2-2012

Comparing the Validity of 2 Physical Activity Questionnaire Formats in African-American and Hispanic Women

Louise C. Mâsse

Janet E. Fulton

Kathleen B. Watson

Susan Tortolero

Harold W. Kohl III

See next page for additional authors

Publication Info

Published in *Journal of Physical Activity and Health*, Volume 9, Issue 2, 2012, pages 237-248.
Mâsse, L. C., Fulton, J. E., Watson, K. B., Tortolero, S., Kohl III, H. W., Meyers, M. C., ... Wong, W. W. (2012). Comparing the validity of 2 physical activity questionnaire formats in African-American and Hispanic women. *Journal of Physical Activity and Health, 9*(2), 237-248.

Author(s)
Louise C. Mâsse, Janet E. Fulton, Kathleen B. Watson, Susan Tortolero, Harold W. Kohl III, Michael C. Meyers, Steven N. Blair, and William W. Wong

Journal of Physical Activity and Health, 2012, 9, 237-248

Comparing the Validity of 2 Physical Activity Questionnaire Formats in African-American and Hispanic Women

Louise C. Mâsse, Janet E. Fulton, Kathleen B. Watson, Susan Tortolero, Harold W. Kohl III, Michael C. Meyers, Steven N. Blair, and William W. Wong

Background: The purpose of this study was to compare the validity of 2 physical activity questionnaire formats—one that lists activities (Checklist questionnaire) and one that assesses overall activities (Global questionnaire) by domain. ***Methods:*** Two questionnaire formats were validated among 260 African-American and Hispanic women (age 40–70) using 3 validation standards: 1) accelerometers to validate activities of ambulation; 2) diaries to validate physical activity domains (occupation, household, exercise, yard, family, volunteer/church work, and transportation); and 3) doubly-labeled water to validate physical activity energy expenditure (DLW-PAEE). ***Results:*** The proportion of total variance explained by the Checklist questionnaire was 38.4% with diaries, 9.0% with accelerometers, and 6.4% with DLW-PAEE. The Global questionnaire explained 17.6% of the total variance with diaries and about 5% with both accelerometers and with DLW-PAEE. Overall, associations with the 3 validation standards were slightly better with the Checklist questionnaire. However, agreement with DLW-PAEE was poor with both formats and the Checklist format resulted in greater overestimation. Validity results also indicated the Checklist format was better suited to recall household, family, and transportation activities. ***Conclusions:*** Overall, the Checklist format had slightly better measurement properties than the Global format. Both questionnaire formats are better suited to rank individuals.

Keywords: questionnaire design, validation study, diary, accelerometer, Doubly Labeled Water methodology

Questionnaires remain the most practical method to assess physical activity in large epidemiological studies where feasibility, cost, and participant burden are paramount.[1] Most epidemiological studies on physical activity and health use questionnaires to show that regular participation in moderate-intensity physical activity lowers the risk of developing coronary heart disease, hypertension, type 2 diabetes, certain cancers, and premature mortality.[1] Questionnaires are therefore useful to investigate the relationship between physical activity and chronic disease development and are also useful to monitor physical activity at the population level.[2,3] Developing valid physical activity questionnaires is therefore essential for large-scale epidemiologic studies.

Mâsse is with the Dept of Pediatrics, School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada. Fulton and Watson are with the Division of Nutrition, Physical Activity, and Obesity, Centers for Disease Control and Prevention, Atlanta, GA. Tortolero is with the School of Public Health, University of Texas–Houston. Kohl is with the Dept of Kinesiology and Health Education, University of Texas, Austin, TX. Meyers is with the Dept of Health and Human Development, Montana State University, Bozeman, MT. Blair is with the Arnold School of Public Health, University of South Carolina, Columbia, SC. Wong is with the Dept of Pediatrics, Baylor College of Medicine, Houston, TX.

Many physical activity questionnaires have been developed for adults,[4] although few[5,6] examine the multiple domains (eg, leisure, household, family, transportation, occupation) in which physical activity often occurs. Relying on questionnaires that assess physical activity in one domain (eg, leisure time physical activity) may lead to underestimation of physical activity among some population subgroups.[7] When a comprehensive assessment of physical activity is needed, it may be important to assess multiple domains—especially among groups that participate less frequently in traditional leisure-time physical activities.[7] Although many physical activity questionnaires have been developed, few have specifically targeted the physical activity behaviors of minority women. Developing a physical activity questionnaire that focuses specifically on African-American and Hispanic women and considers the sociocultural differences among racial/ethnic groups is particularly timely given that low levels of physical activity are reported by these women.[1,2] The disproportionate disease burden among African-American and Hispanic women, for several chronic conditions that may be improved through participation in physical activity,[8] makes it important and timely to develop and validate questionnaires for these populations.

Questionnaires vary in their approach to recall physical activity performed in multiple domains.[9,10] One main difference in designing questionnaires reflects either listing specific activities or assessing overall activities

within domains. When a list of specific activities is provided, participants are asked to recall the frequency (days per week) and duration (time per session) of several activities listed individually (eg, walking, bicycling, gardening). An alternative consists of asking participants to recall overall activities by domain, although example of activities may be provided, participants are asked to cognitively aggregate the amount of time spent in specific domains of physical activity. The listing approach often results in a more detailed and lengthy questionnaire than the overall assessment approach. Most physical activity questionnaire validation designs, attempt to validate a single questionnaire using either approach. As a result, little is known about differences in validity between questionnaires that list activities or assess overall activities to recall physical activity by domains.

The purpose of this study, therefore, was to compare the validity of 2 questionnaires approaches, namely listing specific activities versus assessing overall activities by domains to assess physical activity among women. As there is no one perfect physical activity validation standard,[4,10] comprehensive validation was carried out by triangulating the findings with the following 3 validation standards: 1) accelerometers to validate activities of ambulation; 2) activity diaries to validate physical activity domains (ie, occupation, household, exercise, yard, family, volunteer/church work, and transportation); and 3) Doubly Labeled Water (DLW) methodology to validate total and physical activity energy expenditures.[11] A triangulated approach to validation was selected because questionnaires often perform multiple functions; for example, they can estimate energy expenditure, time spent walking, or time spent in domain-specific moderate-intensity activities.

