

# Physical Activity Questionnaire Comprehension: Lessons from Cognitive Interviews

ANDREA ALTSCHULER<sup>1</sup>, TERESA PICCHI<sup>1</sup>, MELISSA NELSON<sup>1</sup>, JOHN D. ROGERS<sup>2</sup>, JAN HART<sup>1</sup>, and BARBARA STERNFELD<sup>1</sup>

<sup>1</sup>Division of Research, Kaiser Permanente Northern California, Oakland, CA; and <sup>2</sup>Public Research Institute, San Francisco State University, San Francisco, CA

## ABSTRACT

ALTSCHULER, A., T. PICCHI, M. NELSON, J. D. ROGERS, J. HART, and B. STERNFELD. Physical Activity Questionnaire Comprehension: Lessons from Cognitive Interviews. *Med. Sci. Sports Exerc.*, Vol. 41, No. 2, pp. 336–343, 2009. **Purpose:** To determine whether respondents share researchers' understandings of concepts and questions frequently used in the assessment of usual physical activity (PA) behavior. **Methods:** As part of On the Move, a study aimed at reducing measurement error in self-reported physical activity (PA), we conducted cognitive interviews with 19 men and 21 women, ages 45–65, regarding their responses to the PA questionnaires used in two large, population-based studies, Life After Cancer Epidemiology and California Men's Health study. One questionnaire asks about the frequency, the duration, and the perceived intensity of a range of specific activities in several different domains over the past 12 months. The second questionnaire asks about frequency and duration of specific, mostly recreational activities, grouped by intensity (i.e., moderate or vigorous) over the past 3 months. We used verbal probing techniques to allow respondents to describe their thought processes as they completed the questionnaires. All interviews were tape-recorded and transcribed, and the transcripts were then analyzed using standard qualitative methods. **Results:** Cognitive interviews demonstrated that a sizable number of respondents understood "intensity" in terms of emotional or psychological intensity rather than physical effort. As a result, the perceived intensity with which a participant reported doing a specific activity often bore little relationship to the MET value of that activity. Additionally, participants often counted the same activity more than once, overestimated work-related PA, and understood activities that were grouped together in a single category to be definitive lists rather than examples. **Conclusion:** Cognitive interviews revealed significant gaps between respondents' interpretations of some PA questions and researchers' assumptions about what those questions were intended to measure. Some sources of measurement error in self-reported PA may be minimized by additional research that focuses on the cognitive processes required to respond to PA questionnaires. **Key Words:** MEASUREMENT ERROR IN SELF-REPORTED PHYSICAL ACTIVITY, QUESTIONNAIRE DESIGN, QUESTIONNAIRE VALIDITY, RESPONDENT COGNITION

A large body of epidemiological literature demonstrates that regular physical activity (PA) leads to numerous health benefits, including reduced risk of cardiovascular disease, type 2 diabetes, colon cancer, breast cancer, and osteoporotic fractures, and improved mental health and physical and cognitive function (16,17,19,21, 24,32,35,38). However, assessment of PA has been based largely on self-report, which is always subject to recall error (18). Although many PA recalls and questionnaires rank individuals reasonably well and have an acceptable degree of established validity (13), they suffer from some amount

of measurement error that may lead to misclassification. As a result, the magnitude of the observed protective associations between PA and health outcomes is likely to be underestimated (6).

During the past decade, significant efforts have been undertaken to improve assessment of PA. In addition to the rapid development of objective approaches to PA measurement (i.e., accelerometers, pedometers, and various types of physiological monitoring), attention has been focused on improving the specificity of data collected with self-report instruments. For example, it is now widely recognized that many early PA surveys were not comprehensive enough to measure PA accurately in specific subgroups, including women, elderly, and racial/ethnic minorities (1,22). The addition of survey items relevant to household and caregiving activities or to culturally specific activities was intended to reduce the amount of misclassification and to promote greater precision in estimates of association (4). Similarly, the emphasis on assessing moderate "lifestyle" PA, such as walking and other activities related to transportation or daily routine (32), allows for more

Address for correspondence: Barbara Sternfeld, Ph.D., Division of Research, 2000 Broadway, Oakland, CA 94612; E-mail: bxs@dor.kaiser.org.

Submitted for publication February 2008.

Accepted for publication July 2008.

0195-9131/09/4102-0336/0

MEDICINE & SCIENCE IN SPORTS & EXERCISE®

Copyright © 2009 by the American College of Sports Medicine

DOI: 10.1249/MSS.0b013e318186b1b1

accurate observation of interindividual variability, particularly among the large segment of the population that does not engage in exercise *per se* (31).

Less attention has been paid to respondents' comprehension of PA questionnaires, and there is a relative paucity of evidence that PA questionnaire items are understood by respondents in the way intended by researchers. In the current study, we conducted cognitive interviews with participants as one component of a pilot study for On the Move, a project designed to quantify measurement error in PA questionnaires used in two large cohort studies. The specific aims of the current study were to document respondent comprehension and interpretation of PA survey questions and, based on these data, to design improved PA questionnaires that are less susceptible to response errors.

