JECH Online First, published on April 11, 2018 as 10.1136/jech-2017-209703

What they say and what they do: comparing physical activity across the USA, England and the Netherlands

Arie Kapteyn,[1] James Banks,[2] Mark Hamer,[3] James P Smith,[4] Andrew Steptoe,[5] Arthur van Soest,[6] Annemarie Koster,[7] Saw Htay Wah[1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jech-2017-209703).

[1]Center for Economic and Social Research, University of Southern California, Los Angeles, California, USA
[2]University of Manchester and Institute for Fiscal Studies, London, UK
[3]School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK
[4]RAND Corporation, Santa Monica, California, USA
[5]Department of Epidemiology, University College London, London, UK
[6]Department of Econometrics, Tilburg University, Tilburg, the Netherlands
[7]Department of Social Medicine, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, Limburg, the Netherlands

Correspondence to
Professor Arie Kapteyn, University of Southern California, Los Angeles, CA 9007, USA; kapteyn@usc.edu

Received 10 July 2017
Revised 6 January 2018
Accepted 10 January 2018

To cite: Kapteyn A, Banks J, Hamer M, *et al*. *J Epidemiol Community Health* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jech-2017-209703

ABSTRACT

Background Physical activity (PA) is important for maintaining health, but there are fundamental unanswered questions on how best it should be measured.

Methods We measured PA in the Netherlands (n=748), the USA (n=540) and England (n=254), both by a 7 day wrist-worn accelerometer and by self-reports. The self-reports included a global self-report on PA and a report on the frequency of vigorous, moderate and mild activity.

Results The self-reported data showed only minor differences across countries and across groups within countries (such as different age groups or working vs non-working respondents). The accelerometer data, however, showed large differences; the Dutch and English appeared to be much more physically active than Americans h (For instance, among respondents aged 50 years or older 38% of Americans are in the lowest activity quintile of the Dutch distribution). In addition, accelerometer data showed a sharp decline of PA with age, while no such pattern was observed in self-reports. The differences between objective measures and self-reports occurred for both types of self-reports.

Conclusion It is clear that self-reports and objective measures tell vastly different stories, suggesting that across countries people use different response scales when answering questions about how physically active they are.

INTRODUCTION

Physical activity (PA) is a prime component of health behaviour, and accurate measurement is necessary for a better understanding of what drives differences in PA and how PA influences health. Most large-scale studies of population PA including PA at older ages have used self-report questionnaires.[1 2] Valuable though self-report data may be, there are severe limitations to their use. First, responses to questionnaires may suffer from incomplete recall, impaired cognitive ability at older ages and the influence of socially desirable answers, which may vary across place and time. Second, there may be important differences across socioeconomic and demographic groups and places in what is considered PA and vigorous PA in particular. Third, with self-reports alone, it may be difficult to assess PA in different intensities since light-intensity PA is especially hard to measure with questionnaires. With accelerometers, it is possible to better assess PA in different intensities.

Different respondents may also attach different subjective ratings to a given PA level.[3 4] This is particularly likely when using cross-country data as individuals from different countries may have culturally influenced thresholds demarcating response categories commonly found in subjective questions such as 'inactive', 'mildly active', 'moderately active', 'active' and 'very active'.[5]

Self-assessments involve several cognitive processes, including understanding the question asked, recall of relevant information and translation of information into response alternatives offered by the survey administrator.[6] For instance, when asked to report their PA on a five-point scale from 1 being 'inactive' to 5 being 'very active', individuals will first assess their true PA and then translate the assessed activity into what it means to them to be above or below a given threshold (such as 'active' or 'very active'). Individuals may attach different labels to describe the same situation, making it difficult to determine how much of the variation is attributable to true differences and how much is attributable to variations in their subjective thresholds.

One promising way to address the measurement issues inherent in self-reports of PA is to rely on wearable devices such as accelerometers that can objectively measure PA. Accelerometers have been used in several population studies including NHANES[7] and the Health Survey for England,[8] and provide unique information about limitations of self-reported activity data. For example, although patterns across age and gender are qualitatively similar in the USA between the self-reports and accelerometer data, the accelerometer data indicate a much lower adherence to generally recommended PA levels than the self-reports.[7] Similarly, in the UK,[9] activity levels of older people rarely reach generally recommended levels when measured with accelerometers, but do when measured with self-reports.

