Journal of Physical Activity and Health, 2012, 9(Suppl 1), S19-S28

A Systematic Approach to Selecting an Appropriate Measure of Self-Reported Physical Activity or Sedentary Behavior

Barbara Sternfeld and Lisa Goldman-Rosas

Context: Numerous instruments to measure self-reported physical activity (PA) exist, but there is little guidance for determining the most appropriate choice. ***Objective:*** To provide a systematic framework for researchers and practitioners to select a self-reported PA instrument. ***Process:*** The framework consists of 2 components: a series of questions and a database of instruments. The questions encourage users to think critically about their specific needs and to appreciate the strengths and limitations of the available options. Instruments for the database were identified through existing literature and expert opinion. ***Findings:*** Ten questions, ranging from study aim and study design to target population and logistical consideration, guide the researcher or practitioner in defining the criteria for an appropriate PA instruments for a given situation. No one question on its own determines the optimal choice, but taken together, they narrow the potential field. The database currently includes 38 different self-reported PA instruments, characterized by 18 different parameters. ***Conclusions:*** The series of questions presented here, in conjunction with a searchable database of self-report PA instruments, provides a needed step toward the development of guiding principles and good practices for researchers and practitioners to follow in making an informed selection of a self-reported PA instrument.

Keywords: measurement of physical activity, questionnaires, recalls, methodology

The extensive base of evidence accumulated over the past 3 decades that demonstrates the influence of physical activity (PA) and sedentary behavior (SB) on a broad range of health outcomes has been built largely with the use of self-report measures.[1,2] Because self-reported PA and SB have many sources of measurement error, researchers have long sought alternative measurement approaches.[2–9] In recent years, technological developments have produced a variety of devices that directly measure human movement, such as step counters, accelerometers, and physiological and spatial monitors. Nevertheless, because these devices have their own set of limitations and sources of measurement error, there are specific times and circumstances in which self-report measure of PA and SB remain the most appropriate approach (see report by Troiano et al in this issue for further discussion of this topic).

Given that self-report measures remain relevant in both research and practice settings, the aim of the current discussion is to provide a framework for researchers and practitioners, who have decided to assess PA and/or SB by self-report, to select, among the many available self-report measures, the one best suited to their particular situation. This aim is not entirely unique. Previous efforts to facilitate appropriate use of self-report PA instruments include the Quality Assessment of Physical Activity Questionnaire (QAPAQ) checklist, which offers a structure for evaluating both the qualitative attributes of a PA questionnaire and its measurement properties.[10] The collection of PA questionnaires gathered together by Kriska et al and published in the 1997 supplement to *Medicine and Science in Sports and Exercise*,[11] and now accessible in expanded form on the interactive Physical Activity Resource Center for Public Health website (www.parcph.org), also represents a valuable tool for researchers and practitioners with minimal knowledge of PA instruments. The framework presented here differs from previous efforts by combining a practical guide for the user to define the important criteria a PA or SB instrument must fulfill, given the particular situation, with an easily searchable reference tool for linking those criteria to available instruments. In contrast to some previous efforts, it is not designed particularly to evaluate the psychometric qualities of the different instruments, a topic discussed thoroughly by Hagstromer et al in this issue.

After briefly describing the current state of self-reported PA assessment, the discussion will proceed to the 10 questions a researcher or practitioner should consider that help define the criteria, and then end with a brief summary of which of the most widely available PA and SB instruments meet which criteria. Although PA and SB are distinct constructs, we address both in this context as they are often measured using the same or related instruments.

Sternfeld is with the Division of Research, Kaiser Permanente, Oakland, CA. Goldman-Rosas is with the Stanford Prevention Research Center, Stanford University, Palo Alto, CA.

