

Practice of Epidemiology

Validity and Relative Validity of Alternative Methods of Assessing Physical Activity in Epidemiologic Studies: Findings From the Men's Lifestyle Validation Study

Claire H. Perna, Andrea K. Chomistek, Junaidah B. Barnett, Kerry Ivey, Laila Al-Shaar, Susan B. Roberts, Jennifer Rood, Roger A. Fielding, Jason Block, Ruifeng Li, Walter C. Willett, Giovanni Parmigiani, Edward L. Giovannucci, Lorelei A. Mucci, and Eric B. Rimm*

* Correspondence to Dr. Eric Rimm, Department of Nutrition, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115 (e-mail: erimm@hsph.harvard.edu).

Initially submitted December 18, 2020; accepted for publication March 10, 2022.

In the Men's Lifestyle Validation Study (2011–2013), we examined the validity and relative validity of a physical activity questionnaire (PAQ), a Web-based 24-hour recall (Activities Completed Over Time in 24 Hours (ACT24)), and an accelerometer by multiple comparison methods. Over the course of 1 year, 609 men completed 2 PAQs, two 7-day accelerometer measurements, at least 1 doubly labeled water (DLW) physical activity level (PAL) measurement ($n = 100$ with repeat measurements), and 4 ACT24s; they also measured their resting pulse rate. A subset ($n = 197$) underwent dual-energy x-ray absorptiometry ($n = 99$ with repeated measurements). The method of triads was used to estimate correlations with true activity using DLW PAL, accelerometry, and the PAQ or ACT24 as alternative comparison measures. Estimated correlations of the PAQ with true activity were 0.60 (95% confidence interval (95% CI): 0.52, 0.68) for total activity, 0.69 (95% CI: 0.61, 0.79) for moderate-to-vigorous physical activity (MVPA), and 0.76 (95% CI: 0.62, 0.93) for vigorous activity. Corresponding correlations for total activity were 0.53 (95% CI: 0.45, 0.63) for the average of 4 ACT24s and 0.68 (95% CI: 0.61, 0.75) for accelerometry. Total activity and MVPA measured by PAQ, ACT24, and accelerometry were all significantly correlated with body fat percentage and resting pulse rate, which are physiological indicators of physical activity. Using a combination of comparison methods, we found the PAQ and accelerometry to have moderate validity for assessing physical activity, especially MVPA, in epidemiologic studies.

accelerometry; Activities Completed Over Time in 24 Hours (ACT24); doubly labeled water; dual-energy x-ray absorptiometry; measurement error; physical activity questionnaires; validity

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DLW, doubly labeled water; DXA, dual-energy x-ray absorptiometry; ICC, intraclass correlation coefficient; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PAEE, physical activity energy expenditure; PAL, physical activity level; PAQ, physical activity questionnaire; RPR, resting pulse rate; TDEE, total daily energy expenditure.

Physical activity is associated with lower risk of cardiovascular disease, type 2 diabetes, and some cancers, in addition to disease risk factors such as obesity and high blood pressure (1). Because of its substantial role in health, physical activity is widely studied in epidemiology, requiring assessments that are scalable to large populations and, ideally, repeated over time. Such studies primarily use questionnaires for their ease of administration and capacity to

represent usual or long-term behavior, which is of primary interest in epidemiologic studies (2, 3). Validation studies are needed to evaluate how well assessments reflect true levels of physical activity.

Physical activity has multiple dimensions and domains, each of which may involve distinct or similar biological changes (4, 5). Epidemiologic studies demonstrate associations with health outcomes for total activity volume (6) and

specific activity types (7–9). Since various physical activity assessments capture different activity constructs, no single measure is agreed to be a “gold standard.”

Physical activity validation studies use biomarkers, device-based methods, and self-report assessments as comparison methods (10). Doubly labeled water (DLW), a gold standard measure for assessment of total energy expenditure (11, 12), is often considered optimal for validating self-reported physical activity, although it is costly and one cannot distinguish among its dimensions. Accelerometers measure body movement and can serve as a comparison method, but they have limitations for some activities (e.g., swimming, biking). With declining costs, accelerometers can also be used as a primary activity measure in large-scale studies. Self-report recalls of recent physical activity may also be used as a primary method for physical activity assessment or as a comparison method in validation studies, particularly if repeated measures capture seasonal variation (13). Physical activity induces changes in body composition (14–16) and can lower resting pulse rate (RPR) (17). Thus, DLW and dual-energy x-ray absorptiometry (DXA) body composition assessment and RPR offer indirect, physiologically relevant indicators of physical activity.

In the Health Professionals Follow-up Study, physical activity was assessed biennially beginning in 1986 using a modified Paffenbarger physical activity questionnaire (PAQ) that assesses long-term, habitual activity over the past year. Since the previous validation study (13), the PAQ has been modified, with several specific activities and intensity options added. Our primary objective was to examine the validity of the updated PAQ by comparing it with DLW-determined physical activity level (PAL), accelerometer measures, and multiple Web-based Activities Completed Over Time in 24 Hours (ACT24) recalls. We additionally aimed to assess the validity of the ACT24 and accelerometry as alternative measures of physical activity. We applied the method of triads to estimate the correlation between physical activity measured by each method and the latent true activity level. Because body composition and pulse rate are biological consequences of physical activity, we used DXA and DLW adiposity measures and RPR to assess the relative validity of the PAQ, ACT24, and accelerometer assessments.

METHODS

Study population

The Men’s Lifestyle Validation Study is one of 3 National Cancer Institute–funded validation studies that comprise the Multi-Cohort Eating and Activity Study for Understanding Reporting Error (MEASURE), with the purpose of evaluating the validity of dietary and physical activity assessments (18). Briefly, the Men’s Lifestyle Validation Study included a subset of Health Professionals Follow-up Study participants and men from Greater Boston, Massachusetts, who were members of the Harvard Pilgrim Health Care insurance plan. In total, 671 men consented to participate and completed the study protocol. This analysis included 609 men after excluding those without 2 complete PAQs

($n = 17$) and those without at least 1 complete ACT24 measure ($n = 27$), RPR ($n = 1$), DLW PAL ($n = 13$), and accelerometry ($n = 4$) (total number excluded = 62). The study was approved by the human subjects research committees at Partners Healthcare and Harvard T.H. Chan School of Public Health (Boston, Massachusetts).

Data collection procedures

Data collection for each participant occurred over 12–15 months, between 2011 and 2013. Participants were randomly assigned to one of 4 groups to vary measurement order across four 3-month phases (Figure 1). Participants were asked to maintain their regular PALs throughout the study. The PAQ represents habitual activity over the past year and was administered at baseline and at the end of the study, approximately 1 year later. To capture variation during the year, the ACT24 was administered once per season (4 times in total). Accelerometer measures were obtained twice approximately 6 months apart. To avoid misleadingly high correlations between methods due to proximity in time (19, 20), administration was designed to maintain a 2-week minimum spacing between PAQ, ACT24, and accelerometer measurements. DLW PAL was measured once, at random during the year, with a subset of participants ($n = 100$) providing a repeat measure 9–15 months later. DXA was performed in a subset of participants ($n = 197$), and 99 participants completed a second DXA 12 months later.

PAQ and ACT24

On the PAQ, participants reported the average amount of time (hours/week) they had engaged in specific activities during the past year, using 13 response categories ranging from “none” to “40+ hours” (21). Activities included walking to work or for exercise (including golf), jogging (>10 minutes/mile), running (≤ 10 minutes/mile), bicycling (including a stationary machine), lap swimming, tennis, squash or racquetball, other aerobic exercise (e.g., exercise classes), lower-intensity exercise (e.g., yoga, stretching, or toning), moderate outdoor work (e.g., yardwork or gardening), heavy outdoor work (e.g., digging or chopping), weightlifting (including machines), and standing or walking around at work and at home. Participants indicated activity intensity (low, medium, or high) for swimming, biking, and tennis. Participants also reported their usual walking pace outdoors and the number of flights of stairs climbed daily. Sedentary behaviors included sitting at work or while commuting, sitting at home while watching television/videocassettes/digital video discs, and other sitting at home.

