

## Recommendations to Improve the Accuracy of Estimates of Physical Activity Derived From Self Report

Barbara E. Ainsworth, Carl J. Caspersen, Charles E. Matthews, Louise C. Mâsse, Tom Baranowski, and Weimo Zhu

**Context:** Assessment of physical activity using self-report has the potential for measurement error that can lead to incorrect inferences about physical activity behaviors and bias study results. **Objective:** To provide recommendations to improve the accuracy of physical activity derived from self report. **Process:** We provide an overview of presentations and a compilation of perspectives shared by the authors of this paper and workgroup members. **Findings:** We identified a conceptual framework for reducing errors using physical activity self-report questionnaires. The framework identifies 6 steps to reduce error: 1) identifying the need to measure physical activity, 2) selecting an instrument, 3) collecting data, 4) analyzing data, 5) developing a summary score, and 6) interpreting data. Underlying the first 4 steps are behavioral parameters of type, intensity, frequency, and duration of physical activities performed, activity domains, and the location where activities are performed. We identified ways to reduce measurement error at each step and made recommendations for practitioners, researchers, and organizational units to reduce error in questionnaire assessment of physical activity. **Conclusions:** Self-report measures of physical activity have a prominent role in research and practice settings. Measurement error may be reduced by applying the framework discussed in this paper.

**Keywords:** survey, questionnaire, measurement error

All instruments, including self-report questionnaires, have some level of measurement error that, if unchecked, can lead to uncertainty of measures and flawed inference. Measurement error can lead to biased estimates of a behavior and may attenuate or overestimate relationships between a physical activity score with other variables. If severe enough, the attenuation may limit the ability of a study to detect actual relationships between the variables. Thus, it is important to understand the types and sources of measurement error in self-report instruments and to identify ways to reduce and/or correct for measure error.

---

Ainsworth is with the Program in Exercise and Wellness, School of Nutrition and Health Promotion, Arizona State University, Phoenix, AZ. Caspersen is with the Epidemiology and Statistics Branch, Division of Diabetes Translation, Centers for Disease Control and Prevention, Atlanta, GA. Matthews is with the Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD. Mâsse is with the Dept of Pediatrics, School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada. Baranowski is with the Children's Nutrition Research Center, Dept of Pediatrics, Baylor College of Medicine, Houston, TX. Zhu is with the Dept of Kinesiology and Community Health, College of Applied Health Sciences, University of Illinois at Urbana-Champaign.

Two sources of error are common in a self-report instrument: random error and systematic error.<sup>1</sup> Random error exists in all questionnaires. They are generally attributable to unreliable reporting and natural variation in behaviors over time. While random error generally does not affect the mean value for a group, it is unpredictable in how it can affect the variation in scores. Investigators try to reduce random error by standardizing the testing conditions, by enhancing study design (ie, incorporating replicate measures), and by assuring a sufficient number of participants are included in a study to minimize the variability of scores around the mean. Systematic error, on the other hand, is caused by factors that uniformly change the mean value of a score across a study sample. Systematic error may be caused by reporting biases or a mismatch between the scope of a questionnaire relative to a reference measure chosen to estimate the true score. Systematic errors may have equally damaging effects on the study as random errors and can cause the measured mean scores to appear higher or lower than the true scores. Investigators try to reduce systematic error by assuring instruments are valid and reliable in the study population measured.

To improve instrument accuracy in this context, it is clear that one should identify and combat the underlying sources of measurement error and/or correct for it when it cannot be averted. This effort can be accomplished across many steps in the physical activity assessment process

and will involve a variety of individuals performing different roles across the steps.

### Objective

This paper provides an overview of presentations and workgroup discussions that address ways to reduce sources of error in self-report physical activity instruments. The charge of the workgroup was to provide an overview of the characteristics of self-report instruments affecting the accuracy of recall which can result in under- or over-reporting of behavior, and to propose essential ways to reduce bias compromising the accuracy of physical activity derived from self report questionnaires.

### Process

This paper provides a compilation of perspectives shared by the authors of this paper and of comments expressed by members who attended the workgroup (identified in the Appendix).

### Findings

The workgroup discussions summarized in this paper focused on 1) recommendations for researchers, practitioners, and organizations to minimize errors with self-report questionnaires; 2) identifying information that questionnaire users should know and/or should ask to reduce measurement error; and 3) providing suggestions to enhance the use of questionnaires to assess physical activity behavior in research and practice settings.