## METHODS

**Study sample.** The sampling frame for the pilot study was composed of male and female members of the Northern California Kaiser Permanente Medical Care Program (KPNC) between the ages of 45 and 65 yr and living in reasonable proximity to the research clinic in Oakland, CA. KPNC is the nation's oldest and largest non-profit-integrated health care delivery system and provides care to over 3.2 million people in Northern California (approximately 30% of the population). Individuals were randomly selected from the sampling frame, screened for eligibility, and recruited into the study until the targeted sample size for the cognitive interviews ( $N = 40$ ) was reached. Out of 74 individuals successfully contacted by telephone, 6 were excluded due to a language barrier, 5 were excluded for medical conditions that might preclude participation in normal levels of PA, and 23 (36.5% of eligible individuals) declined participation. The remaining 40 (19 men and 21 women) completed the cognitive interviews. Participants were offered \$35 for completion of the study, and all signed an informed consent document. Study protocols were approved by the KPNC Institutional Review Board.

**Cognitive interviews.** Cognitive interviewing is a methodology for eliciting the thought processes behind respondents' answers to survey questions that involves respondents completing a questionnaire and discussing their answers and the ways in which they arrived at them (37). According to Willis (37), there are two primary methods of cognitive interviewing, "think aloud" and "verbal probing." In the former method, respondents are asked to think out loud as they are answering a particular question so as to relay the processes of their thinking in real time. With verbal probing, respondents answer a single question or a series of questions and are then immediately asked about how they arrived at their answers. With both approaches, interviews are usually tape-recorded and transcribed and then analyzed using qualitative methods.

For this study, we used verbal probing techniques that included several predetermined, structured questions, such

as "How hard was it for you to answer this question?" and "How sure are you of your answer?" that were asked of the participants at the end of each page of the questionnaire and that referred back to each of the questions on that page (structured questions are available at [www.dor.kaiser.org/studies/otm/index.shtml](http://www.dor.kaiser.org/studies/otm/index.shtml)). Three experienced interviewers familiar with PA questionnaires were trained by an investigator (AA) with expertise in cognitive interviewing techniques and qualitative analysis to ask the structured questions and also to create and use probes to elicit additional details. Particular attention was focused on questions that required participants to judge the intensity of their activities and questions that asked them to quantify PA. Half of the participants were randomly assigned to complete the interview for one questionnaire, whereas the remaining half completed the interview for the second questionnaire. Each interview lasted between 1 and 1.5 h.

**Source questionnaires.** The two questionnaires that were evaluated were the Life After Cancer Epidemiology (LACE) physical activity questionnaire (PAQ) and the PA questions from the California Men's Health (CMH) study survey. LACE was funded by the National Cancer Institute to examine behavioral factors and breast cancer recurrence, and the CMH study was funded by the California Cancer Research Program to investigate etiologic factors related to prostate cancer. Both questionnaires may be viewed at [www.dor.kaiser.org/studies/otm/index.shtml](http://www.dor.kaiser.org/studies/otm/index.shtml).

The LACE PA questionnaire, which is a 10-page scannable form consisting of 56 items that takes about 15–20 min to complete, is formatted like a food frequency questionnaire. It is modeled loosely after the Arizona Activity Frequency Questionnaire, which was validated against doubly labeled water (28). Respondents are asked to select, from a long list of specific activities, those activities in which they participated at least once a month over the past 12 months. They are also asked to indicate, with categorical responses, the frequency, the duration, and the intensity with which they engaged in each activity. The response categories for frequency range from "never or less than one time per month" to "more than five times per week," and the response categories for duration range from "less than 15 min" to "61–90 min." To assess intensity, respondents are asked, "When you did this activity, did your heart rate and breathing increase?," and the response categories are "not at all or very little," "a medium amount," or "a large amount." The activities are grouped by domain (work-related activities, home/caregiving activities, recreational activities, and transportation), allowing for calculation of domain- and/or intensity-specific summary variables in units of MET-hours per week. METs are measures of absolute intensity that are independent of body weight (1 MET is approximately equal to  $1 \text{ kcal}\cdot\text{kg}^{-1}\cdot\text{h}^{-1}$ ).

The CMH questionnaire, which is a four-page scannable form consisting of 22 items that takes less than 10 min to complete, assesses mostly sports and exercise over the past 3 months with questions adapted from the CARDIA Physical

Activity History, an instrument that has reasonable indirect validity by showing the expected relationships with aerobic capacity and percent body fat (13,14) and strong inverse relation with most cardiovascular risk factors (27). Activities are categorized as either moderate (3–6 METs) or vigorous (>6 METs) intensity, and activities with similar MET values (e.g., softball, volleyball, and shooting baskets) are grouped together. Running or jogging, road or mountain biking, and swimming laps are listed as vigorous activities, and “leisurely” jogging, biking, or swimming are listed as a single category under moderate activities. For each activity or group of activities, respondents indicate the frequency and the duration of participation, and summary scores in MET-hours per week are derived for total recreational activity, vigorous recreational activity, and moderate recreational activity by multiplying assigned MET values by duration and frequency and summing over all activities. The response categories for frequency and duration are the same as those on the LACE questionnaire. The CMH questionnaire also includes two questions about hours per day of sedentary behavior and seven items related to occupational activity taken from the Baecke Physical Activity Questionnaire (7).

