American Journal of Epidemiology
Copyright © 2006 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved; printed in U.S.A.

Vol. 163, No. 10
DOI: 10.1093/aje/kwj112
Advance Access publication March 8, 2006

Practice of Epidemiology

Reliability and Validity of the Past Year Total Physical Activity Questionnaire

Christine M. Friedenreich[1], Kerry S. Courneya[2], Heather K. Neilson[1], Charles E. Matthews[3], Gordon Willis[4], Melinda Irwin[5], Richard Troiano[4], and Rachel Ballard-Barbash[4]

[1] Division of Population Health and Information, Alberta Cancer Board, Calgary, Alberta, Canada.
[2] Faculty of Physical Education and Recreation, University of Alberta, Edmonton, Alberta, Canada.
[3] Division of General Internal Medicine, Department of Medicine, Vanderbilt-Ingram Cancer Centre, Vanderbilt University School of Medicine, Nashville, TN.
[4] Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD.
[5] Department of Epidemiology, School of Public Health, Yale University, New Haven, CT.

Received for publication April 29, 2005; accepted for publication December 21, 2005.

The authors determined the validity and reliability of their Past Year Total Physical Activity Questionnaire (PYTPAQ), which assesses the frequency, duration, and intensity of occupational, household, and recreational activities performed over the past year. The PYTPAQ was completed twice at baseline, 9 weeks apart (on average), by 154 healthy Canadian men and women aged 35–65 years for assessment of reliability. The PYTPAQ was completed again 1 year later as a self-administered questionnaire. Four times during the year, participants wore an accelerometer for 7 days and completed 7-day physical activity logs. The authors assessed validity by comparing PYTPAQ summary values with 1-year averages of the physical activity logs and accelerometer data and with physical fitness and anthropometric data measured at baseline and 1 year. Spearman correlations for reliability (metabolic equivalent-hours/week) were 0.64 for total activity, 0.70 for occupational activity, 0.73 for recreational activity, and 0.65 for household activity. For total activity, the intraclass correlation coefficient for correlation between the PYTPAQ and the 7-day physical activity logs was 0.42 (95% confidence interval: 0.28, 0.54), and for the accelerometer data it was 0.18 (95% confidence interval: 0.03, 0.32). Spearman correlations between PYTPAQ hours/week of vigorous activity and maximal oxygen uptake were 0.37 and 0.32 at baseline and follow-up, respectively. In general, the PYTPAQ has acceptable reliability and validity for measurement of past-year physical activity that is comparable to that of similar questionnaires.

data collection; exercise; physical fitness; questionnaires; reproducibility of results

Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient; MET, metabolic equivalent; PYTPAQ, Past Year Total Physical Activity Questionnaire; VO_{2max}, maximal oxygen uptake.

Questionnaires are often the only feasible method of assessing habitual physical activity in large populations (1), because they are easy to administer, relatively inexpensive, and noninvasive. Accurate and reliable assessment of habitual physical activity has been challenging, particularly for activities that are of low intensity, not done routinely, or not salient for the study respondent. Furthermore, relatively few questionnaires have undergone thorough pretesting or examination of their psychometric properties, thereby bringing into question the validity of disease-exposure relations assessed with these questionnaires.

As part of a program of research on physical activity and cancer outcomes (2), we developed a self-administered questionnaire for assessment of all types of physical activity

Reprint requests to Dr. Christine M. Friedenreich, Division of Population Health and Information, Alberta Cancer Board, 1331 29 Street NW, Calgary, Alberta T2N 4N2, Canada (e-mail: chrisf@cancerboard.ab.ca).

performed during the previous 12 months. This questionnaire was based on a questionnaire we previously developed and tested, the Lifetime Total Physical Activity Questionnaire (3), which we have used in case-control studies of lifetime physical activity and risk of cancers of the breast (4–6) and prostate (7). The Lifetime Total Physical Activity Questionnaire measures all types (i.e., occupational, household, and recreational) and all parameters (i.e., frequency, duration, and intensity) of physical activity engaged in from childhood to the time of the interview. It is an interviewer-administered questionnaire that uses recall calendars and memory-probing techniques known as cognitive interviewing methods (8) to assist respondents in reporting their lifetime history of physical activity. The purpose of this study was to evaluate the reliability and validity of the past-year component of our Lifetime Total Physical Activity Questionnaire, entitled the Past Year Total Physical Activity Questionnaire (PYTPAQ). Our ultimate aim is to use this questionnaire in large-scale cohort studies of various disease outcomes.

MATERIALS AND METHODS

Past Year Total Physical Activity Questionnaire

The PYTPAQ has an open table format, rather than specific questions, and is separated into three sections that assess occupational (including transportation to and from work), household, and recreational activity in the previous 12 months. It includes a description of the type of activity as well as the frequency (months/year, days/week), duration (hours/day), and perceived intensity of the activity. Definitions of each level of intensity (1 = sedentary, 2 = light, 3 = moderate, and 4 = heavy) are provided in the questionnaire, along with examples. For household and recreational activities, respondents are asked to report only those activities they engaged in while they were at least standing, whereas for occupational activity, all types of activity were reported, including sedentary activity. A list of recreational activities is given at the end of the questionnaire as a prompt for completion of that section. Responses are coded and entered into Blaise survey software (Statistics Netherlands; available from Westat, Inc., Rockville, Maryland) before being transferred into SAS (SAS Institute, Inc., Cary, North Carolina) for analysis.

The physical activity outcome measurements include the number of hours per week spent in each type of activity (i.e., occupational, household, and recreational) and the total amount of physical activity (i.e., the sum of the three types of activity) for the past year. The total hours per week spent in each activity are multiplied by the estimated metabolic cost of each activity (metabolic equivalent (MET) value) as determined from the Compendium of Physical Activities (9, 10). Hence, the main variable for comparison with other measures of activity and physical fitness is total activity expressed as MET-hours/week. In assessing validity and reliability, we excluded sedentary occupational activity from all summary activity values in order to evaluate activity (rather than inactivity) reporting.

