

Practice of Epidemiology

Reproducibility, Validity, and Relative Validity of Self-Report Methods for Assessing Physical Activity in Epidemiologic Studies: Findings From the Women's Lifestyle Validation Study

Laila Al-Shaar, Claire H. Pernar, Andrea K. Chomistek, Eric B. Rimm, Jennifer Rood, Meir J. Stampfer, A. Heather Eliassen, Junaidah B. Barnett, and Walter C. Willett*

* Correspondence to Dr. Walter C. Willett, Department of Nutrition and Department of Epidemiology, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115 (e-mail: wwillett@hsph.harvard.edu).

Initially submitted April 8, 2021; accepted for publication December 1, 2021.

Among 683 participants in the Women's Lifestyle Validation Study (2010–2012), we evaluated the performance of a self-administered physical activity questionnaire (PAQ) and Web-based 24-hour recalls (Activities Completed Over Time in 24 Hours (ACT24)) using multiple comparison methods. Two PAQs, 4 ACT24s, two 7-day accelerometer measurements, 1 doubly labeled water (DLW) physical activity level (PAL) measure (repeated; $n = 90$), and 4 resting pulse rate measurements were collected over 15 months. The deattenuated correlation between the PAQ and DLW PAL was 0.41 (95% confidence interval (CI): 0.33, 0.49) for total physical activity (PA) and 0.40 (95% CI: 0.31, 0.48) for moderate-to-vigorous PA (MVPA). These correlations were similar when using accelerometry as the comparison method. Single and averaged ACT24 measurements had lower correlations with DLW and accelerometry as comparison methods. The PAQ showed inverse correlations with DLW body fat percentage and resting pulse rate. Using the method of triads, the estimated correlation of the PAQ with true total PA was 0.54 (95% CI: 0.47, 0.62) and that with true MVPA was 0.60 (95% CI: 0.52, 0.69). For averaged ACT24, the estimated correlations were 0.50 (95% CI: 0.43, 0.59) for total PA and 0.47 (95% CI: 0.39, 0.58) for MVPA, and for averaged accelerometry, these estimated correlations were 0.72 (95% CI: 0.64, 0.81) and 0.62 (95% CI: 0.53, 0.71), respectively. The PAQ provided reasonable validity for total PA and MVPA.

accelerometry; ACT24; doubly labeled water; physical activity questionnaires; reproducibility; validity

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; BMI, body mass index; CI, confidence interval; DLW, doubly labeled water; 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; TEE, total energy expenditure.

Physical activity has many health benefits, including lower risks of coronary heart disease (1), stroke (2), type 2 diabetes (3), hypertension (4), and obesity (5), some specific types of cancer (6), and cognitive decline (7). Some inconsistencies across studies may be related to the instrument used for measuring physical activity. Measurement errors inherent to these assessment tools could introduce bias in estimating the associations, depending on the type and magnitude of these errors. Self-reported measurements taken via physical activity questionnaires (PAQs) or 24-hour recalls are commonly used in prospective cohort studies because of their relatively low cost and ability to provide data on different types and

domains of physical activity (8), but they are inevitably imperfect because they rely on participants' perception and memory (9). After accounting for other determinants of energy expenditure, doubly labeled water (DLW) has been considered a reference method for measuring total energy expenditure (TEE), but it is expensive and cannot determine the type, intensity, frequency, or duration of activities. DLW is also considered an alloyed gold standard, since it includes both analytical and physiological errors (9). Alternatively, daily records or diaries have been widely used as reference methods, but these require several weeks of recording to capture seasonal variation and long-term physical activity

patterns, and participants may change their physical activity because of their data recording (10). Measures of fitness such as resting pulse rate (RPR) have also been used to validate PAQs (11), but fitness also has a genetic component. The use of accelerometers as a reference method has increased due to technological advances, but there is still a debate about the appropriate metrics for analyzing accelerometer data, and some forms of activity are poorly detected by these devices (12). Given these challenges, comparisons of PAQs and 24-hour recalls with biomarkers and device-based measures of physical activity can provide useful data on the validity and relative validity of each method.

We aimed to evaluate the validity and reproducibility of our PAQ and the self-administered Activities Completed Over Time in 24 Hours (ACT24) assessment tool among participants in the Women's Lifestyle Validation Study (2010–2012), by comparing them with physical activity levels (PALs) measured via DLW and 7-day accelerometry. We also examined physical activity assessed by different methods in relation to RPR and body fat mass percentage, which are biological consequences of physical activity, to assess the relative validity of these methods.

METHODS

Study population

The Women's Lifestyle Validation Study is one of the 3 components of the Multi-Cohort Eating and Activity Study for Understanding Reporting Error (MEASURE) (13). The study was conducted within the Nurses' Health Study and Nurses' Health Study II between 2010 and 2012. Recruitment details are provided in the Web Appendix (available at <https://doi.org/10.1093/aje/kwab294>) and Web Figure 1. This study was approved by the human subjects committees of the Harvard T.H. Chan School of Public Health and Brigham and Women's Hospital (Boston, Massachusetts).

Physical activity assessments

To represent the 1-year time frame of the PAQ, data were collected over approximately 15 months, and participants were randomly assigned to one of 4 different groups with the order of measurements varied (Figure 1). This included 2 self-administered PAQs, 4 ACT24s, 1 DLW measurement (repeated among 90 women), two 1-week accelerometer measurements, and 4 RPR measurements (see Web Appendix). The study period was divided into phases, with each phase representing a 3-month interval. Groups 1 and 3 had the same data collection timeline, with the exception that DLW was measured in phase 1 for group 1 and phase 3 for group 3. Groups 2 and 4 also had the same data collection timeline, with the exception that DLW was measured in phase 2 and phase 4, respectively. A subgroup of group 1 participants was randomly selected for a second DLW measurement (DLW2), which was taken at month 9, 12, or 15. In all groups, accelerometer and DLW measurements were taken within the same phase but not during the same week (Figure 1).

Physical activity questionnaire. The PAQ is an updated version of a previously validated self-administered modified Paffenbarger questionnaire (14) that assesses the average amount of time spent per week on 13 specific activities in the previous year (15). Response categories range from 0 to >40 hours/week. Activities included walking to work or for exercise (including golf), jogging (>10 minutes/mile), running (≤10 minutes/mile), bicycling (including a stationary bicycle), lap swimming, tennis, squash or racquetball, other aerobic exercise (exercise classes, etc.), lower-intensity exercise (e.g., yoga, stretching, or toning), moderate outdoor work (e.g., yard work or gardening), heavy outdoor work (e.g., digging or chopping), weightlifting (including use of machines), and standing or walking around at work and at home. Participants indicated the intensity of activity (low, medium, or high) for swimming, biking, and tennis, and they reported their walking pace and number of flights of stairs climbed daily. Data on sedentary behaviors included amounts of time spent sitting at work or while commuting; sitting at home watching television, videocassettes, or digital video discs; and other sitting at home.

A weekly energy expenditure score (metabolic equivalent of task (MET)-hours/week) was computed for each participant, at baseline and at the end of data collection, about 12 months later, after summing up the number of MET-hours per day associated with each activity (16). Activities were grouped by intensity as follows: vigorous physical activity, ≥6.0 METs; moderate physical activity, 3.0–5.9 METs; and moderate-to-vigorous physical activity (MVPA), ≥3.0 METs (Web Appendix). Total activity included light physical activity (<3.0 METs) and MVPA.