**Data analysis.** All interviews were tape-recorded and transcribed, and transcripts were coded and analyzed using standard qualitative methods (8,20). First, each transcript was reviewed to develop major themes and subthemes relevant to cognitive processes required to respond to survey questions, namely, comprehension and interpretation of questions, recall of relevant information, and quantification and synthesis of recalled information into appropriate responses. The questions we initially asked respondents were designed to elucidate these processes for issues related specifically to definitions of intensity, understandings of differences and similarities among specific physical activities, and estimations of frequency and duration of daily, weekly, and seasonal activities. Individual codes were then developed by assessing commonalities among respondents’ answers within each of the themes and subthemes, and all transcripts were coded accordingly. In the small number of cases in which coding disagreements arose, we resolved them by coming to consensus among the two study investigators (AA and BS).

## RESULTS

Respondents represented a range of sociodemographic characteristics but were predominantly African American (27%) or white (57%), used more than 20 h·wk<sup>-1</sup>, economically stable, and well educated (Table 1).

Analysis of the cognitive interviews revealed several problems with the PA questionnaires, including 1) definitions of intensity, 2) estimation of work-related PA, 3) inclusion of the same activity in different domains, 4) generalizing from examples of specific activities, and 5) use of a reference group. These problems are described in more detail below.

TABLE 1. Participant demographics.

	Men (n = 19)	Women (n = 21)	Total (N = 40)
<b>Race</b>			
White	68% (13)	48% (10)	57% (23)
African American	21% (4)	33% (7)	27% (11)
Chinese	5% (1)	5% (1)	5% (2)
Mexican/Central American	0	5% (1)	2% (1)
Declined to state	5% (1)	9% (2)	7% (3)
<b>Education</b>			
College degree or higher	84% (16)	43% (9)	62% (25)
Some college or vocational/ technical school	10% (2)	38% (8)	25% (10)
High school	0	9% (2)	5% (2)
Declined to state	5% (1)	9% (2)	7% (3)
<b>Employment</b>			
≥20 h·wk <sup>-1</sup>	89% (17)	62% (13)	75% (30)
Retired	5% (1)	24% (5)	15% (6)
Declined to state	5% (1)	14% (3)	10% (4)
<b>Paying for basics (food, housing)</b>			
Not hard	84% (16)	76% (16)	80% (32)
Somewhat	10% (2)	9% (2)	10% (4)
Hard	0	5% (1)	2% (1)
Declined to state	5% (1)	9% (2)	7% (3)
<b>Self-rated health</b>			
Very good or excellent	47% (9)	52% (11)	50% (20)
Good	42% (8)	33% (7)	37% (15)
Fair or poor	5% (1)	5% (1)	5% (2)
Declined to state	5% (1)	9% (2)	7% (3)

**Definitions of intensity.** Although increased heart rate and respiration are commonly used as cues for estimating intensity of PA and are explicitly used to describe intensity on the LACE questionnaire, some respondents did not define intensity in this way. Many respondents volunteered sweating, fatigue, and/or muscle soreness as more meaningful indicators of physical intensity. For example, in defining intensity, one woman responded, “I just think of myself all sweaty and putting all the energy out there.” Another woman defined vigorous activity as causing her to pant and become exhausted. Almost all the men mentioned sweating as an indicator of vigorous physical exertion. For respondents who included increased heart rate and respiration as markers of intensity, they often only responded that way after being queried by the interviewer.

For most participants, there was little distinction between moderate and leisurely activities. Two respondents, a man and a woman, even rated leisurely activity as more strenuous than moderate activity. An interchange between the interviewer and the male respondent went as follows:

Q: How do you define intensity if you’re walking to get somewhere, as opposed to walking for exercise?

A: I define it as leisurely. I wouldn’t normally use “moderate” as part of my vocabulary, because, to me, “moderately” is slower and more guarded than “leisurely.”

The most surprising finding related to intensity was that seven respondents (17.5%) interpreted the term almost exclusively in terms of psychological intensity or the sense of pleasure they derived from the activity. As a result, these women and men rated an activity that typically has low physical intensity, such as board games or attending a

concert, as having “high intensity.” For example, one woman stated:

Q: Where you indicated intensity, it says, “When you did this activity, did your heart rate and breathing increase?” You put “Yes, a large amount” for sewing and for reading and for all these activities.

A: Right.

Q: How did you get to this answer?

A: Because it’s something I enjoy. You know what I mean?

Q: So when you’re thinking of intensity, you’re thinking more of how much you *enjoy* it? Is that what you’re thinking? (emphasis in original)

A: Which also increases your heart rate and all of that, because you perk up.