A metabolic equivalent of task (MET) value was assigned to each activity (see Web Table 1, available at <https://doi.org/10.1093/aje/kwac051>) (22). We derived MET-hours per day for each activity by multiplying the MET value by the participant-reported average amount of time using the category midpoint. Total physical activity was defined as the sum of specific MET-hours per week for each activity, including all intensity levels. Activities were also grouped by intensity: vigorous physical activity (≥ 6.0 METs:

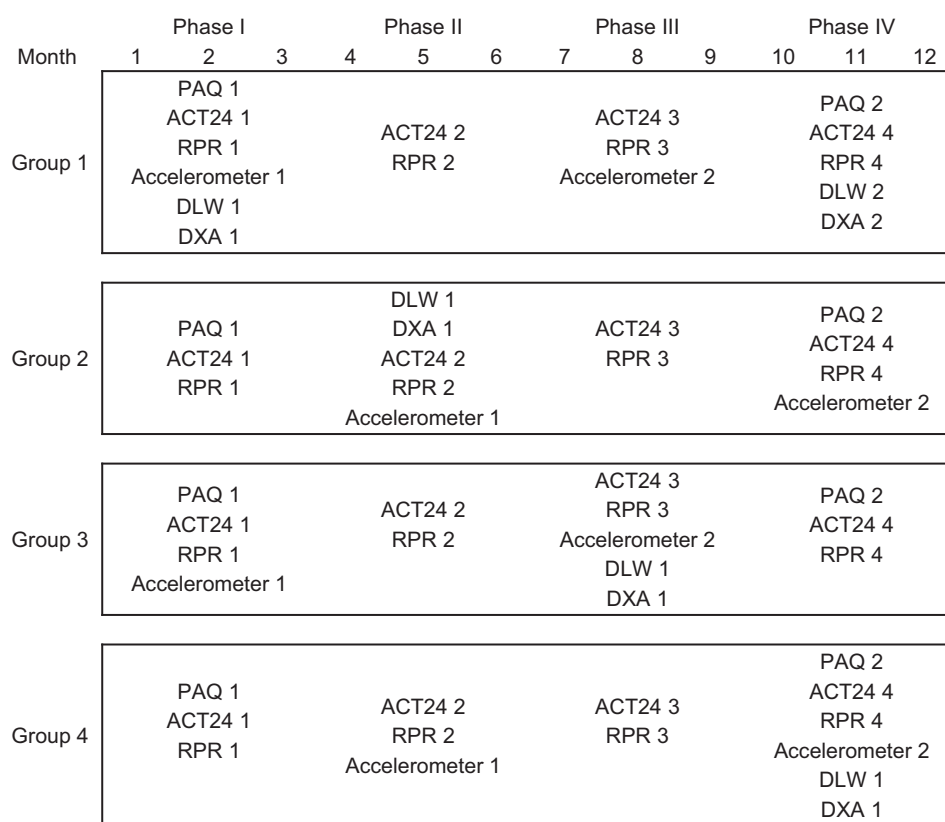


Figure 1. Study timeline and activities according to study group in the Men's Lifestyle Validation Study, Boston, Massachusetts, 2011–2013. The doubly labeled water (DLW), accelerometer, physical activity questionnaire (PAQ), and Activities Completed Over Time in 24 Hours (ACT24) measurements were designed to occur during the same 3-month window (phase) during the study but not during the same week, to avoid artificially high correlations. A subgroup ($n = 100$) of participants in group 1 completed a second DLW measurement at approximately 9, 12, or 15 months. Dual-energy x-ray absorptiometry (DXA) was conducted in participants residing in the Boston area, where scans were performed and scheduled to occur between DLW measurements 1 and 2, or as close as possible to this range. DXA was repeated once in a subset ($n = 99$) of participants in group 1 in month 12. RPR, resting pulse rate.

walking at a pace of ≥ 4 miles/hour (6.4 km/hour), jogging, running, climbing stairs, playing squash/racquetball, and high-intensity bicycling, lap swimming, and tennis), moderate physical activity (3.0–5.9 METs: walking at a pace of 2.0–3.9 miles/hour (3.2–6.2 km/hour), lower-intensity exercise, moderate/heavy outdoor work, other aerobic exercise, weightlifting, and low- to moderate-intensity bicycling, lap swimming, and tennis), and combined moderate-to-vigorous physical activity (MVPA; ≥ 3.0 METs). Sedentary behavior was defined as sitting with a low energy expenditure (< 2.0 METs).

The ACT24 is a Web-based tool for assessing physical activity over the previous 24-hour period (23). Participants reported amounts of time spent sleeping and in active and sedentary behaviors. Participants selected among more than 200 possible activities in four 6-hour blocks and reported the starting and stopping times of activities in ≥ 5 -minute increments. Active behaviors were defined as those that were performed standing upright or that required a higher energy expenditure level (≥ 2.0 METs). Sedentary behaviors were those performed while sitting, reclining, or

lying down and that required a low energy expenditure level (< 2.0 METs). Active behaviors were categorized according to intensity in MET-hours per day using thresholds consistent with the PAQ.

Accelerometry

Participants were provided with an accelerometer (ActiGraph GT3X; ActiGraph Corporation, Pensacola, Florida), detailed instructions for its use, and a wear-time diary. Accelerometer data collection and analytical methods have been described previously (18). We used accelerometer data based on the triaxial vector magnitude. Total activity counts per day were used to represent total physical activity volume. We used a threshold of < 200 counts/minute for sedentary behavior (24), 2,690–6,166 counts/minute for moderate-intensity activity (3.00–5.99 METs), and $\geq 6,167$ counts/minute for vigorous-intensity activity (≥ 6.00 METs) (25). We derived the amount of time spent in each intensity category by summing all minutes spent per day at each intensity level. For sedentary behavior, we also examined

Objective																																																																																																																
Reproducibility	Intraclass Correlation Coefficients																																																																																																															
	<table border="1" style="width: 100%; text-align: center;"> <thead> <tr> <th colspan="2"></th> <th colspan="12">Month</th> </tr> <tr> <th></th> <th></th> <th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th> </tr> </thead> <tbody> <tr> <td>DLW</td> <td>1</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>DLW 2</td> </tr> <tr> <td>DXA</td> <td>1</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>DXA 2</td> </tr> <tr> <td>PAQ</td> <td>1</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>PAQ 2</td> </tr> <tr> <td>ACT24</td> <td>1</td> <td></td><td></td><td></td><td></td><td>ACT24 2</td><td></td><td></td><td>ACT24 3</td><td></td><td></td><td></td><td>ACT24 4</td> </tr> <tr> <td>Accel</td> <td>1</td> <td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>Accel 2</td><td></td><td></td><td></td><td></td> </tr> <tr> <td>RPR</td> <td>1</td> <td></td><td></td><td></td><td></td><td>RPR 2</td><td></td><td></td><td>RPR 3</td><td></td><td></td><td></td><td>RPR 4</td> </tr> </tbody> </table>			Month														1	2	3	4	5	6	7	8	9	10	11	12	DLW	1												DLW 2	DXA	1												DXA 2	PAQ	1												PAQ 2	ACT24	1					ACT24 2			ACT24 3				ACT24 4	Accel	1								Accel 2					RPR	1					RPR 2			RPR 3			
		Month																																																																																																														
		1	2	3	4	5	6	7	8	9	10	11	12																																																																																																			
DLW	1												DLW 2																																																																																																			
DXA	1												DXA 2																																																																																																			
PAQ	1												PAQ 2																																																																																																			
ACT24	1					ACT24 2			ACT24 3				ACT24 4																																																																																																			
Accel	1								Accel 2																																																																																																							
RPR	1					RPR 2			RPR 3				RPR 4																																																																																																			
Validity	Comparison																																																																																																															
	Deattenuated Spearman Correlation Coefficients																																																																																																															
	<table border="1" style="width: 100%; text-align: center;"> <tbody> <tr> <td>PAQ</td> <td>versus</td> <td>DLW PAL Accelerometer ACT24</td> </tr> <tr> <td>ACT24</td> <td>versus</td> <td>DLW PAL Accelerometer</td> </tr> </tbody> </table>	PAQ	versus	DLW PAL Accelerometer ACT24	ACT24	versus	DLW PAL Accelerometer																																																																																																									
PAQ	versus	DLW PAL Accelerometer ACT24																																																																																																														
ACT24	versus	DLW PAL Accelerometer																																																																																																														
Method of Triads																																																																																																																
<table border="1" style="width: 100%; text-align: center;"> <tbody> <tr> <td>PAQ</td> <td>versus</td> <td>True PAL</td> </tr> <tr> <td>ACT24</td> <td></td> <td></td> </tr> <tr> <td>Accelerometer</td> <td></td> <td></td> </tr> <tr> <td>DLW PAL</td> <td></td> <td></td> </tr> </tbody> </table>	PAQ	versus	True PAL	ACT24			Accelerometer			DLW PAL																																																																																																						
PAQ	versus	True PAL																																																																																																														
ACT24																																																																																																																
Accelerometer																																																																																																																
DLW PAL																																																																																																																
Relative Validity	Deattenuated Spearman Correlation Coefficients																																																																																																															
	<table border="1" style="width: 100%; text-align: center;"> <tbody> <tr> <td>PAQ</td> <td>versus</td> <td>DXA Body Fat %</td> </tr> <tr> <td>ACT24</td> <td></td> <td>RPR</td> </tr> <tr> <td>Accelerometer</td> <td></td> <td></td> </tr> <tr> <td>DLW PAL</td> <td></td> <td></td> </tr> </tbody> </table>	PAQ	versus	DXA Body Fat %	ACT24		RPR	Accelerometer			DLW PAL																																																																																																					
	PAQ	versus	DXA Body Fat %																																																																																																													
ACT24		RPR																																																																																																														
Accelerometer																																																																																																																
DLW PAL																																																																																																																
<table border="1" style="width: 100%; text-align: center;"> <tbody> <tr> <td>PAQ</td> <td>versus</td> <td>DLW Body Fat %</td> </tr> <tr> <td>ACT24</td> <td></td> <td></td> </tr> <tr> <td>Accelerometer</td> <td></td> <td></td> </tr> </tbody> </table>	PAQ	versus	DLW Body Fat %	ACT24			Accelerometer																																																																																																									
PAQ	versus	DLW Body Fat %																																																																																																														
ACT24																																																																																																																
Accelerometer																																																																																																																