International comparisons of PA have mostly used standardised questionnaires, such as the International PA Questionnaire.[10 11] When comparing activity levels across European countries, the highest PA level was found in the Netherlands and Germany, while lower PA levels were found for Belgium and Sweden.[12]

The goal of this paper is to compare PA across countries and across different Socio-economic groups within countries, using both self-reports and accelerometry for the same people.

METHODS

Participants and cohorts

Participants were recruited from three nationally representative cohorts; the Dutch Longitudinal

Table 1 Subjective and objective distribution of physical activities: ages 50+

	Subjective			Objective		
	The Netherlands (LISS)	The USA (UAS)	England (ELSA)	The Netherlands (LISS)	The USA (UAS)	England (ELSA)
Inactive (%)	8	10	5	20	38	21
Mildly active (%)	21	32	23	20	22	18
Moderately active (%)	42	34	43	20	14	17
Active (%)	25	20	26	20	11	27
Very active (%)	3	5	3	20	14	17
Observations (n)	447	279	248	447	283	254
	$\chi^2(4)$	P value		$\chi^2(4)$	P value	
LISS vs UAS	13.54	0.0089		35.57	3.6e-07	
LISS vs ELSA	2.29	0.681		5.74	.2194	
UAS vs ELSA	13.42	0.010		32.6855	1.38e-06	

ELSA, English Longitudinal Study of Ageing; LISS, Longitudinal Internet Studies for the Social Sciences; UAS, Understanding America Study.

Internet Studies for the Social Sciences (LISS) panel (https://www.lissdata.nl/lissdata/Home), the Understanding America Study (UAS) (https://uasdata.usc.edu) and the English Longitudinal Study of Ageing (ELSA) (http://www.elsa-project.ac.uk/). The Dutch and US panels interview respondents over the internet. To maintain national representativeness, respondents without prior internet access are provided with a computer or tablet and an internet subscription. ELSA is a nationally representative panel of English respondents aged 50 years and older who are interviewed face-to-face. Respondents from all cohorts were randomly selected to participate in the accelerometer study. The final analytical sample contained 748 participants from LISS, 540 from UAS and 254 from ELSA. Informed consent was obtained from participants.

Physical activity data

PA was assessed objectively with a wrist-worn accelerometer (Geneactiv, UK). We collected data at 50Hz in England and the USA and at 60 Hz in the Netherlands.[i] The data used in the analysis are based on 60 s epochs. Data were stored in gravity (g) units (1 g=9.81 m/s^2). The Euclidean norm of the three raw signals minus 1 g, with negative numbers rounded to zero, was used to quantify acceleration related to the movement registered and expressed in milligravity units.[13] Participants were required to wear the device for 24 hours a day for a minimum of seven consecutive days. Participants who failed to wear the device for at least 10 hours/day were excluded. There was little variation in average wear time across the three countries (23.66 hours/day in LISS, 23.26 hours/day UAS, 23.26 hours/day ELSA). Respondents were asked to report what time they woke up and went to bed, if and how long they were bicycling, and whether they had taken off the accelerometer and for how long.

Participants were also asked to self-report their PA ('Overall, how would you describe your level of PA?'), rated on a five-point scale: inactive, mildly active, moderately active, active and very active. This question was not specifically designed to refer to the period of accelerometer wear, but elicits a general self-evaluation of PA. Furthermore, all three surveys asked for frequency of PA at three levels of intensity. For vigorous, moderate and mild activity, respondents are asked if they engage in them 'hardly ever, or never', 'one to three times a month', 'once a week' or 'more than once a week'.

Covariates

All three datasets contained information on gender, age, marital status, work status, education and ethnicity.