Study population and recruitment

Eligible participants were between 35 and 65 years of age, resided in the Calgary Health Region of Alberta, Canada, were able to read and write in English, had no previous history of cancer (except nonmelanoma skin cancer), were not pregnant or planning to become pregnant in the next year, and were able to perform a submaximal physical fitness treadmill test at the beginning of the study. We chose this population to mirror the participants in an ongoing cohort study in Alberta for which the questionnaire was designed. The University of Calgary and the Alberta Cancer Board ethics review committees approved the study, and each participant provided informed consent prior to enrollment.

The study population was identified through random digit dialing using a stratified random sampling method. The sample was stratified by sex and age (35–49 years and 50–64 years) to ensure approximately equal distributions of participants across these strata. Study subjects were also asked two questions on their occupational and recreational activity and were stratified by level of physical activity (low, medium, high) according to their responses. We used this stratified random sampling method to ensure an equal distribution of study subjects across all levels of physical activity. A total of 1,190 people were screened for eligibility, of which 371 were eligible and 283 agreed to receive a study package; 177 completed baseline measurements and signed the informed consent form, and 154 completed all aspects of the study and were included in the analysis.

Study design and data collection methods

The overall study sequence is depicted in figure 1. A 1-year cohort study design was used to assess physical activity at baseline and for the year following enrollment. At baseline, study participants visited the Human Performance Laboratory at the University of Calgary, completed questionnaires on their medical, health, and lifestyle history, and completed the modified Physical Activity Readiness Questionnaire for determination of their eligibility for the physical fitness appraisal (11). They were provided with the PYTPAQ to report their physical activities for the preceding 12 months (PYTPAQ1) without any oral instructions or assistance, to simulate the completion of this questionnaire in a self-administered, unassisted study setting.

For determination of the reliability of the PYTPAQ, participants completed a second PYTPAQ (PYTPAQ2) 9 weeks (on average) after the baseline visit. During the year of study following baseline assessment, participants wore an accelerometer for four 1-week periods, each followed immediately by a week during which they completed 7-day physical activity logs. These four 2-week data collection periods were completed 12 weeks apart during the year to cover all seasons of the year. Study data collection was conducted between February 2002 and 2003. At the end of 12 months of data collection, the study participants were invited to return for follow-up measurements. They completed a third PYTPAQ for the 12 months of the study period (PYTPAQ3)

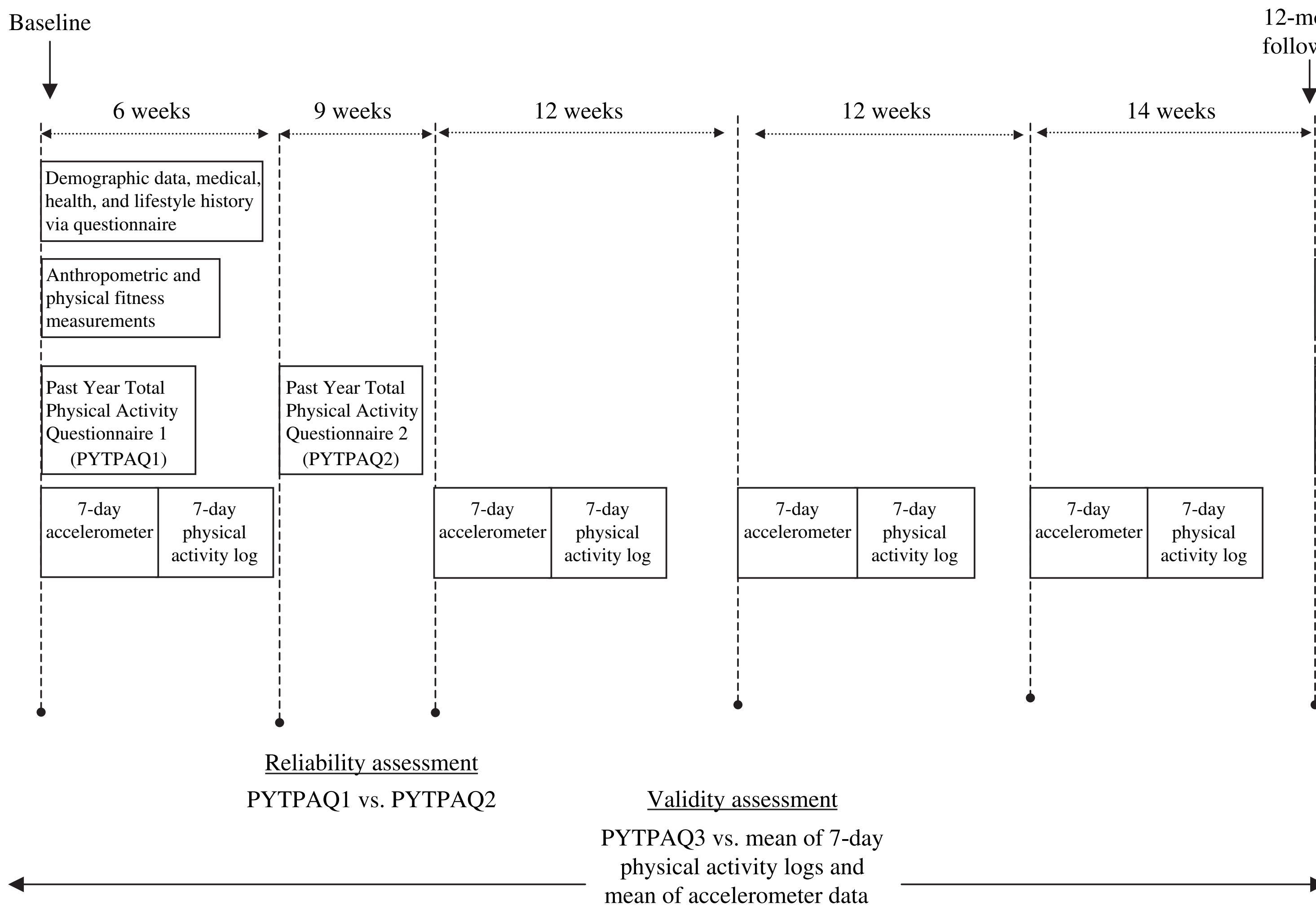

FIGURE 1. Study design for the Past Year Total Physical Activity Questionnaire (PYTPAQ) reliability and validation study, Calgary, Alberta, Canada, 2002–2003.