A conceptual framework for reducing errors with self-report assessment of physical activity is presented to summarize the workgroup discussions (Figure 1). The framework identifies 6 steps of the physical activity assessment process among which efforts might be made to reduce measurement errors with self-report. The steps are identifying the need to measure physical activity, selecting an instrument, collecting data, analyzing data, developing a summary score, and interpreting data. Underlying the first 4 steps are behavioral parameters that comprise the primary aspects to be assessed in physical activity questionnaires. The first 4 parameters pertain to

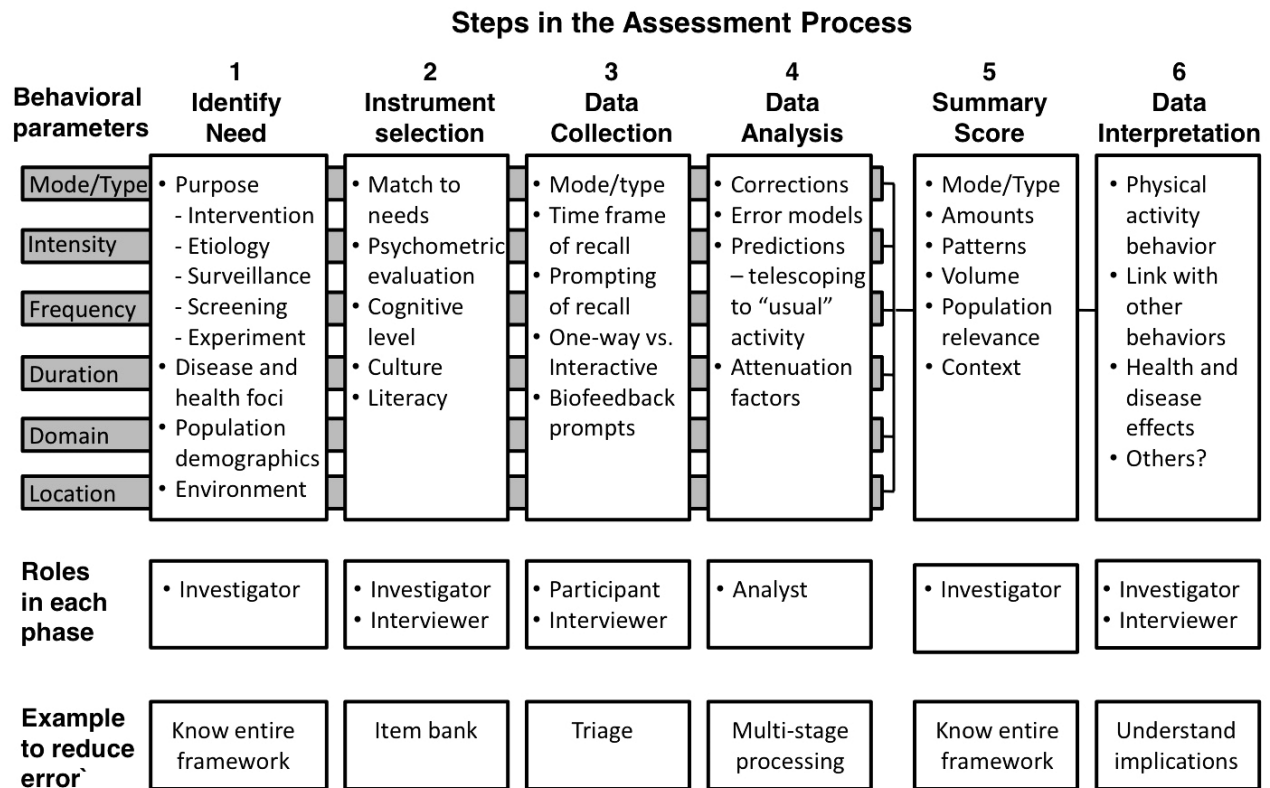


Figure 1 — Conceptual framework for reducing errors across 6 steps of self-report physical activity assessment.

physical activity: type, intensity, frequency, and duration, followed by the domain and location of activity participation. A research team and/or investigator must have a thorough understanding of all steps in the framework to minimize errors related to self-reported physical activity assessment. As the workgroup endorsed the overarching focus of the framework, the findings are presented across the 6 steps of assessment process.

### Step 1: Identifying Need

The goal of this step is to reduce investigator error by having a clear understanding of the purpose of the questionnaire application, the setting where the behaviors occur, and knowing which behaviors should be assessed. Researchers should have a clear idea of the physical activity behavioral focus of their study, whether they intend to relate the assessment with a particular outcome(s), the target population for assessment, and the characteristics of the physical activity environment in which the behavior can be measured.<sup>2</sup>

The type of study performed and the study aims should dictate the questionnaire to be employed. Surveillance systems aim to identify what proportion of a population is meeting a set of parameters, most often national recommendations for health-enhancing physical activity. Intervention studies have a goal to determine the effectiveness of a particular strategy in increasing physical activity. Studies may seek to understand how physical activity behaviors modify physiological mechanisms involved in the disease process (experiment), or in altering the risks for developing incident diseases or adverse health outcomes, such as diabetes or disabilities (etiology). Screening is used to triage persons in the clinical setting, for eligibility for research studies, or to identify persons who need specific interventions to modify their physical activity behaviors. Specific recommendations for identifying the need to assess physical activity behaviors were (a) matching the questionnaire format with a viable and valid interpretation of the resulting summary score of that questionnaire, and (b) being clear about the types of behaviors measured and as they are expected to relate to the outcome measures of interest. Appropriate questions to ask are