Figure 2. Study objectives and analytical methods used in the Men’s Lifestyle Validation Study, Boston, Massachusetts, 2011–2013. Accel, accelerometer; ACT24, Activities Completed Over Time in 24 Hours; DLW, doubly labeled water; DXA, dual-energy x-ray absorptiometry; PAQ, physical activity questionnaire; RPR, resting pulse rate.

time spent in ≥15-minute bouts. Accelerometer measures were averaged over the valid number of days worn for each week of wear (Web Appendix).

Biomarker assessments

Total daily energy expenditure (TDEE) was measured from urine samples using the DLW method, as previously described (18). DLW PAL was estimated by dividing DLW TDEE by resting metabolic rate (26). Physical activity energy expenditure (PAEE) was estimated by subtracting resting metabolic rate and the thermic effect of food (10% of total energy) from DLW TDEE (27). Resting metabolic rate was predicted as described by Mifflin et al. (28).

Whole-body DXA was used to assess total body fat percentage, excluding bone mineral density, in a subset of men using a Hologic Discovery QDR x-ray bone densitometer (Hologic, Inc., Bedford, Massachusetts). Scans were per-

formed and analyzed at Tufts Medical Center. Scans were reviewed for quality control and analyzed using Hologic Discovery software (Hologic, Inc.) to obtain body composition measures.

RPR was self-reported 4 times (mean = 3.8 times), approximately once per season, on the PAQ (phases 1 and 4) and using a Web-based survey (phases 2 and 3). Participants reported their RPR in 10 categories ranging from <55 beats/minute to ≥100 beats/minute after sitting for 5 minutes.

Statistical analysis

We calculated mean values and standard deviations for physical activity and body composition variables at each administration. To assess reproducibility, we calculated intraclass correlation coefficients (ICCs) among participants with repeated measures (Figure 2). Data for all measures except body fat percentage were log-transformed to increase

normality. We defined PAQ validity as the Spearman correlation coefficient, with DLW PAL, accelerometer measurements, and the ACT24 used as comparison methods. DLW TDEE and DLW PAEE were used as comparison methods in secondary analyses. The second PAQ was used as the primary comparison to best approximate the period during which comparison methods were administered. We also assessed the validity of the first PAQ and an average of both. To evaluate ACT24 validity, we also calculated correlations of ACT24 with DLW PAL and accelerometry. We assessed individual ACT24s to represent a typical measurement in an epidemiologic study and the average of up to 4 (mean = 3.5) ACT24s to represent activity throughout the year. The validity of the accelerometer was assessed using method-of-triads analysis. To assess the relative validity of the PAQ, ACT24, and accelerometer to capture biological responses to physical activity, we used correlations of these methods with RPR and DXA and DLW body fat percentage. Individual accelerometer measures and the average of up to 2 (mean = 1.8) measures were used. Men who reported taking antihypertensive medication were excluded ($n = 130$) from RPR analyses, as some of these medications affect pulse rate.

Spearman correlations were adjusted for age across all comparisons by using residuals from linear regression on age. Accelerometer measures were additionally adjusted for wear time. We obtained deattenuated correlation coefficients to adjust for random within-person variation in the comparison methods by using information from individuals with repeated DLW measurements ($n = 100$ with 2 replicates), DXA ($n = 99$ with 2 replicates), RPR ($n = 479$ with up to 4 replicates), accelerometry ($n = 516$ with 2 replicates), and ACT24 ($n = 578$ with up to 4 replicates) (Web Appendix (29)). For the comparison with DLW PAL, we examined subgroups of age (<70 years, ≥ 70 years) and body mass index (weight (kg)/height (m)²; <25.0, ≥ 25.0) using median values as cutoffs.

We applied the method of triads to estimate correlations with the latent, true physical activity variable for each method using validity coefficients and 95% confidence intervals (CIs) as described previously (Web Appendix) (30, 31). We used data from 3 pairwise correlations between the questionnaire (second PAQ or average of ACT24s), the mean of 2 accelerometer measures, and DLW PAL. The method of triads assumes that errors in the 3 methods are uncorrelated, which is a reasonable assumption because the technologies are completely different and rely on different types of information. In addition, we ensured by design that the different methods were administered at different times over the 1-year period to avoid spuriously high correlations. We further assumed positive linear correlations between the assessment methods and the true PALs (30).

RESULTS

Descriptive statistics

In the Men's Lifestyle Validation Study overall, participants had a mean age of 68.1 (standard deviation, 7.6) years and a mean baseline body mass index of 26.1 (standard deviation, 3.7) (Table 1). Participants who completed at least

1 DXA scan and repeated DLW measurements were younger on average, had a higher DLW PAL, and were more likely to be Harvard Pilgrim Health Care enrollees than participants overall. Harvard Pilgrim enrollees were younger on average and more likely to be African-American (Web Table 2).

The distributions of energy expenditure, body composition, and physical activity variables overall (Table 2) and by age and body mass index subgroup (Web Table 3) are presented. On average, PAQ measures were lower than ACT24 measures. Compared with accelerometry, activity was higher according to the ACT24 and PAQ for moderate activity, vigorous activity, and MVPA. Accelerometer-determined sedentary time was more similar to PAQ when considering ≥ 15 -minute bouts and more similar to ACT24 when including every minute assessed.

Reproducibility

PAQ measurements representing the past year administered approximately 12 months apart had high reproducibility, with the highest ICC being 0.75 for vigorous activity and the lowest being 0.50 for sedentary time (Table 2). Two 7-day accelerometer measurements and 2 DLW measurements had similarly high reproducibility, with ICCs of 0.71–0.79. DXA body fat percentage assessed 12 months apart demonstrated the highest reproducibility, with an ICC of 0.96. ACT24 recalls of the past day, completed up to 4 times over 12 months, had the lowest reproducibility, with ICCs of 0.22–0.43.

Validity

DLW as the comparison method. Deattenuated correlation coefficients for correlation between the age-adjusted second PAQ and DLW PAL were 0.44 (95% CI: 0.36, 0.51) for total activity, 0.47 (95% CI: 0.38, 0.54) for MVPA, 0.39 (95% CI: 0.31, 0.47) for vigorous activity, 0.27 (95% CI: 0.18, 0.36) for moderate activity, and -0.16 (95% CI: -0.25 , -0.07) for sedentary time (Table 3). These correlations were similar when comparing PAQ with DLW PAEE (Web Table 4). Correlations of PAQ total activity and MVPA with DLW PAL were stronger among men under age 70 years than among older men and were similar between body mass index subgroups (Web Tables 5 and 6). Correlations between age-adjusted averaged ACT24s and DLW PAL ranged from 0.27 to 0.42 for active categories, similar to the PAQ. For sedentary time, we observed a correlation of age-adjusted averaged ACT24 with DLW PAL of -0.23 (95% CI: -0.32 , -0.14), slightly stronger than for PAQ. The second PAQ, capturing recalled activity over the same time period as the comparison methods, performed similarly to the baseline PAQ capturing the previous year.