TABLE 1. Baseline characteristics of participants in a reliability and validation study of past-year physical activity assessment, Calgary, Alberta, Canada, 2002–2003 (*n* = 154)

Characteristic	Females (*n* = 79)			Males (*n* = 75)			Total (*n* = 154)		
	Mean (SD*)	No.	%†	Mean (SD)	No.	%†	Mean (SD)	No.	%†
Age (years)	48.9 (8.0)			48.5 (7.4)			48.7 (7.7)		
35–44		24	30.4		28	37.3		52	33.8
45–54		34	43.0		30	40.0		64	41.6
55–65		21	26.6		17	22.7		38	24.7
Body mass index‡	27.0 (5.4)			27.4 (3.6)			27.2 (4.6)		
Waist:hip ratio	0.8 (0.1)			0.9 (0.1)			0.8 (0.1)		
Percentage of body fat	33.9 (5.3)			21.9 (4.5)			28.1 (7.8)		
Predicted maximal oxygen uptake (ml/kg/minute)	26.5 (5.5)			32.8 (6.1)			29.6 (6.6)		
Self-rated general health									
Excellent		10	12.7		10	13.3		20	13.0
Very good		38	48.1		29	38.7		67	43.5
Good		25	31.7		30	40.0		55	35.7
Fair or poor		6	7.6		5	6.7		11	7.1
Usual frequency of alcohol consumption									
Never used alcohol		13	16.5		11	14.7		24	15.6
<1 drink/week		26	32.9		21	28.0		47	30.5
1–6 drinks/week		29	36.7		32	42.7		61	39.6
≥1 drink/day		10	12.7		11	14.7		21	13.6
Smoking habits									
Never smoker		38	48.1		43	57.3		81	52.6
Ex-smoker		28	35.4		24	32.0		52	33.8
Current smoker		12	15.2		8	10.7		20	13.0
Marital status									
Ever married or living common-law		71	89.9		70	93.3		141	91.6
Single, never married		8	10.1		5	6.7		13	8.4
Educational level									
High school or less		8	10.1		9	12.0		17	11.0
At least some technical school/college		29	36.7		21	28.0		50	32.5
At least some university		42	53.2		45	60.0		87	56.5
Annual household income									
<$50,000		17	21.5		13	17.3		30	19.5
$50,000–$89,999		28	35.4		29	38.7		57	37.0
≥$90,000		26	32.9		30	40.0		56	36.4

* SD, standard deviation.
† Percentages may not add up to 100% because of missing values.
‡ Weight (kg)/height (m)^2.

and had their anthropometric measures and physical fitness reassessed with the same methods as those used at baseline.

Anthropometric measures and physical fitness assessment

At the initial visit, certified physical fitness appraisers assessed the participants' anthropometric characteristics, body composition, and physical fitness using standardized methods. These measures included height, weight, waist and hip circumferences, and the sum of five skinfold thickness measurements. The multistage physical work capacity test developed by Sjostrand (12) was used on a mechanically braked cycle ergometer with a 12-minute length. A minimum of a 3-minute workload was used to estimate the participant's physical working capacity at a given heart rate.

TABLE 2. Physical activity as estimated by different assessment methods, in average hours/week and MET*-hours/week, by type of activity†, Calgary, Alberta, Canada, 2002–2003 (n = 154)

Physical activity assessment	Hours/week		MET-hours/week	
	Mean (SD*)	Median (IQR*)	Mean (SD)	Median (IQR)
Total activity				
Year 1				
PYTPAQ1*	50.5 (23.0)	51.0 (29.1)	133.3 (70.0)	119.1 (85.8)
PYTPAQ2	53.1 (18.5)	54.3 (23.6)	136.6 (66.0)	124.6 (64.8)
Year 2				
PYTPAQ3	52.1 (19.4)	53.7 (24.2)	133.7 (61.8)	121.4 (72.5)
7-day physical activity log	50.6 (17.6)	50.3 (26.4)	132.6 (51.0)	126.2 (62.7)
Accelerometer‡	43.1 (9.1)	42.1 (10.2)	135.2 (30.1)	131.9 (31.8)
Occupational activity				
Year 1				
PYTPAQ1	30.0 (20.9)	35.0 (35.6)	62.0 (59.8)	51.6 (58.3)
PYTPAQ2	31.9 (19.8)	37.5 (28.5)	65.1 (59.6)	57.3 (48.6)
Year 2				
PYTPAQ3	30.0 (19.4)	37.0 (32.3)	59.8 (51.6)	54.8 (55.4)
Household activity				
Year 1				
PYTPAQ1	14.2 (11.5)	12.5 (10.7)	40.4 (33.1)	34.8 (33.4)
PYTPAQ2	15.1 (11.2)	13.9 (11.5)	41.5 (30.2)	35.4 (33.1)
Year 2				
PYTPAQ3	15.8 (14.1)	12.2 (14.1)	43.4 (39.4)	33.4 (40.0)
Recreational activity				
Year 1				
PYTPAQ1	5.9 (5.7)	5.1 (6.4)	29.7 (27.9)	25.4 (31.7)
PYTPAQ2	5.7 (4.7)	5.0 (6.3)	28.8 (26.1)	23.6 (29.8)
Year 2				
PYTPAQ3	6.1 (4.9)	4.8 (6.8)	29.5 (28.2)	23.9 (29.4)

* MET, metabolic equivalent; SD, standard deviation; IQR, interquartile range; PYTPAQ, Past Year Total Physical Activity Questionnaire.

† 7-day physical activity log and accelerometer data were not collected according to type of activity (i.e., occupational, household, or recreational).

‡ Accelerometer totals include activities entered into the physical activity logs while the monitors were turned off. Measured and logged activities were combined, and activity levels of less than 150 counts/minute were censored (see Materials and Methods).