Current State of Self-Reported PA Assessment

Even a cursory review of the instruments used to assess PA by self-report reveals a confusing array of choices. This is exemplified by the 33 different approaches to self-reported PA measurement specified in 87 grant proposals funded by the National Cancer Institute between 2005 and 2009 (personal communication, H Bowles). This degree of heterogeneity hinders comparability of findings across different studies, thus limiting knowledge and understanding of how different dimensions of PA and SB affect health.[3]

Despite the broad spectrum of choices that exist, self-reported PA instruments can generally be classified into 5 categories, differentiated by the amount of detail requested and the time frame of report: (1) diaries; (2) logs; (3) short-term recalls; (4) long-term recalls, often referred to as semiquantitative or quantitative questionnaires; and (5) global or general surveys.[6,8,12–15] As shown in Table 1, a diary usually assesses a greater range of activities than a log, but both diaries and logs require varying degrees of real-time recording. Recalls may also target the same time frame as a diary or a log (ie, past 24 hours), but the time frame of some recalls, such as the 7 Day Physical Activity Recall[16,17] may be as long as a week. They differ from logs or diaries by explicitly asking respondents to *remember* all activities in which they engaged, rather than to record them as they occur. In contrast, longer-term recalls or quantitative questionnaires are designed to elicit more of an estimate of usual activity, rather than recalling specific episodes of activity. They also generally refer to a longer time frame (1 month, 1 year, lifetime). Global or general surveys provide a more qualitative, relative rating of usual activity. Although these categories are not entirely distinct, they provide a useful classification scheme for thinking about the different types of self-report PA instruments. They could also apply to different types of measures of SB, even though the range of those measures is currently quite limited.

Irrespective of the category of self-report measure, there is general consensus regarding the basic criteria for any measure of PA/SB, including self-reported measures.[4,8] First, the measure must be reliable; it must provide reasonably similar estimates of PA or SB with repeated measurement of the same individual, assuming a minimal learning effect and minimal change in behavior over time. Secondly, it must be valid, in that it actually measures the construct that it intends to measure. Given the absence of a gold standard for measurement of PA and SB,[4] validity usually refers to face validity, criterion validity (ie, agreement with other measures of PA/SB) or indirect validity (ie, expected association with characteristics known to be related to PA/SB). Thirdly, the measure must be practical, given available resources and characteristics of the population. Finally, any measure of PA or SB must be nonreactive and not interfere with the behavior that it intends to measure. Each of the standard, commonly used self-report measures of PA meet these criteria to one degree or another, although actual reliability, validity and practicality differ depending on the specific instrument and the specific situation in which it is used.

However, neither the general categorization of type of self-reported PA/SB instrument nor the criteria for acceptability provide systematic guidance for the researcher or practitioner needing to choose a specific instrument. In this paper, we begin to fill that gap by describing a systematic framework for finding the self-report PA/SB instrument that most meets the needs of a particular situation. Essentially, the framework consists of 2 components: a series of questions for researchers

Table 1 Traditional Categorization of Approaches to Self-Reported Physical Activity

Category	Level of detail	Time frame	Common uses
Diaries	High: recording of duration of all activities, sedentary behaviors, and sleep throughout a 24-hour period	Real time, often maintained for several days or a week	Validation of less burdensome measurement approaches
Logs	High: recording of requested details of specific activities of interest	Real time, past day, may be maintained for many weeks	Adherence to interventions
Recalls	Medium: recall of varying levels of detail (duration, frequency, intensity) of specific activities or categories of activities	Past 24 hours, past 7 days	Effects of interventions, assessment of current activity
Quantitative questionnaires	Medium: estimate varying levels of detail (duration, frequency, intensity) of specific activities or categories of activities	Usual week, past month, past year, lifetime	Assessment of longer-term activity, usual activity
Global surveys	Low: general categorization of activity level	Current, not specified	Ranking of usual activity

and practitioners to think about before they select an instrument, and a database of self-report instruments that can be searched to identify an appropriate measure based on responses to the questions. This framework is not intended to be proscriptive; it cannot make the decision for the user. Rather, it is meant to support a process of independent decision-making and to be as widely applicable and flexible as possible. Ultimately, the framework is intended to enable users to appreciate the strengths and limitations of the different options so that they can make an informed and appropriate choice.

Ten Questions to Guide Selection of a Self-Reported PA/SB Measure

The following 10 questions are designed to guide the researcher or practitioner in selecting an optimal PA instrument for a given situation. The responses to the questions help the user to determine the specific criteria an instrument must fulfill. No one question on its own will allow the user to select the single most appropriate PA/SB instrument. However, taken together, the questions will be useful in narrowing the potential choices.