Methods and Procedures

Participants

Participants in the validation study were 130 African-American women and 130 Hispanic women (predominantly of Mexican decent). Women residing in the greater metropolitan area of Houston, Texas were eligible to participate if they self-identified as being African-American or of Hispanic decent, were 40 to 70 years of age, literate in English or Spanish, not pregnant, not suffering from any health conditions that would preclude them being active, and successfully completed and returned a 1-day physical activity diary (administered with the screening protocol). Participants were recruited through advertisements and community presentations. Of the 656 women who expressed an interest in the study, 260 enrolled in the validation study (recruitment protocol described elsewhere).[12] Demographic characteristics of the participants are summarized in Table 1.

Study Protocol

This study was approved by the Institutional Review Boards for the Protection of Human Participants of the

Table 1 Demographic Characteristics of the 260 Women Participating in the Women on the Move Study

Characteristics	
Age (yrs)	49.2 ± 7.0
Body composition	
Height (cm)[a]	160.2 ± 6.5
Weight (kg)[a]	76.9 ± 17.3
BMI (kg/m2)[a]	30.0 ± 6.3
Education[a]	
Elementary	7.8%
High school/GED	17.5%
Some college	42.0%
Graduate school	32.7%
Household income[a]	
<$14,999	22.2%
$15,000–$24,999	21.3%
$25,000–$34,999	17.8%
≥$35,000	38.7%
Ethnicity	
African-American	50%
Hispanic	50%

[a] Difference between African-American and Hispanic women significant at $P < .001$.

University of Texas-Houston Health Science Center and the Centers for Disease Control and Prevention. Data collected as part of the 2-week observational protocol are reported herein and described elsewhere.[12]

Participants were initially screened by telephone and eligible participants were scheduled for an in-person meeting where they were briefed about the study, signed a consent form, completed a demographic questionnaire, and were given a 1-day practice activity diary to complete and return by prepaid mail. In addition, participants had their height and weight (without shoes) recorded using a Seca alpha-digital scale (QuickMedical, Snoqualmie, WA) and a portable Accustat stadiometer (Genentech, San Francisco, CA) with the mean of 2 measurements used to compute participants' body mass index (BMI). Those who returned and successfully completed the 1-day practice diary were enrolled in the study.

The protocol started with the administration of questionnaires that assessed demographic characteristics and physical activity over the past 7 days with administration of the physical activity questionnaires randomly assigned but balanced. Before the administration of the questionnaires, research staff walked with participants at a pace of 3 to 4 mph to demonstrate brisk walking. Walking speed was controlled by having the participant walk a set distance within a given time frame which was set by the research staff (monitored with a stopwatch). All research staff were graduate students in public health and all staff received a 1-day training before administering the questionnaires.

To estimate energy expenditure from DLW, a baseline urine sample was collected followed by the ingestion of a known dose of DLW as $^{2}H_{2}{}^{18}O$ at 100mg/kg of body mass (Isotec, Miamisburg, OH). Participants were responsible for collecting urine on days 3, 5, 7, 10, 12, and 14. They were instructed not to collect the first void of the morning, to record the date and time of collection, and to store the samples in the freezer. On day 7, participants had another in-person meeting where they provided a urine sample, completed a battery of physical activity questionnaires, were fitted with an accelerometer, and were asked to wear the accelerometer and keep a physical activity diary for the following 7 days. Finally, on day 14, participants had another in-person meeting where they provided a urine sample, completed another battery of physical activity questionnaires, and reviewed their diary with a research assistant to ensure completeness and clarify any entries. Resting metabolic rate (RMR) was measured after the 2-week period.

Measures

Doubly Labeled Water Method. The DLW method was used to estimate Total Energy Expenditure (TEE) over the 2-week period under free-living conditions.[13] Gas-isotope-ratio mass spectrometry was used to analyze the baseline and postingestion urine samples for deuterium and ^{18}O isotopic enrichments. Hydrogen isotope ratio measurement was assessed by reducing the water in 10 µl of urine to hydrogen gas with 200 mg of zinc reagent at 500°C for 30 min[14] and the $^{2}H/^{1}H$ isotope ratios of the hydrogen gas were measured with a Finnigan Delta-E gas-isotope-ratio mass spectrometer (Finnigan MAT, San Jose, CA). The oxygen-isotope-ratio measurement was made by equilibrating 100µl of urine with 300 mbar of CO_2 of known ^{18}O content at 25°C for 10 hours using a VG ISOPREP-18 water-CO_2 equilibration system and the $^{18}O/^{16}O$ ratios of the CO_2 were measured with a VG SIRA-12 gas-isotope-ratio mass spectrometer (VG Isogas, Cheshire, England). The Weir equation served to compute TEE (kcal/day).[15] The multipoint method was employed to calculate the 7- and 14-day TEE[16] and the second week 7-day TEE was used for the analyses.

Resting Metabolic Rate (RMR). Participants were asked to refrain from strenuous activities 48 hours before RMR testing. On the day of the measurement, fasting RMR was measured between 6 and 8 AM. Participants were provided a hospital gown and rested for 20 minutes in a supine position before the RMR measurement. Measurement was taken in a thermo-regulated environment with minimal light and noise. Expiratory gas exchange was assessed by indirect calorimetry using a SensorMedics Vmax 229 ventilated open-hood system (SensorMedics, Yorba Linda, CA). RMR measurement was assessed for 40 minutes and participants were supervised to ensure they remained awake during the measurement. Oxygen variation of ±25 ml/min was the criterion to determine whether data collection was successful.[17]

Physical Activity Questionnaires. Following a comprehensive literature review, consultation with 53 experts[18] and 11 focus groups,[19] and building on the strengths of existing questionnaires, the research team developed 2 questionnaires—one that list specific activities by domains (referred to as the Checklist questionnaire) and another one that assessed overall physical activity by domains (referred to as the Global questionnaire).

The Checklist questionnaire is a 64-item, self-administered and partially interviewer-administered questionnaire that assessed the frequency and duration of physical activities performed in the previous 7 days (see online supplemental material A). The instrument is based on the Minnesota Leisure Time Physical Activity Questionnaire (MLTPAQ)[9,20,21] modified to assess the following domains of physical activity: household (24 items), yard (10 items), family (6 items), church/volunteer work (11 items), and transportation (2 items). For each activity listed, participants indicated whether the activity was performed during the past week (Yes/No), and the frequency (days/week) and duration (total minutes per week) of these activities. Assessment of the occupation and exercise domains were based on the Baecke questionnaire.[22] For occupation, the participant listed the 5 most frequently performed activities at work; whereas for exercise, they listed all exercises in which they participated on the previous 7 days.