ACT24 assessment. The ACT24 is an online tool for assessing physical activity over the past 24 hours (17). Participants choose among more than 200 activities in four 6-hour time periods and report the starting and stopping times for each activity in ≥5-minute increments (Web Appendix). Active behaviors were classified on the basis of energy expenditure level: light activity (<3.0 METs), moderate activity (3.0–5.9 METs), vigorous activity (≥6.0 METs), or MVPA (≥3.0 METs).

Accelerometry. Participants were mailed an accelerometer (ActiGraph GT3X; ActiGraph LLC, Pensacola, Florida), detailed instructions, and a wear-time diary. The methods for deriving accelerometer data have been previously described (12) (Web Appendix). Data on daily total activity counts based on the triaxial vector magnitude were used to determine the total volume of physical activity. Intensity was defined as <200 counts/minute for sedentary behavior (18), 200–2,689 counts/minute for light-intensity activity, 2,690–6,166 counts/minute for moderate-intensity activity, and ≥6,167 counts/minute for vigorous-intensity activity (19). For sedentary behavior, every minute and bouts of ≥15 minutes were examined. The mean amount of time spent in activity of different intensity categories daily was calculated across all valid wear days. To predict METs from triaxial counts, we used the equation $\text{METs} = 0.000863 \times (\text{activity counts from all 3 axes}) + 0.668876$ (19).

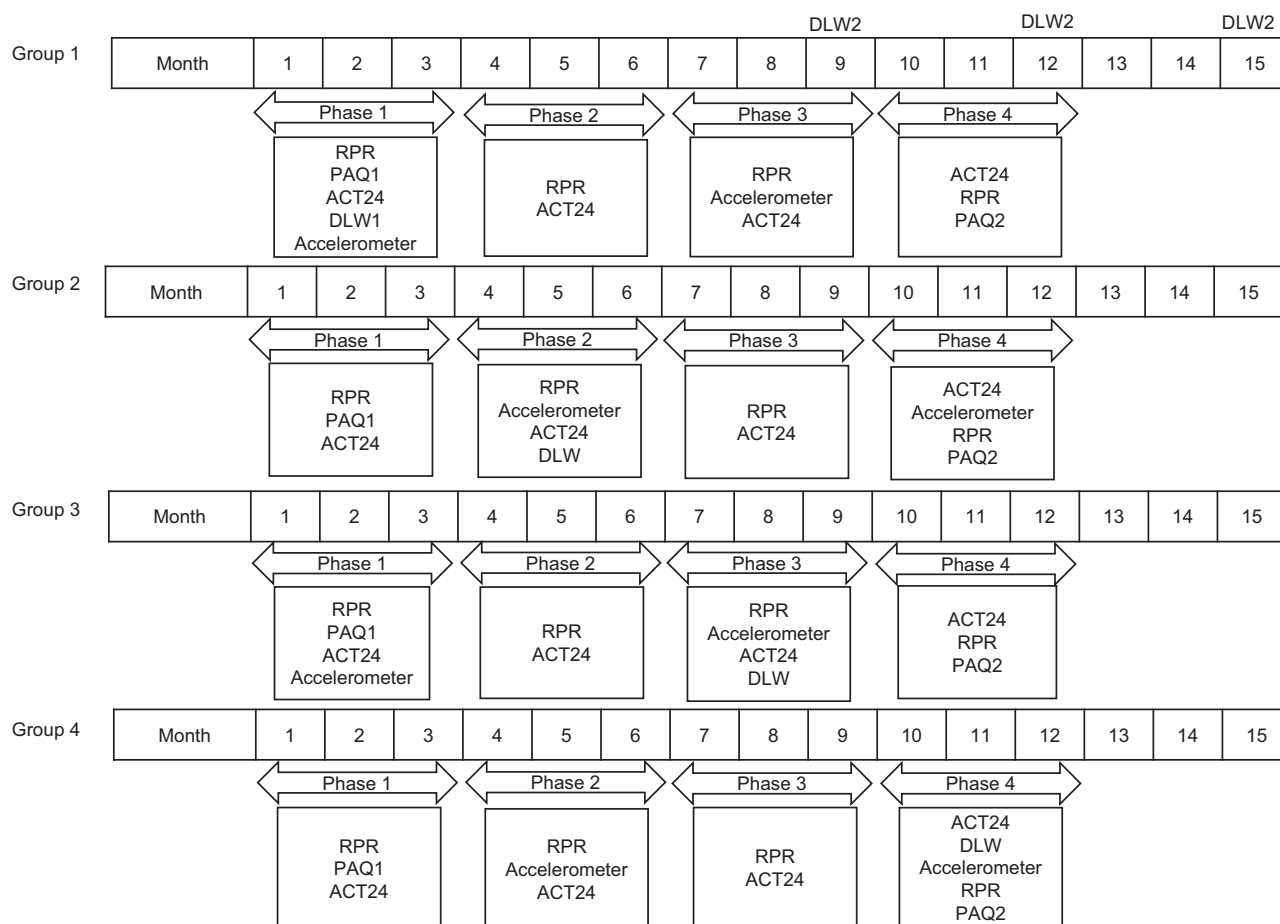


Figure 1. Measurement schedule for assessing physical activity over a 12- to 15-month period in Women's Lifestyle Validation Study groups 1–4, United States, 2010–2012. Participants were randomly selected into 4 different groups. Groups 1 and 3 had the same data collection timeline, with the exception that doubly labeled water (DLW) was measured in phase 1 for group 1 and phase 3 for group 3. Groups 2 and 4 also had the same data collection timeline, with the exception that DLW was measured in phase 2 and phase 4, respectively. A subgroup of group 1 participants was randomly selected for a second DLW measurement (DLW2) which was taken at month 9, 12, or 15. In all groups, accelerometer and DLW measurements were taken within the same phase but not during the same week. ACT24, Activities Completed Over Time in 24 Hours; PAQ1, first physical activity questionnaire; PAQ2, second physical activity questionnaire; RPR, resting pulse rate.

Doubly labeled water. TEE (kcal/day) was obtained from DLW data, a widely used standard for estimating TEE in free-living people (12). We also calculated physical activity energy expenditure (PAEE) by subtracting the resting metabolic rate and thermic effect of food (10% of TEE) from TEE. Resting metabolic rate was calculated using the Mifflin prediction equation (20). PAL was calculated by dividing the TEE by resting metabolic rate (21). DLW body fat percentage was calculated by dividing DLW fat mass by DLW body weight, multiplied by 100.

Resting pulse rate. RPR was self-reported 4 times during the study, once per phase. RPR measurements were collected as part of the PAQ in phases 1 and 4 and using an online tool in phases 2 and 3. Participants were given instructions for measuring their RPR and report from 10 choices in increments of 5, ranging from <55 beats/minute to ≥ 100 beats/minute.

Statistical analyses

Absolute physical activity measurements were characterized as mean values and standard deviations. Because most variables were not normally distributed, they were log-transformed after changing 0 values to 0.0001, except for PAL and DLW body fat percentage. Subsequent analyses were based on the ranks of log-transformed data.