- *Is the questionnaire needed for surveillance, interventions, or to uncover associations between physical activity and disease?*
- *Who is the population to be measured?*
- *Are there characteristics of the environment that need to be considered, such as weather or location?*

### Step 2: Instrument Selection

The goal of this step is to reduce errors made by the investigator and interviewers by assuring the questionnaire is appropriate for the abilities and interests of the target population. Most importantly, is there evidence that the questionnaire is a valid assessment instrument

for the population of interest? In selecting an instrument, researchers should consider factors that may influence the ability of the respondents to answer the questions and the relevance of the questionnaire to the study question.

Before use, every physical activity questionnaire should be evaluated for its psychometric properties. This includes selecting a questionnaire that supports the interpretation sought by the study (ie, making proper inferences in light of the validity of the questionnaire). This evaluation includes an assessment of existing validity evidence to determine the constructs of physical activity that the questionnaire captures (construct validity evidence), how the questionnaire information converges with other sources of data shown to reflect different types of physical activity (convergence validity evidence), how well the types of activities included in a questionnaire reflect the types of physical activity of interest (content validity evidence),<sup>3</sup> and if a questionnaire is able to capture changes in physical activity behaviors over time, also referred to as sensitivity to change.<sup>4,5</sup> The reliability evidence, defined as the consistency of a measurement from one time to another, also must be known before selecting a physical activity questionnaire as it also influences accuracy of a score.<sup>6</sup> The questionnaire or instrument reliability should not be confused with the stability of respondents' physical activity behaviors.

A questionnaire should be appropriate for the survey population and specific for the behaviors assessed. Workgroup members proposed 2 ways to identify appropriate questionnaires to assess physical activity behaviors identified for a study purpose: 1) item banks and 2) questionnaire clearinghouses. An item bank is a particularly attractive idea as researchers could select from a collection of questions organized by the content of each item as well as the measurement characteristics of the item, such as validity, reliability, recall period, cognitive difficulty, and sharing the same scale. An item bank may be useful for computerized administration of self-report instruments inserted into a decision tree format based on a respondent's reported physical activity behaviors. If the questions in an item bank are independently validated, a bank of items may be created for use in specific settings or to answer specific questions, such as a desire to identify the types of behaviors associated with increased obesity risks. An example of an item bank for chronic disease is the PROMIS network designed by the National Institutes of Health for research about patient-reported health status involving physical, mental, and social well-being.<sup>7</sup> In contrast, questionnaire clearinghouses may provide validated and reliable self-report instruments tailored for specific populations and settings. Clearinghouses may serve to prevent the unneeded development of new questionnaires to assess physical activity and encourage researchers to evaluate the psychometrics of existing questionnaires for any study purpose. The Physical Activity Resource Center for Public Health provides a clearinghouse for physical activity instruments and intervention.<sup>8</sup> A major limitation of a clearinghouse of existing questionnaires is that comparing scores across different questionnaires

is difficult unless score equivalencies are established, an issue that has yet to be solved.

Questionnaires are used in a variety of study settings. Questionnaires used for screening should be short, focused on the types of behaviors to be assessed, and correctly and efficiently classify persons at varying levels of activity, or, perhaps the time they spent in specific types of physical activities. The workgroup members recommended that surveillance questionnaires should reflect the types of behaviors to be tracked, include information that maximized understanding of the questionnaire, and if the behaviors can be tracked using short, easily recalled questionnaires, ask short questions and focus on activities that are easy to recall (ie, vigorous intensity and structured activities), allow respondents to report individual activities performed as opposed to grouping types or intensities of activities together into a single question, and to use words that best capture behaviors of interest. The questionnaires also should have introductions that maximize an understanding of a questionnaire's objectives and also captivates the respondent's interest and commitment to answer the entire questionnaire.

Questionnaires used in intervention studies need to be able to characterize behaviors that differentiate persons with and without a particular health outcome. They also should be sensitive to changes in physical activity behaviors, although few questionnaires are truly able to do so.<sup>9,10</sup> The workgroup members noted that more research is needed to identify questionnaires that are sensitive to change in behaviors. As it currently stands, many questionnaires are used for purposes that differ from what they were developed for and without validity evidence to support the inference made with the instrument. For example, the International Physical Activity Questionnaire (IPAQ) was developed for use in surveillance settings, but has been used to assess changes in physical activity behaviors without tests for its sensitivity to change.<sup>11</sup>

Questionnaires should be culturally relevant and address the cognitive abilities of the respondent population (eg, children). For persons with lower reading and numeracy levels, questions should avoid wording that combines multiple activities and/or behaviors into a single question that ask respondents to average time spent in multiple activities or across many days. Questionnaire items should relate to the types of physical activities that respondents' value and perform. The lead investigator and interviewers should be familiar with the types of questionnaires created that will match their study needs with the respondent's abilities to complete the questionnaire as designed. Because there is no questionnaire design that is ideal for all study populations, nor will there ever be an ideal questionnaire, use of a questionnaire/item bank of valid and reliable questionnaires/questions may be useful to reduce measurement error.