**Estimation of work-related PA.** Both questionnaires ask about the amount of time spent sitting, standing, walking, lifting heavy loads, use of heavy equipment, and stooping or bending during the work day. Doing heavy manual labor was asked on the LACE questionnaire, and sweating from exertion was asked on the CMH questionnaire. On the CMH questionnaire, respondents were asked if they did each of the above activities “never, seldom, sometimes, often, or always.” On the LACE PAQ, respondents were asked if they did these activities “never or less than 1 h·d<sup>-1</sup>, 1–2 h, more than 2–4 h, more than 4–6 h, more than 6–8 h, and more than 8 h.”

Respondents often found these questions confusing, in part due to the difficulty in quantifying the amount of time they spend on each of these activities on a typical workday. In addition, some respondents had difficulty understanding the activities as distinct; they pointed out that walking cannot be done without standing. It also was confusing for individuals whose work involves walking, but not walking that they thought could reasonably be interpreted as exercise. For example, a male teacher said:

“You’re standing and walking [in the classroom]. There’s a little something in between there, too. But not like I’m doing an aerobic walk down the road. Doing a power walk is different than walking around in a classroom.”

Finally, many sedentary office workers appeared to overestimate the amount of time they spent walking or standing because the walking they described was to the copy machine or to a colleague’s office.

#### **Inclusion of the same activity in different domains.**

Respondents often double- or triple-counted the amount of time they spent walking and cycling and occasionally running or jogging. This was because the LACE questionnaire included walking in several different domains (e.g., walking the dog in caregiving activities, walking for exercise/pleasure at a brisk pace and walking for exercise/pleasure at a leisurely pace in recreational activities, and walking for transportation). Both questionnaires also included walking in the questions about work activities. The LACE PAQ also listed biking twice, once under recreational/sports activities, and once under transportation, as did the

CMH questionnaire, once as “leisurely biking” under moderate activities and once as “road or mountain biking, stationary biking, or spinning” under vigorous activities. Nearly a third of the respondents told us that although they were aware that they had already reported their walking or biking in a previous category, they would report it again if the category seemed to describe their situation. Occasionally, respondents expressed frustration or confusion with the repetition.

#### **Generalizing from examples of specific activities.**

On the original LACE PAQ, several items were grouped together into a general category and then described in more detail by a series of examples (e.g., “light yard work” was exemplified as planting, pruning, weeding, etc.). The intent was to provide a sense of the activities that were included in the categories but not to provide an exclusive, comprehensive list. However, almost half of the LACE respondents thought these lists were too long, and, occasionally, that the activities were exclusive rather than a series of examples. Sometimes, individuals tended not to notice all the listed items because the examples were too numerous. In contrast, some respondents wanted to report behaviors that were not specifically listed and were unsure whether, for example, sewing could be included with arts and crafts projects or plumbing included with carpentry. Respondents were confused by these omissions and unclear if these activities “counted.”

**Use of a reference group.** The CMH questionnaire included the following standard global question: “In comparison with other people your own age and sex, do you think your work for pay or as a volunteer is physically: much heavier, heavier, about the same as, lighter, much lighter?” Despite the documented ability of this question to rank individuals in terms of known health and demographic correlates of PA (29), some women and men compared themselves to people in general, regardless of gender, whereas others simply limited themselves to coworkers or people in the same profession/job category or workplace. Several respondents were simply baffled by the question and did not know how to answer. One man stated:

“I have no idea how to answer the last question. *No* idea (emphasis in original). Do you need me to put an answer there?”

Still another man compared himself to his wife and notably discussed the heaviness of his work in psychological terms:

A: I think of the number of hours I work, and I think of the daily grind. Like, maybe you don’t go and lift so much each day, but when I think of the daily grind of the work.... I mean, everybody has a hard job, and you feel funny when you think that your workload is much heavier. But I’m thinking of the grind of having to do it *every single day*, (emphasis in original) and occasionally get a day off. That’s why I said that.

Q: So you said much heavier. And were you thinking of the lifting?

A: No.

Q: You're thinking more the length of hours that you work?

A: Yes. But I think probably, if I was honest, everybody feels that their job is very difficult.

**Differentiation of walking, hiking, jogging, and running.** This was one area in which the respondents' definitions were largely, although not totally, congruent with the definitions used by researchers. Almost all respondents understood "brisk walking" to mean walking fast and often fast enough to increase heart rate and respiration and cause sweating or muscle fatigue, which constituted "exercise" for many respondents. For some people, walking at a pace of 3–4 mph was a meaningful statement in terms of defining brisk walking, but for many it was not. For one woman, brisk walking was walking at 2 mph. On the other hand, some regular walkers were very strict in their definitions; one woman defined brisk walking as a 15- to 16-min mile, and she was aware that she walks a 17-min mile. Leisurely walking, in contrast, often was defined comparatively as slower than brisk walking. It also was defined by some as not constituting exercise and taking place when there is no rush to get anywhere or not having any kind of goal in mind other than socializing and conversation.