Question #1: What Is the Primary Aim of the Study or Program? Clarifying *why* it is important to measure PA or SB is an essential first step to any study, evaluation or program. Although specifying the study aims or the goals for the evaluation or program will aid in narrowing the possible choices of instruments, in most cases, there will still be several possible instrument choices. For instance, the International Physical Activity Questionnaire (IPAQ), which was developed by an international group of researchers and extensively evaluated for psychometric properties,[18,19] is an appropriate choice for a range of different aims, including defining patterns of PA/SB in the population, informing public health policy and making cross-cultural comparisons. On the other hand, the IPAQ is not the only instrument that can accomplish these aims. The Four-Week History,[20] the Modifiable Activity Questionnaire,[21] and many other quantitative activity frequency questionnaires can as well.

Nevertheless, not all self-reported measures are suitable for all purposes. Understanding the purpose for which the original instrument was developed may be a useful guide when considering its use for a different purpose. The IPAQ, for instance, was developed for surveillance, not for assessment of behavior change in response to an intervention. Therefore, despite the extensive psychometric testing, it was not evaluated for sensitivity to change.[22] Similarly, a global question, such as "How physically active are you relative to others of your age and gender?" would not be appropriate to assess dose response relations with a health outcome, and using an instrument that does not assess walking or biking to work would obviously be a poor choice for evaluating an intervention directed at active transport.

In general, long-term quantitative questionnaires are useful for assessing usual activity over a specified time frame in studies investigating the dose response relationship of activity or SB with a particular health outcome. They are also useful in studies assessing the impact of an intervention or other program, as long as the questionnaire includes the target activities and has been shown to be sensitive to change. In contrast, global questions and general surveys are commonly used to rank participants by usual activity levels and are most useful when resources are limited and there is little need for quantifying PA. Short-term activity recalls are frequently used to study intervention effects, but may also be used to study associations with health outcomes, if the time frame of the recall is reasonably representative of usual activity in that population. Activity logs are commonly used to assess adherence to interventions, given the focus on specific activities, and diaries, which are thought to be impractical for large epidemiological studies, given their high level of detail and participant burden, are most often used as validation criteria,,

Question #2: What Is the Study Design? Thinking about the study design is especially important for choosing the appropriate time frame and the necessary level of detail required of the instrument. In a case-control study, for example, instruments that assess current PA or SB, such as a diary or a short-term recall, are not appropriate choices because the timeframe of the exposure does not precede the disease outcome.[23] Similarly, an instrument that assesses PA or SB in the distant past is not an optimal choice for evaluating cross-sectional determinants of current PA or SB behavior.

Thinking about study design also helps to define the level of detail that is needed. In surveillance surveys or evaluations of public health interventions, where PA or SB on the population level is the variable of interest, accounting for intraindividual variability in PA/SB behavior is not critical. In this case, assessing PA/SB with a single 24-hour activity recall is an adequate choice. In contrast, when PA or SB on the individual level is of interest, such as in etiological or intervention studies, the single 24-hour activity recall would not be appropriate although multiple administrations could be.

Question #3: What Are the Study Hypotheses; Is PA or SB an Independent Variable, Dependent Variable, or a Covariate? Specifying PA or SB in relation to the study hypotheses can clarify the degree of precision with which PA/SB needs to be assessed. If PA or SB is the primary independent or dependent variable, then it needs to be measured as thoroughly and precisely as possible, and the instrument must be capable of yielding the desired summary measure (eg, MET-hours/week, meeting PA recommendations, active transport). This may require a relatively comprehensive, detailed and quantitative instrument such as a log, multiple recalls or a detailed long-term quantitative questionnaire. In contrast, when PA or SB will be treated as a confounder or effect modifier, and is not the primary relation of interest, then it may be reasonable to use a shorter, more general instrument that provides a relative ranking of individuals, rather than quantitative estimates of

PA or SB behavior, such as a global question or a less comprehensive usual activity questionnaire.

Question #4: What Is the PA or SB Construct to Be Measured? As described in the paper by Pettee Gabriel and Morrow in this issue, PA is a multidimensional complex behavior that involves human movement. The behavior, which is influenced by characteristics of the individual and the environment, occurs within all domains of life, it may be discretionary or nondiscretionary, and results in energy expenditure and variable degrees of physical fitness, both of which may have short-term and long-term health consequences. The parameters by which PA is most frequently described are mode or type, frequency, duration, and intensity. Since no single measure of PA is capable of capturing all of the constructs implied in this model of PA, each researcher or practitioner must consider the PA constructs most relevant to his/her particular situation, and then find the instrument(s) that best captures those constructs.