Specific recommendations for selecting a questionnaire were to 1) match the questionnaire with the study purpose (eg, ensure the instrument can validly and specifically correspond to the outcome(s) under investigation),

2) develop clearinghouses of valid and reliable items/questionnaires for use in various settings, and 3) assess and adapt questionnaires as needed for use in different cultures and languages using standard translation methods.<sup>12</sup> Thus, appropriate questions to ask are

- *Is the questionnaire a good match for the participants relative to the types of activities they perform and find relevant?*
- *Is the questionnaire interviewer or self-administered? (A questionnaire designed for interviewer administration may not have the same validity and reliability properties when it is self-administered)*
- *Are resources needed to use an interviewer-administered questionnaire?*
- *Has the questionnaire been subjected to a thorough evaluation of the psychometrics regarding the validity and reliability evidences in populations similar to the target population?*
- *Does the validity data support the interpretation that is sought by the study (ie, making proper inferences based on the validity of the questionnaire)?*
- *What are the cognitive demands of the questionnaire relative to the recall of information?*
- *Does the literacy level of the participants match the literacy and cognitive demands of the questionnaire?*
- *Are the questions to be asked appropriate for the cultural experiences of the target population?*
- *Has the validity of the questionnaire changed following translation to a different language?*

### Step 3: Data Collection

The goal of this step is to reduce reporting errors made by the participants and to assure that interviewers, if there are any, are effective and consistent in their questionnaire delivery methods. To assure fidelity at this step, researchers should identify elements of the data collection process that may increase measurement error. These include using the wrong time frame for the recall of physical activity, using recall prompts with jargon that respondents do not understand, providing inappropriate examples, and using incorrect modes of administration (eg, using self-report methods when questionnaire instructions require interviewer-administered methods). Interviewers need to be trained and tested, and the consistency of their interviewing monitored.

These elements have the potential to increase inter- and intraparticipant error that result in wide variability of summary scores within the group, and inconsistency of responses within an individual over repeated measures. The workshop participants noted that reducing error in the data collection step may be extremely difficult. In fact, some have argued that it may not be possible to reduce the error substantially and that self-reports are often regarded as being better suited for ranking individuals.<sup>13</sup>

Although, workshop participants noted that measurement error associated with data collection is not an

issue that is unique to the field of physical activity, they recommended that it was important to address some of the data collection issues to help reduce errors and improve the accuracy of self-report in estimating true behavior. One way questionnaire developers have tried to reduce the variability in data collection is to have trained interviewers administer the questionnaire using a standardized format, while clarifying for the respondent terminology or a questionnaire item's intent when misunderstanding arises with a question. For example, an interviewer may be able to differentiate between light, moderate, and vigorous intensity physical activity if a respondent appears to be confused by the definitions provided in the questionnaire. Standardized wording and use of prompts may help respondents understand the intention of the questions. A prompt may be an example of an intensity of activity or an example of types of activities surveyed.

Many self-report questionnaires are used to identify if people meet national guidelines for physical activity<sup>14</sup> and Healthy People objectives<sup>15</sup> to increase moderate- and vigorous-intensity physical activity. Most of the validation data for questionnaires has been accumulated through correlating the questionnaire with a standard, such as an accelerometer or doubly labeled water. This only serves to indicate that the self-report provides a valid ranking; not that the questionnaire is able to assess true behaviors.<sup>16</sup> The need to assess meeting a guideline increases the difficulty in recalling the frequency and duration of physical activity behaviors when the respondent is asked to consider simultaneously all activities that s/he performs. For example, when the goal is to assess if respondents are meeting physical activity recommendations/objectives that require a daily or weekly goal for a given activity intensity (ie, 5×30 min/day or 150 min/week of moderate-intensity activities), they are challenged with the need to recall the frequency and duration of many behaviors at such selected intensities.<sup>17</sup> As a potential solution, scoring algorithms can create an index to approximate the objective or guideline from a larger battery of questions, although the goal of using only 1 question for assessment will be lost.