Analyses

We rescaled objective PA measures on a similar five-point scale as used for self-reports. For this purpose, we took the Dutch data as a basis. To translate the objective measure into categorical responses on a five-point scale, we first computed each respondent's average of measured acceleration conditional on having valid wear days over the 7-day period and assigned the label 'inactive' if the 7-day average fell below the 20th percentile, 'mildly active' if the average was between 20th and 40th percentile, 'moderately active' between 40th and 60th percentile, 'active' between 60th and 80th percentile, and 'very active' >80th percentile of the distribution. By construction, 20% of Dutch respondents fell in each of the five activity categories. We compared differences across groups in both PA measures using χ^2 tests. Separate linear regression models were estimated to explore how covariates explained subjective and objective activity measures for all three countries. Linear regressions were redone using ordered probit and ordered logit analyses to check for sensitivity of results to the analysis method adopted. The set of respondent characteristics included marital status (not married is the reference group); gender (male is the reference group); ethnicity (non-white is the reference group) or immigrant status (in the Netherlands, non-native Dutch is the reference group); two dummy variables indicating educational attainment levels (low is the reference group); a dummy variable for working in the labour market (not working is the reference group) and a dummy for being 65 years or older (ages 50–64 are the reference group). To facilitate comparability, virtually all analyses presented in this paper were restricted to the 50 years+ population.

RESULTS

Sixty per cent of our sample was married, and almost half were female. Compared with the Dutch and English samples, the US sample share of the low education group was smaller while that of the high education group was higher (online supplementary table S1).

Compared with the US sample, the Dutch and especially English samples are relatively old: 36% of Americans are among

[i]We used the GENEActiv PC procedure to generate the data. Under this procedure, the individual data points are summed to create the epochs. Since the Dutch data are based on 60Hz, we multiplied the Dutch epochs by 5/6, to make outcomes comparable across the three samples.

the 65+ group, which compares with 48% for Dutch respondents and 62% for English respondents. As a result of these differences in age composition, far fewer respondents in the ELSA sample are working than in either the Dutch or the US sample (30% in ELSA, 34% in LISS and 48% in UAS). The English sample has a high share of respondents with a low education and a low share of high education, which may be partly due to differences in definition of education categories.

Global self-reports and accelerometry

Results from table 1, showing distributions for the 50+ samples, suggest that the Dutch and English are more inclined to stay near the middle of the scale in self-reports: they are somewhat less likely to call themselves either inactive or very active compared with the Americans, but overall differences are modest. In contrast, the right three columns of table 1 show large differences between objectively measured PA of the Dutch and English, compared to the American sample: 38% of Americans would be in the bottom 20% of the Dutch distribution (and 21% of the English). The χ^2 statistics at the bottom provide tests of three bivariate comparisons. The comparisons of self-reports show that the level of PA between the UK and the Netherlands is not significantly different, but both are significantly different from the USA.[ii] In terms of statistical significance, the comparison of the objective data shows a similar pattern, but the p-values of the tests comparing the USA with the Netherlands or England are several orders of magnitude smaller, and indeed the patterns are notably different between the USA and the two European countries.

Next consider breakdowns by salient background characteristics. The breakdown by age in table 2 presents once again a stark contrast between the conclusions based on self-reports and objective measures. In this table, to highlight the age patterns we provide the more complete age distribution statistics for the Dutch and American samples. In England and the Netherlands self-reports on activity do not show an appreciable relation with age, while the American data suggest both a modest increase in the number of inactive respondents and decrease in the number of very active individuals with age. The objective data make clear however that PA drops sharply at older ages in all three countries, while the age patterns in the USA are much steeper than for the self-reported data.

Table 3 shows no appreciable difference in self-reported PA between working and non-working respondents[iii], while the objective measures show highly significant differences in all three countries; working respondents are significantly more active. Once again, individuals' subjective standards of what constitutes PA may vary by their work status with a higher threshold for PA among workers.

This finding may be partly related to the pattern found for different age groups, as older people are less likely to be working and are also less likely to be physically active. To disentangle the effect of the various background characteristics, we therefore turn to regression analysis. Redoing the analysis with ordered probit or ordered logit leads to qualitatively similar conclusions.