Heart rate was measured every minute and blood pressure at the end of each stage. Pedal revolutions were counted for the last 2 minutes of each workload to determine revolutions per minute (60–70 revolutions/minute). A linear regression of power output against heart rate was used to predict maximal oxygen uptake (VO_{2max}).

Physical activity logs

At the baseline visit, subjects were taught how to record their physical activity four times during the year using 7-day physical activity logs based on a modified version of the Bouchard Physical Activity Record (13). The participants coded their activity for each 15-minute interval of every hour of the day for 7 days consecutively. There were six available codes that participants could use to describe their activity (lying, sitting, and light, moderate, heavy, and very heavy activity). Examples of relevant types of activities were provided for each code. These data, in hours/week, were transformed into MET-hours/week by assigning a midpoint-intensity MET value to each category (2.0 for light activity, 3.5 for walking, 4.0 for moderate activity, 6.0 for heavy activity, and 8.0 for very heavy activity). We excluded time spent lying or sitting from all summary measures, for comparison with PYTPAQ estimates that also excluded sedentary activity. For inclusion in the data analysis, at least 5 days per week of physical activity log data were required. All 154 participants had at least 5 days of physical activity log data available for each week of data collection.

TABLE 3. Reliability of physical activity reports as assessed by Past Year Total Physical Activity Questionnaires 1 and 2 (PYTPAQ1 and PYTPAQ2), by type of activity, in average MET†-hours/week, Calgary, Alberta, Canada, 2002–2003 (n = 154)

Physical activity assessment comparison	MET-hours/week			
	Median difference (PYTPAQ1 − PYTPAQ2)	Spearman rank correlation	Intraclass correlation coefficient	95% confidence interval
Total activity				
Total population (n = 154)	−4.72 (39.72)‡	0.64**	0.66	0.56, 0.74
Sex				
Male (n = 75)	−5.18 (34.30)	0.71**	0.68	0.53, 0.78
Female (n = 79)	−2.99 (59.37)	0.56**	0.49	0.31, 0.64
Age (years)				
<50 (n = 75)	−5.18 (45.46)	0.68**	0.63	0.48, 0.75
≥50 (n = 79)	−4.47 (38.79)	0.59**	0.56	0.39, 0.70
Physical activity level§				
High (n = 51)	−4.47 (35.47)	0.68**	0.72	0.56, 0.83
Low/moderate (n = 103)	−4.98 (50.81)	0.58**	0.55	0.40, 0.67
Body mass index¶				
<25.0 (n = 55)	−9.52 (40.59)	0.66**	0.68	0.51, 0.80
≥25.0 (n = 99)	−2.96 (47.02)	0.63**	0.60	0.45, 0.71
Occupational activity				
Total population (n = 154)	−1.76 (28.39)	0.70**	0.58	0.47, 0.68
Sex				
Male (n = 75)	−2.13 (40.45)	0.72**	0.43	0.22, 0.60
Female (n = 79)	−1.65 (23.52)	0.71**	0.69	0.55, 0.79
Age (years)				
<50 (n = 75)	0.00 (27.70)	0.75**	0.54	0.36, 0.69
≥50 (n = 79)	−2.46 (28.71)	0.66**	0.63	0.47, 0.74
Physical activity level				
High (n = 51)	0.00 (30.11)	0.78**	0.78	0.64, 0.87
Low/moderate (n = 103)	−2.33 (26.30)*	0.66**	0.48	0.32, 0.62
Body mass index				
<25.0 (n = 55)	0.00 (25.38)	0.71**	0.61	0.41, 0.75
≥25.0 (n = 99)	−2.13 (30.60)	0.68**	0.56	0.40, 0.68

Table continues

Accelerometer

At the baseline visit, subjects were taught how to wear and record their activity using the MTI actigraph (Manufacturing Technology, Inc., Fort Walton Beach, Florida), formerly called the CSA uniaxial accelerometer (Computer Science and Applications, Inc., Shalimar Florida), which captures minute-by-minute observations of whole-body motion when worn on the waist. This device has been reported to be a valid and reliable tool for the measurement of dynamic physical activities, such as walking and running (14–18). Participants were instructed to wear the accelerometer during waking hours for 7 days and to record the times at which they turned the device on and off each day. If the accelerometer was removed during the day, the activity that was performed during this time (e.g., bathing, swimming) was recorded. A MET value was applied to each activity using the Compendium of Physical Activities (9, 10), permitting adjustment of the accelerometer data for periods of nonwear.

Data collected by the accelerometer were a series of activity counts representing the intensity and duration of motion in the sampling interval (i.e., counts/minute). The basic activity count data were summarized in terms of hours/week and, using the equation of Swartz et al. (19), as total energy expenditure (MET-hours/week) after censoring of activity counts below 150 counts/minute. Duration (hours/week) of time spent in activity of specific levels of intensity was estimated using the following thresholds: 150–759 counts/minute for light activity, 760–5,724 counts/minute for moderate activity, and ≥5,725 counts/minute for vigorous activity. Days with fewer than 10 hours of monitor wear, as