The Global questionnaire is a brief, 8-item, interviewer-administered survey that was developed to assess moderate to vigorous intensity activities for 6 domains of physical activity [occupation, household, yard, family, church/volunteer work, and transportation (see online supplemental material B)]. Unlike the Checklist, the Global questionnaire did not probe for specific activities—instead participants had to recall activities from a given domain that met the criteria of being of moderate intensity. Unlike the Checklist, participants were provided with examples instead of a complete list of activities. For the exercise domain, participants were instructed to list their activities and indicate the total minutes spent doing each of the activities.

The Global format was selected as it is most often employed with self-report or interviewer-administered questionnaires. In contrast, the Checklist format was selected to improve recall of nonleisure time physical activity. The exercise domain for both questionnaires used the Baecke format given that it has been shown to provide adequate reliability (0.90) and validity (ranging from 0.32–0.57 with objective assessments and 0.71 with diaries)[4,9,22,23] and based on the recommendation from the experts to integrate previous knowledge.

Physical Activity Diary. Participants were asked to record all activities performed for at least 10 minutes in a 7-day physical activity diary designed for this study (described elsewhere[24]). Participants recorded starting and ending times for each activity, a description of the activity, and type of activity performed (occupation, household, yard, child care, pet care, exercise, volunteer work, transportation, walking, personal care, and other activities). Participants also recorded sleep time. Each diary was double-coded by trained research assistants using a standardized source.[25] Coding discrepancies were adjudicated by a third research assistant and the principal investigator (LCM).

Accelerometer. Participants wore a uniaxial accelerometer—the CSA model 7164 WAM (Computer Science Applications, Inc., Shalimar, FL)—during the same week they recorded their activities in the diary. Participants were instructed to wear the accelerometer (weight: 39.8 g, dimension: 5.1 × 4.9 × 1.6 cm) over the right hip, snug against the body, and preferably under their clothes, during all waking hours except when in contact with water (eg, showering or swimming). Counts per minute were recorded for a period of 7 days.

Data Processing

DLW Physical Activity Estimation. DLW physical activity energy expenditure (PAEE) in kcal/day was calculated as .9 multiplied by TEE minus RMR. Previous validation studies have employed the physical activity level (PAL) calculated as TEE/RMR, the PAL was not reported in this study because investigators have shown the PAL does not accurately correct for the effect of body mass on energy cost of activities and is thus not suited for comparison with questionnaires.[26]

Questionnaires and 7-Day Diary Physical Activity Estimation. Total and domain-specific MET-min/day (where 1 MET is equivalent to the energy expended at rest) were computed by multiplying minutes of activity by the corresponding MET value using a standard source.[25] For the Global questionnaire, MET-min were computed by assessing a fixed MET of 4.0 for domains that recalled overall moderate to vigorous activities. Moderate to vigorous intensity physical activity scores were computed by summing activities ≥ 3 METs. Total MET-min/day were computed by adding sleep time (0.9 METs) and unaccounted time (1.2 METs) and multiplying total MET-min/day by weight (kg) divided by 60. Similar to the DLW methodology, PAEE was computed as .9 multiplied by TEE minus RMR using a standard prediction equation.[27] Energy expenditure estimates are presented for participants with at least 5 valid days of data (n = 227).

Accelerometer Estimation. Published cut-points were employed to identify activities ≥ 3 METs.[28,29] Counts/day and accelerometer minutes ≥ 3 METs were compared with the questionnaires. Note that bouts of activities were compared with the questionnaire data, where the bout of activity had to be at least 10-minutes in duration but could allow a 1 minute interruption anywhere in the bout (eg, to account for stopping at a light). Participants who did not wear the accelerometer for ≥ 10 hours/day and for ≥ 5 days were excluded (further processing information provided elsewhere).[30]

Statistical Analysis

All data were analyzed using SPSS for Windows, Release 16.0 (Chicago, IL). Outliers were identified by reviewing all values with ≥ 3 standard deviations and bivariate regression plots. Pearson or Spearman correlations were calculated to assess the relationship between the questionnaires and the validation standards—with Spearman correlations reported for nonnormal data. Agreements between methods were examined using Bland and Altman's methods comparison technique.[31] To account for the effect of body mass on energy cost estimated with DLW, the method used by Mâsse et al was employed[30] to generate a random normal variable with the mean and standard deviation equal to the MET-min/day distribution of the DLW (TEE or PAEE). This random variable was then multiplied by body mass and served as a covariate in the multiple regression analyses. This method isolated the effect of body mass on the associations with the validation standard without eliminating the association that exists between the questionnaire and the validation standard.[30] For the regression analyses, the body mass covariate was first entered into the model followed by the questionnaire. All associations were considered meaningful if the r-square or incremental r-square was ≥ 5%.

Results

Diary as the Validation Standard

Checklist Questionnaire. Results comparing the Checklist with diary estimates are presented in Table 2. Correlations for all domains ranged from 0.32 to 0.67 and the correlation for all moderate intensity activities was 0.62. At the group level, the Checklist overestimated activities reported in household, yard, family, and volunteer/church domains (ranging from 12%–112%); whereas it underestimated the activities in the transportation domain (–40% difference). In contrast, the percentage difference for the occupation and exercise domains was less than 5% (–2% for occupation and –5% for exercise). When all moderate intensity activities were combined, the Checklist overestimated the MET-min/day reported in the diary by 68%. At the individual level, agreement between methods was poor as reflected by the large standard deviation of the difference (Table 2). In addition, for the occupational, yard, volunteer/church, and transportation domains the disagreement was systematically biased [see Table 2—the correlations of the difference against their means were significant ($P < .05$)]. As shown in Figure 1, in all cases, the scatter of the difference showed that the disagreement increased as participants' reported more activities. Although Figure 1 presents limits of agreement for select domains, this pattern was similar for all domains (ie, less disagreement among those who did little exercise and more disagreement among those who reported doing more exercise).

Global Questionnaire. Results comparing the Global questionnaire with the diary estimates are presented in Table 3. Four of the domains had correlations greater than 0.30: occupation, exercise, yard, and volunteer/church activities (ranging from 0.36–0.73). The correlation between the Global questionnaire and diary for all moderate to vigorous intensity activities equaled 0.42. At the group level, the Global questionnaire noticeably overestimated or underestimated min/day reported in the diary, except for yard, family, and volunteer/church activities. When all moderate intensity activities were combined, the Global questionnaire overestimated the min/day reported in the diary by 209%. At the individual level, the results for the Global questionnaire are similar to the Checklist—poor agreement between methods as observed by the large standard deviation of the difference (Table 3). In addition, disagreement between methods was systematically biased; it increased as the minutes of activities increased (see Figure 1).