Table 4 presents results for separate regressions explaining the subjective and objective activity measures for all three countries, and once again restricting the samples to 50+. Using the Dutch

Table 2 Subjective and objective distributions by age: age 18+

	The Netherlands (LISS)							The USA (UAS)							England (ELSA)					
	Subjective				Objective			Subjective				Objective			Subjective			Objective		
	18–39	40–50	51–64	65+	18–39	40–50	51–64	65+	18–39	40–50	51–64	65+	18–39	40–50	51–64	65+	51–64	65+	51–64	65+
Inactive (%)	5	10	7	9	9	11	15	40	8	7	10	11	21	39	39	60	6	5	22	32
Mildly active (%)	29	20	24	18	9	23	22	24	19	34	39	20	17	12	21	20	24	22	14	20
Moderately active (%)	38	47	41	43	25	20	21	12	37	30	29	42	16	10	13	12	42	44	16	27
Active (%)	25	17	25	26	27	23	18	16	24	21	18	22	22	16	14	3	26	25	26	15
Very active (%)	3	6	3	4	30	23	24	8	12	9	4	5	24	22	13	5	1	5	22	6
Observations (n)	154	145	234	213	155	146	234	213	141	116	180	99	141	116	182	101	95	153	97	157
	χ^2(12)	P value			χ^2(12)	P value			χ^2(12)	P value			χ^2(12)	P value			χ^2(4)	P value	χ^2(4)	P value
	14.72	0.257			111.53	3.0e-18			29.12	0.0038			60.7	1.68e-08			2.82	0.58	20.84	0.00034

ELSA, English Longitudinal Study of Ageing; LISS, Longitudinal Internet Studies for the Social Sciences; UAS, Understanding America Study.

[ii]This conclusion remains true when we correct significance levels for multiple hypothesis testing using Holm-Bonferroni critical values.

[iii]The seemingly significant effect for the USA goes away once we correct for multiple hypothesis testing.

Table 3 Subjective and objective distributions broken down by work status: ages 50+

	The Netherlands (LISS)				The USA (UAS)				England (ELSA)			
	Subjective		Objective		Subjective		Objective		Subjective		Objective	
	Non-working	Working	Non-working	Working	Non-working	Working	Non-working	Working	Non-working	Working	Non-working	Working
Inactive (%)	9	7	27	7	16	4	48	28	6	3	25	13
Mildly active (%)	22	20	23	13	31	33	22	22	23	22	22	8
Moderately active (%)	42	41	18	24	33	34	12	17	45	40	19	13
Active (%)	25	27	18	23	16	23	9	14	23	32	25	33
Very active (%)	3	5	14	33	4	6	10	19	3	4	10	33
Observations (n)	295	152	295	152	143	136	145	138	175	73	178	76
χ^2 (4) (P value)	1.75	(0.78)	46.84	(1.6e-9)	10.82	(0.02)	14.75	(0.005)	3.08	(0.54)	29.20	(7.11e-06)

ELSA, English Longitudinal Study of Ageing; LISS, Longitudinal Internet Studies for the Social Sciences; UAS, Understanding America Study.

sample to illustrate, Dutch respondents show very different patterns in the regression analyses for the subjective and physical measures of PA. The objective measure demonstrates a pronounced decline in PA with age, which is not present at all in the subjective measure. A very similar pattern is seen in the USA and England with the English data showing a significant increase in self-reported PA with age.

The estimates suggest some other country differences that differ depending on whether we are using a subjective or an objective measure of PA. For instance, in the Dutch sample, being female is positively associated with self-reported PA, while in the USA the association is negative. This association disappears in all three countries when we use the objective measure of activity. The relationship of PA with working is stronger with the objective measure than the subjective measure in all three countries. The objective measures do not indicate a relation of PA with education in the Netherlands or the USA. The self-reports suggest a positive association between high education and PA in the Netherlands. In England both self-reports and objective data suggest a positive association with high education, but the effect is twice as large in the objective data.