Reproducibility was assessed using rank intraclass correlation coefficients (ICCs) calculated using random-effects analysis of variance (22) between the first PAQ (PAQ1) and the second PAQ (PAQ2) (collected 12 months apart), the first and second accelerometer measurements (collected 6 months apart), and the 4 ACT24 and RPR measurements (collected once every 3 months over the 12-month period). Spearman rank correlation coefficients and their 95% confidence intervals (CIs) were calculated to assess validity. Deattenuated correlation coefficients and their 95% CIs were

computed to account for within-person variation and number of repeated measurements of the reference method (23–25) (Web Appendix). We considered PAQ2 the primary comparison, since it would best represent the physical activity spent in the same period in which the comparison methods were administered. We also assessed the PAQ1 and the average of the 2 PAQs (PAQ_{avg}) as secondary comparisons. When studying the validity of the ACT24, we used the average of up to 4 measurements of ACT24 recalls (ACT24_{avg}) to represent activity throughout the year, and we also report the results of individual ACT24s to represent a typical measurement in an epidemiologic study.

All Spearman correlation coefficients were age-adjusted; accelerometer measurements were also adjusted for wear time to reduce spurious variation. Correlations of 0.4 or higher were considered to represent adequate validity. Analyses were also conducted after stratifying by age (≤ 60 years, > 60 years), body mass index (BMI) (weight (kg)/height (m)²; < 25 , ≥ 25), and cohort (Nurses' Health Study, Nurses' Health Study II).

To estimate the correlation (validity coefficient) (26) and 95% CI (27) between true physical activity and activities estimated via the PAQ and the ACT24, we used the method of triads with data from the 3 pairwise correlations between PAQ2 (or the mean ACT24), the mean of 2 accelerometer measurements, and DLW PAL (or PAEE). It is reasonable to assume that these methods have independent errors because they rely on different technologies, and the assessments were spaced in time by design (Web Appendix).

Out of 796 enrolled participants, 795 completed the baseline PAQ (PAQ1); of those participants, 755 completed PAQ2, 695 completed at least 1 ACT24 (272 completed all 4 ACT24s; the average was 3 ACT24s per participant), 684 had at least 1 DLW measurement, and 683 had at least 1 accelerometer measurement. In the analyses of RPR, women who reported use of antihypertensive medication were excluded ($n = 220$).

RESULTS

Descriptive statistics

On average, participants were 61 (standard deviation, 10) years of age at enrollment and had a mean BMI of 26.4 (standard deviation, 5.3); 98% were nonsmokers, and 91% were White (Table 1). Similar distributions were observed when data were stratified by cohort, except that Nurses' Health Study II participants were younger and 87% were White (Table 1). The mean DLW-determined TEE was 2,194 (standard deviation, 360) kcal/day.

Distributions of energy expenditures and physical activity measurements as assessed by different methods are presented for the overall sample (Table 2, Web Tables 1 and 2) and stratified by age, BMI, and cohort (Web Tables 1, 3, and 4). In general, total physical activity and time spent in moderate activity, MVPA, and sedentary activities assessed by PAQ were lower than those estimated by ACT24. Times spent in moderate activity and MVPA assessed by accelerometer were similar to those estimated by PAQ and lower than those estimated by ACT24. For sedentary behaviors, self-

reported time by PAQ was similar to accelerometer-assessed time when bouts greater than or equal to 15 minutes were included. Sedentary time assessed by ACT24 was similar to accelerometer-assessed time when every minute was included. Similar distributions were observed when we stratified by age, BMI, and cohort (Web Tables 1–4).

Reproducibility

High reproducibility was observed among the different physical activities assessed by PAQs, accelerometers, and DLW-derived measures: PAL, TEE, PAEE, and body fat mass percentage (Table 2, Web Table 5). ICCs for PAQ-assessed physical activities had a median value of 0.64 and ranged from 0.35 for sedentary time to 0.71 for MVPA. The ICCs for accelerometer measurements had a median value of 0.76, ranging between 0.59 for vigorous activity and 0.78 for total physical activity. The ICCs for DLW-derived measures were 0.74 (95% CI: 0.66, 0.81) for TEE, 0.57 (95% CI: 0.44, 0.68) for PAL, and 0.42 (95% CI: 0.28, 0.58) for PAEE. In contrast, ACT24 showed lower reproducibility, and the median ICC was 0.21, with ICCs ranging between 0.17 for moderate activity and 0.41 for total activity. The reproducibility of the PAQ was slightly higher among participants with BMI ≥ 25 than among those with BMI < 25 (median ICC, 0.66 vs. 0.51), among those aged ≤ 60 years than among those aged > 60 years (median ICC, 0.73 vs. 0.57), and among Nurses' Health Study II participants than among Nurses' Health Study participants (median ICC, 0.69 vs. 0.59). The reproducibility of each of the accelerometer and DLW-determined TEE, PAL, and body fat mass percentage measurements was comparable across the different subgroups, unlike the DLW-determined PAEE, which showed higher reproducibility in the older subgroup (age > 60 years, Nurses' Health Study) and those with BMI ≥ 25 . The RPR measurements were highly reproducible in the overall sample and across all subgroups (Table 2, Web Table 5).

Validity

DLW as the reference method. Comparing activity assessed via PAQ2 with that assessed by DLW PAL, the median value of Spearman correlations excluding sedentary time was 0.31, and the coefficient became stronger after accounting for within-person variation in DLW measurements (median $r = 0.37$; Table 3). Sedentary time assessed by PAQ2 was not correlated with DLW PAL. Similar deattenuated multivariate-adjusted correlations were observed with DLW-determined PAEE (Web Table 6). PAQ1 and PAQ_{avg} showed comparable correlations with DLW PAL (Table 3) and after stratifying by BMI, age, and cohort type (Web Tables 7–9).

Relatively lower correlations were observed between activities assessed by ACT24 and DLW PAL (for active categories, median deattenuated $r = 0.29$). Correlations were consistently lower when using a single ACT24 versus ACT24_{avg} (for active behaviors measured by the second ACT24, median deattenuated $r = 0.19$ vs. $r = 0.29$ for

Table 1. Baseline Characteristics (Mean (Standard Deviation)) of Participants in the Women's Lifestyle Validation Study (*n* = 683), by Body Mass Index, Age, and Cohort, United States, 2010–2012