The same process should guide the researcher or practitioner wanting to assess SB. SB is also complex and multidimensional, occurs in all domains of life and may be discretionary or nondiscretionary. It is characterized by lack of volitional movement and little to no energy expenditure beyond the resting metabolic rate, and is most often described in terms of duration and posture.

To illustrate this process, we will draw on 3 specific examples in which self-reported PA or SB needs to be measured. The first example is a currently funded, ongoing multicenter, observational study of natural menopause (SWAN) in which a multiethnic cohort of initially premenopausal women have been followed for the past 12 years. The overall aim of SWAN is to understand the biological, physiological, social and cultural factors that affect the menopausal transition and the health status of midlife women.[24] The second example, a project now in the planning stage, is an obesity intervention in Mexican American adolescents that uses a community-based participatory research approach. The third example is hypothetical: a national surveillance survey of trends in SB in the adult population in the US.

Each of these examples focuses on somewhat different constructs of PA (or SB). In SWAN, with its many potential outcomes of interest, the focus was on moderate and vigorous physical activity (MVPA) across several different domains. This is a construct that is reasonably well assessed by many standard, self-report instruments such as the Paffenbarger Alumni Study questionnaire[25] the Modifiable Activity Questionnaire,[21] and the Baecke Physical Activity Survey.[26] For the obesity intervention in Mexican American youth, the focus is the school environment, and the instruments under consideration include the Physical Activity Questionnaire for Children, which assesses specific activities common among elementary and middle school children[27,28] or the Modifiable Activity Questionnaire for Adolescents,[29] which includes questions about participation on sports teams. For surveillance of sedentary trends, the focus would be on discretionary SB, such as television viewing and on-line social networking. Here the choices of available self-report instruments are more limited since, with the notable exception of the 24 PAR,[30,31] a new on-line 24 hour PA recall, few existing instruments comprehensively assess SB. As seen in these 3 examples, specifying the construct to be assessed shortens the list of potential instruments that could be used.

Question #5: What PA Domains Need to Be Measured? PA may occur in a recreational context, during housework and care-giving, with volunteer or paid work, or in transport or other daily routines. Often the same movement, such as walking, can occur in multiple contexts, although often with different duration, frequency and intensity. Considering the 3 examples above, researchers with SWAN needed to assess various PA domains, since it is well established that much of women's PA occurs, not in the recreational domain, but in the household and caregiving domain.[32] This consideration was critical in the selection of the Kaiser Physical Activity Survey[33,34] as the assessment tool, since that instrument was developed specifically to assess women's physical activities in multiple domains. In contrast, the domains of activity most important to assess for the obesity intervention, which will most likely involve physical education curriculum development and teacher training as well as access to school facilities outside of regular school hours, are physical education classes and organized after-school activities. For the sedentary trends survey, assessment of discretionary behaviors related to transport and recreation might be of highest priority.

Question #6: What Parameters of PA or SB Need to Be Assessed? Deciding what parameters to assess, such as the type or mode of activity, frequency, duration, and/or intensity, is relevant for all study aims and study designs and determines the summary measure(s) of PA or SB that will be available for analysis. In terms of mode, many of the widely used long-term quantitative questionnaires can be dichotomized into those that inquire about specific activities, through open-ended questions, or lists of individual items, and those that ask respondents to categorize specific activities according to level of intensity (further discussed in Question #7). This difference results in differences in how PA can be expressed in a summary variable. For example, in an etiologic study aimed at establishing a dose response relationship between PA and some health outcome, collecting data on frequency, duration and intensity is highly desirable whereas it may not be necessary to document specific individual activities. In contrast, in other studies, the exposure of interest could be a specific activity, such as walking, resistance exercise or yoga, or television viewing making it important to collect data on these specific activities but perhaps not on the intensity.