TABLE 3. Continued

Physical activity assessment comparison	MET-hours/week			
	Median difference (PYTPAQ1 – PYTPAQ2)	Spearman rank correlation	Intraclass correlation coefficient	95% confidence interval
Household activity				
Total population (n = 154)	−1.31 (24.49)	0.65**	0.57	0.46, 0.67
Sex				
Male (n = 75)	−0.75 (18.44)	0.50**	0.53	0.35, 0.68
Female (n = 79)	−1.33 (30.30)	0.58**	0.43	0.23, 0.60
Age (years)				
<50 (n = 75)	−0.27 (21.61)	0.61**	0.52	0.34, 0.67
≥50 (n = 79)	−3.12 (27.94)	0.66**	0.62	0.47, 0.74
Physical activity level				
High (n = 51)	−4.33 (24.65)	0.64**	0.54	0.32, 0.71
Low/moderate (n = 103)	0.00 (25.20)	0.64**	0.57	0.43, 0.69
Body mass index				
<25.0 (n = 55)	−0.58 (27.14)	0.67**	0.52	0.30, 0.69
≥25.0 (n = 99)	−3.12 (23.45)	0.63**	0.61	0.47, 0.72
Recreational activity				
Total population (n = 154)	0.00 (14.74)	0.73**	0.64	0.53, 0.72
Sex				
Male (n = 75)	0.00 (21.42)	0.67**	0.45	0.25, 0.61
Female (n = 79)	0.00 (13.56)	0.78**	0.83	0.75, 0.89
Age (years)				
<50 (n = 75)	0.00 (13.68)	0.68**	0.57	0.40, 0.71
≥50 (n = 79)	0.00 (15.95)	0.75**	0.69	0.55, 0.79
Physical activity level				
High (n = 51)	1.30 (13.44)	0.84**	0.66	0.47, 0.79
Low/moderate (n = 103)	0.00 (19.98)	0.66**	0.62	0.49, 0.73
Body mass index				
<25.0 (n = 55)	0.10 (14.34)	0.72**	0.66	0.48, 0.78
≥25.0 (n = 99)	0.00 (18.10)	0.74**	0.63	0.50, 0.74

* $p < 0.05$; ** $p < 0.0001$.
† MET, metabolic equivalent.
‡ Numbers in parentheses, interquartile range.
§ Level of activity according to 4 weeks of accelerometer data; the highest tertile was compared with the lowest and middle tertiles in MET-hours/week.
¶ Weight (kg)/height (m)^2.

determined by self-report, were excluded from analysis, as were weeks that had fewer than 3 days of data meeting the 10 hours/day criterion. Seven days of monitoring have been shown to capture more than 80 percent of the interindividual variation in adult physical inactivity and activity (20). Out of a possible 4,312 days of monitor data (154 participants × 28 days each), we excluded 15 days because of monitor malfunctions, 305 days because monitors had been worn for less than 10 hours, and five sessions each containing fewer than 3 full days of data (8 days in total). The average number of days and average duration of monitor wear were 23.2 days (standard deviation, 4.6) and 14.9 hours/day (standard deviation, 1.5), respectively.

Data analysis

The distribution of the physical activity data was examined prior to analysis, and since some data distributions were nonnormal, nonparametric tests were used. Wilcoxon signed-rank tests and Spearman rank correlations were used to assess questionnaire reliability and validity. Intraclass correlation coefficients (ICCs) were also calculated. We examined all assumptions for correlation before performing the analyses, and we chose the most appropriate results for presentation; that is, when necessary, the log-transformed ICCs were estimated and are presented here. We examined reliability by comparing the first two administrations of the questionnaire (PYTPAQ1 vs. PYTPAQ2).

As a measure of construct validity, we compared PYTPAQ3 with the mean values from the 4 weeks of 7-day physical activity logs and the accelerometer data. We used PYTPAQ3 for validity testing in order to reflect the appropriate temporal sequence between the physical activity logs and accelerometer measurements. Finally, we compared PYTPAQ1 and PYTPAQ3 with the measures of physical fitness and anthropometry taken at baseline and follow-up, respectively. The ICC models used for the PYTPAQ1-versus-PYTPAQ2 comparisons of reliability were one-way random-effects models. The ICC models used to compare PYTPAQ3 with the 7-day physical activity logs and accelerometer data were two-way mixed-effects models appropriate for the assessment of validity (21–23), where allowance is made for a difference in data collection methods. The latter ICC calculations were based on single measurements rather than average measurements and absolute agreement definitions (i.e., the measures produced by the various assessment instruments share a common metric (hours/week or MET-hours/week) and could therefore be directly compared rather than simply correlated). The analyses were stratified by gender, age ($<$50 years and $\geq$50 years), body mass index (weight (kg)/height (m)^2; $<$25.0 and $\geq$25.0), and level of activity according to the 4 weeks of accelerometer data (highest tertile vs. lowest and medium tertiles, in MET-hours/week).

RESULTS

The study population was fairly evenly divided into males ($n = 75$) and females ($n = 79$), and the average age was the same for both sexes (48 years), indicating that the sampling procedure for this study achieved an appropriate balance by gender and age as desired. Assessment of the participants' baseline sociodemographic characteristics, lifestyle risk factors, and health and medical profiles (table 1) showed that this population was moderately overweight (mean body mass index = 27.2) and had a relatively low predicted VO_{2max}. Over half of the study participants rated their own health as very good or excellent. There was a higher prevalence of ever or current smoking among women, while alcohol consumption levels were higher among men. A high percentage of this population was married or living common-law, was well-educated, and had a high household income.

The PYTPAQ is designed to estimate physical activity for each type of activity separately and for all activities combined (i.e., total activity) in terms of duration of activity (hours/week) and energy expenditure (MET-hours/week). When we examined the reliability of the estimates, the second measurement of physical activity obtained from PYTPAQ2 was slightly higher than that from PYTPAQ1, with the exception of recreational activity (table 2). The overall activity data obtained from the 7-day physical activity logs approximated those of the self-administered questionnaire (PYTPAQ3). The lowest estimates of activity, in hours/week, were from the accelerometer data; however, when the estimates were considered in MET-hours/week, the lowest median estimates were for the self-administered questionnaire.

We quantified the reliability of the PYTPAQ questionnaire by first comparing estimates for total activity and type of activity (table 3). For the overall study population, when we compared the first and second administrations of the PYTPAQ, there was a non-statistically-significant difference of just 1 hour of total activity per week (data not shown) or 4.72 MET-hours/week, and the Spearman rank correlation was 0.64. Somewhat higher Spearman correlations were found for total activity for males, persons under age 50 years at enrollment, persons with a lower body mass index, and high-activity participants. We also assessed reliability by examining each type of activity separately. The highest Spearman correlation for the total population was found for recreational activity (0.73), followed by occupational (0.70) and household (0.65) activity. Females had higher correlations than males for household activity (0.58 vs. 0.50) and recreational activity (0.78 vs. 0.67) and nearly equal correlations for occupational activity (0.71 vs. 0.72). Persons who were more active reported their occupational (0.78 vs. 0.66) and recreational (0.84 vs. 0.66) activity more reliably than did those with lower levels of activity. Correlations for occupational activity were also slightly higher for the younger ($<$50 years) study subjects and those with normal body mass indexes ($<$25). Less active persons reported significantly more occupational activity on the second PYTPAQ than on the first (median difference (PYTPAQ1 – PYTPAQ2) = −2.33 MET-hours/week). In further detailed analyses of the reliability of recall by intensity of total activity (data not shown), significant Spearman correlations were observed for all intensities, with the highest correlation being seen for vigorous activity (0.64) and lower correlations for moderate (0.56) and light (0.52) activity.