Baseline Characteristic	Total (<i>n</i> = 683)	Body Mass Index Category ^a				Age Group, years		Cohort	
		<25 (<i>n</i> = 322)		≥25 (<i>n</i> = 361)		≤60 (<i>n</i> = 335)	>60 (<i>n</i> = 348)	NHS (<i>n</i> = 292)	NHS II (<i>n</i> = 391)
Age, years	61 (10)	61 (10)	61 (9)	53 (5)	69 (5)	71 (4)	54 (5)		
Height, m	1.64 (0.07)	1.65 (0.08)	1.63 (0.06)	1.65 (0.07)	1.63 (0.07)	1.62 (0.06)	1.65 (0.07)		
Weight, kg	71 (15)	61 (7)	81 (14)	73 (17)	69 (13)	69 (13)	73 (16)		
Weight change ^b , kg	-0.2 (2.9)	-0.1 (1.8)	-0.3 (3.6)	-0.3 (2.8)	-0.2 (3.0)	-0.3 (3.0)	-0.2 (2.8)		
Body mass index	26.4 (5.3)	22.3 (1.8)	30.1 (4.6)	26.7 (5.8)	26.2 (4.8)	26.2 (4.8)	26.7 (5.7)		
Current smoking (yes) ^c	2	2	3	4	1	1	3		
White race/ethnicity ^c	91	94	87	86	95	96	87		
Physical activity questionnaire									
Total activity, MET-hours/day ^d	8.2 (5.0)	9.1 (5.2)	7.4 (4.7)	8.7 (4.7)	7.7 (5.2)	7.4 (4.3)	8.8 (5.4)		
MVPA, MET-hours/day	4.3 (4.5)	5.0 (4.8)	3.7 (4.1)	4.4 (4.1)	4.2 (4.8)	4.1 (3.7)	4.5 (4.9)		
Sedentary time, hours/day	3.1 (1.3)	3.1 (1.2)	3.1 (1.3)	3.2 (1.3)	2.9 (1.2)	2.9 (1.2)	3.2 (1.3)		
ACT24 assessment tool									
Total activity, MET-hours/day	16.1 (9.1)	17.3 (9.3)	15.0 (8.8)	17.6 (10.1)	14.7 (7.9)	15.1 (7.9)	16.8 (9.9)		
MVPA, MET-hours/day	6.7 (8.5)	7.5 (8.9)	6.1 (8.1)	8.2 (9.7)	5.3 (6.8)	5.5 (7.0)	7.6 (9.4)		
Sedentary time, hours/day	7.1 (3.2)	6.7 (3.0)	7.5 (3.3)	7.0 (3.3)	7.2 (3.1)	7.1 (3.0)	7.1 (3.4)		
Accelerometer									
Total activity									
Activity counts/day	588,069 (179,367)	634,888 (175,115)	546,308 (172,906)	635,071 (176,452)	542,823 (170,522)	533,800 (165,588)	628,598 (178,745)		
MET-hours/day	29.8 (2.2)	30.4 (2.2)	29.3 (2.1)	30.4 (2.2)	29.3 (2.0)	29.2 (2.0)	30.3 (2.2)		
MVPA, hours/day	0.7 (0.4)	0.8 (0.5)	0.6 (0.4)	0.8 (0.4)	0.6 (0.4)	0.5 (0.4)	0.8 (0.5)		
Sedentary time, hours/day ^e	3.7 (1.4)	3.4 (1.3)	4.0 (1.5)	3.4 (1.4)	4.0 (1.4)	4.1 (1.4)	3.5 (1.4)		
Doubly labeled water									
Physical activity level ^f	1.7 (0.2)	1.8 (0.2)	1.7 (0.2)	1.7 (0.2)	1.7 (0.2)	1.7 (0.2)	1.7 (0.2)		
Total energy expenditure, kcal/day	2,194 (360)	2,067 (304)	2,308 (368)	2,324 (365)	2,070 (307)	2,045 (301)	2,306 (360)		
Total physical activity expenditure, kcal/day	709 (241)	695 (233)	722 (247)	758 (244)	662 (228)	657 (232)	748 (240)		
Resting pulse rate, beats/minute	69 (9)	67 (8)	70 (9)	69 (9)	68 (8)	68 (8)	69 (9)		

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; NHS, Nurses' Health Study; NHS II, NHS, Nurses' Health Study II.

^a Weight (kg)/height (m)².

^b Weight change was estimated over the 12-month study period.

^c Values are expressed as percentages.

^d Includes active behavior.

^e Including bouts of ≥15 minutes.

^f Physical activity level is the ratio of total energy expenditure to resting metabolic rate.

Table 2. Distributions (Mean (Standard Deviation)) of and Intraclass Correlation Coefficients (95% Confidence Interval) for Different Physical Activity Measurements in the Women's Lifestyle Validation Study ($n = 683$), United States, 2010–2012

Physical Activity Measurement	No. of Participants	Mean (SD)	ICC	95% CI
DLW physical activity level ^a			0.57	0.44, 0.68
Measurement 1	683	1.7 (0.2)		
Measurement 2	87	1.7 (0.3)		
DLW-determined PAEE, kcal/day			0.42	0.28, 0.58
Measurement 1	683	709 (241)		
Measurement 2	87	706 (285)		
DLW-determined TEE, kcal/day			0.74	0.66, 0.81
Measurement 1	683	2,194 (360)		
Measurement 2	87	2,158 (390)		
DLW body fat mass, %			0.83	0.76, 0.87
Measurement 1	683	38 (8)		
Measurement 2	87	38 (7)		
Total physical activity (not including sedentary time)				
PAQ, MET-hours/day			0.58	0.53, 0.63
PAQ1	683	8.2 (5.0)		
PAQ2	683	8.1 (4.8)		
ACT24, MET-hours/day			0.41	0.36, 0.47
Phase 1	683	16.1 (9.1)		
Phase 2	616	16.6 (9.5)		
Phase 3	501	16.2 (8.4)		
Phase 4	268	16.1 (8.1)		
Accelerometer				
Total activity counts/day			0.78	0.75, 0.81
Phase 1	683	588,069 (179,367)		
Phase 2	662	585,181 (173,407)		
MET-hours/day			0.77	0.74, 0.80
Phase 1	683	29.8 (2.2)		
Phase 2	662	29.8 (2.1)		
Moderate-to-vigorous physical activity				
PAQ, MET-hours/day			0.71	0.67, 0.74
PAQ1	683	4.3 (4.5)		
PAQ2	683	4.3 (4.2)		
ACT24, MET-hours/day			0.18	0.14, 0.24
Phase 1	683	6.7 (8.5)		
Phase 2	616	7.5 (9.3)		
Phase 3	501	7.4 (8.0)		
Phase 4	268	7.2 (7.4)		
Accelerometer, hours/day			0.77	0.74, 0.80
Phase 1	683	0.7 (0.4)		
Phase 2	662	0.7 (0.4)		
Vigorous activity				
PAQ, MET-hours/day			0.67	0.62, 0.71
PAQ1	683	1.6 (2.6)		
PAQ2	683	1.6 (2.6)		

Table continues

Table 2. Continued

Physical Activity Measurement	No. of Participants	Mean (SD)	ICC	95% CI
ACT24, MET-hours/day			0.21	0.17, 0.26
Phase 1	683	0.9 (2.8)		
Phase 2	616	1.4 (4.3)		
Phase 3	501	1.4 (3.9)		
Phase 4	268	1.4 (3.6)		
Accelerometer, hours/day			0.59	0.54, 0.64
Phase 1	683	0.04 (0.1)		
Phase 2	662	0.03 (0.1)		
Moderate activity				
PAQ, MET-hours/day			0.64	0.60, 0.68
PAQ1	683	2.7 (2.9)		
PAQ2	683	2.7 (2.8)		
ACT24, MET-hours/day			0.17	0.13, 0.22
Phase 1	683	5.8 (8.1)		
Phase 2	616	6.1 (8.3)		
Phase 3	501	5.9 (7.1)		
Phase 4	268	5.8 (6.9)		
Accelerometer, hours/day			0.76	0.73, 0.79
Phase 1	683	0.6 (0.4)		
Phase 2	662	0.6 (0.4)		
Sedentary time				
PAQ, hours/day			0.35	0.29, 0.42
PAQ1	681	3.1 (1.3)		
PAQ2	681	3.1 (1.3)		
ACT24, hours/day			0.35	0.29, 0.41
Phase 1	683	7.1 (3.2)		
Phase 2	616	7.3 (3.1)		
Phase 3	501	7.2 (3.0)		
Phase 4	268	7.5 (3.1)		
Accelerometer, hours/day				
Including every minute			0.69	0.65, 0.73
Phase 1	683	8.0 (1.4)		
Phase 2	662	7.9 (1.4)		
Including bouts of ≥ 15 minutes			0.71	0.67, 0.74
Phase 1	683	3.7 (1.4)		
Phase 2	662	3.7 (1.4)		
Resting pulse rate, beats/minute			0.60	0.56, 0.63
Measurement 1	677	68.7 (8.5)		
Measurement 2	671	69.2 (7.7)		
Measurement 3	641	68.5 (7.9)		
Measurement 4	675	68.6 (8.3)		

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; DLW, doubly labeled water; ICC, intraclass correlation coefficient; MET, metabolic equivalent of task; PAEE, physical activity energy expenditure; PAQ, physical activity questionnaire; TEE, total energy expenditure.