Accelerometry as the comparison method. Correlations between adjusted accelerometer measures and the second PAQ were 0.43 (95% CI: 0.36, 0.50) for total activity, 0.41 (95% CI: 0.34, 0.48) for MVPA, 0.35 (95% CI: 0.26, 0.42) for vigorous activity, and 0.37 (95% CI: 0.31, 0.44) for moderate activity (Table 3). Correlations were similar when comparing averaged ACT24 active categories with

Table 1. Characteristics of Participants, Overall and by Dual-Energy X-Ray Absorptiometry and Doubly Labeled Water Subset, in the Men's Lifestyle Validation Study, Boston, Massachusetts, 2011–2013

Variable	Total (n = 609)		Completed 1 DXA Scan ^a (n = 197)		Completed Repeat DLW Measurements ^b (n = 100)	
	Mean (SD)	%	Mean (SD)	%	Mean (SD)	%
Age, years	68.1 (7.6)		61 (8.1)		62.4 (9.1)	
Height, m	1.8 (0.1)		1.8 (0.1)		1.8 (0.1)	
Weight, kg	81.7 (12.3)		83.1 (12.4)		83 (12.8)	
Weight change, kg ^c	−0.04 (2.67)		0.15 (2.94)		0.0 (3.4)	
Body mass index ^d	26.1 (3.7)		26.5 (3.7)		26.5 (3.7)	
Current smoker ^e		1.2		1.0		1.0
African-American race/ethnicity		2.3		4.6		4.0
Harvard Pilgrim Health Care member		28.2		85.3		72.0
Second PAQ						
Total activity, MET-hours/day	10.4 (6.6)		9.7 (6.9)		9.4 (7.3)	
MVPA, MET-hours/day	7.4 (5.5)		6.9 (5.5)		6.9 (5.7)	
Sedentary time, hours/day	4.6 (2.9)		5.5 (3.2)		5.2 (3)	
First DLW PAL	1.7 (0.2)		1.8 (0.3)		1.8 (0.3)	

Abbreviations: DLW, doubly labeled water; DXA, dual-energy x-ray absorptiometry; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PAL, physical activity level; PAQ, physical activity questionnaire; SD, standard deviation.

^a DXA was conducted in a subset of participants residing in the Boston area (n = 197).

^b A subgroup (n = 100) of participants in group 1 completed a second DLW measurement at approximately 9, 12, or 15 months.

^c Weight change was calculated as the difference in reported weight between administration of the first and second PAQs; data were missing for 1 participant.

^d Weight (kg)/height (m)².

^e Data on current smoking were missing for 5 men.

accelerometry, except that the vigorous activity correlation of 0.13 (95% CI: 0.05, 0.22) was much lower. For sedentary time, correlations of PAQ 2 and averaged ACT24 with accelerometry were 0.24 (95% CI: 0.16, 0.32) and 0.37 (95% CI: 0.29, 0.43), respectively. Correlations were stronger for the averaged ACT24s than for a single measurement (Web Tables 7 and 8).

ACT24 as the comparison method. When ACT24 served as the comparison method, correlations were markedly stronger after deattenuation to account for within-person variation in repeated ACT24 assessments (Table 4). Deattenuated correlations between the age-adjusted second PAQ and ACT24 were highest in magnitude for vigorous activity, at 0.72, and lowest for sedentary time, at 0.26 (Table 4).

Relative validity using body fat percentage and RPR as biological responses. Total activity, MVPA, and vigorous activity measured by the PAQ, ACT24, and accelerometer were significantly predictive of lower DLW and DXA body fat percentage and lower RPR (Table 5; Web Tables 9 and 10). Compared with other methods, DLW PAL was most strongly correlated with DXA body fat percentage. Accelerometer-measured sedentary time was more strongly predictive of body fat than the PAQ and ACT24, while

PAQ-measured MVPA predicted body fat more strongly than accelerometry and ACT24. Vigorous activity had the highest inverse correlations with RPR, and correlations were strongest for the second PAQ ($r = -0.33$, 95% CI: -0.40 , -0.24), followed by accelerometry ($r = -0.26$, 95% CI: -0.34 , -0.17) and averaged ACT24s ($r = -0.21$, 95% CI: -0.29 , -0.11).

Estimated correlations with true PAL. Using the method of triads with DLW PAL serving as a biomarker and accelerometry as the reference method, validity coefficients comparing PAQ-measured activity with true activity were 0.60 for total activity, 0.69 for MVPA, 0.76 for vigorous activity, and 0.52 for moderate activity (Table 6). Validity coefficients comparing accelerometry with true activity were 0.68 for total activity, 0.53 for MVPA, 0.38 for vigorous activity, and 0.64 for moderate activity. Validity coefficients comparing DLW PAL with true activity were 0.70 for total activity, 0.64 for MVPA, 0.49 for vigorous activity, and 0.50 for moderate activity. Correlations were similar when DLW PAEE and body-weight-adjusted DLW PAEE served as biomarkers (Web Tables 11 and 12). ACT24 performance as compared with true activity was similar to the PAQ, with slightly lower correlations for MVPA and vigorous activity.

Table 2. Mean Values and Intraclass Correlation Coefficients for Physical Activity and Body Composition Variables in the Men's Lifestyle Validation Study, Boston, Massachusetts, 2011–2013^a

Variable and Measurement No.	No. of Participants	Mean (SD)	ICC	95% CI
Predicted RMR, kcal/day ^b				
RMR 1	609	1,589 (155)		
RMR 2	100	1,626 (153)		
Accelerometer wear time, hours/day				
Accelerometer 1	609	15.1 (1.2)		
Accelerometer 2	516	15.1 (1.2)		
Physical activity level				
DLW 1	609	1.7 (0.2)	0.72	0.63, 0.79
DLW 2	100	1.8 (0.3)		
PAEE, kcal/day				
DLW 1	609	902 (335)	0.74	0.66, 0.81
DLW 2	100	1,036 (388)		
TDEE, kcal/day				
DLW 1	609	2,767 (429)	0.79	0.73, 0.85
DLW 2	100	2,957 (482)		
Body fat, %				
DLW				
DLW 1	609	27.9 (6.6)	0.91	0.88, 0.93
DLW 2	100	26.8 (6.1)		
DXA				
DXA scan 1	197	23.2 (5.2)	0.96	0.94, 0.97
DXA scan 2	99	23.1 (4.9)		
Total physical activity ^c				
PAQ, MET-hours/day				
PAQ 1	609	10.6 (11.2)	0.62	0.57, 0.67
PAQ 2	609	10.4 (6.6)		
Average PAQ ^d	609	10.5 (7.6)		
ACT24, MET-hours/day				
ACT24 1	540	18.8 (11.6)	0.43	0.39, 0.48
ACT24 2	524	16.5 (9.8)		
ACT24 3	528	15.3 (9.9)		
ACT24 4	540	16.2 (9.9)		
Average ACT24 ^e	609	17.0 (7.9)		
Accelerometer, TAC/day				
Accelerometer 1	609	596,032 (193,831)	0.76	0.73, 0.80
Accelerometer 2	516	574,118 (186,905)		
Average accelerometer ^f	609	587,021 (181,709)		

Table continues

DISCUSSION

Our study examined the reproducibility and validity of self-reported physical activity assessed by questionnaire (PAQ) using DLW, accelerometry, ACT24 recalls, RPR, and DXA as comparison methods in a cohort of men aged 46–

82 years. Repeated measurements allowed adjustment for random within-person variation of each comparison method. We observed moderate correlations of PAQ-measured total activity and MVPA with DLW PAL, ACT24, and accelerometer measures. The 3 alternative methods—PAQ, ACT24, and accelerometry—all demonstrated moderate validity in