Additional assessments of the PYTPAQ's validity for total activity are reported in table 4. The median difference between PYTPAQ3 and the 7-day physical activity logs for the total population was only −1.48 MET-hours/week and was not statistically significant. Spearman correlation coefficients and ICCs for this comparison were 0.41 and 0.42 (95 percent confidence interval (CI): 0.28, 0.54), respectively. Slightly higher ICCs were found for males (ICC = 0.50, 95 percent CI: 0.31, 0.65), high-activity participants (ICC = 0.49, 95 percent CI: 0.24, 0.68), persons with a lower body mass index (ICC = 0.45, 95 percent CI: 0.21, 0.63), and persons under age 50 years (ICC = 0.51, 95 percent CI: 0.32, 0.66). Finally, the comparisons between PYTPAQ3 and the accelerometer data showed an overall underestimation of total activity in PYTPAQ3 relative to the accelerometer of 8.21 MET-hours/week that was not statistically significant. The Spearman correlations between the PYTPAQ3 and accelerometer data were the lowest estimated in this study and were statistically significant only for the total population, males, persons with a lower body mass index, and participants under age 50 years (0.26, 0.39, 0.38, and 0.43, respectively). In detailed analyses (data not shown), Spearman correlations between PYTPAQ3 and the accelerometer data for activity duration were significant for moderate (0.26) and vigorous (0.34) activity but not for light-intensity activity (−0.08).

The final method of assessing PYTPAQ validity was to compare it with measures of physical fitness and anthropometry. The anthropometric factors body mass index and

TABLE 4. Validity of total physical activity as estimated by different assessment methods, in average MET†-hours/week, Calgary, Alberta, Canada, 2002–2003 (n = 154)

Physical activity assessment comparison	MET-hours/week			
	Median difference (PYTPAQ3† − x‡)	Spearman rank correlation	Intraclass correlation coefficient	95% confidence interval
PYTPAQ3 vs. 7-day physical activity log				
Total population (n = 154)	−1.48 (80.01)§	0.41***	0.42	0.28, 0.54
Sex				
Male (n = 75)	4.15 (82.51)	0.46***	0.50	0.31, 0.65
Female (n = 79)	−3.37 (76.59)	0.36*	0.26	0.04, 0.45
Age (years)				
<50 (n = 75)	5.31 (78.23)	0.53***	0.51	0.32, 0.66
≥50 (n = 79)	−9.15 (80.80)	0.26*	0.31	0.09, 0.49
Physical activity level¶				
High (n = 51)	−13.68 (66.36)*	0.47**	0.49	0.24, 0.68
Low/moderate (n = 103)	13.81 (78.07)	0.30*	0.24	0.05, 0.42
Body mass index#				
<25.0 (n = 55)	−3.33 (78.01)	0.55**	0.45	0.21, 0.63
≥25.0 (n = 99)	−1.23 (81.52)	0.32*	0.30	0.11, 0.47
PYTPAQ3 vs. accelerometer				
Total population (n = 154)	−8.21 (64.31)	0.26*	0.18	0.03, 0.32
Sex				
Male (n = 75)	−0.02 (63.69)	0.39**	0.30	0.08, 0.49
Female (n = 79)	−14.24 (61.72)	0.14	0.10	−0.11, 0.31
Age (years)				
<50 (n = 75)	−5.67 (70.63)	0.43**	0.32	0.10, 0.51
≥50 (n = 79)	−9.18 (59.12)	0.05	0.01	−0.21, 0.23
Physical activity level				
High (n = 51)	−18.84 (68.74)*	0.16	0.14	−0.09, 0.38
Low/moderate (n = 103)	−5.01 (64.90)	0.15	0.05	−0.14, 0.24
Body mass index				
<25.0 (n = 55)	−10.75 (83.05)	0.38*	0.33	0.07, 0.55
≥25.0 (n = 99)	−5.67 (57.40)	0.19	0.16	−0.04, 0.34

* $p < 0.05$; ** $p < 0.001$; *** $p < 0.0001$.
† MET, metabolic equivalent; PYTPAQ, Past Year Total Physical Activity Questionnaire.
‡ x = comparison method.
§ Numbers in parentheses, interquartile range.
¶ Level of activity according to the average of 4 weeks of accelerometer data; the highest tertile was compared with the lowest and middle tertiles in MET-hours/week.
# Weight (kg)/height (m)^2.

waist:hip ratio were not consistently correlated with the PYTPAQ data for vigorous activity, but VO_{2max} was moderately correlated (table 5). Comparable results were obtained when we compared the anthropometric and physical fitness data measured at baseline and at follow-up with the measures of vigorous activity taken at these two time points.

DISCUSSION

In this reliability and validation study of a self-administered questionnaire on all types of past-year physical activity that was evaluated among men and women aged 35–65 years, reliability was acceptable overall. It tended to be higher for recreational and occupational activity than for household and total activity. In terms of total activity, somewhat higher reliability was seen among males, persons in the most physically active tertile of the population, persons with a lower body mass index, and persons who were under 50 years of age at recruitment. The validity of the PYTPAQ as expressed in MET-hours/week when the PYTPAQ was compared with 7-day physical activity logs and accelerometer data was modest, but it was sufficiently strong to suggest that the

TABLE 5. Correlation of past-year vigorous physical activity with baseline and follow-up physical fitness and anthropometric factors, in average hours/week, Calgary, Alberta, Canada, 2002–2003 (n = 154)

Comparison	Hours/week	
	No. of participants	Spearman rank correlation
PYTPAQ1† with:		
Body mass index‡ at baseline	154	−0.07
Waist:hip ratio at baseline	154	0.04
VO_{2max}† at baseline	154	0.37**
PYTPAQ3 with:		
Body mass index at follow-up	150	−0.22*
Waist:hip ratio at follow-up	150	−0.06
VO_{2max} at follow-up	146	0.32**

* $p < 0.05$; ** $p < 0.0001$.