Also, social desirability among respondents to be viewed as "physically active" may create over-reporting of physical activity behaviors, producing a potential "intensity bias" wherein respondents define intensities differently than what the investigators plan. Another form of intensity bias may occur when a questionnaire asks about intensities in absolute terms (ie, using the same MET value for each activity) while respondents reply to intensities relative to their personal characteristics, such as sex, age, physiological capacities, and movement experiences. This bias provides an incomplete ascertainment of physical activity behaviors and makes it very difficult to compare results between accelerometry and self-report methods of physical activity.<sup>18</sup>

Another concern in collecting data are how long is the questionnaire time frame in terms of past day, week, month, year, or a lifetime. Knowing that recall

error increases with the duration of recall time and that prospective studies are time consuming and carry an extensive participant burden for multiple recalls, it is useful to know what is the shortest recall period to characterize usual physical activity behaviors analogous to studies that identify minimum wear times for pedometers<sup>19</sup> and accelerometers.<sup>20</sup> Such an approach may be useful to answer the question of how far into the past a respondent can recall his or her physical activity and the number of assessments needed in a prospective study to obtain accurate and reliable responses to assess a stable physical activity exposure. The optimal methods to make these estimates are evolving.<sup>21</sup> By knowing the minimum number of days needed to obtain a reliable estimate of physical activity, accuracy may be improved and the respondent recall burden lessened. This often requires one to monitor physical activity with records, logs, or previous day recalls. While some types of physical activities are routine and performed daily, such as self care activities, work, and selected household activities, others are intermittent and may depend on the day of the week, month, or season during a year.<sup>22</sup> In addition, the frequency, duration, intensity, and types of activities performed will likely vary by age, sex, culture, and geographic residence, making generalizations difficult to identify the minimum recall period to reflect usual physical activity behaviors of specific segments of a population.

Compounding the potential for measurement error is the challenge of recalling hours, minutes, and types of physical activity behaviors that are performed intermittently. Some workshop participants thought it may be better to just identify questionnaire items that provide an index of "moving about" as was done in the Yale Physical Activity Survey.<sup>23</sup> Such an index may be more accurate in classifying physical activity levels as inactive, low active, active, highly active than in computing summary scores that rely on summing estimates of minutes at varying intensities.<sup>27</sup> The Stanford Brief Activity Survey (SBAS)<sup>24-26</sup> uses a similar approach to classify patterns of activity at work and in leisure settings. However, since this method does not assess the minutes of physical activity performed, one cannot assess national activity goals and intervention aims based on a minutes/week goal.

A specific recommendation for data collection was to conduct more qualitative research and cognitive interviews<sup>28</sup> in diverse sex, age, race/ethnicity segments of the population to understand how respondents understand the physical activity concepts measured. Appropriate questions to ask are

- *What types of physical activities does the questionnaire contain?*
- *What are the recall time frames?*
- *Will the recall have prompts to encourage recall of general and specific activities?*
- *Is the questionnaire format interactive in providing feedback to the participant about their response or one-way without feedback?*

## Step 4: Data Analysis

The goal of this step is to reduce error in data analysis by using correct scoring procedures and analytic methods. Elements of this step include applying statistical methods to identify and correct for measurement errors that occur in completing physical activity questionnaires, which may require the use of measurement error correction models and estimating attenuation factors, using information derived from the models.<sup>29</sup>

The goal in the data analysis phase is to identify, reduce and control errors using some quantitative and qualitative methods as appropriate. A common method to reduce measurement error in dietary analyses is by using analytic error correction techniques. This method is originally based on the premise that nutrient intake is measured with error when using recall instruments, such as a 24-hour recall or a food frequency questionnaire. By conducting a validation study using criterion or reference measures (eg, doubly-labeled water, dietary records) over the same period as the recall questionnaire, it is possible to use regression models to determine the sources of the recall error for selected nutrients and use this information to estimate attenuation factors that allow for the correction of these errors in assessing the relation between physical activity and the outcome of interest.<sup>30</sup> In assessing diet or physical activity behaviors, when one uses a questionnaire to recall a nutrient or physical activity level, the impact of that exposure on a disease risk is generally believed to be attenuated by the recall error, although under certain circumstances the reverse could be true (ie, errors could induce an association). Measurement error methods offer an approach to estimate in a quantitative way the magnitude of the various sources of error that may influence results of physical activity and health studies. This approach has been applied by Ferrari and colleagues<sup>31</sup> and described by Nusser<sup>29</sup> to quantify specific sources of measurement error and to estimate attenuation factors associated with measurement error in self-reports of physical activity. More work in this area is needed to further develop this methodology for physical activity studies.

If physical activity questionnaires are to be comparable between studies, they must use the correct scoring protocols. There are many physical activity questionnaires that can be accessed on the web, in books, and in publications. However, rarely are the scoring methods identified in sufficient detail to provide the user guidance in how to score the questionnaire as was done with the Collection of Physical Activity Questionnaires.<sup>32</sup> Having a clearinghouse for physical activity questionnaires modeled after the National Human Genome Research Institute's consensus measures for Phenotypes and Exposures (PhenX; [www.phenX.org](http://www.phenX.org)), the National Cancer Institute's (NCI) Grid Enabled Measures Database (GEM; [www.gem-beta.org](http://www.gem-beta.org)), and the National Collaborative on Childhood Obesity Research (NCCOR; <http://tools.nccor.org/measures>) would be useful. These websites