^a DLW physical activity level was calculated as the ratio of DLW-determined TEE to resting metabolic rate.

Table 3. Age-Adjusted and Deattenuated Spearman Correlation Coefficients (95% Confidence Interval) for Comparisons of a Physical Activity Questionnaire and 24-Hour Recalls With Doubly Labeled Water-Determined Physical Activity Level, Accelerometry, and Resting Pulse Rate Measurements in the Women's Lifestyle Validation Study (n = 683), United States, 2010–2012

Physical Activity Measurement	DLW PAL			Accelerometer ^a			Resting Pulse Rate ^b		
	Age-Adjusted	Deattenuated	95% CI	Age-Adjusted	Deattenuated	95% CI	Age-Adjusted	Deattenuated	95% CI
Total physical activity, MET-hours/day									
PAQ1	0.32	0.39	0.31, 0.47	0.39	0.43	0.36, 0.49	-0.19	-0.20	-0.29, -0.11
PAQ2	0.34	0.41	0.33, 0.49	0.40	0.43	0.37, 0.50	-0.25	-0.23	-0.32, -0.13
PAC _{avg}	0.35	0.43	0.35, 0.51	0.43	0.46	0.39, 0.52	-0.25	-0.25	-0.34, -0.15
ACT24 (phase 2)	0.22	0.28	0.18, 0.37	0.31	0.32	0.24, 0.39	-0.24	-0.16	-0.25, -0.06
ACT24 _{avg}	0.32	0.39	0.30, 0.47	0.37	0.40	0.33, 0.46	-0.17	-0.18	-0.27, -0.09
MVPA, MET-hours/day									
PAQ1	0.31	0.39	0.30, 0.47	0.39	0.43	0.36, 0.49	-0.25	-0.26	-0.35, -0.17
PAQ2	0.33	0.40	0.31, 0.48	0.39	0.42	0.35, 0.48	-0.31	-0.27	-0.36, -0.18
PAC _{avg}	0.34	0.41	0.33, 0.49	0.41	0.45	0.38, 0.51	-0.29	-0.30	-0.37, -0.21
ACT24 (phase 2)	0.17	0.21	0.11, 0.30	0.28	0.26	0.18, 0.34	-0.12	-0.12	-0.22, -0.03
ACT24 _{avg}	0.26	0.31	0.22, 0.39	0.32	0.35	0.28, 0.41	-0.11	-0.11	-0.20, -0.02
Vigorous activity, MET-hours/day									
PAQ1	0.20	0.25	0.16, 0.34	0.35	0.41	0.33, 0.48	-0.24	-0.23	-0.32, -0.14
PAQ2	0.22	0.26	0.17, 0.35	0.35	0.40	0.33, 0.48	-0.27	-0.23	-0.32, -0.14
PAC _{avg}	0.21	0.26	0.17, 0.35	0.37	0.44	0.37, 0.51	-0.27	-0.25	-0.34, -0.15
ACT24 (phase 2)	0.07	0.08	-0.01, 0.17	0.20	0.23	0.14, 0.32	0.00	0.02	-0.08, 0.11
ACT24 _{avg}	0.15	0.19	0.10, 0.27	0.27	0.31	0.23, 0.39	-0.07	-0.07	-0.16, 0.02
Moderate activity, MET-hours/day									
PAQ1	0.27	0.33	0.24, 0.41	0.35	0.38	0.32, 0.45	-0.15	-0.16	-0.25, -0.07
PAQ2	0.28	0.34	0.25, 0.42	0.35	0.38	0.31, 0.44	-0.19	-0.17	-0.26, -0.08
PAC _{avg}	0.29	0.36	0.27, 0.44	0.38	0.41	0.34, 0.47	-0.19	-0.19	-0.28, -0.11
ACT24 (phase 2)	0.13	0.16	0.07, 0.25	0.20	0.20	0.12, 0.28	-0.13	-0.14	-0.23, -0.04
ACT24 _{avg}	0.21	0.26	0.16, 0.35	0.27	0.29	0.22, 0.37	-0.06	-0.06	-0.15, 0.03

Table continues

Table 3. Continued

Physical Activity Measurement	DLW PAL			Accelerometer ^a			Resting Pulse Rate ^b		
	Age-Adjusted	Deattenuated	95% CI	Age-Adjusted	Deattenuated	95% CI	Age-Adjusted	Deattenuated	95% CI
Sedentary time, hours/day									
PAQ1	-0.04	-0.06	-0.14, 0.03	0.08	0.08	0.00, 0.16	-0.02	-0.03	-0.11, 0.07
PAQ2	-0.01	-0.02	-0.11, 0.07	0.12	0.13	0.05, 0.21	-0.01	-0.01	-0.11, 0.09
PAQ _{avg}	-0.03	-0.04	-0.12, 0.06	0.12	0.12	0.05, 0.20	-0.03	-0.04	-0.12, 0.06
ACT24 (phase 2)	-0.11	-0.14	-0.23, -0.05	0.23	0.25	0.17, 0.33	0.06	0.06	-0.04, 0.16
ACT24 _{avg}	-0.12	-0.15	-0.24, -0.06	0.25	0.28	0.20, 0.35	0.00	0.00	-0.10, 0.10

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; ACT24_{avg}, average of up to 4 ACT24 measurements; CI, confidence interval; DLW, doubly labeled water; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PAL, physical activity level; PAQ, physical activity questionnaire; PAQ_{avg}, average of PAQ1 and PAQ2.

^a Accelerometer measurements were based on the triaxial vector magnitude; accelerometer measures of MVPA, vigorous activity, and moderate activity included ≥ 1 -minute bouts; sedentary time was based on ≥ 15 -minute bouts. Values from accelerometer were adjusted for age (in years) and accelerometer wear time (hours/day).

^b Among nonhypertensive participants ($n = 474$). 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. (20).

ACT24_{avg}) (Table 3, Web Table 6). Stronger correlations were noted between ACT24 and DLW-derived measurements among older subgroups (age >60 years, Nurses' Health Study) and those with BMI ≥ 25 (Web Tables 7–9).

Accelerometry as the reference method. Using the accelerometer as the comparison method, the median Spearman correlation coefficient for activity assessed by PAQ2 was 0.35 (Table 3); the correlation became slightly stronger after accounting for within-person variation in the accelerometer measurements (median deattenuated $r = 0.40$, ranging between 0.13 for sedentary time and 0.43 for total activity). Similar correlations were observed with PAQ1 and PAQ_{avg} (Table 3). Correlations were stronger for all active behaviors in the younger group (age ≤ 60 years, Nurses' Health Study II) and for sedentary time among those with BMI < 25 (Web Table 10).

We observed slightly lower correlations between ACT24 and accelerometry as compared with the PAQ, except for sedentary activity (Table 3, Web Table 11). Correlations were stronger for vigorous and sedentary activities among those aged ≤ 60 years and Nurses' Health Study II participants, and for moderate activities among those older than 60 years (Web Table 10).