Seasonality, an important source of intraindividual variability in PA behavior,[35–37] may also be a parameter to consider. Few self-report PA instruments collect this type of detail. Without such detail, those instruments that use the past year as the time frame of reference place a fairly difficult cognitive burden on respondents, requiring them to average out what may be variable levels of

participation in different activities. This may result in considerable over-reporting.[37] Seasonality is also an important consideration for studies that use a PA instrument with a short period of reporting (eg, past month, past 3 months) but enroll respondents and collect data from them over a longer period of time. In that case, seasonality will contribute importantly to interindividual variability, which cannot be accounted for if it is not assessed.

The parameter of intensity also deserves special attention. Absolute intensity is defined by standard MET values or caloric costs of specific activities while relative intensity is defined more by the respondent's perception of whether a given activity is of light, moderate, or vigorous intensity.[2,6,38] Assessing relative intensity accounts for the fact that the intensity of the same activity varies depending on an individual's age, fitness, health status, or other factors, but assessing absolute intensity allows for comparability across studies. This issue is discussed in the paper by Troiano et al in this issue, but is worth describing in this context in terms of the 3 different examples presented above.

In any large, population-based cohort study, such as SWAN, where there is a high degree of heterogeneity in the sample, using an instrument that assesses relative intensity might be a sensible choice. In SWAN, both relative and absolute intensity were assessed by assigning standard intensity codes to specific activities, but also by asking respondents to indicate whether each of the sports and exercises they listed in open-ended questions was performed with a little, moderate or a lot of effort. Whether this type of approach reduces the measurement error, or adds to it, has not yet been adequately addressed. In contrast to SWAN, the obesity intervention will need to consider absolute intensity, since actual energy expenditure, and not perceived effort, is relevant for evaluating the success of the intervention. Finally, the sedentary trends survey does not need to consider intensity at all since, by definition, the focus in on behaviors with MET values of 1 to 2.

Question #7: Do Specific Activities Need to Be Assessed or Can They Be Categorized? Whether activities should be assessed individually or in groups of similar activities will depend on the characteristics of the participants and the goals of the study or program. Many cohort studies, including SWAN, which examine PA as an exposure in relation to a wide variety of outcomes, have tended to ask participants about specific activities, either with open-ended questions or with activity lists. Asking respondents to name the activities in which they participate in an open-ended fashion, an approach used in the Harvard Alumni Study[25,39] or the Baecke survey,[26] has the obvious advantage of being able to capture the broad array of activities in the population. The primary disadvantage to this approach is that activities must then be coded in a standardized way, which is labor intensive, and, activities in which few people participate often end up pooled together in the analysis. To overcome this problem, many PA questionnaires, such as the Modifiable Activity Questionnaire,[21] the California Teachers' Physical Activity Questionnaire,[40] and the Cross-Cultural Activity Participation Study (CAPS) questionnaire[41] include lists of specific activities that account for the large proportion of activities, or at least moderate-vigorous physical activities, engaged in by the population. In some cases (eg, CAPS), the lists are derived from qualitative research conducted in the population of interest to ensure relevance of the activity lists to that population. Although this approach has benefits, one limitation is that the questionnaire can become relatively long and burdensome, and may still not ask about some activities in which some proportion of respondents engages.

On the other hand, instruments that ask respondents to report on categories of activities, either by type (eg, sports and exercise) or by intensity (light, moderate, vigorous) may be more comprehensive since they allow for the inclusion of all activities. They may also be less burdensome in terms of respondent time, but more burdensome in terms of the cognitive demand to sum level of participation over all relevant activities. Respondents may also mistake examples of activities within each category for definitive lists.[42]

For intervention studies, such as the obesity intervention for Mexican American youth that will target activities in Physical Education classes and recreational team sports, assessing targeted behaviors will be essential for determining the effectiveness of the intervention. Thus, an appropriate instrument might only need to include detailed information on these specific activities. For the sedentary trends survey, collecting data about individual SB, such as driving, television viewing, and computer time, would be more informative than assessing SB in aggregate.