Table 2. Continued

Variable and Measurement No.	No. of Participants	Mean (SD)	ICC	95% CI
MVPA				
PAQ, MET-hours/day			0.68	0.63, 0.72
PAQ 1	609	7.4 (9.3)		
PAQ 2	609	7.4 (5.5)		
Average PAQ	609	7.4 (6.4)		
ACT24, MET-hours/day			0.28	0.24, 0.33
ACT24 1	540	12.6 (11.7)		
ACT24 2	524	10.3 (10.1)		
ACT24 3	528	9.5 (9.9)		
ACT24 4	540	10.0 (9.9)		
Average ACT24	609	10.7 (7.7)		
Accelerometer, hours/day ^g			0.76	0.72, 0.79
Accelerometer 1	609	0.85 (0.51)		
Accelerometer 2	516	0.79 (0.48)		
Average accelerometer	609	0.83 (0.47)		
Vigorous activity				
PAQ, MET-hours/day			0.75	0.71, 0.78
PAQ 1	609	2.6 (3.6)		
PAQ 2	609	2.7 (3.5)		
Average PAQ	609	2.6 (3.3)		
ACT24, MET-hours/day			0.26	0.21, 0.31
ACT24 1	540	4.2 (8.9)		
ACT24 2	524	3.2 (7.0)		
ACT24 3	528	2.6 (6.7)		
ACT24 4	540	2.6 (6.1)		
Average ACT24	609	3.3 (5.2)		
Accelerometer, hours/day ^g			0.73	0.69, 0.77
Accelerometer 1	609	0.07 (0.14)		
Accelerometer 2	516	0.07 (0.13)		
Average accelerometer	609	0.07 (0.13)		
Moderate activity				
PAQ, MET-hours/day			0.54	0.48, 0.59
PAQ 1	609	4.5 (4.2)		
PAQ 2	609	4.7 (4.2)		
Average PAQ	609	4.6 (3.7)		
ACT24, MET-hours/day			0.22	0.17, 0.27
ACT24 1	540	8.4 (8.2)		
ACT24 2	524	7.1 (7.9)		
ACT24 3	528	6.8 (7.9)		
ACT24 4	540	7.3 (8.0)		
Average ACT24	609	7.4 (5.6)		
Accelerometer, hours/day ^g			0.74	0.70, 0.78
Accelerometer 1	609	0.78 (0.46)		
Accelerometer 2	516	0.72 (0.44)		
Average accelerometer	609	0.76 (0.43)		

Table continues

Table 2. Continued

Variable and Measurement No.	No. of Participants	Mean (SD)	ICC	95% CI
Sedentary time, hours/day				
PAQ			0.50	0.44, 0.56
PAQ 1	609	4.8 (3.0)		
PAQ 2	609	4.6 (2.9)		
Average PAQ	609	4.7 (2.6)		
ACT24			0.38	0.33, 0.44
ACT24 1	540	8.9 (3.4)		
ACT24 2	524	9.0 (3.2)		
ACT24 3	528	9.1 (3.2)		
ACT24 4	540	8.8 (3.3)		
Average ACT24	609	8.9 (2.6)		
Accelerometer				
Including bouts \geq 1 minute			0.71	0.67, 0.75
Accelerometer 1	609	8.5 (1.5)		
Accelerometer 2	516	8.6 (1.5)		
Average accelerometer	609	8.5 (1.4)		
Including bouts \geq 15 minutes			0.73	0.69, 0.77
Accelerometer 1	609	4.5 (1.6)		
Accelerometer 2	516	4.7 (1.6)		
Average accelerometer	609	4.6 (1.5)		
RPR, beats/minute ^h				
RPR 1	464	63.7 (8.6)	0.67	0.63, 0.70
RPR 2	469	63.0 (7.9)		
RPR 3	407	63.3 (8.3)		
RPR 4	468	63.6 (8.5)		
Average RPR	479	63.4 (7.2)		

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DXA, dual-energy x-ray absorptiometry; ICC, intraclass correlation coefficient; MET, metabolic equivalent of task; PAEE, physical activity energy expenditure; RMR, resting metabolic rate; RPR, resting pulse rate; PAQ, physical activity questionnaire; SD, standard deviation; TAC, total activity count; TDEE, total daily energy expenditure.

^a The ICCs represent correlations between administrations of the PAQ, the ACT24, and the accelerometer that reflect activity performed over the past year, the past 24 hours, and the current week, respectively.

^b RMR was predicted on the basis of age, sex, height, and weight as described by Mifflin et al. (28).

^c Total activity determined by PAQ and ACT24 was based on active MET-hours/day (i.e., not including sedentary behavior).

^d Average PAQ indicates the average of the first and second questionnaires.

^e Average ACT24 indicates the average of up to four 24-hour recalls.

^f Average accelerometer indicates the average of up to 2 accelerometer measurements.

^g MVPA, vigorous activity, and moderate activity were calculated on the basis of bouts \geq 1 minute in duration.

^h Men who reported taking antihypertensive medication at baseline were excluded.

comparison with true activity levels, with correlations ranging from 0.53 to 0.71. DLW PAL, PAQ, ACT24, and accelerometer-measured activity significantly predicted DXA body fat percentage and RPR, with PAQ correlations at least as high as those for the other methods. Additionally, the method of triads demonstrated moderate validity of PAQ-measured total activity, MVPA, and vigorous activity compared with true activity, having estimated correlations with true activity that were similar to or greater

than those for methods that have been considered gold standards.

Our validation study uniquely used 6 different methods to assess physical activity and body composition in a large population of men with heterogeneous age and body mass index. The high reproducibility of the PAQs observed in our study is consistent with other studies (32–34). A systematic review showed that only a limited number of existing PAQs had acceptable validity when compared with accelerometry

Table 3. Spearman Correlation Coefficients Comparing a Physical Activity Questionnaire and the ACT24 Web-Based 24-Hour Recall With Doubly Labeled Water–Determined Physical Activity Level and Corresponding Accelerometer-Determined Activity in the Men's Lifestyle Validation Study ($n = 609$), Boston, Massachusetts, 2011–2013^a

Variable and Measurement No.	DLW-Determined PAL ^b			Accelerometer ^c		
	Age-Adjusted <i>r</i>	Deattenuated <i>r</i>	95% CI	Age-Adjusted <i>r</i>	Deattenuated <i>r</i>	95% CI
Total activity, MET-hours/day						
PAQ 1	0.35	0.41	0.33, 0.49	0.35	0.38	0.31, 0.45
PAQ 2	0.38	0.44	0.36, 0.51	0.40	0.43	0.36, 0.50
Average PAQ ^d	0.40	0.46	0.37, 0.54	0.41	0.44	0.37, 0.51
ACT24 2	0.29	0.33	0.24, 0.41	0.34	0.37	0.29, 0.44
Average ACT24 ^e	0.36	0.42	0.34, 0.50	0.39	0.43	0.36, 0.49
MVPA, MET-hours/day						
PAQ 1	0.38	0.45	0.37, 0.53	0.32	0.36	0.28, 0.43
PAQ 2	0.40	0.47	0.38, 0.54	0.37	0.41	0.34, 0.48
Average PAQ	0.42	0.50	0.41, 0.57	0.38	0.41	0.34, 0.48
ACT24 2	0.28	0.33	0.23, 0.41	0.29	0.32	0.23, 0.39
Average ACT24	0.37	0.42	0.34, 0.50	0.32	0.36	0.28, 0.42
Vigorous activity, MET-hours/day						
PAQ 1	0.34	0.40	0.32, 0.48	0.27	0.31	0.23, 0.38
PAQ 2	0.33	0.39	0.31, 0.47	0.30	0.35	0.26, 0.42
Average PAQ	0.35	0.41	0.33, 0.49	0.30	0.34	0.26, 0.42
ACT24 2	0.18	0.21	0.11, 0.31	0.06	0.08	−0.02, 0.16
Average ACT24	0.30	0.34	0.26, 0.42	0.11	0.13	0.05, 0.22
Moderate activity, MET-hours/day						
PAQ 1	0.19	0.22	0.12, 0.31	0.31	0.33	0.26, 0.40
PAQ 2	0.23	0.27	0.18, 0.36	0.34	0.37	0.30, 0.44
Average PAQ	0.23	0.27	0.18, 0.36	0.35	0.39	0.31, 0.45
ACT24 2	0.17	0.20	0.11, 0.29	0.25	0.27	0.18, 0.35
Average ACT24	0.23	0.27	0.18, 0.36	0.27	0.29	0.21, 0.37
Sedentary time, hours/day						
PAQ 1	−0.11	−0.13	−0.21, −0.04	0.23	0.25	0.17, 0.33
PAQ 2	−0.14	−0.16	−0.25, −0.07	0.22	0.24	0.16, 0.32
Average PAQ	−0.14	−0.16	−0.25, −0.08	0.26	0.28	0.20, 0.36
ACT24 2	−0.15	−0.17	−0.27, −0.08	0.26	0.28	0.20, 0.37
Average ACT24	−0.20	−0.23	−0.32, −0.14	0.34	0.37	0.29, 0.43

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DLW, doubly labeled water; MVPA, moderate-to-vigorous physical activity; PAL, physical activity level; PAQ, physical activity questionnaire.

^a $n = 524$ for the second ACT24 recall.