Relative validity of assessment methods using biological responses. Among participants who were not taking anti-hypertensive medication, PAQ- and accelerometer-assessed active behaviors had comparable inverse correlations with RPR measurements, but ACT24_{avg} had lower correlations (median deattenuated r : for PAQ2, $r = -0.23$; for accelerometry, $r = -0.26$; and for ACT24_{avg}, $r = -0.09$) (Table 3, Web Table 11).

Compared with DLW body fat percentage, the PAQs, both single and averaged, showed the highest correlations for total activity, MVPA, vigorous activity, and moderate activity, followed by the averaged accelerometer measurements. Lower correlations were observed with single and averaged ACT24s (Table 4).

ACT24 as the reference method. When the ACT24 was used as the comparison method, correlations with PAQ became stronger after accounting for within-person variation in ACT24 measurements, except for sedentary activities (Table 5). For example, the age-adjusted Spearman correlations of active behaviors estimated by PAQ2 with those estimated by ACT24 had a median value of 0.39, which became stronger after accounting for within-person variation in ACT24 (median deattenuated $r = 0.59$). These correlations were slightly different across the different subgroups. Correlations for total activity and MVPA were stronger among those with BMI ≥ 25 (Web Table 12).

Estimated correlations with true physical activity. Using the method of triads with assessments unlikely to have correlated errors, the estimated correlations between PAQ2 and true physical activity were calculated and are presented in Table 6. The correlation coefficients were 0.54 (95% CI: 0.47, 0.62) for total physical activity and 0.60 (95% CI: 0.52, 0.69) for MVPA. The correlation coefficients between MVPA assessed by PAQ2 and true physical activity were

Table 4. Age-Adjusted and Deattenuated Spearman Correlation Coefficients (95% Confidence Interval) for Comparisons of Physical Activity Questionnaires, 24-Hour Recalls, and Accelerometer Measurements With Body Fat Mass Percentage Calculated from Doubly Labeled Water in the Women's Lifestyle Validation Study ($n = 683$), United States, 2010–2012

Physical Activity Measurement	DLW Body Fat Mass %		
	Age-Adjusted	Deattenuated	95% CI
Total activity, MET-hours/day			
PAQ1	−0.35	−0.37	−0.44, −0.31
PAQ2	−0.34	−0.37	−0.44, −0.30
PAQ _{avg}	−0.37	−0.40	−0.47, −0.34
ACT24 (phase 2)	−0.26	−0.28	−0.36, −0.20
ACT24 _{avg}	−0.29	−0.31	−0.37, −0.24
Accelerometer _{avg}	−0.33	−0.35	−0.42, −0.28
MVPA, MET-hours/day			
PAQ1	−0.37	−0.40	−0.46, −0.33
PAQ2	−0.38	−0.41	−0.47, −0.34
PAQ _{avg}	−0.40	−0.43	−0.49, −0.36
ACT24 (phase 2)	−0.25	−0.27	−0.34, −0.19
ACT24 _{avg}	−0.25	−0.27	−0.35, −0.20
Accelerometer _{avg}	−0.27	−0.29	−0.37, −0.22
Vigorous activity, MET-hours/day			
PAQ1	−0.27	−0.29	−0.37, −0.22
PAQ2	−0.23	−0.25	−0.33, −0.18
PAQ _{avg}	−0.25	−0.27	−0.35, −0.20
ACT24 (phase 2)	−0.11	−0.12	−0.19, −0.04
ACT24 _{avg}	−0.17	−0.18	−0.25, −0.11
Accelerometer _{avg}	−0.23	−0.24	−0.32, −0.17
Moderate activity, MET-hours/day			
PAQ1	−0.31	−0.34	−0.40, −0.27
PAQ2	−0.34	−0.36	−0.43, −0.30
PAQ _{avg}	−0.35	−0.38	−0.44, −0.31
ACT24 (phase 2)	−0.20	−0.21	−0.29, −0.13
ACT24 _{avg}	−0.21	−0.22	−0.30, −0.15
Accelerometer _{avg}	−0.24	−0.26	−0.33, −0.18
Sedentary time, hours/day			
PAQ1	0.05	0.06	−0.02, 0.13
PAQ2	−0.01	−0.02	−0.10, 0.06
PAQ _{avg}	0.02	0.02	−0.06, 0.10
ACT24 (phase 2)	0.10	0.10	0.02, 0.18
ACT24 _{avg}	0.10	0.10	0.03, 0.18
Accelerometer _{avg} (≥ 1 -minute bouts)	0.28	0.30	0.23, 0.37
Accelerometer _{avg} (≥ 15 -minute bouts)	0.26	0.28	0.20, 0.35

Abbreviations: Accelerometer_{avg}, average of two 7-day accelerometer measurements; ACT24, Activities Completed Over Time in 24 Hours; ACT24_{avg}, average of up to 4 ACT24 measurements; CI, confidence interval; DLW, doubly labeled water; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PAQ, physical activity questionnaire; PAQ_{avg}, average of PAQ1 and PAQ2.

very similar to those between true physical activity and each of the DLW PAL and accelerometer measurements (Table 6). Validity coefficients were comparable when using

DLW PAEE instead of DLW PAL (Web Table 13) and after further adjusting DLW PAEE for body weight (Web Table 14).

Table 5. Age-Adjusted and Deattenuated Spearman Correlation Coefficients (95% Confidence Interval) for Comparisons of Physical Activity Estimated by Questionnaire With the Average of Up to Four 24-Hour Activity Recalls in the Women's Lifestyle Validation Study ($n = 683$), United States, 2010–2012

Physical Activity Measurement	ACT24		
	Age-Adjusted	Deattenuated	95% CI
Total activity, MET-hours/day			
PAQ1	0.47	0.62	0.53, 0.69
PAQ2	0.50	0.61	0.51, 0.68
PAQ _{avg}	0.53	0.67	0.58, 0.74
MVPA, MET-hours/day			
PAQ1	0.45	0.60	0.48, 0.69
PAQ2	0.43	0.61	0.49, 0.70
PAQ _{avg}	0.47	0.63	0.51, 0.72
Vigorous activity, MET-hours/day			
PAQ1	0.33	0.38	0.31, 0.45
PAQ2	0.32	0.37	0.29, 0.44
PAQ _{avg}	0.33	0.38	0.31, 0.45
Moderate activity, MET-hours/day			
PAQ1	0.37	0.55	0.41, 0.67
PAQ2	0.35	0.56	0.42, 0.67
PAQ _{avg}	0.39	0.59	0.45, 0.70
Sedentary time, hours/day			
PAQ1	0.06	0.08	−0.01, 0.17
PAQ2	0.05	0.07	−0.03, 0.16
PAQ _{avg}	0.06	0.08	−0.01, 0.17

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; CI, confidence interval; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PAQ, physical activity questionnaire; PAQ_{avg}, average of PAQ1 and PAQ2.

For ACT24_{avg}, the estimated correlation was 0.50 (95% CI: 0.43, 0.59) for true total physical activity and 0.47 (95% CI: 0.39, 0.58) for true MVPA. As for sedentary behavior, the correlation with true behavior was 0.31 (95% CI: 0.21, 0.44). Heywood cases (correlations > 1) (26) were observed in these comparisons, and correlations did not change if 1-minute bouts versus ≥ 15 -minute bouts were included.