Question #8: What Is the Desired Summary Measure of PA or SB? Since most self-report PA instruments consist of multiple questions, individual items are generally summarized into 1 or more variables that are used in subsequent analyses. Depending on the aims of the study and the detail of the individual items, the summary variable may be dichotomous (eg, active vs. nonactive; meeting recommendations vs. not meeting recommendations), ordinal categorical (eg, low, medium, high), or continuous (eg, MET-hours/week, time spent in MVPA/week, time spent in a specific activity). Continuous measures, which tend to be skewed, are frequently categorized into quantiles, using cut points based either on the population of interest or meaningful amounts of activity. Some continuous measures, such as the summary variable of Exercise Units obtained from the CARDIA Physical Activity History[43] or the scores obtained from the PASE,[44] Baecke,[26] and KPAS,[34] are not directly translatable into meaningful amounts of PA behavior, although most are. In selecting a self-report PA instrument, researchers and practitioners should confirm that the instruments they consider can yield the desired summary variables.

In terms of the 3 examples discussed in this paper, researchers with SWAN were initially motivated by the desire to rank women in some quantitative fashion according to their activity level, and were less concerned about expressing that level in terms of actual behavior. Over

time, this initial choice has limited the ability to translate study findings into explicit clinical and public health messages. With the obesity intervention, the PA instrument needs to measure change in PA energy expenditure or time spent in specific activities (kcals/week, MET-hrs/week, minutes of active time in PE class) in response to the intervention, which means that the instrument for assessing PA should collect data on at least duration and frequency, and probably intensity of those specified activities. With the sedentary trends survey, a categorical measure may be adequate, although a more quantitative measure of time in SB per day or week would provide greater information for informing public health policy.

Question #9: Who Is the Target Population? Not all self-report PA instruments are equally appropriate for all populations, both in terms of domains and types of activities that are assessed as well as cognitive demands on the respondents. Characteristics of the population that should influence the choice of a PA instrument include age, gender, race/ethnicity, educational level, language, and cultural context. Consideration should also be given to the qualitative attributes of the instrument and whether it is comprehensible and meaningful to the population of interest. Establishing face validity of an instrument within the target population through cognitive interviews is desirable.[42]

In SWAN, where the target population was midlife women of diverse race/ethnicity and socioeconomic status, including both English and non-English speakers, the need was for an instrument that assessed domains (eg, household/caregiving, transport) and types of activities (eg, dancing, gardening, vacuuming) relevant to women, that allowed for assessment of a wide range of activities relevant to different cultural backgrounds, that did not require difficult cognitive processing, and that could be easily translated. The KPAS met all of these criteria. An instrument designed for use with adolescents, such as the Previous Day Physical Activity Recall, which has a short time frame of recall[45] will be selected for the obesity intervention, while the sedentary trends survey could consider using the 24PAR,[31] the new on-line 24-hour physical activity recall that targets numerous SB.

Question #10: What Are the Important Logistical Constraints? The assessment of PA and SB does not occur in a vacuum and is always subject to budgetary constraints, as well as constraints related to respondent and staff burden. These constraints will impact the feasibility of different modes of administration including interviewer administered, in person or over the telephone, self-administered, either in person (where study staff can check responses for completeness) or by mail, and on-line surveys. They will also impact how much time and effort can be expended by both participants and staff. In addition, there are often competing needs that have to be balanced. For instance, in large prospective studies, such as SWAN, PA is only one of many exposures that investigators wish to measure. This limits the length and complexity of the PA instrument that will be acceptable to both participants and study investigators. Although these practical, feasibility issues are critical for any project, they are too often the determining factor in the choice of a self-reported PA instrument. This question was deliberately placed last in the series of questions to encourage researchers and practitioners to give no more weight to this question than to the other 9 questions discussed above.

Building a Database of Self-Report Instruments

The second component of a systematic approach to selecting a self-reported PA instrument that we describe here is a comprehensive database of PA measures, including information on measures of SB, where possible. As mentioned above, several compilations of self-reported PA instruments have existed for a number of years, most notably the 1997 supplement to *Medicine and Science in Sports and Exercise*,[11] edited by Dr. Andrea Kriska of the University of Pittsburgh, as well as the Physical Activity and Public Health Resource Center for Public Health website (www.parcph.org), also maintained by Dr. Kriska and colleagues. Building on these prior efforts, we developed a database that includes, to date, 38 different self-reported PA instruments, characterized by 18 different parameters or fields. Our strategy for identifying instruments for inclusion in the database started with the MSSE Supplement and the Physical Activity and Public Health Resources website, and then expanded to a survey of experts in the field, including the authors of the other papers in this supplement, and a search of the personal files of the authors and the reference lists in various evaluation and review papers.[3,20,46–51]

The parameters used to describe each instrument are listed in Table 2, along with their definitions. They range from general information, such as title, primary reference, type of instrument and general description to very specific information, such as the summary measures that can be derived, the population for which it was developed, and the domains of activity that it assesses.