† PYTPAQ, Past Year Total Physical Activity Questionnaire; VO_{2max}, maximal oxygen uptake.

‡ Weight (kg)/height (m)^2.

instrument would be able to rank-order adults by their activity levels.

We observed levels of reliability and validity for past-year physical activity which were comparable to those of previous studies that have examined these psychometric properties for self-administered activity questionnaires (24–29). In these previous studies, the reliability coefficients were generally higher than the validity coefficients, as can be expected, since diaries, logs, and accelerometers are not perfect ''gold standard'' measures of activity. Correlations on the order of 0.5–0.9 were observed for reliability, while the validity correlations were between 0.03 and 0.5. The correlations for the PYTPAQ were equal to or greater than those we observed in the reference-year (past-year) assessment of our Lifetime Total Physical Activity Questionnaire (3). In that assessment, we observed reliability correlations of 0.50 for total activity, 0.70 for occupational activity, 0.50 for household activity, and 0.57 for recreational activity (3).

One of the strengths of the PYTPAQ as compared with previously developed physical activity questionnaires is that it captures information about all *types* of physical activity, as well as frequency, intensity, and duration of activity. It also has an open table format that permits respondents to report the specific activities in which they regularly engage, including descriptions of each activity and individual coding of the intensity of the activity. Prior to this validation study, we undertook extensive pretesting and pilot-testing of the questionnaire using cognitive interviewing methods to ensure that the questions were understood and that the respondents could retrieve information appropriately and make judgements and estimations of the responses required (8). Other strengths of this validation study include the relatively large stratified random sample of men and women of a wide age range, the use of several validation methods, the low subject dropout rate during the year-long follow-up, and the rigorous study conduct.

The limitations of this study must be considered when reviewing these results. Three methods were used to assess the validity of this instrument: 7-day physical activity logs, accelerometers, and measurements of physical fitness and body composition. None of these methods are true ''gold standard'' measures.

The first validation method, the 7-day physical activity logs, provided 1-week measures of activity that were averaged to approximate the amount and intensity of activity the respondents performed over the course of the year. The advantage of physical activity logs is that they are complete records of activity performed during a fixed time period that are comprehensive and less subject to the types of recall and memory errors associated with long-term recall. These records, however, were completed for only four 1-week periods and thus may not be representative of the total 12 months. However, we believe that this sampling plan for the physical activity logs served to minimize variability in the measures due to both season and intraindividual variation in physical activity (30). Furthermore, the physical activity logs needed to be relatively easy to complete in order to decrease the respondent burden; therefore, participants were asked to record their activities in 15-minute blocks of time for seven general categories of activity intensity. Consequently, data collection methods for this instrument may have decreased the accuracy and precision of the data to some degree.

The second method of validation, accelerometers, provided a measure of all activity performed when the participant was not in water (bathing, swimming), asleep, or not wearing the accelerometer for some other reason. Hence, the amount of time during the day that was covered by the accelerometer was less than that for the PYTPAQs and the 7-day physical activity logs. While complete 24-hour data were recorded for the 7-day physical activity logs, only 10 hours per/day of accelerometer data were required for inclusion in the analysis. An attempt was made to include activities that were performed when the accelerometer was not worn, to increase comparability between these methods. However, the coding of these activities was less precise, since the respondents provided less detail on the exact type and intensity of the activity they were engaging in when the accelerometer was not being worn. In addition, many common light activities, such as housework, may not be captured reliably by an accelerometer because of the device's limitations in capturing energy expenditure from activities that entail little physical motion.

Measurements of physical fitness and anthropometry, used as the third method of evaluating the psychometric properties of the PYTPAQ, are actually indirect methods of validation, since the determinants of fitness and anthropometric factors are multidimensional and differ from those for physical activity. They provide an additional means of evaluating the properties of the questionnaire but cannot be used as the sole method of questionnaire validation. Nonetheless, we found positive and significant associations between physical fitness and vigorous physical activity and mainly negative associations between anthropometric factors and vigorous activity. Hence, the direction of the correlations was the direction expected, despite the fact that some correlations were only poor to moderate.

Despite the recognized limitations of these three comparison measures in validating physical activity, all are considered acceptable methods with which to validate questionnaires (26, 31–34). We believe that the consistency of effect across each measure reinforces the notion that the PYTPAQ is a valid method of assessing habitual physical activity in Canadian adults. Furthermore, each of the comparison measures we selected for this study has an error structure that is conceptually independent from errors in the PYTPAQ. When errors are uncorrelated, as may be the situation between our questionnaire and the comparison measures, the correlations between the objective measures and the PYTPAQ tend to be underestimated (35).

In conclusion, this study has shown that the PYTPAQ has an acceptable level of reliability and validity for measuring all types of self-reported physical activity in the past year in a middle-aged population of healthy men and women—a level comparable to that of other, similar questionnaires. The PYTPAQ was shown generally to have very good reliability and moderate validity for all subgroups of the study population examined, but reliability and validity were somewhat better for men, persons under age 50 years, and persons with a lower body mass index and for assessing recreational activity in terms of MET-hours/week. These patterns could be partly explained by that fact that these study population subgroups also had the highest levels of vigorous physical activity. Specifically, women tended to engage in more light activity—a type of activity for which assessment was generally less reliable and valid—than vigorous activity. Estimation of light activity, which also dominates physical activity in daily life, is recognized as challenging and is a limitation of most self-report physical activity assessment methods. The PYTPAQ was shown to have significant rank-order correlations, suggesting that it would be appropriate for epidemiologic studies in which the participants are classified into quantiles of physical activity. Additional research is needed to increase our understanding of the effect of measurement error on the correlations observed.