DISCUSSION

We evaluated the reproducibility and validity of our self-administered PAQ and ACT24 by comparing the reported activity with that derived from accelerometer measurements and DLW. We also evaluated the relative validity of different methods by their correlations with RPR and adiposity. The correlations between PAQ and activity assessed by accelerometry or DLW PAL were moderate. Using the method triads to estimate correlations with true activity, the PAQ had good validity for active behaviors, and these estimated correlations were similar to those using DLW PAL or accelerometer methods. The ACT24 tool had low reproducibility, and a single measurement may not be adequate to

capture day-to-day and seasonal variation; using the average of up to 4 days, the correlation coefficients with DLW PAL and accelerometry were moderate but slightly lower than those for the PAQ, except for sedentary behavior. Correlations were consistently lower when using single ACT24 data versus ACT24_{avg} data. When ACT24 was used as a reference method, the correlations with PAQ were moderate, and they became stronger after adjusting for within-person variation of ACT24. Using correlations with RPR and DLW body fat percentage to evaluate relative validity, PAQ2 had higher validity than accelerometer measurements, and single ACT24 and ACT24_{avg} had the lowest validity.

We consider this study an expanded update to our previous validation study of our PAQ (14), conducted among 231 Nurses' Health Study II participants in 1989–1991 using 4 past-week activity recalls and four 7-day activity diaries completed over the course of a year as reference methods. In addition, we evaluated the validity of the ACT24. The reproducibility of our PAQ for total activity, assessed by ICC, was 0.58 in the current study versus 0.59 in the previous study. The validity of the PAQ was lower in the current study, potentially because of different comparison methods; the previous study used 28-day

Table 6. Validity Coefficients Showing Correlations With True Physical Activity for Total Physical Activity and Activity Intensity Categories as Estimated From the Final Physical Activity Questionnaire, 24-Hour Recalls, Accelerometry, and Doubly Labeled Water-Determined Physical Activity Level Among Participants in the Women's Lifestyle Validation Study (*n* = 662), United States, 2010–2012^a

Physical Activity Measurement ^b	Spearman's Correlation Coefficient				Validity Coefficient					
	PAQ2 or ACT24 _{avg} and Accelerometer (<i>r</i> _{QR})	PAQ2 or ACT24 _{avg} and DLW PAL (<i>r</i> _{QM})	Accelerometer and DLW PAL (<i>r</i> _{RM})	PAQ2 or ACT24 _{avg} and True PA	DLW PAL and True PA	Accelerometer and True PA	VC _{QT}	VC _{MT}	VC _{RT}	95% CI
				VC _{QT}	95% CI	VC _{MT}	95% CI	VC _{RT}	95% CI	
Q = PAQ2										
Total activity, MET-hours/day	0.39	0.34	0.46	0.54	0.47, 0.62	0.64	0.56, 0.72	0.72	0.64, 0.81	
MVPA, MET-hours/day	0.37	0.33	0.35	0.60	0.52, 0.69	0.56	0.48, 0.66	0.62	0.53, 0.71	
Vigorous activity, MET-hours/day	0.31	0.22	0.17	0.64	0.51, 0.81	0.34	0.26, 0.45	0.49	0.39, 0.61	
Moderate activity, MET-hours/day	0.33	0.28	0.33	0.53	0.44, 0.63	0.53	0.44, 0.63	0.62	0.53, 0.74	
Q = ACT24 _{avg}										
Total activity, MET-hours/day	0.36	0.32	0.46	0.50	0.43, 0.59	0.64	0.56, 0.73	0.72	0.63, 0.81	
MVPA, MET-hours/day	0.30	0.26	0.35	0.47	0.39, 0.58	0.54	0.45, 0.65	0.64	0.54, 0.76	
Vigorous activity, MET-hours/day	0.25	0.15	0.17	0.48	0.35, 0.65	0.32	0.23, 0.46	0.52	0.38, 0.69	
Moderate activity, MET-hours/day	0.25	0.20	0.33	0.39	0.30, 0.51	0.51	0.42, 0.63	0.64	0.53, 0.78	

Abbreviations: ACT24, Activities Completed Over Time in 24 Hours; ACT24_{avg}, average of up to 4 ACT24 measurements; CI, confidence interval; DLW, doubly labeled water; MET, metabolic equivalent of task; MVPA, moderate-to-vigorous physical activity; PA, physical activity; PAL, physical activity level; PAQ2, second physical activity questionnaire; VC, validity coefficient.

^a M stands for DLW PAL; Q stands for PAQ2 or ACT24_{avg}; R stands for accelerometer; and T stands for true PA.

^b All measurements were adjusted for age only, except for accelerometer measurements, which were further adjusted for wear time.

activity diaries (for MVPA, deattenuated $r = 0.56$), while the current study used accelerometry (for MVPA, deattenuated $r = 0.42$) and DLW measurements (deattenuated r of MVPA = 0.40). However, the accelerometer and DLW methods may have systematic within-person errors (person-specific biases) that would have underestimated the validity of the PAQ beyond that due to random errors that were accounted for by deattenuation. The DLW method is subject to analytical error in the laboratory measurement of changes in isotope ratios, variations in energy balance, differences in dietary macronutrient intake within and between participants, homeostatic control of energy intake, and biological differences in absorption and metabolism. The resting metabolic rate, used to calculate DLW PAL and PAEE, was not measured but estimated using the Mifflin prediction equation (20), which may have underestimated the correlation between the PAQ- and DLW-derived measures. However, in a study conducted by our research group, the correlation between calorimetry-measured resting metabolic rate and that predicted by the Mifflin equation ranged from 0.66 to 0.70 (Dr. Claire Pernar, Harvard T.H. Chan School of Public Health, unpublished data, 2021). In addition, the correlations between the accelerometer measurements and DLW-derived measures were moderate (Web Table 15) (12). Accelerometer assessments also have multiple potential sources of error, including being less sensitive to some activities such as swimming or cycling. Lastly, the Compendium of Physical Activities, which classifies activities according to their cost of energy expenditure (16), could also have affected the validity of the PAQ because the Compendium estimates rely on group averages and may not apply to all individuals. These errors may have similarly applied to the ACT24.

The method of triads can be valuable when no method is a perfect reference for assessing validity; a key assumption is that errors for the 3 methods are uncorrelated, which is reasonable in this case. Using the method of triads, the estimated correlation between our PAQ and true underlying physical activity was stronger than the deattenuated correlations when either DLW or accelerometry was used as the reference method, indicating that PAQ validity was not seriously overestimated in our previous study. This estimated correlation with true physical activity can be viewed as the upper boundary of validity (because correlated errors between methods could overestimate validity), and the deattenuated correlation with DLW or accelerometry as its lower boundary (because nonrandom errors in the comparison method could lead to an underestimate of validity). The modest differences between these lower and upper limits suggest that the validity of the PAQ was not seriously overestimated. Its correlations with RPR, which reflects cardiac fitness, and with DLW body fat percentage are consistent with its ability to robustly predict cardiometabolic endpoints (28–30). In addition, these correlations, which were similar to or higher than those for DLW PAL and accelerometry with these biological measurements, were consistent with the relative strength of correlations in the method-of-triads analyses.

On the other hand, the ACT24 had slightly lower correlations with true total physical activity and MVPA even

after using the average of up to 4 ACT24 assessments. However, this suboptimal performance of ACT24 was not observed among AARP participants who completed up to 6 ACT24 recalls, spaced 2 months apart (17). Because month-to-month variability in physical activity is large, more than 4 ACT24 recalls may be needed over a year to estimate true long-term physical activity for adequate measurement of true physical activity. Thus, the ACT24 has the potential to be used as a comparison method in future validation studies if adequate adjustment is made for within-person variation.