Theoretically, a user could narrow his/her choice of PA instruments by using keywords to search the parameters of the database to find those that are consistent with the various combinations of responses given to the 10 questions described above. However, to be fully functional, this database needs several enhancements: 1) additional entries to be as comprehensive as possible; 2) integration into a searchable, relational database; 3) comprehensive testing and evaluation of "smart" algorithms for matching responses to questions with different instruments; and 4) ongoing maintenance and support.

Nevertheless, even without those enhancements, the database is still useful as a static table that can be used as a reference tool and is included as Online Appendix A. By simply counting the number of instruments that meet specific criteria determined by the responses to the series of 10 questions listed above, the user can identify a subgroup of instruments from which to choose. For

Table 2 A Database of Self-Reported Physical Activity Instruments: Database Fields and Definitions

Database field	Definition
Instrument	Common name(s) by which instrument is known
Overview	Brief description of the instrument, including type or category of instrument, such as diary, log, recall, semiquantitative questionnaire, quantitative questionnaire or global or general survey
Primary purpose	Typical study design for which instrument has been used
Mode of administration	The mode of administration such as interviewer-administered, self-administered, telephone-administered, on-line
Original population	A description of the population(s) in which the instrument has been validated or commonly used, including the age, gender, and race/ethnicity
Time frame	The time frame of activity the instrument assesses, such as usual past 24 hours, past week, past 3 months, past year, historical
Domain	The domain(s) of activity the instrument assesses such as housework, occupation, leisure time.
Summary variables	The summary variable(s) that can be derived from the instrument including the units and whether it is continuous or categorical
List- or category-based activities	A description of how instrument asks about mode of activity (ie, lists of specific activities or groups of activities or aggregation of activities into categories
Intensity: absolute or relative	Whether the instrument assesses absolute intensity (ie, standard MET values) or relative intensity (ie, respondent's perception)
Frequency	Whether the instrument assesses the frequency of the activity (Yes/No)
Duration	Whether the instrument assesses the duration of activity (Yes/No)
Seasonality	Whether the instrument specifically assesses seasonal variation
Walking	Whether the instrument includes an assessment of walking (Yes/No)
Strength	Whether the instrument includes an assessment of resistance exercises (Yes/No)
Sedentary behavior	Whether the instrument includes an assessment of sedentary activity (Yes/No)
Flexibility/balance	Whether the instrument includes an assessment of activities that target flexibility or balance (Yes/No)
Relation to other instruments	Comparison of how the instrument relates to other existing instruments, such as whether it is an adaptation or segment of a previous instrument
Primary reference	The original or most common reference for the instrument

example, if the sedentary trends study were to be implemented, examination of Online Appendix A shows that 9 of 24 instruments designed for adult populations assess SB, but only 5 of those are self-administered and 1 of those is web-based, which may not be appropriate yet for a population survey.

Conclusions

Over the past 25 years, we have learned many valuable lessons about the assessment of PA by self-report. The first is that there is no perfect self-report measure. As with other types of measures, self-reported PA and SB reflects underlying "true" constructs, but those constructs are measured with error. The sources and the magnitude of the errors are not the same for all self-report PA or SB methods and may vary with any one method depending on the setting in which it is used. Therefore, it is essential to understand the properties of any given method thoroughly before using it: what it actually measures, what its potential sources of error are, for what type of target population was it developed, how it can be interpreted, what resources it takes to use it correctly, and how it compares to other methods. Secondly, no given measure is the "correct" measure to use under all circumstances. The approach to selecting a self-reported PA or SB method described in this paper will, hopefully, encourage people to think critically from the start about why they need to assess PA or SB, what precisely measuring PA or SB means to them, and how their particular situation influences how they can and should measure PA or SB. Finally, there is no need to be apologetic about assessing PA or SB with self-report. The self-reported methods are not "second choice" methods when so-called objective methods cannot be used. Rather, they are valid methods in their own right and the optimal choice in certain circumstances.