DLW as the Validation Standard

DLW TEE. Results comparing the Checklist and Global questionnaires with DLW TEE are presented in Table 4 (see models 2 and 3). Overall, the correlations with DLW TEE and the questionnaires ranged from 0.54 to 0.62. Adjusting these associations for the effect of body

Table 2 Summary Statistics Comparing the Checklist and Diary Estimates of MET-min·day^{-1} in Moderate-to-Vigorous Intensity Physical Activity Using the Bland and Altman Methods Comparison Technique

		MET-min·day^{-1}		Correlations[b]	Bland-Altman measures of agreement		
		Diary	Checklist		Difference	% Difference	Correlation of the difference against their means[e]
Domains	n[a]	Mean (SD)	Mean (SD)	Diary vs Checklist	Mean (SD)[c]	Mean (SD)[d]	
Occupation	218	43 (95)	32 (84)	0.50[g]	–11 (84)	–5 (198)	–0.39[g]
Household	220	41 (47)	52 (74)	0.52[g]	13 (63)	51 (259)	0.12
Exercise	220	37 (57)	35 (60)	0.51[g]	0 (48)	–2 (118)	0.00
Yard	218	5(15)	9 (22)	0.67[g]	3 (16)	12 (127)	0.32[g]
Family	218	5 (14)	12 (30)	0.32[g]	7 (29)	112(1688)	0.30[g]
Volunteer/church work	225	3 (15)	11 (35)	0.34[g]	8 (35)	1 (50)	0.60[g]
Transportation	221	33 (46)	14 (25)	0.39[g]	–19 (40)	–40 (59)	–0.53[g]
Total moderate or vigorous activities	218	173 (150)	203 (197)	0.62[g]	31 (174)	68 (244)	0.18[f]

[a] Outliers were eliminated for the analyses and defined as being 3 standard deviations (SD) from the mean and through bivariate plots.

[b] Spearman correlations are presented.

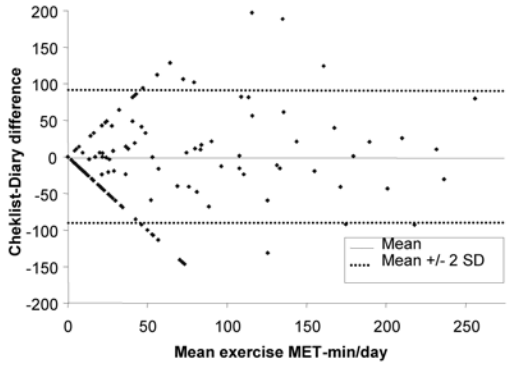
[c] Difference computed as the Checklist minus the diary.

[d] Percentage difference computed as the difference divided by the diary estimate × 100%.

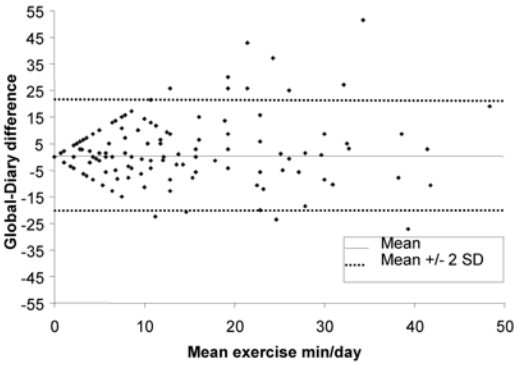
[e] Represents the correlation between the difference against the mean of the diary and Checklist.

[f] Significant at $P < .05$.

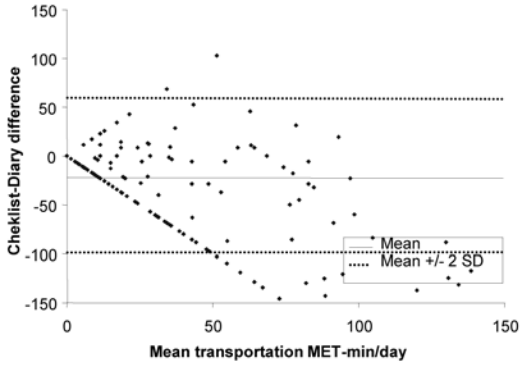
[g] Significant at $P < .001$.



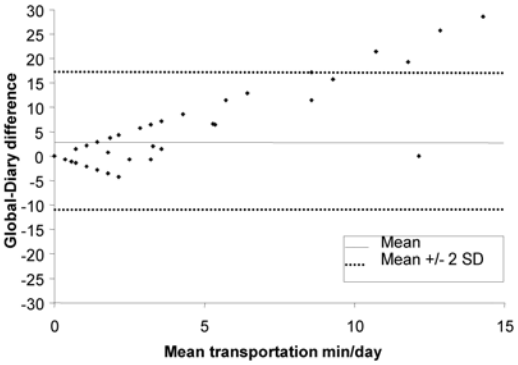
1a: Difference between Checklist and diary for exercise activities



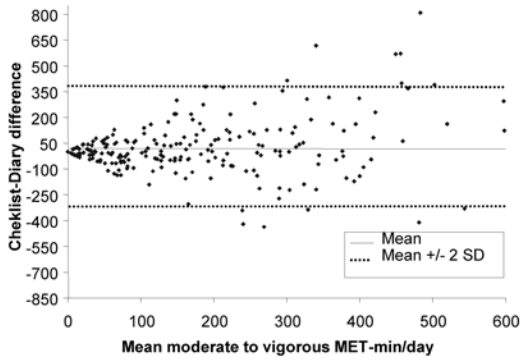
1e: Difference between Global and diary for exercise activities



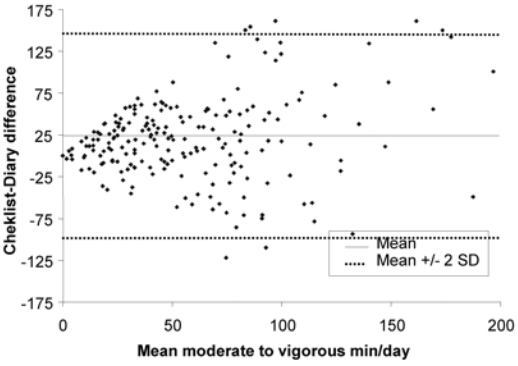
1b: Difference between Checklist and diary for transportation activities



1f: Difference between Global and diary for transportation activities



1c:Difference between Checklist and diary for all moderate and vigorous physical activities



1g: Difference between Global and diary for all moderate and vigorous physical activities

Figure 1 — Difference in MET-min/day between the Checklist (1a to 1c) and Global (1d to 1f) questionnaires with the diary for the exercise and transportation domains and for all moderate to vigorous activities.