Respondents generally were able to differentiate walking from hiking and jogging from running. Nearly all the respondents described walking as taking place on paved surfaces and hiking taking place on unpaved terrain. Hiking also typically was seen as more strenuous because it requires maneuvering around obstacles, more hilly terrain, and more careful footwork. Hiking was also often seen as taking more time than walking. Running and jogging were almost as distinct in respondents' minds as were walking and hiking, with almost all participants perceiving running to be faster than jogging. However, two individuals said that running and jogging are basically the same or that the terms could be used interchangeably.

**Questionnaire revisions as a result of cognitive interviews.** As a result of the feedback we received from respondents, we substantially revised many items on the LACE PAQ and in the CMH questions. Although data collection for both LACE and CMH were completed several years before the current study, the revised versions of the questionnaires are currently in use in the On the Move study and are available for use by other researchers. Both revised questionnaires may be viewed at [www.dor.kaiser.org/studies/otm/index.shtml](http://www.dor.kaiser.org/studies/otm/index.shtml). Tables 2 and 3 below summarize the changes for CMH and LACE, respectively.

## DISCUSSION

The cognitive interviews reported on in this study strongly suggest that some questions and wording frequently used in PA questionnaires may be understood by respondents in ways unintended by researchers. Respondents typically expressed difficulty with definitions of intensity, estimation

of work-related PA, differentiation of similar activities in different domains, understanding lists of activities as examples rather than definitive categories, and comparison of their own behavior to a reference group. The confusion and the misunderstanding expressed by the respondents, especially in such key areas as intensity, frequency, and duration, support the importance of using cognitive interviewing techniques in the design and/or revision of PAQs.

The one area in which respondents' did not experience difficulty was defining walking, hiking, and running. This may be because these activities are so commonly experienced that their meaning is widely shared and broadly accessible.

TABLE 2. Changes made to CMH questionnaire.

Domain	Original Wording	Changes
Intensity	Groups of activities defined by the word "intensity" and further defined as "moderate" and "vigorous" or "strenuous" and "nonstrenuous."	<ul style="list-style-type: none"> <li>activities no longer defined as either moderate or vigorous or strenuous or nonstrenuous</li> <li>word "intensity" replaced with "physical effort"</li> <li>for each activity, respondents are asked "how hard were you working?" with response categories of "not hard," "somewhat hard," and "hard"</li> <li>standard MET values may be adjusted accordingly.</li> </ul>
Estimation of work-related physical activity	Response categories for frequency of occupational activity (sit, stand, walk, lift heavy loads, use heavy equipment, sweat from exertion) were "never, seldom, sometimes, often, always"	<ul style="list-style-type: none"> <li>occupational activities collapsed into "sitting/standing, mostly walking, heavy labor"</li> <li>response categories for frequency changed to "less than 1 h·d<sup>-1</sup>, 1–2 h·d<sup>-1</sup>, more than 2 h·d<sup>-1</sup>"</li> </ul>
Duplicate counting of the same activity	Walking for exercise, transportation, or pleasure was listed twice, once as brisk walking and once as leisurely walking.	<ul style="list-style-type: none"> <li>collapsed into single item</li> <li>respondents indicated how hard the physical exertion was.</li> </ul>
	Cycling was listed under "moderate activities" as "leisurely biking," and under "vigorous activities" as "road or mountain biking, stationary biking, spinning."	<ul style="list-style-type: none"> <li>cycling listed only once as "road or mountain biking, stationary biking, spinning."</li> <li>respondents indicated how hard the physical exertion was.</li> </ul>
Reference groups	Original question was "In comparison with other people your own age and sex, do you think your work for pay or as a volunteer is physically: much heavier/heavier/about the same as/lighter/much lighter."	<ul style="list-style-type: none"> <li>eliminated question</li> </ul>

TABLE 3. Changes made to LACE questionnaire.