provide detailed protocols, including scoring methods, for many measures used in research settings. The NCI also provides a list of 102 physical activity questionnaires and their related references, and details of 72 physical activity questionnaire validation studies on their website at <http://appliedresearch.cancer.gov/tools/paq>. An optimal website would include a list of questionnaires with downloadable materials to identify the questionnaire's intended purpose, literacy requirements, detailed scoring protocols with available scoring syntax, recommended strategies to identify bias and correction methods for measurement error, how to best analyze the questionnaire data for different purposes, a list of references of the validity and reliability studies and also publications using the questionnaire. In addition, having available existing SAS or SPSS codes needed to analyze physical activity questionnaire data, including application of the error correction models, would help to provide standardized questionnaire scores. A model for this resource is the SAS code provided by the National Cancer Institute used to score ActiGraph accelerometer data found in the National Health and Nutrition Examination Survey ([http://riskfactor.cancer.gov/tools/nhanes\\_pam](http://riskfactor.cancer.gov/tools/nhanes_pam)).

Specific recommendations to reduce data analysis error were to: a) put more effort into developing standardized statistical methodologies to adjust for reporting errors in physical activity questionnaires, and b) provide a resource website to provide access to and educate the correct ways to score physical activity questionnaires, identify and correct for measurement error, and provide guidance in preparing data for subsequent reporting of results. Appropriate questions to ask to reduce data analysis error are *How do I identify error in my data?* and *How do I select and apply error correction models?*

## Step 5: Summary Scores

A goal of this step is to select the correct summary score for the study aim and to score the questionnaire correctly which is critical to the success of a study. The summary score can be presented in various ways to include minutes or hours per day or week; MET-minutes or MET-hours per day or week; kcal; kcal per kg per day; points; and so forth. A questionnaire might be used only for classification purposes as employed with the Behavioral Risk Factor Surveillance System<sup>33</sup> or the SBAS,<sup>24</sup> especially to assess progress in meeting physical activity guidelines.<sup>34,35</sup> The summary score also can add to measurement error. If minutes are the raw units reported, transforming the units into MET-minutes (by multiplying the minutes reported by a MET value) may potentially increase the measurement error if the MET value applied to estimated intensity differs from a MET value measured using indirect calorimetry.

To maintain the validity of a questionnaire, investigators must use the scoring methods developed for the questionnaire, ideally using a standardized physical activity intensity value. The Compendium of Physical

Activities was developed to reduce the variability in assigning MET values between questionnaires.<sup>36,37</sup> Three versions of the Compendium have been published with the 2011 version using an evidence-based approach to report measured MET values from published sources<sup>38</sup> (<https://sites.google.com/site/compendiumofphysicalactivities>). Also, error in estimation of time spent in different intensities can be reduced by using common MET cut-off points for absolute intensity ranges. For adults, these include sedentary (<1.5 METs), light intensity (1.6–2.9 METs), moderate intensity (3.0–5.9 METs), and vigorous intensity ( $\geq 6.0$  METs), recognizing that such a scale uses a healthy, younger adult as the reference standard.<sup>14,39</sup> Such practices allow for direct comparison of the questionnaire metrics between studies when absolute intensity is the desired frame of reference. On occasion the Compendium provides multiple MET values for the same activity. Investigators need to generate a rationale for using the one they select.

Specific recommendations to reduce error in selecting summary scores were to a) use the scoring methods developed for the questionnaire, b) use the same MET value for the same activities, and c) use common age-appropriate cut-off points to assess relative intensity levels. An appropriate question to ask is *What summary score do I need to address my study aims?*

## Step 6: Data Interpretation

The goal of this step is to reduce error by interpreting the data correctly. The investigator selecting a questionnaire must have a clear understanding of how the summary scores will be applied to the outcome of interest and, in collaboration with the data analyst, must understand the implications for success of the study aims in using one analytic method over another. Since physical activity data are used in many settings, including linking with other behavioral data, such as dietary intake; assessing health status; identifying dose response relationships; determining the prevalence of meeting recommendations; and identifying behavior changes, investigators and data analysts also must understand how the analytic methods were selected and how uncontrolled measurement error, confounding and effect modification can influence the interpretation of study findings.

Specific recommendations to avoid error in interpreting the data were a) to have a clear idea how the questionnaire data will be used to answer the study aims, and b) to know how the study framework can assist in making correct interpretations of study outcomes. Appropriate questions to ask are *Am I interpreting the physical activity scores correctly?* and *Is there unaccounted error that may influence the interpretation of the results?*

The value and utility of self-report measures is currently misunderstood because many researchers have employed self-report incorrectly (eg, employing a measure in an intervention context when the measure has not been demonstrated to be valid to detect change in behavior). The workgroup participants recommended

that self-report measures have their place in the battery of assessments of PA but researchers need to carefully select the instrument that meets the purpose of their study and one that has validity evidence supporting the interpretation that they seek.