ACKNOWLEDGMENTS

This study was funded by grants from the Alberta Cancer Board's New Initiatives Program and the Canadian Cancer Society's Sociobehavioural Cancer Research Network. While conducting this research, Dr. Christine Friedenreich was supported by a Canadian Institutes for Health Research New Investigator Award and an Alberta Heritage Foundation for Medical Research Health Scholar Award. Dr. Kerry Courneya was supported by the Canada Research Chairs Program and a Research Team Grant from the National Cancer Institute of Canada. Dr. Charles Matthews was supported by a Career Development Award (K07 CA109363) from the US National Cancer Institute.

The authors acknowledge Nicole Slot for study coordination and data collection, Aleata Rhyorchuk for data cleaning, and Xuehao Chen for preliminary data analyses.

A copy of the Past Year Total Physical Activity Questionnaire and the coding manual can be obtained from Dr. Christine Friedenreich, Division of Population Health and Information, Alberta Cancer Board, 1331 29 Street NW, Calgary, Alberta T2N 4N2, Canada (e-mail: chrisf@cancerboard.ab.ca).

Conflict of interest: none declared.

REFERENCES

1. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003;37:197–206.
2. Courneya KS, Friedenreich CM. Framework PEACE: an organizational model for examining physical exercise across the cancer experience. Ann Behav Med 2001;23:263–72.
3. Friedenreich CM, Courneya KS, Bryant HE. The Lifetime Total Physical Activity Questionnaire: development and reliability. Med Sci Sports Exerc 1998;30:266–74.
4. Friedenreich CM, Bryant HE, Courneya KS. Case-control study of lifetime physical activity and breast cancer risk. Am J Epidemiol 2001;154:336–47.
5. Friedenreich CM, Courneya KS, Bryant HE. Influence of physical activity in different age and life periods on the risk of breast cancer. Epidemiology 2001;12:604–12.
6. Friedenreich CM, Courneya KS, Bryant HE. Relation between intensity of physical activity and breast cancer risk reduction. Med Sci Sports Exerc 2001;33:1538–45.
7. Friedenreich CM, McGregor SE, Courneya KS, et al. Case-control study of lifetime total physical activity and prostate cancer risk. Am J Epidemiol 2004;159:740–9.
8. Willis GB. Cognitive interviewing: a tool for improving questionnaire design. Thousand Oaks, CA: Sage Publications, 2005.
9. Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of Physical Activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000;32(suppl):S498–504.
10. Ainsworth BE, Haskell WL, Leon AS, et al. Compendium of Physical Activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 1993;25:71–80.
11. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992;17:338–45.
12. Sjostrand T. Changes in respiratory organs of workmen at an ore melting works. Acta Med Scand 1947;196:687–95.
13. Bouchard C, Tremblay A, Leblanc C, et al. A method to assess energy expenditure in children and adults. Am J Clin Nutr 1983;37:461–7.
14. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998;30:777–81.
15. Janz KF. Validation of the CSA accelerometer for assessing children's physical activity. Med Sci Sports Exerc 1994;26:369–75.
16. Matthews CE. Calibration of accelerometer output for adults. Med Sci Sports Exerc 2005;37(suppl):S512–22.
17. Nichols JF, Morgan CG, Chabot LE, et al. Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. Res Q Exerc Sport 2000;71:36–43.
18. Welk GJ, Schaben JA, Morrow JR Jr. Reliability of accelerometry-based activity monitors: a generalizability study. Med Sci Sports Exerc 2004;36:1637–45.

19. Swartz AM, Strath SJ, Bassett DR Jr, et al. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Med Sci Sports Exerc 2000;32(suppl):S450–6.
20. Matthews CE, Ainsworth BE, Thompson RW, et al. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc 2002;34:1376–81.
21. Armstrong BK, White E, Saracci R. Validity and reliability studies. In: Principles of exposure measurement in epidemiology. Oxford, United Kingdom: Oxford University Press, 1994:78–112.
22. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods 1996;1:30–46.
23. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: which coefficient should be used? Qual Life Res 2004;13:571–86.
24. Chasan-Taber L, Erickson JB, Nasca PC, et al. Validity and reproducibility of a physical activity questionnaire in women. Med Sci Sports Exerc 2002;34:987–92.
25. 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;7:81–6.
26. Jacobs DR Jr, Ainsworth BE, Hartman TJ, et al. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc 1993;25:81–91.
27. Matthews CE, Shu XO, Yang G, et al. Reproducibility and validity of the Shanghai Women's Health Study physical activity questionnaire. Am J Epidemiol 2003;158:1114–22.
28. Wolf AM, Hunter DJ, Colditz GA, et al. Reproducibility and validity of a self-administered physical activity questionnaire. Int J Epidemiol 1994;23:991–9.
29. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. Int J Epidemiol 2002;31:168–74.
30. Matthews CE, Hebert JR, Freedson PS, et al. Sources of variance in daily physical activity levels in the Seasonal Variation of Blood Cholesterol Study. Am J Epidemiol 2001; 153:987–95.
31. Ainsworth BE, Bassett DR Jr, Strath SJ, et al. Comparison of three methods for measuring the time spent in physical activity. Med Sci Sports Exerc 2000;32(suppl): S457–64.
32. Conway JM, Seale JL, Jacobs DR Jr, et al. Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. Am J Clin Nutr 2002;75:519–25.
33. Freedson PS, Miller K. Objective monitoring of physical activity using motion sensors and heart rate. Res Q Exerc Sport 2000;71(suppl):S21–9.
34. Leenders NY, Sherman WM, Nagaraja HN. Comparisons of four methods of estimating physical activity in adult women. Med Sci Sports Exerc 2000;32:1320–6.
35. Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a new dietary measurement error model for estimation of relative risk: application to four calibration studies. Am J Epidemiol 1999;150:642–51.