Many validation studies have been conducted for PAQs using DLW as a reference method (31), and the vast majority did not show adequate validity in estimating usual physical activity. As shown by the method-of-triads analyses, the correlation between DLW PAL and true physical activity was moderate; thus, using it as the reference method would underestimate the validity of other assessment methods. The moderate reproducibility of DLW-determined physical activity was also seen among women in other studies (17).

Consistent with our study, some PAQ validation studies have found low validity in measuring sedentary activities (32) because PAQs are not designed to capture the full spectrum of sedentary behaviors or lower-intensity activities, and sedentary and activity hours do not add up plausibly. On the other hand, our PAQ showed acceptable validity when accelerometry was used as the reference method. The validity was comparable to several validation studies that used the ActiGraph accelerometer as a reference method, with comparable thresholds of counts per minute (32–35).

To our knowledge, this was the largest and most comprehensive validation study on physical activity assessment methods among women to date, because of the multiple methods used for comparison over a period of 12–15 months. However, the study had several limitations. The results may not be generalizable to the general population, persons with chronic diseases, or men. The similar results in our Men's Lifestyle Validation Study support generalizability to male health professionals and a more general middle-aged population. Inclusion of participants from the Nurses' Health Study and Nurses' Health Study II cohorts provided a wide range of ages. However, our findings may not be generalizable to other groups defined by race/ethnicity, education, or motivation or for different physical activity behaviors. In addition, RPR measurements were self-reported; however, the participants were all nurses, and the ICCs were comparable across age, BMI, and cohort subgroups.

In conclusion, our findings indicate that our self-administered PAQ has good validity for measuring physical activity but not for measuring sedentary time. Based on the results of the method of triads and correlations with multiple comparison methods, the PAQ appears to have validity similar to that for methods using DLW, accelerometers, or repeated ACT24 assessments. The far lower cost of the PAQ, which facilitates repeated assessments over time, makes this or similar tools particularly suitable for epidemiologic studies. If multiple ACT24 assessments are collected over the year and within-person variation is taken into account, the ACT24 may be considered as a method for assessing physical activity in epidemiologic studies.

ACKNOWLEDGMENTS

Author affiliations: Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States (Laila Al-Shaar, Andrea K. Chomistek, Eric B. Rimm, Junaidah B. Barnett, Meir J. Stampfer, Walter C. Willett); Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, Pennsylvania, United States (Laila Al-Shaar); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States (Claire H. Perna, Eric B. Rimm, Meir J. Stampfer, A. Heather Eliassen, Walter C. Willett); Department of Epidemiology and Biostatistics, School of Public Health, Indiana University, Bloomington, Indiana, United States (Andrea K. Chomistek); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States (Eric B. Rimm, Meir J. Stampfer, A. Heather Eliassen, Walter C. Willett); Pennington Biomedical Research Center, Louisiana State University, Baton Rouge, Louisiana, United States (Jennifer Rood); and Nutritional Immunology Laboratory, Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts, United States (Junaidah B. Barnett).

This research was supported by the National Institutes of Health (grants UM1 CA186107, U01 CA176726, and P01CA055075-18S1). L.A.-S. received research support from National Institutes of Health training grant T32 HL 098048.

The data set analyzed in the current study is not publicly available but can be accessed upon request.

We thank the participants and staff of the Women's Lifestyle Validation Study for their valuable contributions.

This work was presented at the 2021 Annual Meeting of the Society for Epidemiologic Research (held virtually), June 22–25, 2021.

Conflict of interest: none declared.

REFERENCES

- Sattelmair J, Pertman J, Ding EL, et al. Dose response between physical activity and risk of coronary heart disease: a meta-analysis. *Circulation*. 2011;124(7):789–795.
- Li J, Siegrist J. Physical activity and risk of cardiovascular disease—a meta-analysis of prospective cohort studies. *Int J Environ Res Public Health*. 2012;9(2):391–407.
- Wahid A, Manek N, Nichols M, et al. Quantifying the association between physical activity and cardiovascular disease and diabetes: a systematic review and meta-analysis. *J Am Heart Assoc*. 2016;5(9):e002495.
- Diaz KM, Shimbo D. Physical activity and the prevention of hypertension. *Curr Hypertens Rep*. 2013;15(6):659–668.
- Carbone S, Del Buono MG, Ozemek C, et al. Obesity, risk of diabetes and role of physical activity, exercise training and cardiorespiratory fitness. *Prog Cardiovasc Dis*. 2019;62(4):327–333.
- Kerr J, Anderson C, Lippman SM. Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. *Lancet Oncol*. 2017;18(8):e457–e471.
- Erickson KI, Hillman C, Stillman CM, et al. Physical activity, cognition, and brain outcomes: a review of the 2018 physical activity guidelines. *Med Sci Sports Exerc*. 2019;51(6):1242–1251.
- Sattler MC, Ainsworth BE, Andersen LB, et al. Physical activity self-reports: past or future? *Br J Sports Med*. 2021;55(16):889–890.
- Willett WC. *Nutritional Epidemiology*. 3rd ed. New York, NY: Oxford University Press; 2013.
- Hu FB. *Obesity Epidemiology*. New York, NY: Oxford University Press; 2008.
- 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.
- 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.
- Yuan C, Spiegelman D, Rimm EB, et al. Validity of a dietary questionnaire assessed by comparison with multiple weighed dietary records or 24-hour recalls. *Am J Epidemiol*. 2017;185(7):570–584.
- 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.
- Nurses' Health Study. Questionnaires. <https://www.nurseshealthstudy.org/participants/questionnaires>. Published 2016. Accessed June 9, 2021.
- 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, Kozey Keadle S, 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.
- 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.
- Westertep KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Front Physiol*. 2013;4:90.
- Chavarro JE, Rosner BA, Sampson L, et al. Validity of adolescent diet recall 48 years later. *Am J Epidemiol*. 2009;170(12):1563–1570.
- Rosner B, Willett WC. Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol*. 1988;127(2):377–386.
- Perisic I, Rosner B. Comparisons of measures of interclass correlations: the general case of unequal group size. *Stat Med*. 1999;18(12):1451–1466.
- 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.
27. 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.
 28. Li TY, Rana JS, Manson JE, et al. Obesity as compared with physical activity in predicting risk of coronary heart disease in women. *Circulation.* 2006;113(4):499–506.
 29. Hu FB, Stampfer MJ, Colditz GA, et al. Physical activity and risk of stroke in women. *JAMA.* 2000;283(22):2961–2967.
 30. Al-Shaar L, Li Y, Rimm EB, et al. Physical activity and mortality among male survivors of myocardial infarction. *Med Sci Sports Exerc.* 2020;52(8):1729–1736.
 31. 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.
 32. Helmerhorst HJF, 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. Brown WJ, Burton NW, Marshall AL, et al. Reliability and validity of a modified self-administered version of the Active Australia physical activity survey in a sample of mid-age women. *Aust N Z J Public Health.* 2008;32(6):535–541.
 34. Hurtig-Wennlöf A, Hagströmer M, Olsson LA. The International Physical Activity Questionnaire modified for the elderly: aspects of validity and feasibility. *Public Health Nutr.* 2010;13(11):1847–1854.
 35. Pettee Gabriel K, McClain JJ, Lee CD, et al. Evaluation of physical activity measures used in middle-aged women. *Med Sci Sports Exerc.* 2009;41(7):1403–1412.