## Summary Recommendations

The workgroup discussion was summarized into recommendations that address the needs of practitioners, researchers, and organizational units to reduce error in questionnaire assessment of physical activity behaviors.

### Practitioners

Practitioners need to identify appropriate types of questionnaires for use in various settings such as surveillance systems, clinical settings, and identifying correlates of physical activity behaviors. Summary recommendations were

- Match the purpose of the assessment with the type of questionnaire
- Use the questionnaire as intended
- Use an appropriate questionnaire for the study population, when possible
- Use the correct scoring and analytic methods
- Avoid using questionnaires that aggregate activity domains of more complex physical activity behaviors that are difficult to recall.

### Researchers

Through innovation, researchers can increase knowledge about ways to minimize bias in data collection and data analysis of physical activity questionnaires. Summary recommendations were

- Understand how to apply error correction models
- Identify sources of variability in self-report questionnaires that lead to systematic (eg, reporting bias) and random (eg, unreliable reporting, true intraindividual variability in behavior) errors
- Identify ways to reduce the error structure in measuring physical activity
- Determine how many days are needed to estimate usual behavior and the domains of physical activity using questionnaires
- Identify how and which questionnaire(s) is (are) able to assess true behaviors
- Identify how and which questionnaires are sensitive to change for use in intervention studies and are appropriate to a subgroup in the population.

### Organizational Units

Organizations have the resources and capacity to help researchers and practitioners select appropriate

questionnaires to assess physical activity and apply recommendations to the development, delivery, and analysis of physical activity questionnaire. Summary recommendations were

- Provide a clearing house of questionnaires, educational information, and analytic strategies to help users select the appropriate questionnaire for their intended purpose, deliver and score the questionnaire as intended, identify and correct for errors in the data, use the correct analytic methods, and report the data correctly
- Conduct workshops in how to analyze and interpret physical activity data
- Develop and provide open-platform software to upload and analyze data using standardized methods
- Offer ways to minimize and correct for measurement error in questionnaire assessment of physical activity behaviors
- Develop a web-based certificate program in physical activity assessment for data collectors.

## Conclusions

There are many sources of error in the self report of physical activity behaviors. The purpose of this paper was to describe ways to reduce self-report error. We provided a 6-step process which researchers and practitioners can use as a guide to reduce error in self-report physical activity assessment. As well, we provided summary recommendations relevant to practitioners, researchers, and organizations to advance self-report of physical activity. This report is the first step in developing best practices for using self-report measures for physical activity.

## Acknowledgments

We acknowledge the members of the workgroup for their input to the ideas presented in this paper: Barbara Ainsworth, Tom Baranowski, Wendy Brown, Matthew Buman, Carl Caspersen, Bronwyn Clark, Christine Friedenreich, Lauren Grieco, Stephen Herrmann, NiCole Keith, Bill Kohl, Susan Krebs-Smith, Brigid Lynch, Louise C. Mâsse, Charles Matthews, James McClain, James Morrow, Sarah Nusser, Jill Reedy, John Sirard, Donna Spruijt-Metz, John Staudenmayer, Jeff Whitfield, Gordon Willis, and Weimo Zhu. We thank Jane Senior from the American College of Sports Medicine and Heather Bowles from the National Cancer Institute for their tireless efforts to host this workshop. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

## References

1. Trochim W. Measurement Error. *Research Methods Knowledge Base* 2006; 2:<http://www.socialresearchmethods.net/kb/>. Accessed April 4, 2011.
2. LaPorte R, Montoye H, Caspersen C. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep.* 1985;100(2):131–146.
3. Trochim W. *Measurement validity types*. Research Methods Knowledge Base 2006; <http://www.socialresearchmethods.net/kb/measval.php>. Accessed April 3, 2011.
4. Stratford P, Binkley J, Riddle D. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* 1996;76(10):1109–1123.
5. Fritz J. Sensitivity to change. *Phys Ther.* 1999;79(4):420–422.
6. Trochim W. *Types of reliability*. Research Methods Knowledge Base 2006; <http://www.socialresearchmethods.net/kb/relytypes.php>. Accessed April 3, 2011.
7. Domain Frameworks-Physical Health AIR. *Patient Reported Outcomes Measurement Information System (PROMIS)* 2009; <http://www.nihpromis.org/measures/instrumentoverview>. Accessed April 3, 2011.
8. Kriska A. Physical Activity Resource Center for Public Health. 2011; <http://www.parcph.org/>. Accessed April 3, 2011.
9. Stewart A, Mills K, King A, Haskell W, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Med Sci Sports Exerc.* 2001;33(7):1126–1141.
10. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW, Blair SN. Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: a randomized trial. *JAMA.* 1999;281(4):327–334.
11. Bauman A, Ainsworth B, Bull F, et al. Progress and pitfalls in the use of the International Physical Activity Questionnaire (IPAQ) for adult physical activity surveillance. *J Phys Act Health.* 2009;6(suppl 1):S5–S8.
12. Arredondo EM, Mendelson T, Holub C, Espinoza N, Marshall S. Cultural adaptation of physical activity self-report instruments. *J Phys Act Health.* 2012;9(Suppl 1):S37–S43.
13. Sallis JF, Saelens B. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport.* 2000;71(2 suppl):S1–S14.
14. United States Department of Health and Human Services. 2008 Physical Activity Guidelines for Americans. 2008; <http://www.health.gov/paguidelines/default/asp>. Accessed May 23, 2011.
15. United States Department of Health and Human Services. Healthy People 2020. 2010; <http://www.healthypeople.gov/hp2020>. Accessed May 23, 2011.
16. Shephard R. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med.* 2003;37(3):197–206.
17. Paterson D, Jones G, Rice C. Ageing and physical activity: evidence to develop exercise recommendations for older adults. *Appl Physiol Nutr Metab.* 2007;32(S2E):S69–S108.
18. Troiano R, Berrigan D, Dodd K, Mâsse L, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40:181–188.
19. Kang M, Bassett DJ, Barreira T, et al. How many days are enough? A study of 365 days of pedometer monitoring. *Res Q Exerc Sport.* 2009;80(3):445–453.
20. Matthews C, Ainsworth B, Thompson R, Bassett DJ. Sources of variance in daily physical activity levels as measured by an accelerometer. *Med Sci Sports Exerc.* 2002;34(8):1376–1381.
21. Baranowski T, Mâsse L, Ragan B. How many days was that? We're still not sure, but we're asking the question