^b PAL was estimated from DLW total daily energy expenditure divided by resting metabolic rate. Resting metabolic rate was predicted on the basis of age, sex, height, and weight as described by Mifflin et al. (28).

^c Measurements were based on the triaxial vector magnitude. Total activity was based on total activity counts per day; accelerometer measures of MVPA, vigorous, and moderate activity included ≥ 1 -minute bouts; sedentary time was based on ≥ 15 -minute bouts. Values from accelerometry were adjusted for age (years) and accelerometer wear time (hours/day).

^d Average PAQ indicates the average of the first and second questionnaires.

^e Average ACT24 indicates the average of up to four 24-hour recalls.

or DLW energy expenditure (32). Of 65 included studies, there was a median validity correlation of 0.30 among adults and 0.40 among the elderly. These correlations are similar to

what we observed when using a single comparison method, but the previous studies have generally not considered errors in the comparison methods. Consistent with our study,

Table 4. Spearman Correlation Coefficients Comparing Physical Activity Measured by a Physical Activity Questionnaire With the Average of Four 24-Hour Recalls of Activity in the Men's Lifestyle Validation Study ($n = 609$), Boston, Massachusetts, 2011–2013

Variable and Measurement No.	ACT24 Recall ^a		
	Age-Adjusted r	Deattenuated Age-Adjusted r	95% CI
Total activity, MET-hours/day			
PAQ 1	0.44	0.53	0.44, 0.60
PAQ 2	0.43	0.52	0.43, 0.59
Average PAQ ^b	0.49	0.58	0.49, 0.65
MVPA, MET-hours/day			
PAQ 1	0.47	0.58	0.48, 0.66
PAQ 2	0.43	0.56	0.46, 0.64
Average PAQ	0.49	0.62	0.52, 0.69
Vigorous activity, MET-hours/day			
PAQ 1	0.41	0.68	0.51, 0.81
PAQ 2	0.42	0.72	0.55, 0.83
Average PAQ	0.44	0.74	0.57, 0.86
Moderate activity, MET-hours/day			
PAQ 1	0.38	0.50	0.39, 0.59
PAQ 2	0.37	0.51	0.40, 0.60
Average PAQ	0.41	0.56	0.45, 0.64
Sedentary time, hours/day			
PAQ 1	0.30	0.36	0.28, 0.44
PAQ 2	0.22	0.26	0.17, 0.35
Average PAQ	0.30	0.36	0.28, 0.44

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; PAQ, physical activity questionnaire.

^a Up to 4 ACT24 measurements were used.

^b Average PAQ indicates the average of the first and second questionnaires.

previous validation studies found lower validity correlations for sedentary behavior compared with active behavior assessed by PAQ (32).

Previously, in a subset of 238 Health Professionals Follow-up Study participants, our research group conducted a validation study of PAQs compared with 4 weekly physical activity diaries across different seasons, and they observed correlations for inactivity, nonvigorous activity, and vigorous activity of 0.41, 0.28, and 0.58, respectively (13). These correlations for activity were lower than those observed in the current study comparing the PAQ with ACT24s. Despite potential for overstating validity due to correlated errors between the PAQ and 24-hour recall methods, the similar deattenuated correlations observed when comparing the PAQ with ACT24s or when using the method of triads to estimate correlations with true activity suggested that correlated errors were not substantial. Furthermore, these methods also have different sources of error, given that they rely on different types of memory and involve different procedures. Nevertheless, some degree of

correlated errors cannot be excluded, and the method of triads is sensitive to such errors (35).

When comparing the PAQ and ACT24 with DLW PAL, we observed the strongest correlations for total activity and MVPA. This was consistent with our expectations, since PAL represents all energy expenditure due to physical activity rather than any one particular type. In a previous study, Matthews et al. (23) found a high level of accuracy when comparing PAEE estimated from multiple ACT24s with that from DLW. Correlations of PAQ-measured MVPA with DLW PAL in the present study were similar to those of accelerometry with DLW energy expenditure in a prior analysis in the Men's Lifestyle Validation Study (18).

When comparing PAQ and ACT24 with accelerometry, correlations were similar to those with DLW PAL as the comparison method. The strongest correlations were observed for total activity and MVPA. An advantage of accelerometers is their ability to capture short activity increments, including ≥ 1 -minute bouts. PAQ- and accelerometer-determined sedentary time were more similar when using

Table 5. Spearman Correlation Coefficients Comparing a Physical Activity Questionnaire, Multiple ACT24 Recalls, and Accelerometer Measurements With Resting Pulse Rate and Body Fat Percentage Measured by Doubly Labeled Water in the Men's Lifestyle Validation Study, Boston, Massachusetts, 2011–2013

Variable and Measurement No.	RPR (<i>n</i> = 479) ^a			DLW Body Fat % (<i>n</i> = 609) ^a		
	Age-Adjusted <i>r</i> ^b	Deattenuated Adjusted <i>r</i> ^b	95% CI	Age-Adjusted <i>r</i> ^b	Deattenuated Adjusted <i>r</i> ^b	95% CI
DLW PAL 1	−0.19	−0.21	−0.30, −0.11			
Total activity						
PAQ 1	−0.21	−0.22	−0.31, −0.13	−0.27	−0.27	−0.35, −0.20
PAQ 2	−0.19	−0.18	−0.27, −0.09	−0.27	−0.28	−0.35, −0.20
Average PAQ ^c	−0.20	−0.21	−0.30, −0.12	−0.29	−0.29	−0.37, −0.22
ACT24 2	−0.11	−0.11	−0.21, −0.02	−0.16	−0.17	−0.25, −0.09
Average ACT24 ^d	−0.15	−0.15	−0.24, −0.07	−0.21	−0.22	−0.30, −0.14
Average accelerometer ^e	−0.19	−0.20	−0.29, −0.11	−0.34	−0.35	−0.42, −0.28
MVPA						
PAQ 1	−0.25	−0.26	−0.35, −0.17	−0.32	−0.33	−0.39, −0.25
PAQ 2	−0.28	−0.26	−0.35, −0.17	−0.33	−0.34	−0.40, −0.27
Average PAQ	−0.27	−0.28	−0.37, −0.19	−0.35	−0.35	−0.42, −0.28
ACT24 2	−0.09	−0.09	−0.19, 0.00	−0.13	−0.14	−0.22, −0.05
Average ACT24	−0.18	−0.18	−0.26, −0.09	−0.20	−0.21	−0.28, −0.12
Average accelerometer	−0.20	−0.21	−0.3, −0.11	−0.29	−0.30	−0.37, −0.22
Vigorous activity						
PAQ 1	−0.28	−0.29	−0.37, −0.20	−0.37	−0.38	−0.44, −0.31
PAQ 2	−0.31	−0.33	−0.40, −0.24	−0.35	−0.37	−0.43, −0.30
Average PAQ	−0.32	−0.34	−0.41, −0.25	−0.37	−0.39	−0.45, −0.32
ACT24 2	−0.05	−0.05	−0.15, 0.05	−0.11	−0.11	−0.19, −0.03
Average ACT24	−0.20	−0.21	−0.29, −0.11	−0.23	−0.24	−0.31, −0.16
Average accelerometer	−0.25	−0.26	−0.34, −0.17	−0.27	−0.28	−0.36, −0.20
Moderate activity						
PAQ 1	−0.05	−0.05	−0.13, 0.05	−0.12	−0.12	−0.20, −0.04
PAQ 2	−0.61	−0.02	−0.11, 0.07	−0.16	−0.16	−0.24, −0.09
Average PAQ	−0.06	−0.04	−0.12, 0.06	−0.14	−0.15	−0.23, −0.07
ACT24 2	0.01	0.02	−0.08, 0.11	−0.03	−0.03	−0.11, 0.05
Average ACT24	−0.17	−0.05	−0.14, 0.04	−0.09	−0.09	−0.17, −0.01
Average accelerometer	−0.16	−0.17	−0.26, −0.08	−0.26	−0.27	−0.35, −0.19
Sedentary time						
PAQ 1	0.03	0.04	−0.06, 0.12	0.03	0.03	−0.05, 0.11
PAQ 2	0.07	0.07	−0.02, 0.16	0.05	0.05	−0.03, 0.12
Average PAQ	0.00	0.00	−0.10, 0.09	0.01	0.01	−0.08, 0.09
ACT24 2	0.04	0.04	−0.06, 0.13	0.06	0.07	−0.02, 0.15
Average ACT24	0.08	0.08	−0.01, 0.17	0.10	0.10	0.02, 0.18
Average accelerometer	0.11	0.09	0.00, 0.18	0.31	0.32	0.25, 0.39

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DLW, doubly labeled water; MVPA, moderate-to-vigorous physical activity; PAL, physical activity level; PAQ, physical activity questionnaire; RPR, resting pulse rate.