Table 3 Summary Statistics Comparing the Global and Diary Estimates of MET-min·day^{-1} in Moderate-to-Vigorous Intensity Physical Activity Using the Bland and Altman Methods Comparison Technique

		MET-min·day^{-1}		Correlations[b]	Bland-Altman measures of agreement		
		Diary	Global		Difference	% Differrence	Correlation of the difference against their means[e]
Domains	n[a]	Mean (SD)	Mean (SD)	Diary vs Global	Mean (SD)[c]	Mean (SD)[d]	
Occupation	218	12 (27)	10 (21)	0.51[g]	–2 (27)	–78 (587)	–0.11
Household	215	12 (14)	19 (21)	0.20[g]	7 (23)	135 (570)	0.18
Exercise	219	7 (11)	9 (12)	0.73[g]	2 (10)	19 (109)	0.11
Yard	218	1 (3.9)	1 (3)	0.51[g]	–0 (4)	–6 (34)	–0.12
Family	218	1 (3.5)	2 (6)	0.10	1 (7)	72 (1219)	0.00
Volunteer/church work	222	1 (4.8)	2 (5)	0.36[g]	1 (6)	1 (37)	0.27
Transportation	218	0 (01)	4 (7)	0.09	3 (7)	10 (107)	0.80
Total moderate or vigorous activities	216	48 (41)	73 (59)	0.42[g]	26 (59)	209 (608)	–0.22

[a] Outliers were eliminated for the analyses and defined as being 3 standard deviations (SD) from the mean and through bivariate plots.

[b] Spearman correlations are presented.

[c] Difference computed as the Global minus the diary.

[d] Percentage difference computed as the difference divided by the diary estimate × 100.

[e] Represents the correlation between the difference against the mean of the diary and Global.

[f] Significant at $P < .05$.

[g] Significant at $P < .001$.

mass (model 1) indicated that the Checklist questionnaire explained about 23% of the total variance and the Global questionnaire explained about 13% of the total variance in DLW TEE.

As shown in Figure 2, the 95% limits of agreement for the 2 questionnaires reflect poor precision with DLW TEE. At the individual level, both questionnaires had smaller proportions of participants who underestimated DLW TEE as the majority overestimated their DLW TEE (ie, at the group level the overestimation was 12% for the Checklist and 6%for the Global). The significant ($P < .05$) correlations between the differences against their mean (Table 4) suggested a systematic bias ($P < .05$) among methodologies; as shown in Figure 2 the scatter of the difference showed that across values of DLW TEE, there is less agreement as DLW TEE increases.

DLW PAEE. Results comparing the 2 questionnaires against DLW PAEE are also presented in Table 4. The correlations among DLW PAEE with the questionnaires ranged from 0.23 to 0.26 (see models 5 and 6). Correcting these associations for the effects of body mass (model 4) showed that the Checklist questionnaire explained 6.4% of the total variance in DLW PAEE and the Global questionnaire explained 5%. Overestimation at the group level, although higher for the Checklist than the Global with DLW PAEE (67% versus 47%), was poor as well as agreement at the individual level (115% versus 115%). The correlations of the difference against their means with DLW PAEE were not significant, except for about 10 participants, the scatter of the difference appeared normally distributed (Figure 2).

Accelerometery as the Validation Standard

The MET-min/day of moderate to vigorous intensity activities from the 2 questionnaires were compared with accelerometer counts and total min ≥ 3 MET intensity. Associations between the Checklist and total accelerometer counts were 0.30 ($P < .05$) and 0.23 for accelerometer min ≥ 3 METs ($P < .05$). Associations with the Global questionnaire were 0.23 for total accelerometer counts ($P < .05$) and 0.22 for accelerometer min ≥ 3 METs ($P < .05$).

Discussion

Overview of Findings

As questionnaires will remain a useful tool to assess physical activity levels in large scale epidemiologic studies, this study sought to examine the validity of 2 physical activity questionnaire formats—one that lists specific activities by domains (Checklist questionnaire) and one that assesses overall physical activity by domains (Global questionnaire). Our findings show that both the Checklist and Global approaches had measurement properties consistent with other validated questionnaires.[4] Previous reviews and studies have noted that, in adults, correlations between past week self-report of physical activity and accelerometers have typically ranged from 0.22 to 0.39 with a median of 0.30 whereas with diaries the correlations have ranged from –0.02 to 0.52.[4–6,9] Our findings showed associations between both questionnaires

Table 4 Summary Statistics Comparing the 2 Questionnaires (Checklist and Global) With Estimates of Total Energy Expenditure (TEE) and Physical Activity Energy Expenditure (PAEE) From Doubly Labeled Water (DLW) Using the Bland and Altman Methods Comparison Technique

	T	Association between DLW and the questionnaire[a]		Bland-Altman measure of agreement between the DLW and the questionnaire		
				Difference	% Difference	Correlation of the difference against their means[d]
	Mean (SD)	r	Incremental r^2	Mean (SD)[b]	Mean (SD)[c]	
DLW TEE (kcal/day) (n = 207)	2283 (436)					
Model 1: Covariate[a]		.401[g]				
Model 2: Covariate + Checklist (kcal/day)	2531 (564)	.621[g]	22.5%	247 (451)	12 (20)	.207[g]
Model 3: Covariate + Global (kcal/day)	2380 (577)	.539[g]	13.0%	97 (509)	6 (23)	.318[g]
DLW PAEE (kcal/d) (n = 180)	689 (309)					
Model 4: Covariate[a]		.045				
Model 5: Covariate + Checklist (kcal/day)	939 (385)	.256[g]	6.4%	248 (411)	67 (116)	.111
Model 6: Covariate + Global (kcal/day)	821 (420)	.229[g]	5.0%	130 (459)	47 (118)	.101

[a] Random normal deviate using the mean and standard of the DLW in MET-min/day multiplied by the participants weight. This served to correct for the effect of body mass without eliminating the correlation that exists between body mass and physical activity.