- better! *Med Sci Sports Exerc.* 2008;40(7, suppl):S544–S549.
22. Eason K, Mâsse L, Kelder S, Tortolero S. Dairy days needed to estimate activity among older African-American and Hispanic women. *Med Sci Sports Exerc.* 2002;34(8):1308–1315.
  23. DiPietro L, Caspersen C, Ostfeld A, Nadel E. A survey for assessing physical activity among older adults. *Med Sci Sports Exerc.* 1993;25(5):628–642.
  24. Taylor-Piliae R, Norton L, Haskell W, et al. Validation of a new brief physical activity survey among men and women aged 60–69 years. *Am J Epidemiol.* 2006;164(6):598–606.
  25. Taylor-Piliae R, Haskell W, Iribarren C, et al. Clinical utility of the Stanford Brief Activity Survey in men and women with early-onset coronary artery disease. *J Cardiopulm Rehabil Prev.* 2007;27(4):227–232.
  26. Taylor-Piliae R, Fair J, Haskell W, et al. Validation of the Stanford Brief Activity Survey: examining psychological factors and physical activity levels in older adults. *J Phys Act Health.* 2010;7(1):87–94.
  27. Caspersen C, Merritt R, Stephens T. International physical activity patterns: A methodological perspective. In: Dishman R, ed. *Advances in exercise adherence.* Champaign, IL: Human Kinetics Publishers; 1994:71–108.
  28. Willis GB. *Cognitive interviewing: a tool for improving questionnaire design.* 1st ed. Thousand Oaks, CA: Sage Publications; 2005.
  29. Nusser SM, Beyler NK, Welk GJ, et al. Modeling errors in physical activity recall data. *J Phys Act Health.* 2012;9(Suppl 1):S56–S67.
  30. Rosner B, Gore R. Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am J Epidemiol.* 2001;154(9):827–835.
  31. Ferrari P, Friedenreich C, Matthews C. The role of measurement error in estimating levels of physical activity. *Am J Epidemiol.* 2007;166(7):832–840.
  32. Kriska A, Caspersen C. Introduction to a collection of physical activity questionnaires. *Med Sci Sports Exerc.* 1997;29:5–9.
  33. Macera C, Ham S, Yore M, et al. Prevalence of physical activity in the United States: Behavioral Risk Factor Surveillance System, 2001. *Prev Chronic Dis.* 2005;2(2).
  34. Kruger J, Ham A, Kohl HI. Trends in leisure-time physical inactivity by age, sex, and race/ethnicity—United States, 1994–2004. *MMWR Morb Mortal Wkly Rep.* 2005;54(39):991–994.
  35. Carlson S, Fulton J, Galuska D, Lobelo F, Loustalot F. Prevalence of self-reported physically active adults—United States, 2007. *MMWR Morb Mortal Wkly Rep.* 2008;57(48):1297–1300.
  36. Ainsworth B, Haskell W, Leon A, et al. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exerc.* 1993;25:71–80.
  37. Ainsworth B, Haskell W, Whitt M, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc.* 2000;32(9, Suppl):S498–S504.
  38. Ainsworth B, Haskell W, Herrmann S, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc.* 2011;43(8):1575–1581.
  39. Pate R, O’Neill J, Lobelo F. The evolving definition of “sedentary”. *Exerc Sport Sci Rev.* 2008;36(4):173–178.