Self-reported frequencies and accelerometry

Rather than the global self-report question analysed so far, many studies use a more quantitative approach by asking how often one engages in vigorous, moderate or mild activity. In particular, this question asks respondents to report how often they take part in sports or activities that are vigorous, moderate and mild; answers to each question are (1) hardly ever, or never; (2) once to three times a month; (3) once a week and (4) more than once a week. Responses to these questions reveal that among

Table 4 Cross-country regressions on activity: ages 50+

	The Netherlands (LISS)		The USA (UAS)		England (ELSA)	
	Subjective	Objective	Subjective	Objective	Subjective	Objective
Married	0.303	0.221	–0.131	–0.155	0.007	0.019
	(0.0922)***	(0.1318)*	(0.1355)	(0.188)	(0.1240)	(0.183)
Female	0.377	0.009	–0.302	–0.067	–0.061	0.123
	(0.0908)***	(0.1298)	(0.1333)**	(0.185)	(0.1248)	(0.183)
White/Dutch†	–0.050	–0.023	0.286	0.451	0.045	–0.040
	(0.141)	(0.201)	(0.1763)	(0.248)*	(0.329)	(0.491)
Medium	0.042	–0.088	0.037	–0.078	0.043	0.009
	(0.111)	(0.159)	(0.1633)	(0.227)	(0.144)	(0.210)
High	0.418	0.161	0.161	0.003	0.263	0.460
	(0.110)***	(0.157)	(0.1611)	(0.223)	(0.155)*	(0.226)**
Working	0.197	0.602	0.421	0.438	0.319	0.888
	(0.118)*	(0.168)***	(0.1341)***	(0.186)**	(0.148)**	(0.218)***
Age 65+	0.165	–0.555	0.343	–0.531	0.308	–0.034
	(0.113)	(0.161)***	(0.1421)**	(0.196)***	(0.139)**	(0.206)
Constant	2.338	2.918	2.331	2.151	2.627	2.629
	(0.188)***	(0.269)***	(0.2474)***	(0.345)***	(0.347)***	(0.516)***
N	442	442	271	283	247	253
R^2	0.089	0.133	0.084	0.082	0.041	0.106

SEs are in parentheses.
*P<0.10, **P<0.05, ***P<0.01.
†See text for explanation.
E LSA, English Longitudinal Study of Ageing; LISS, Longitudinal Internet Studies for the Social Sciences; UAS, Understanding America Study .

Table 5 Distribution of responses by level of activity (50+)

	Vigorous activity			Moderate activity			Mild activity		
	The Netherlands (LISS)	The USA (UAS)	England (ELSA)	The Netherlands (LISS)	The USA (UAS)	England (ELSA)	The Netherlands (LISS)	The USA (UAS)	England (ELSA)
Hardly ever, or never (%)	45	41	22	22	17	62	17	5	84
One to three times a month (%)	9	17	9	15	20	15	7	11	8
Once a week (%)	18	11	12	22	13	7	15	26	1
More than once a week (%)	29	31	58	41	49	16	60	57	6
	χ^2(3)	P value		χ^2(3)	P value		χ^2(3)	P value	
LISS vs UAS	17.57	0.00054		13.43	0.0038		34.79	1.3e-7	
LISS vs ELSA	59.91	6.1e-13		124.51	8.2e-27		323.05	1.0e-69	
UAS vs ELSA	45.39	7.6e-10		117.48	2.7e-25		353.21	3.0e-76	

ELSA, English Longitudinal Study of Ageing; LISS, Longitudinal Internet Studies for the Social Sciences; UAS, Understanding America Study.

respondents 50+ Americans and Dutch report rather similar frequency and intensity of PA (table 5). However, the English self-reports are quite different. In particular, the English are much more likely to say that they frequently engage in vigorous PA. Consistent with this, the English are much less likely to say that they frequently participate in moderate or mild activity.

The results of OLS regressions of the measured weekly average of accelerometer readings on a set of dummies representing the categories listed in table 5 are shown in online supplementary table S2. First considering the coefficients of the Dutch variables we note that the self-reported frequency of activities does not appear to have a consistent relation with measured PA. For example for vigorous PA, the effect of doing that one to three times a month is about equal to the effect of engaging in vigorous PA more than once a week. For moderate activity, we observe a somewhat larger effect for engaging in moderate activity more than once a week compared with the lower frequencies. For mild activities, there is no clear pattern, although we find a significant negative effect of doing mild activities once a week on measured PA.