[b] Pearson correlations are presented for DLW TEE; whereas Spearman correlations are presented for all other associations to account for the distributional properties of the data.

[c] Difference computed as the questionnaire minus the DLW.

[d] Percent difference computed as the difference divided by the DLW estimate.

[e] Represents the correlation between the difference against the mean of the questionnaire and DLW.

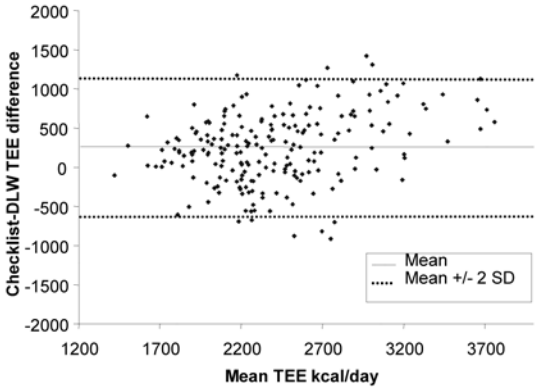
[f] Significant at $P < .05$.

[g] Significant at $P < .001$.

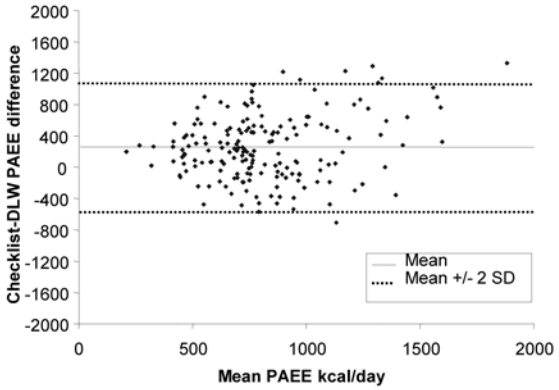
and accelerometers were within previously observed ranged—with the Checklist being at the median of the range (0.30) and the Global being closer to the lower end of the range (0.23). Associations with the diary differed slightly than previous studies as the Checklist association was higher than previously observed (0.62) whereas the Global was closer to the upper end of the range (0.42). Finally, comparing the DLW findings with previous studies is difficult given that previous studies have rarely isolated the effect of body mass using the methodology employed in this study.[32–41] For comparison purposes, when we validated accelerometers against DLW PAEE using the same methodology, accelerometers explained about 5% of the variance in DLW PAEE, after accounting for the effect of body mass.[30] Interestingly, both questionnaires explained about the same amount of variance in DLW PAEE than with accelerometry (ie, the 6.4% and about 5.0% for the Checklist and Global, respectively). Considering that we have triangulated the validity of the Checklist and Global questionnaires across 3 validation standards and that our findings are comparable to other instruments, our results support the acceptability of the measurement properties of both the Checklist and Global questionnaires.

Comparing the Psychometric Properties of the Checklist and Global Questionnaires

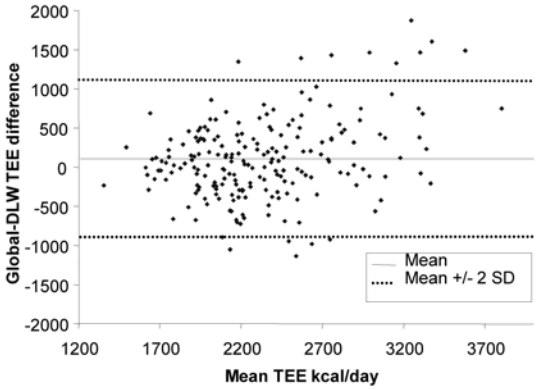
Our findings suggest the Checklist had slightly better psychometric properties than the Global across the validation standards. The amount of variance the Checklist explained was 38.4% with the diaries, 9.0% with accelerometers, and 6.4% with DLW PAEE whereas the Global explained 17.6% of the variance in the diaries and about 5% of the variance in accelerometers and DLW PAEE. The Global questionnaire differs from the Checklist in the following ways: 1) the Checklist provides a list of activities whereas the Global requires participant to recall which activities were of moderate intensity based on the walking demonstration; 2) the Global requires more mental computation (eg, adding minutes of activities per week and per domain of activities) than the Checklist; and 3) the Global does not distinguish moderate and vigorous intensity activities within a domain as a result there are assigned the same MET values whereas the Checklist activities are coded using the Compendium of Physical Activities.[25] Certainly, these main differences may explain the differences in the psychometric



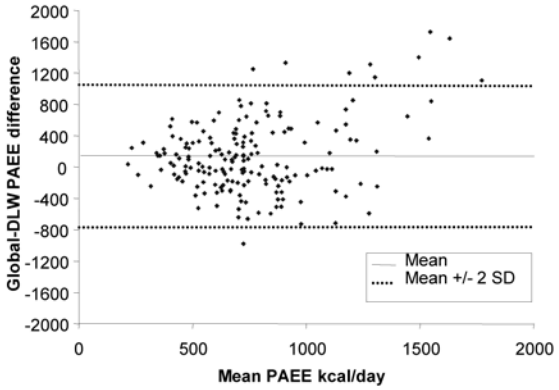
2a: Difference between the Checklist and DLW TEE (kcal/day)



2c: Difference between the Checklist and DLW PAEE (kcal/day)



2b: Difference between the Global and DLW TEE (kcal/day)



2d: Difference between the Global and DLW PAEE (kcal/day)

Figure 2 — Difference between the Checklist and Global total with Doubly Labeled Water (DLW) total energy expenditure (kcal/day) and with DLW physical activity energy expenditure (PAEE) (kcal/day).

properties between the 2 questionnaires. In addition, observing slightly inferior psychometric properties with the Global questionnaire may suggest that when women are asked to recall activities for specific domains, they may be unclear as to what specific activities to recall.[42] Conversely, probing for time spent in activities with a list, as done with the Checklist, may help women more accurately recall the activities in which they participate. An alternative explanation may be that since the Checklist approach requires less mental computation (ie, adding minutes across multiple activities) it may have resulted in fewer inaccuracies.[42] The latter explanation seems less plausible in light of the associations observed by domains, where the Global was found to inappropriately recall activities for the household, family, and transportation domains—further suggesting a list format may be better to improve recall in some domains as it may help participants recall activities they were not thinking of.