^a For ACT24 2, *n* = 414 in the RPR analysis and *n* = 524 in DLW analysis.

^b Accelerometer measures additionally adjusted for wear time (hours/day).

^c Average PAQ indicates the average of the first and second questionnaires.

^d Average ACT24 indicates the average of up to four 24-hour recalls.

^e Average accelerometer indicates the average of up to 2 accelerometer measurements.

Table 6. Validity Coefficients^a Showing Correlations With True Physical Activity for Physical Activity as Estimated From the Second Physical Activity Questionnaire, the ACT24, Accelerometry, and Doubly Labeled Water–Determined Physical Activity Level in the Men's Lifestyle Validation Study ($n = 516$), Boston, Massachusetts, 2011–2013^b

Variable	Spearman's Correlation Coefficient ^c			Validity Coefficient					
				Q-T		R-T		M-T	
	r_{QR}	r_{QM}	r_{RM}	VC _{QT}	95% CI	VC _{RT}	95% CI	VC _{MT}	95% CI
Q = PAQ 2									
Total activity	0.40	0.42	0.47	0.60	0.52, 0.68	0.68	0.61, 0.75	0.70	0.62, 0.78
MVPA	0.37	0.45	0.34	0.69	0.61, 0.79	0.53	0.45, 0.63	0.64	0.55, 0.75
Vigorous activity	0.29	0.37	0.18	0.76	0.62, 0.93	0.38	0.29, 0.48	0.49	0.38, 0.63
Moderate activity	0.34	0.26	0.32	0.52	0.43, 0.64	0.64	0.53, 0.78	0.50	0.40, 0.62
Q = average ACT24									
Total activity	0.38	0.35	0.47	0.53	0.45, 0.63	0.71	0.63, 0.80	0.67	0.58, 0.76
MVPA	0.32	0.36	0.34	0.58	0.49, 0.69	0.55	0.46, 0.66	0.62	0.53, 0.73
Vigorous activity	0.12	0.28	0.18	0.42	0.29, 0.62	0.28	0.19, 0.43	0.65	0.46, 0.92
Moderate activity	0.27	0.24	0.32	0.45	0.36, 0.58	0.60	0.48, 0.75	0.53	0.42, 0.67

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DLW, doubly labeled water; MVPA, moderate-to-vigorous physical activity; PAL, physical activity level; PAQ, physical activity questionnaire; VC, validity coefficient.

^a Q = second PAQ adjusted for age or the average of up to 4 ACT24 recalls adjusted for age; R = accelerometer-determined activity adjusted for age and wear time; M = DLW-determined PAL adjusted for age; T = true physical activity.

^b This analysis was restricted to men with 2 accelerometer measurements.

^c Coefficient for correlation between the PAQ 2 or average ACT24 and the accelerometer (r_{QR}), the PAQ 2 or average ACT24 and DLW PAL (r_{QM}), or the accelerometer and DLW PAL (r_{RM}).

accelerometer bouts of ≥ 15 minutes. This suggests that the PAQ may better capture longer bouts of sedentary time as compared with 1-minute bouts. Comparing PAQ and ACT24, we observed higher absolute activity based on ACT24s. The ACT24 provides more activity response options and assesses activities of daily living more broadly than the PAQ. We observed relatively low but statistically significant correlations when comparing PAQ, ACT24, DLW, and accelerometer measures with body fat and RPR; these are useful for evaluation of relative validity for assessment of physical activity that is physiologically relevant. Our results are consistent with previous studies showing that MVPA but not sedentary time is associated with body fat percentage (36, 37). In our study, PAQ-measured MVPA tended to predict body fat and RPR more strongly than accelerometer-measured activity, while DLW PAL was most strongly predictive of DXA body fat. This is in contrast with a UK Biobank study finding that associations of accelerometer-measured activity with adiposity were 2-fold larger than those of questionnaire-measured activity (38). However, the adapted version of the International Physical Activity Questionnaire used in the UK Biobank study only assessed 3 intensities of activity.

Our method of evaluating validity was based on rankings of PAL, not absolute measures. Our finding that the PAQ tended to provide higher absolute values of physical activity time than accelerometer measures should be considered in making specific recommendations regarding activity times.

However, thresholds for activity levels using accelerometer counts have varied across studies. Furthermore, the method-of-triads analyses suggested that accelerometer-measured vigorous activity did not perform well in comparison with true activity. Future study would be needed to calibrate the PAQ or ACT24 to absolute levels of physical activity with consideration of the scale of measurement and the optimal comparison method. While calibration is desirable, it would not affect statistical power or change the ability to observe associations using the PAQ or ACT24.

Multiple factors have been proposed to explain discrepancies between PAQs and comparison methods (12, 39). First, different methods may not capture activity over the same time periods. Our study design included careful timing of assessments to represent the same period of physical activity as much as possible and to avoid correlated errors due to administration of multiple short-term measures close in time. Nonetheless, as reflected in our reproducibility estimates, appreciable within-person variation in DLW PAL and accelerometer measures was present, and neither directly measures activity over the past year as the PAQ does. Second, correlations are influenced by MET values assigned to activities reported by both PAQ and ACT24, which could potentially induce correlated errors. For accelerometry, intensity-level thresholds used to categorize data may influence observed correlations with this method. Third, neither the PAQ nor the accelerometer captures the full range of physical activities. Our current PAQ

includes 15 active behaviors and 3 sedentary behaviors, primarily in the leisure domain. The waist-worn ActiGraph GT3X may not fully capture activities such as cycling, carrying objects, or arm movements, and it cannot be worn during water activities. Correlations with DLW are affected by errors due to variations in energy balance and homeostatic control mechanisms, variation in the dietary carbohydrate:fat ratio, biological differences in absorption and metabolism, and technical laboratory measurements (20, 40).

This validation study expanded substantially on our previous study (13) in that it included a larger population and combined multiple types of assessments, including measures of energy expenditure and body composition that are considered gold standards (20) and that can be reasonably considered to have no correlated errors with self-reported activity. However, errors in the comparison methods are to be expected, and correlated errors between methods cannot be ruled out. The multiple assessment methods used in this study enabled us to utilize 2 approaches to estimate the validity of the PAQ as compared with true activity levels: 1) Spearman correlations adjusting for random error in a single comparison method and 2) validity coefficients using the method of triads. In the first approach, nonrandom errors in the comparison method could lead to an underestimate of validity. In the second approach, correlated errors between methods could lead to an overestimate of validity. Therefore, we consider the correlations derived using the method of triads an upper “boundary” of validity and the correlation with DLW or accelerometry alone a lower “boundary.” The method-of-triads analyses suggested that DLW PAL and accelerometer measures of physical activity have appreciable error and may not be substantially superior to the PAQ or ACT24 for assessing physical activity. Implications include the possibility that the validity of physical activity may have been underestimated in previous studies using these methods alone as comparisons.

Another limitation of this study is that these methods may perform differently in other contexts or study populations. Therefore, our findings may not be generalizable to women or to populations with a different distribution of age, educational level, occupation, or race/ethnicity.