The US and England columns show interactions with the dummies for these countries. For both for the USA and England the coefficients for vigorous activity frequency are mostly large and negative. This implies that for the same frequency of reported vigorous activity the objectively measured level of activity is mostly much lower in the USA and England compared with the Netherlands. For moderate activity, the English interactions are once again large and negative. The interactions with the US dummy are not significantly different from zero for this level of activity. For mild activity, the England dummies are not significant, while for the USA they are significant and positive. Altogether, the relation between self-reports and measured PA is unclear, confirming once again the difficulty of measuring differences in PA based on self-reports. The most significant outcome in table S2 is the large negative US dummy, indicating that at the same level of self-reported PA, measured PA is considerably lower.

DISCUSSION

The GENEactive model was chosen on grounds of cost, ease of use and available software.[14] In addition, the wrist-worn accelerometer has several advantages over waist-worn actigraphs: it is easier for participants to wear and avoids errors of positioning. It can be worn continuously, so that movements at night can be detected.

The very high compliance rate in all three cohorts reduces the need for imputation for the time that the device was not worn. Since moreover the wear time was not statistically different across countries, we deemed it unnecessary to impute activity for the limited time the device was not worn. The large differences found between patterns of self-reports and of measured activity across countries or across demographic groups within countries are unlikely to be affected materially by the minor differences in compliance across cohorts.

We have noted a number of demographic and socioeconomic differences between the three samples. Although these differences may affect conclusions about population levels of PA, it should be borne in mind that the main aim of the paper is not to compare PA by country or socioeconomic group per se, but rather to analyse how conclusions may change if we adopt different measures of PA.

It is clear that self-reports and objective measures tell vastly different stories. Both across countries and across different

socioeconomic and demographic groups within countries, self-reports vary only moderately or not at all. At the same time, accelerometry indicates large differences across certain groups. We have found a sharp decline of PA with age, a much higher level of activity in the Netherlands and England than in the USA, and a higher level of activity among working than among non-working respondents. The discrepancy between the objective measures and the global self-reports points at reporting standards for PA that vary across groups. Individuals in different environments and in different age groups simply have different standards of what it means to be physically active. Respondents seem to adjust their standards for what it means to be physically active to their own circumstances, such as age.[15 16] Conceivably, standards are set relative to others in the same age bracket or same demographic group, so that standards vary in proportion to the average level of PA in a group.

The global self-report question ('Overall, how would you describe your level of PA? (1) inactive, (2) mildly active (3) moderately active, (4) active and (5) very active.') would seem to be an obvious approach to comparing PA in different countries. Our analysis suggests that this goal is elusive. The global question is an example of a question using 'vague quantifiers'.[17 18] The criticism often made against such a question is their inherent incomparability: for instance, 'moderate activity' may mean very different things to different people. This does not necessarily mean that quantitative questions are more informative if respondents find it difficult to recall the quantitative information with sufficient accuracy. Indeed measures listed with vague quantifiers are more predictive of target variables of interest than answers to quantitative questions.[19] Similarly, if we run regressions with the self-reports on a five-point scale as presented in table 1, the R^2 is 0.250, which is almost twice as high as the R^2 in online supplementary table S2 (0.136).

Thus, the issue is not that simple self-reports of PA are less reliable than the more detailed questions for frequency of various levels of PA. Rather the problem with both types of questions is that they are understood systematically differently by different groups or by respondents in different countries and hence are unsuitable for use in comparisons across these groups. For that purpose, the use of accelerometry appears indispensable.

What is already known on this subject

- Most large-scale population studies have used self-report questionnaires to assess physical activity.
- There may be important differences in reporting bias across socioeconomic and demographic groups.

What this study adds

- The self-reported physical activity data showed only minor differences across countries.
- In contrast, objective data showed dramatic differences.
- Across countries people use different response scales when answering questions about how physically active they are.