Domain	Original Wording	Changes
Intensity	Intensity described in terms of increases in heart rate and breathing	<ul style="list-style-type: none"> <li>• “Intensity” replaced by “physical effort”</li> <li>• Sweating added to heart rate and breathing to describe levels of physical effort</li> </ul>
Estimation of work-related PA	Response categories for frequency of occupational activity (sitting; standing; walking; lifting, carrying or pushing heavy items (20 lb or more); stooping or bending; doing heavy manual work) were “never or less than 1 h; 1–2 h; more than 2–4 h; more than 4–6 h; more than 6–8 h; more than 8 h”	<ul style="list-style-type: none"> <li>• Occupational activities condensed to “sitting/standing, mostly walking, heavy labor”</li> <li>• Responses categories for frequency condensed to “less than 1 h·d<sup>-1</sup>, 1–2 h·d<sup>-1</sup>, more than 2 h·d<sup>-1</sup>”</li> </ul>
Duplicate counting of the same activity	<p>Walking the dog listed under caregiving activities, walking for exercise/pleasure at a brisk pace and walking for exercise/pleasure at a leisurely pace listed under recreational activities, walking for transportation listed under transportation, and walking at work listed under occupational activity</p> <p>Bicycle riding (including stationary bikes), touring or racing, road/mountain bike listed under recreational activity, and riding a bicycle or using rollerblades for transportation listed under transportation*</p>	<ul style="list-style-type: none"> <li>• All walking/running activities consolidated into one domain labeled, “stairs and outside walking, hiking, and running” with the following seven specific items: “walking, slower than 20 min·mile<sup>-1</sup>, walking briskly (20 min·mile<sup>-1</sup> or faster), hiking (walking hilly or uneven terrain), backpacking, jogging, running, climbing up stairs”</li> <li>• Eliminated “stationary” from bicycle riding in recreational domain</li> <li>• Added “stationary bike” to a newly created “aerobic conditioning equipment” item that included treadmill, stationary bike, stair climber, rowing machine, elliptical trainer, cross-country skiing machine</li> </ul>
Generalizing from lists of examples	<p>“Light yard work” included the following examples: weeding, planting, cultivating a garden, pruning or trimming bushes or trees, vacuuming leaves, sweeping outside, raking or edging yard, watering yard or plants”</p> <p>“Heavy yard work” included the following examples: spading, digging, filling in garden, chopping wood, using a nonpower push lawnmower, laying bricks, shoveling snow</p> <p>Single item asked about “doing laundry, ironing”</p> <p>Single item asked about “preparing meals, baking, cleaning up from meals”</p> <p>Single item asked about “taking care of young children (aged 3–5 yr old)”</p> <p>Single item asked about “taking care of elderly or disabled people”</p> <p>Single item asked about “carpentry”</p> <p>“Attending concerts” included in item that asked about attendance at group activities</p>	<ul style="list-style-type: none"> <li>• Separated this item into three individual items: “weeding, planting, cultivating a garden”; “pruning or trimming shrubs and bushes”; and “vacuuming leaves, sweeping outside, raking leaves”</li> <li>• Separated this item into three individual items: “spading or digging in garden”; “chopping wood, laying brick, shoveling snow”; and “mowing lawn with a manual mower”</li> <li>• “Folding clothes” added to this item</li> <li>• “Washing and drying dishes” added to this item</li> <li>• “Bathing, feeding, holding, carrying, and playing” added to this item</li> <li>• Changed wording to “helping elderly or disabled people with personal care (bathing, feeding, dressing, transferring) or pushing a wheelchair”</li> <li>• “Plumbing, electrical work” added to this item</li> <li>• Separate item added that asked about “playing a musical instrument” after several respondents noted that we had asked about listening to, but not playing, music</li> </ul>

To improve the comprehension of various items in the other areas, we attempted to mimic this more naturalistic terminology. For instance, as summarized in Tables 2 and 3, we replaced the word “intensity” and the use of “moderate” and “vigorous” to describe intensity, with the term “physical effort” and, in the CMH questionnaire, with descriptors of effort as either “hard,” “somewhat hard,” or “not at all hard.” We also eliminated descriptors of walking and cycling as either “leisurely or brisk” and “moderate and strenuous” and simply asked about those activities in general, letting the respondent tell us, in addition to frequency and duration of participation, how hard the physical effort was for them (response categories for effort were “not at all hard,” “somewhat hard,” or “very hard”). Although this may introduce an additional source of error due to factors that affect perception of intensity, it avoids presenting respondents with the difficult cognitive task of determining whether the walking or cycling they do is “moderate” or “vigorous” or both and then figuring out how much time is spent in the same activity but at different intensity levels. It also allows researchers the flexibility of either using standard MET values for these activities or adjusting the standard MET value of a given activity (2) either up or down, depending on the participant’s reported intensity with which it was performed. In addition, we revised the wording of intensity

questions on the LACE questionnaire so that more vigorous intensity activities were described according to the cues more commonly used by respondents (i.e., sweating as well as increases in heart rate and breathing).

Problems with the terminology of intensity may arise because intensity can be considered in either relative or absolute terms. As others have discussed (32,36), relative intensity depends on several characteristics of the individual, such as age and fitness level, and an activity that feels hard for one individual may only be perceived of as a moderate activity by another. In contrast, the absolute intensity of an activity standardizes the energy cost of activity and may be more relevant in terms of physiological responses and health outcomes. Although measurement error occurs in the assessment of both relative and absolute intensity, the magnitude of that error may be minimized by using terms to describe intensity that are meaningful to a wide range of people.