In summary, we found that physical activity assessed by our updated questionnaire was moderately correlated with physical activity assessed by DLW or accelerometry, and that it had moderate estimated correlations with true activity levels using a combination of methods. These correlations with true activity were similar to those with multiple 24-hour recalls using the ACT24, suggesting that multiple recalls can serve as a reasonable comparison method in validation studies of PAQs if associations are adjusted for within-person variation. Furthermore, correlations with true activity suggest that the optimal standard may be a combination of methods rather than 1 alone. Although biomarker or device-based measures of physical activity are useful in validation studies, they are far more expensive than self-report measures and do not directly assess activity type. Importantly, questionnaires provide information beyond that gleaned from accelerometry or biomarker assessments, including specific activities and domains. This information is relevant to understanding

behavior and informing public health messaging around physical activity. Our study showed that the PAQ performs particularly well when measuring MVPA, an important factor in many chronic diseases. Given its low cost, acceptability, ease of administration, and ability to capture physical activity repeatedly over the long term, the PAQ is appropriate and provides useful information for large observational studies of chronic disease risk.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Claire H. Pernar, Walter C. Willett, Edward L. Giovannucci, Lorelei A. Mucci, Eric B. Rimm); Department of Epidemiology and Biostatistics, School of Public Health, Indiana University, Bloomington, Indiana, United States (Andrea K. Chomistek); Jean Mayer USDA Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States (Junaidah B. Barnett, Susan B. Roberts, Roger A. Fielding); Department of Nutrition, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Junaidah B. Barnett, Kerry Ivey, Laila Al-Shaar, Ruifeng Li, Walter C. Willett, Edward L. Giovannucci, Eric B. Rimm); South Australian Health and Medical Research Institute, School of Medicine, Flinders University, Adelaide, South Australia, Australia (Kerry Ivey); Department of Nutrition and Dietetics, College of Nursing and Health Sciences, Flinders University, Adelaide, South Australia, Australia (Kerry Ivey); Department of Epidemiology, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania, United States (Laila Al-Shaar); Department of Nutrition, Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts, United States (Susan B. Roberts, Roger A. Fielding); Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, Louisiana, United States (Jennifer Rood); Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts, United States (Jason Block); Channing Division of Network Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts, United States (Walter C. Willett, Edward L. Giovannucci, Lorelei A. Mucci, Eric B. Rimm); Department of Data Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts, United States (Giovanni Parmigiani); and Department of Biostatistics, T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, United States (Giovanni Parmigiani).

This work was funded by the National Institutes of Health (grants T32 ES 007069 and T32 CA 009001 to C.H.P.) and the National Cancer Institute (Cancer Center Support Grant 5P30CA006516-53). The Men’s Lifestyle Validation Study and the Health Professionals Follow-up Study are supported by grants U01CA152904 and U01CA167552 from the National Cancer Institute.

The data set analyzed in the current study is not publicly available, but limited data can be accessed from the corresponding author upon request.

We thank Dr. Heather Bowles for her comments and contributions to this work. We thank the participants and staff of the Men's Lifestyle Validation Study for their valuable contributions.

This work was presented at the 51st Annual Meeting of the Society for Epidemiologic Research, Baltimore, Maryland, June 19–22, 2018.

Conflict of interest: none declared.

REFERENCES

- World Health Organization. *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. Geneva, Switzerland: World Health Organization; 2009.
- Hu FB, Stampfer MJ, Colditz GA, et al. Physical activity and risk of stroke in women. *JAMA*. 2000;283(22):2961–2967.
- Huerta JM, Chirlaque MD, Tormo MJ, et al. Work, household, and leisure-time physical activity and risk of mortality in the EPIC-Spain cohort. *Prev Med*. 2016;85:106–112.
- Pettee Gabriel KK, Morrow JR Jr, Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. *J Phys Act Health*. 2012;9(suppl 1):S11–S18.
- Mann S, Beedie C, Jimenez A. Differential effects of aerobic exercise, resistance training and combined exercise modalities on cholesterol and the lipid profile: review, synthesis and recommendations. *Sports Med*. 2014;44(2):211–221.
- Wolff-Hughes DL, Fitzhugh EC, Bassett DR, et al. Total activity counts and bouts minutes of moderate-to-vigorous physical activity: relationships with cardiometabolic biomarkers using 2003–2006 NHANES. *J Phys Act Health*. 2015;12(5):694–700.
- Manson JE, Hu FB, Rich-Edwards JW, et al. A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. *N Engl J Med*. 1999;341(9):650–658.
- Chomistek AK, Cook NR, Flint AJ, et al. Vigorous-intensity leisure-time physical activity and risk of major chronic disease in men. *Med Sci Sports Exerc*. 2012;44(10):1898–1905.
- Yates T, Zaccardi F, Dhalwani NN, et al. Association of walking pace and handgrip strength with all-cause, cardiovascular, and cancer mortality: a UK Biobank observational study. *Eur Heart J*. 2017;38(43):3232–3240.
- Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37(3):197–206.
- Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *J Nutr*. 1988;118(11):1278–1289.
- Neilson HK, Robson PJ, Friedenreich CM, et al. Estimating activity energy expenditure: how valid are physical activity questionnaires? *Am J Clin Nutr*. 2008;87(2):279–291.
- Chasan-Taber S, Rimm EB, Stampfer MJ, et al. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology*. 1996;7(1):81–86.
- Bann D, Kuh D, Wills AK, et al. Physical activity across adulthood in relation to fat and lean body mass in early old age: findings from the Medical Research Council National Survey of Health and Development, 1946–2010. *Am J Epidemiol*. 2014;179(10):1197–1207.
- den Hoed M, Westerterp KR. Body composition is associated with physical activity in daily life as measured using a triaxial accelerometer in both men and women. *Int J Obes (Lond)*. 2008;32(8):1264–1270.
- Ekelund U, Besson H, Luan J, et al. Physical activity and gain in abdominal adiposity and body weight: prospective cohort study in 288,498 men and women. *Am J Clin Nutr*. 2011;93(4):826–835.
- Carter JB, Banister EW, Blaber AP. Effect of endurance exercise on autonomic control of heart rate. *Sports Med*. 2003;33(1):33–46.
- Chomistek AK, Yuan C, Matthews CE, et al. Physical activity assessment with the ActiGraph GT3X and doubly labeled water. *Med Sci Sports Exerc*. 2017;49(9):1935–1944.
- Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *Am J Epidemiol*. 2014;180(2):172–188.
- Willett W. *Nutritional Epidemiology*. 3rd ed. New York, NY: Oxford University Press; 2013.
- Harvard T.H. Chan School of Public Health. HPFS questionnaires. <http://sites.sph.harvard.edu/hpfs/questionnaires/>. Accessed February 23, 2022.
- Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc*. 2011;43(8):1575–1581.
- Matthews CE, Keadle SK, Moore SC, et al. Measurement of active and sedentary behavior in context of large epidemiologic studies. *Med Sci Sports Exerc*. 2018;50(2):266–276.
- Aguilar-Farías N, Brown WJ, Peeters GM. ActiGraph GT3X+ cut-points for identifying sedentary behaviour in older adults in free-living environments. *J Sci Med Sport*. 2014;17(3):293–299.
- Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport*. 2011;14(5):411–416.
- Westerterp KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Front Physiol*. 2013;4:90.
- Colbert LH, Matthews CE, Havighurst TC, et al. Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc*. 2011;43(5):867–876.
- Mifflin MD, St Jeor ST, Hill LA, et al. A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr*. 1990;51(2):241–247.
- Rosner B, Glynn RJ. Interval estimation for rank correlation coefficients based on the probit transformation with extension to measurement error correction of correlated ranked data. *Stat Med*. 2007;26(3):633–646.
- Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr*. 1997;65(4 suppl):1232S–1239S.
- Al-Shaar L, Yuan C, Rosner B, et al. Reproducibility and validity of a semi-quantitative food frequency questionnaire in men assessed by multiple methods. *Am J Epidemiol*. 2021;190(6):1122–1132.
- Helmerhorst HJ, Brage S, Warren J, et al. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int J Behav Nutr Phys Act*. 2012;9:103.

33. Silsbury Z, Goldsmith R, Rushton A. Systematic review of the measurement properties of self-report physical activity questionnaires in healthy adult populations. *BMJ Open*. 2015;5(9):e008430.
34. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol*. 1994;23(5):991–999.
35. Gormley IC, Bai Y, Brennan L. Combining biomarker and self-reported dietary intake data: a review of the state of the art and an exposition of concepts. *Stat Methods Med Res*. 2020;29(2):617–635.
36. Drenowatz C, Hill JO, Peters JC, et al. The association of change in physical activity and body weight in the regulation of total energy expenditure. *Eur J Clin Nutr*. 2017;71(3):377–382.
37. Bowen L, Taylor AE, Sullivan R, et al. Associations between diet, physical activity and body fat distribution: a cross sectional study in an Indian population. *BMC Public Health*. 2015;15:281.
38. Guo W, Key TJ, Reeves GK. Accelerometer compared with questionnaire measures of physical activity in relation to body size and composition: a large cross-sectional analysis of UK Biobank. *BMJ Open*. 2019;9(1):e024206.
39. Plasqui G, Bonomi AG, Westerterp KR. Daily physical activity assessment with accelerometers: new insights and validation studies. *Obes Rev*. 2013;14(6):451–462.
40. Trabulsi J, Troiano RP, Subar AF, et al. Precision of the doubly labeled water method in a large-scale application: evaluation of a streamlined-dosing protocol in the Observing Protein and Energy Nutrition (OPEN) Study. *Eur J Clin Nutr*. 2003;57(11):1370–1377.