Agreement With the Validation Standards at the Group and Individual Levels

Although the Checklist outperformed the Global approach, both questionnaires overestimated the amount of time participants were active—a common criticism of questionnaires.[4] It is important to note, however, that

the amount of overestimation seemed to be associated with the format employed. To examine this we focused on the agreement between the questionnaires and the DLW results instead of the diaries given that diaries may provide a less objective assessment of physical activity. Overall agreement for both questionnaires was modest with DLW TEE (6 ± 23% to 12 ± 20%) and poor with DLW PAEE (47 ± 116% to 67 ± 118%). Two studies in the literature have reported adequate agreement at the group level with DLW TEE, although the measurement errors at the individual level remain quite high in these studies (ranging from –59% to 27%)[32,41,43] This is not surprising given the DLW methodology itself yields high measurement errors (ranging from –38% to 54%).[11,44] Similar to our findings, no published studies have reported adequate agreement with DLW PAEE and have indicated that questionnaires overestimate physical activity. Overall agreement with DLW TEE was twice as high for the Checklist (12%) as it was with the Global questionnaire (6%), and higher as well with DLW PAEE (67% of the Checklist versus 47% for the Global). Therefore, the Checklist appears to overestimate the DLW results more than the Global, suggesting that probing for specific activities may result in over reporting the amounts of physical activity. Clearly, the Checklist format resulted in a higher amount of overestimation than the Global format—even though the Checklist explained more of the variance in DLW TEE and DLW PAEE than the Global. It is unclear if listing the activities results in greater overestimation because participants reported time in multiple activities and/or because it is difficult to report time spent in concurrent activities when a discrete list of activities is provided.[42] The overestimation noted with the Checklist and Global questionnaires is consistent with other studies[4,42,45] and reinforces the difficulty of using questionnaires to provide accurate assessment of physical activity or to assess prevalence estimates for the population.[46,47]

Questionnaires Valid at Ranking Participants

Since both questionnaires explained a significant amount of variance with all the validation standards, they can be used in epidemiological studies focused on identifying risk factors associated with physical activity and similar questionnaires have been used extensively for this purpose. The Checklist and Global questionnaires were found to provide valid ranking of participants. It is important to note that similar to other questionnaires, the Checklist and Global questionnaires have not been shown to accurately estimate the amount of physical activity at the individual or group levels.[46,47] Similar to other questionnaires,[46,47] the Checklist and Global questionnaires can provide valid ranking of individuals and as such they are better suited to examine associations with risk factors in cross-sectional or longitudinal studies than they are at assessing actual amounts of physical activity performed.

Selecting a Specific Format

Among other considerations, the measurement properties of the Checklist or Global questionnaires should be taken into consideration when choosing a questionnaire format for an epidemiologic study. As previously mentioned, the Checklist outperformed the Global format as both questionnaires explained a modest amount of variance with accelerometry and DLW validation standards and a larger amount of variance with the diaries. Although these findings are comparable to previous validation studies,[4] varying the format to aid recall of moderate to vigorous intensity physical activity appeared to result in small gains in measurement properties. However, if the purpose of the study is to estimate physical activity by domains, the Checklist appears to best at recalling activities within the household, family, and transportation domains and might be better suited in studies where this is the main focus. Finally, it is acknowledged that the length of the Checklist may also be a limiting factor and may in some cases be a deciding factor in selecting a specific format but should be done with caution given that the shorter format may not be suited to answer all research questions.

Strengths and Limitations

Our validation study has a number of strengths which include having multiple validation standards, adjusting the DLW results for the well-known effect of body mass, and providing validity data for African-American and Hispanic women. Our study, however, is not without limitations. This study did not assess the ability of these questionnaires to detect change in behavior. In addition, the questionnaires are limited to English- and Spanish-speaking populations, and as volunteers, our sample is not necessarily generalizable to other African-American and Hispanic women in the United States, as the women in our study had higher education and income levels than U.S. women of similar age. In addition, as with any questionnaire, researchers who intend to use these questionnaires in other population subgroups should determine the cultural suitability of a questionnaire before using it with a different cultural group.[48]

Conclusions

Nevertheless, the major conclusions about the Checklist and Global questionnaires seem reasonable. The questionnaires appear to both provide acceptable validity for ranking participants in term of level of physical activity they perform and appear less ideal to estimate actual amount of physical activity performed. Overall, providing a list of activities seems to yield more accurate ranking of participants' activities than when participants are asked to recall activities by domains, in addition listing activities appears to be best at recalling certain physical activity domains (eg, household, family, and transportation

activities). Overall, neither questionnaires provided a valid estimate of the amount of activities performed as a result these questionnaires might have limited utility to assess population-based levels of physical activity. Finally, it appears that using a Checklist approach increases over-estimation of physical activity more than the Global questionnaire approach.

Acknowledgments

The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. This study was supported by the Women's Health Initiative through the Centers for Disease Control and Prevention (U48/CC609653). Dr. Mâsse received salary support from the Michael Smith Foundation for Health Research and the Child Family and Research Institute of British Columbia, Canada.