We also made revisions to questions about work-related PA. Because most respondents seemed to overestimate the amount and frequency of PA at work, we simplified the response categories to “mostly sedentary (sitting/standing),” “somewhat active (mostly walking),” and “very active (heavy labor).” We also increased the specificity of frequency categories by providing a narrower and more specific set of time intervals (less than 1 h·d<sup>-1</sup>, 1–2 h·d<sup>-1</sup>, more

than  $2 \text{ h} \cdot \text{d}^{-1}$ ) rather than a more comprehensive range (as on the LACE PAQ) or the more general time-based adverbs (“never,” “seldom,” “sometimes,” “often,” and “always”) used in the CMH questions (Tables 2 and 3). Although limiting the response categories in this way prevents accurate estimation of work-related activity in terms of MET-hours per week (the original intent of the occupational questions on the LACE PAQ), it provides respondents with a more comprehensible question and allows for accurate ranking of individuals (the initial intent of the CMH occupational questions).

To eliminate the opportunity to count the same activity more than once, we consolidated questions, particularly those concerning walking and cycling, while still retaining the original response categories for frequency and duration. Finally, we decided to avoid the issue of asking respondents to generalize from a specified list of examples by expanding the number of activities about which we asked and listing them each separately. Although this approach undoubtedly fails to assess all of the activities any given individual may do, it requires less complex cognitive processes and may, therefore, improve the accuracy of the reporting for those activities that are specified (10). However, in a few instances, we actually expanded some categories to include more examples so that the category was better described.

The difficulty some respondents had comparing their own behavior to that of others was not easily remedied. Although global questions asking respondents to rate their level of PA relative to others have been shown to rank people reasonably well in terms of their actual behavior (12,34), evidence suggests that the frame of reference respondents use may be narrowly defined and may not adequately capture interindividual variability in PA level across differing reference groups, such as race/ethnic groups (29). Given this evidence and the difficulty respondents expressed answering this question in the CMH questionnaire, we simply decided to eliminate the question.

The focus of this study, respondents’ comprehension of two PA questionnaires, adds to the methodological literature on PA assessment. Although some researchers have considered this problem using a cognitive model (10), very few have actually explored the content of these issues (15,30,33). In general, methodological studies of self-reported PA have focused more on evaluating reliability (test/retest for the same participants over time) and/or validity (intermethod reliability) (3,5,11,13,23,25,39). Although this body of literature has demonstrated that PA questionnaires are generally repeatable and correlate reasonably well with other self-report measures, the generally

low correlations between self-report and more objective measures of PA (9,26) may be due, in part, to problems respondents have with comprehension and other cognitive processes related to answering PA questionnaires. Redesigning PA questionnaires in ways that minimize these problems may reduce measurement error in self-report and result in higher levels of agreement with more objective methods of assessment.

This study has several limitations that may affect the degree to which findings are generalizable. African Americans and whites were well represented in the sample but not individuals of other race/ethnicities, and most of the sample was relatively well educated. The sample was also restricted to midlife respondents from a small geographic area. In addition, only two specific PA questionnaires were evaluated, and neither of the revised questionnaires was re-evaluated with cognitive interviews, although the test–retest repeatability and the validity of both revised questionnaires against a PA diary and accelerometry are currently being examined in a follow-up study.

Despite these limitations, some of the lessons learned in this study may be relevant to other studies that rely on self-reported PA. Perhaps most important, our findings strongly suggest that the terms that PA researchers commonly use to describe intensity—light, moderate, and vigorous—do not translate well for the public at large. To improve the accuracy of PA questionnaires, researchers might ensure more meaningful responses if they ask about physical effort rather than intensity and avoid grouping activities by objective intensity level. Our findings also suggest that the attempt to improve recall by contextualizing PA in terms of domains may actually increase reporting error and result in overestimation due to double-counting.

Although some in the population may share backgrounds and frames of reference that are similar to those of researchers, many, if not most, typical respondents probably do not answer certain items on PA questionnaire in the ways intended and assumed by researchers. The findings of this study suggest that we could improve our knowledge base in PA and health by more carefully evaluating the design and wording of PA questionnaires. Additional research into respondents’ comprehension of PA questions would help to identify the best ways to redesign PA questionnaires to avoid the cognitive challenges revealed in this study.

The authors thank the women and men who participated in the study. This project was funded by the National Cancer Institute, grant number R01 CA103974. The results of the present study do not constitute endorsement by ACSM.

## REFERENCES

1. Ainsworth BE. Issues in the assessment of physical activity in women. *Res Q Exerc Sport*. 2000;71(2 suppl):S43–S46.
2. 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–S516.
3. Ainsworth BE, Jacobs DR Jr, Leon AS, Richardson MT,