Acknowledgements The authors thank Joris Mulder and Annette Scherpenzeel for designing and managing the data collection among members of the LISS panel and Tania Gutsche, Bas Weerman and Eric Esajian for the design and management of the data collection among members of the Understanding America Study. Margaret Blake, Marta Jackowska and Stephanie Schrempft contributed to the management of data collection in ELSA. The authors also thank Arthur Stone for many helpful comments.

Contributors AK, JB, MH, JPS and AS conceived the study and drafted the paper. SHW and AK carried out the statistical analysis, had full access to the data and take responsibility for the integrity and accuracy of the results. All authors contributed intellectually to refine the study design and to the critical revision of the manuscript.

Funding This research was funded by grants from the National Institute on Aging including R-37AG25529 to JPS at Rand and R01AG20717 to AK at USC. Funding for ELSA was provided by the National Institute of Aging (R01AG017644) and a consortium of UK government departments coordinated by the Economic and Social Research Council.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Ethical approval was gained from the respective ethics committees of each study (ELSA, LISS, UAS).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement ELSA data are publicly available from http://www.data-archive.ac.uk/.

REFERENCES

1 Bauman A, Bull F, Chey T, *et al*. The International Prevalence Study on Physical Activity: results from 20 countries. *Int J Behav Nutr Phys Act* 2009;6:21.
2 Farrell L, Hollingsworth B, Propper C, *et al*. The socioeconomic gradient in physical inactivity: evidence from one million adults in England. *Soc Sci Med* 2014;123:55–63.
3 Hamer M, Kivimaki M, Steptoe A. Longitudinal patterns in physical activity and sedentary behaviour from mid-life to early old age: a substudy of the Whitehall II cohort. *J Epidemiol Community Health* 2012;66:1110–5.
4 Hagströmer M, Troiano RP, Sjöström M, *et al*. Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. *Am J Epidemiol* 2010;171:1055–64.
5 King G, MURRAY CJL, Salomon JA, *et al*. Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am Polit Sci Rev* 2003;97:567–83.
6 Schwarz N. Self-reports: how the questions shape the answers. *Am Psychol* 1999;54:93–105.
7 Troiano RP, Berrigan D, Dodd KW, *et al*. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* 2008;40:181–8.
8 Craig R, Mindell J, Hirani V. *Health Survey for England 2008. Vol 1: Physical activity and fitness*. London, UK: The NHS Information Centre for health and social care, 2009.
9 Harris TJ, Owen CG, Victor CR, *et al*. What factors are associated with physical activity in older people, assessed objectively by accelerometry? *Br J Sports Med* 2009;43:442–50.
10 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.
11 Loyen A, Van Hecke L, Verloigne M, *et al*. Variation in population levels of physical activity in European adults according to cross-European studies: a systematic literature review within DEDIPAC. *Int J Behav Nutr Phys Act* 2016;13:72.
12 Sjöström M, Oja P, Hagströmer M, *et al*. Health-enhancing physical activity across European Union countries: the Eurobarometer study. *J Public Health* 2006;14:291–300.
13 Sabia S, van Hees VT, Shipley MJ, *et al*. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. *Am J Epidemiol* 2014;179:781–90.
14 Esliger DW, Rowlands AV, Hurst TL, *et al*. Validation of the GENEA acelerometer. *Med Sci Sports Exerc* 2011;43:1085–93.
15 Kapteyn A, Smith JP, van Soest AHO. Life satisfaction. In: Diener E, Helliwell JF, Kahneman D, eds. *International differences in well-being*. Oxford: Oxford University Press, 2009:70–104.
16 Baghal TA. Is vague valid? the comparative predictive validity of vague quantifiers and numeric response options. *Survey Research Methods* 2014;8:169–79.
17 Bradburn NM, Miles C. Vague quantifiers. *Public Opin Q* 1979;43:92–101.
18 Schaeffer NC. Hardly ever or constantly? group comparisons using vague quantifier. *Public Opin Q* 1991;55:395–423.
19 Kapteyn A, Smith JP, Soest A. Vignettes and self-reports of work disability in the United States and The Netherlands. *Am Econ Rev* 2007;97:461–73.