References

1. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation*. 2007;116(9):1081–1093.
2. Kruger J, Kohl HW, III, Miles IJ. Prevalence of regular physical activity among adults—United States, 2001 and 2005. *MMWR*. 2007;56:1209–1212.
3. Carlson SA, Fulton JE, Galuska DA, Kruger J, Lobelo F, Loustalot FV. Prevalence of self-reported physical activity adults - United States, 2007. *MMWR*. 2008;57(48):1297–1300.
4. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport*. 2000;71(2, Suppl):S1–S14.
5. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–1395.
6. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health*. 2009;6(6):790–804.
7. Evenson KR, Rosamond WD, Cai J, Pereira MA, Ainsworth BE. Occupational physical activity in the atherosclerosis risk in communities study. *Ann Epidemiol*. 2003;13(5):351–357.
8. The Office of Minority Health. *Data/Statistics.* Hyattsville, MD: US Department of Health and Human Services, The Office of Minority Health; 2008. Available at http://www.omhrc.gov/templates/browse.aspx?lvl=1&lvlid=2
9. Jacobs DR, Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc*. 1993;25(1):81–91.
10. Montoye HJ, Kemper HC, Saris WH, Washburn RA. *Measuring physical activity and energy expenditure*. Champaign, IL: Human Kinetics; 1996.
11. Speakman JR. *Doubly labelled water: theory and practice. Boundary Row*. London: Chapman & Hall; 1997.
12. Escobar-Chaves SL, Tortolero SR, Mâsse LC, Watson KB, Fulton JE. Recruiting and retaining minority women: findings from the Women on the Move study. *Ethn Dis*. 2002;12(2):242–251.
13. Lifson N. Theory of use of the turnover rates of body water for measuring energy and material balance. *J Theor Biol*. 1966;12(1):46–74.
14. Wong WW, Cochran WJ, Klish WJ, Smith EO, Lee LS, Klein PD. In vivo isotope-fractionation factors and the measurement of deuterium- and oxygen-18-dilution spaces from plasma, urine, saliva, respiratory water vapor, and carbon dioxide. *Am J Clin Nutr*. 1988;47(1):1–6.
15. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. *J Physiol*. 1949;109(1-2):1–9.
16. Cole TJ, Coward WA. Precision and accuracy of doubly labeled water energy expenditure by multipoint and two-point methods. *Am J Physiol*. 1992;263(5 Pt 1):E965–E973.
17. Turley KR, McBride PJ, Wilmore JH. Resting metabolic rate measured after subjects spent the night at home vs at a clinic. *Am J Clin Nutr*. 1993;58(2):141–144.
18. Mâsse LC, Ainsworth BE, Tortolero S, et al. Measuring physical activity in midlife, older, and minority women: issues from an expert panel. *J Womens Health*. 1998;7(1):57–67.
19. Tortolero SR, Mâsse LC, Fulton JE, Torres I, Kohl HW, III. Assessing physical activity among minority women: focus group results. *Womens Health Issues*. 1999;9(3):135–142.
20. Folsom AR, Caspersen CJ, Taylor HL, et al. Leisure time physical activity and its relationship to coronary risk factors in a population-based sample. The Minnesota Heart Survey. *Am J Epidemiol*. 1985;121(4):570–579.
21. Folsom AR, Jacobs DR, Jr, Caspersen CJ, Gomez-Marin O, Knudsen J. Test-retest reliability of the Minnesota Leisure Time Physical Activity Questionnaire. *J Chronic Dis*. 1986;39(7):505–511.
22. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36(5):936–942.
23. Pereira MA, FitzerGerald SJ, Gregg EW, et al. A collection of physical activity questionnaires for health-related research. *Med Sci Sports Exerc*. 1997;29(6 Suppl):S1–S205.
24. Eason KE, Mâsse LC, Kelder SH, Tortolero SR. Diary days needed to estimate activity among older African-American and Hispanic women. *Med Sci Sports Exerc*. 2002;34(8):1308–1315.
25. 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(9, Suppl):S498–S504.
26. Schoeller DA, Jefford G. Determinants of the energy costs of light activities: inferences for interpreting doubly labeled water data. *Int J Obes Relat Metab Disord*. 2002;26(1):97–101.
27. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr*. 1990;51(2):241–247.
28. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*. 1998;30(5):777–781.
29. Mâsse LC, Fulton J, Watson K, et al. Detecting bouts of physical activity in a field setting. *Res Q Exerc Sport*. 1999;70(3):212–219.

30. Mâsse LC, Fulton JE, Watson KL, Mahar MT, Meyers MC, Wong WW. Influence of body composition on physical activity validation studies using doubly labeled water. *J Appl Physiol*. 2004;96(4):1357–1364.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
32. Racette SB, Schoeller DA, Kushner RF. Comparison of heart rate and physical activity recall with doubly labeled water in obese women. *Med Sci Sports Exerc*. 1995;27(1):126–133.
33. Staten LK, Taren DL, Howell WH, et al. Validation of the Arizona Activity Frequency Questionnaire using doubly labeled water. *Med Sci Sports Exerc*. 2001;33(11):1959–1967.
34. Conway JM, Seale JL, Jacobs DR, Jr, Irwin ML, Ainsworth BE. Comparison of energy expenditure estimates from doubly labeled water, a physical activity questionnaire, and physical activity records. *Am J Clin Nutr*. 2002;75(3):519–525.
35. Conway JM, Irwin ML, Ainsworth BE. Estimating energy expenditure from the Minnesota Leisure Time Physical Activity and Tecumseh Occupational Activity questionnaires - a doubly labeled water validation. *J Clin Epidemiol*. 2002;55(4):392–399.
36. Irwin ML, Ainsworth BE, Conway JM. Estimation of energy expenditure from physical activity measures: determinants of accuracy. *Obes Res*. 2001;9(9):517–525.
37. Koebnick C, Wagner K, Thielecke F, et al. Validation of a simplified physical activity record by doubly labeled water technique. *Int J Obes (Lond)*. 2005;29(3):302–309.
38. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. *J Am Geriatr Soc*. 2001;49(1):28–35.
39. Bonnefoy M, Kostka T, Berthouze SE, Lacour JR. Validation of a physical activity questionnaire in the elderly. *Eur J Appl Physiol Occup Physiol*. 1996;74(6):528–533.
40. Mahabir S, Baer DJ, Giffen C, et al. Calorie intake misreporting by diet record and food frequency questionnaire compared to doubly labeled water among postmenopausal women. *Eur J Clin Nutr*. 2006;60(4):561–565.
41. Maddison R, Ni MC, Jiang Y, et al. International Physical Activity Questionnaire (IPAQ) and New Zealand Physical Activity Questionnaire (NZPAQ): a doubly labelled water validation. *Int J Behav Nutr Phys Act*. 2007;4:62.
42. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. *Med Sci Sports Exerc*. 2009;41(2):336–343.
43. Washburn RA, Jacobsen DJ, Sonko BJ, Hill JO, Donnelly JE. The validity of the Stanford Seven-Day Physical Activity Recall in young adults. *Med Sci Sports Exerc*. 2003;35(8):1374–1380.
44. Montoye HJ, Kemper HC, Saris WH, Washburn RA. *Measuring physical activity and energy expenditure*. Champaign, IL: Human Kinetics; 1995.
45. Jacobs DRJ, Hahn LP, Haskell WL, Pirie P, Sidney S. Validity and reliability of short physical activity history: cardia and the Minnesota Heart Health Program. *J Cardiopulm Rehabil*. 1989;9:448–459.
46. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37(3):197–206.
47. Westerterp KR. Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol*. 2009;105(6):823–828.
48. Mâsse LC. Reliability, validity, and methodological issues in assessing physical activity in a cross-cultural setting. *Res Q Exerc Sport*. 2000;71(2, Suppl):S54–S58.