responses to the two questions. Moreover, it points to the fundamental limitations of attempting concurrent validity and the need to validate self-report measures with objective measurements using accelerometers.

Several limitations to this study should be noted. First, testing was undertaken using telephone interviews only. It is not clear how reliably the single-item measure will capture physical activity levels when administered in other formats—for example, self-completion, face-to-face interview or via a web interface, all of which have potential value for physical activity assessment and surveillance. The single-item measure was designed to capture activity of sufficient intensity to raise breathing rate but did not distinguish between moderate- and vigorous-intensity activity. In addition, the question specified that the activity must be undertaken for a total of 30 min and will not detect days when respondents undertake moderate-vigorous activity which does not total at least 30 min in duration. As with all self-report measurement tools, it is likely that the single-item has some degree of measurement error. Many factors influence the accuracy of self-reported physical activity levels, including the respondent’s ability to accurately recall the behaviour as well as differences in people’s perceptions and interpretations of what is classified as “physical activity”.

CONCLUSIONS

This new single-item physical activity measure has demonstrated strong repeatability and moderate validity, and overall it performed as well or better than other short or single-item physical activity tools with the added advantage of being tailored to the current national recommendations. Our results suggest that this single-item has potential for screening participants for a range of physical activity interventions. Criterion validity testing, using objective measurements, is recommended to further determine the ability of this single-item to accurately assess true physical activity levels. Further research

is also required to determine the ability of the single-item measure to detect changes in physical activity over time and thus provide a useful instrument for monitoring and programme evaluation.

Acknowledgements The single-item measure was developed by the authors in collaboration with representatives from a number of key agencies, who we would like to acknowledge for their contributions: Peter Ashcroft, Natural England; William Bird, Natural England; Tim Buchanan, Sport England; Nick Cavill, Cavill Associates; Andy Cope, Sustrans; Hugo Crombie, National Institute of Health and Clinical Excellence; Stella Goddard, Natural England; Richard Harry, Sports Council Wales; Matthew Lowther, Scottish Executive Health Department; Elaine McNish, Welsh Assembly Government; Lisa Muller, Sustrans; Veronica Reynolds, Walk England; Nick Rowe, Sport England; Graeme Scobie, NHS Health Scotland; Stacy Sharman, Big Lottery Fund; Paul Stonebrook, Department of Health.

Funding This study was collaboratively funded by the Big Lottery Fund, Natural England, NHS Health Scotland, Scottish Government, Sustrans, Sports Council Wales and the Welsh Assembly Government.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Detail has been removed from this case description/these case descriptions to ensure anonymity. The editors and reviewers have seen the detailed information available and are satisfied that the information backs up the case the authors are making.

REFERENCES

1. **Bauman A**, Phongsavan P, Schoeppe S, *et al*. Physical activity measurement—a primer for health promotion. *Promot Educ* 2006;**13**:92–103.
2. **Sallis JF**, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport* 2000;**71**:S1–14.
3. **Shephard RJ**. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med* 2003;**37**:197–206; discussion 206.
4. **Eaton CB**, Lapane KL, Garber CE, *et al*. Physical activity, physical fitness, and coronary heart disease risk factors. *Med Sci Sports Exerc* 1995;**27**:340–6.
5. **Jackson AW**, Morrow JR Jr, Bowles HR, *et al*. Construct validity evidence for single-response items to estimate physical activity levels in large sample studies. *Res Q Exerc Sport* 2007;**78**:24–31.
6. **Kohl HW**, Blair SN, Paffenbarger RS Jr, *et al*. A mail survey of physical activity habits as related to measured physical fitness. *Am J Epidemiol* 1988;**127**:1228–39.
7. **Li S**, Carlson E, Holm K. Validation of a single-item measure of usual physical activity. *Percept Mot Skills* 2000;**91**:593–602.
8. **Macera CA**, Ham SA, Jones DA, *et al*. Limitations on the use of a single screening question to measure sedentary behavior. *Am J Public Health* 2001;**91**:2010–12.
9. **Schechtman KB**, Barzilai B, Rost K, *et al*. Measuring physical activity with a single question. *Am J Public Health* 1991;**81**:771–3.
10. **Siconolfi SF**, Lasater TM, Snow RC, *et al*. Self-reported physical activity compared with maximal oxygen uptake. *Am J Epidemiol* 1985;**122**:101–5.
11. **Sternfeld B**, Cauley J, Harlow S, *et al*. Assessment of physical activity with a single global question in a large, multiethnic sample of midlife women. *Am J Epidemiol* 2000;**152**:678–87.
12. **Washburn RA**, Goldfield SR, Smith KW, *et al*. The validity of self-reported exercise-induced sweating as a measure of physical activity. *Am J Epidemiol* 1990;**132**:107–13.
13. **Weiss TW**, Slater CH, Green LW, *et al*. The validity of single-item, self-assessment questions as measures of adult physical activity. *J Clin Epidemiol* 1990;**43**:1123–9.
14. **Smith BJ**, Marshall AL, Huang N. Screening for physical activity in family practice: evaluation of two brief assessment tools. *Am J Prev Med* 2005;**29**:256–64.
15. **Iwai N**, Hisamichi S, Hayakawa N, *et al*. Validity and reliability of single-item questions about physical activity. *J Epidemiol* 2001;**11**:211–8.
16. **Marshall AL**, Smith BJ, Bauman AE, *et al*. Reliability and validity of a brief physical activity assessment for use by family doctors. *Br J Sports Med* 2005;**39**:294–7; discussion 294–7.
17. **Rose SB**, Elley CR, Lawton BA, *et al*. A single question reliably identifies physically inactive women in primary care. *NZ Med J* 2008;**121**:U2897.
18. **Marcus BH**, Simkin LR. The transtheoretical model: applications to exercise behavior. *Med Sci Sports Exerc* 1994;**26**:1400–4.
19. Department of Health. *Be active be healthy: a plan for getting the nation moving*. London: Department of Health, 2009.
20. **Craig CL**, Marshall AL, Sjöström M, *et al*. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003;**35**:1381–95.
21. World Health Organisation. *Global Physical Activity Questionnaire (GPAQ)*. Geneva: World Health Organisation, 2006.
22. Information Centre for Health and Social Care. *Health survey for England 2007 latest trends*. London: The NHS Information Centre, 2008.
23. **Haselden L**. *Differences in estimates of welsh language skills*. London: Office for National Statistics, 2003.
24. **Bull FC**, Maslin T, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health* 2009;**6**:790–804.
25. Ipsos MORI. *The active people survey technical report*. London: Sport England, 2007.
26. **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
27. **Sudman S**, Bradburn NM, Schwarz N. *Thinking about answers: the application of cognitive processes to survey methodology*. San Francisco, California, USA: Jossey-Bass, 1996.