- Montoye HJ. Assessment of the accuracy of physical activity questionnaire occupational data. *J Occup Med*. 1993;35:1017–27.
4. Ainsworth BE, Richardson M, Jacobs DR Jr, Leon AS. Gender differences in self-reported physical activity. *J Women Sport Activity*. 1993;23:1–16.
  5. Ainsworth BE, Sternfeld B, Richardson MT, Jackson K. Evaluation of the Kaiser physical activity survey in women. *Med Sci Sports Exerc*. 2000;32(7):1327–38.
  6. Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*. Oxford: Oxford University Press; 1994. p. 70–1.
  7. Baecke JAH, Burema J, Fritjers JER. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36:936–42.
  8. Charmaz K. Grounded theory. In: Smith J, Harre R, Langenhove V, editors. *Rethinking Methods Psychology*. London: Sage; 1995. p. 27–40.
  9. 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):519–25.
  10. Durante R, Ainsworth BE. The recall of physical activity: using a cognitive model of the question-answering process. *Med Sci Sports Exerc*. 1996;28(10):1282–91.
  11. Friedenreich CM, Courneya KS, Bryant HE. The lifetime total physical activity questionnaire: development and reliability. *Med Sci Sports Exerc*. 1998;30(2):266–74.
  12. Godin G, Shepard RJ. A simple method to assess exercise behavior in the community. *Can J Appl Sport Sci*. 1985;10(3):141–6.
  13. 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.
  14. Jacobs DR Jr, 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;9:448–59.
  15. Johnson TP, Cho YI, Holbrook AL, O'Rourke D, Warnecke RB, Chavez N. Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Ann Epidemiol*. 2006;16(9):661–8.
  16. Kannus P. Preventing osteoporosis, falls, and fractures among elderly people. *BMJ*. 1999;318:205–6.
  17. Kritiz-Silverstein D, Barrett-Connor E, Corbeau C. Cross-sectional and prospective study of exercise and depressed mood in the elderly: the Rancho Bernardo study. *Am J Epidemiol*. 2001;153(6):596–603.
  18. LaPorte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep*. 1985;100:131–46.
  19. Lee IM. Physical activity and cancer prevention—data from epidemiologic studies. *Med Sci Sports Exerc*. 2003;35(11):1823–7.
  20. Lofland J, Lofland L. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Wadsworth; 2006. 282 p.
  21. Manson JE, Rimm EB, Stampfer MJ, et al. Physical activity and incidence of non-insulin-dependent diabetes mellitus in women. *Lancet*. 1991;338:774–8.
  22. Masse 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.
  23. Matthews CE, Ainsworth BE, Hanby C, et al. Development and testing of a short physical activity recall questionnaire. *Med Sci Sports Exerc*. 2005;37(6):986–94.
  24. Powell KE, Thompson PD, Caspersen CJ, Kendrick JS. Physical activity and the incidence of coronary heart disease. *Annu Rev Public Health*. 1987;8:253–87.
  25. Richardson MT, Ainsworth BE, Wu H-C, Jacobs DR Jr, Leon AS. Ability of the atherosclerosis risk in communities (ARIC)/Baecke questionnaire to assess leisure-time physical activity. *Int J Epidemiol*. 1995;24:685–93.
  26. Richardson MT, Leon AS, Jacobs DR Jr, Ainsworth BE, Serfass R. Ability of the Caltrac accelerometer to assess daily physical activity levels. *J Cardiopulm Rehabil*. 1995;15:107–13.
  27. Sidney S, Jacobs DR Jr, Haskell WL, et al. Comparison of two methods of assessing physical activity in the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Am J Epidemiol*. 1991;133(12):1231–45.
  28. 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):1959–67.
  29. Sternfeld B, Cauley J, Harlow S, Liu G, Lee M. Assessment of physical activity with a single global question in a large, multi-ethnic sample of midlife women. *Am J Epidemiol*. 2000;152(7):678–87.
  30. Tudor-Locke C, Henderson KA, Wilcox S, Cooper RS, Durstine JL, Ainsworth BE. In their own voices: definitions and interpretations of physical activity. *Womens Health Issues*. 2003;13(5):194–9.
  31. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. *Sports Med*. 2001;31(2):91–100.
  32. US Department of Health and Human Services. *Physical Activity and Health: A Report of the Surgeon General*. Atlanta (GA): US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Control and Prevention; 1996. p. 32–33; 85–172.
  33. Warnecke RB, Johnson TP, Chavez N, et al. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol*. 1997;7(5):334–42.
  34. Weiss TW, Slater CH, Green LW, Kennedy VC, Albright DL, Wun CC. The validity of single-item, self-assessment questions as measures of adult physical activity. *J Clin Epidemiol*. 1990;43:1123–9.
  35. Wendel-Vos GC, Schuit AJ, Feskens EJ, et al. Physical activity and stroke. A meta-analysis of observational data. *Int J Epidemiol*. 2004;33(4):787–98.
  36. Wilcox S, Irwin ML, Addy C, et al. Agreement between participant-rated and compendium-coded intensity of daily activities in a tri-ethnic sample of women ages 40 years and older. *Ann Behav Med*. 2001;23(4):253–62.
  37. Willis G. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications; 2005. 335 p.
  38. Yaffe K, Barnes D, Nevitt M, Lui LY, Covinsky K. A prospective study of physical activity and cognitive decline in elderly women: women who walk. *Arch Intern Med*. 2001;161(14):1703–8.
  39. Yore MM, Ham SA, Ainsworth BE, et al. Reliability and validity of the instrument used in BRFSS to assess physical activity. *Med Sci Sports Exerc*. 2007;39(8):1267–74.