Nevertheless, users of self-reported PA face some unique challenges. First is the issue of over-reporting and the extent to which over-reporting may vary depending on the structure of the instrument itself. Although the magnitude of over-reporting of PA behavior is not well established, instruments that present respondents with lists of specific activities may encourage over-reporting because the respondent is given more opportunity to report activity. If this is the case, then innovative self-report methods, such as electronic diaries or on-line recalls, may only exacerbate this problem. On the other hand, instruments that ask about categories of activities may also result in over-estimation because of the cognitive difficulties respondents have accurately aggregating their different activities.[42,52,53] It is also likely that the magnitude of over-reporting varies both between and within populations, depending on characteristics such as age, education, and race/ethnicity[54] and other factors. An in-depth understanding of the target population is, therefore, necessary to conceptualize and minimize the degree of over-reporting that may occur. Similar challenges may also apply to self-reported SB, although the error may be more toward under-reporting, rather than over-reporting.

A second challenge users face is how small changes in a self-report instrument might change the meaning and therefore the validity of the responses. Part of the reason for the multiplicity of self-reported PA instruments is the tendency of researchers and practitioners to make small "tweaks" to existing instruments. Although this practice is understandable, and perhaps even necessary, given different purposes and different populations, it is not clear whether the new version is actually a different instrument. In some sense, many of the existing instruments are simply subtle variations on each other. The more the field can move in the direction of systematizing these variations into coherent categories, the easier it will be for users to select an appropriate method.

Despite these, and other, challenges, it is clear that self-reported methods of assessing PA and SB continue to be relevant. Given that, the development of good practices and guiding principles and decision-making tools to aid researchers and practitioners in making informed and independent choices about when to use a self-report measure and which one to use is a high priority. The set of 10 questions provided here, in conjunction with a searchable database of self-report PA and SB instruments, is a first step in that direction. By enhancing our ability to systematize measurement of PA or SB, we will, ultimately, enhance our ability to understand more fully the relations between PA (or SB) and health.

References

1. U.S. Department of Health and Human Services OoDPaHP. Physical Activity Guidelines Advisory Committee Report. 2008.
2. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. *Br J Sports Med*. 2003;37(3):197–206.
3. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Res Q Exerc Sport*. 2000;71(2, Suppl):S1–S14.
4. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? *Int J Obes Relat Metab Disord*. 1998;22(Suppl 2):S30–S38.
5. Westerterp KR. Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol*. 2009;105(6):823–828.
6. Lagerros YT, Lagiou P. Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases. *Eur J Epidemiol*. 2007;22(6):353–362.
7. Baranowski T. Crisis and chaos in behavioral nutrition and physical activity. *Int J Behav Nutr Phys Act*. 2006;3:27.
8. LaPorte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep*. 1985;100:131–146.
9. Pols MA, Peeters PH, Kemper HC, Grobbee DE. Methodological aspects of physical activity assessment in epidemiological studies. *Eur J Epidemiol*. 1998;14(1):63–70.
10. Terwee CB, Mokkink LB, van Poppel MN, Chinapaw MJ, van Mechelen W, de Vet HC. Qualitative attributes and measurement properties of physical activity questionnaires: a checklist. *Sports Med*. 2010;40(7):525–537.
11. Kriska AM, Caspersen CJ. A collection of physical activity questionnaires for health related research. *Med Sci Sports Exerc*. 2011;29:s1–s205.

12. Ainsworth BE, Montoye HJ, Leon AS. Assessment of physical activity, fitness, and health. In: Bouchard C, Shepard RJ, Stephens T, eds. *Physical activity, fitness, and health: international proceedings and consensus statement*. Champaign, IL: Human Kinetics Publishers; 1994:146–159.
13. Ainsworth B, Levy S. Assessment of health-enhancing physical activity. Methodological Issues. In: Oja P, Borms J, eds. *Health enhancing physical activity*. Oxford, UK: Meyer & Meyer Sport; 2004:239–270.
14. Lamonte MJ, Ainsworth BE. Quantifying energy expenditure and physical activity in the context of dose response. *Med Sci Sports Exerc*. 2001;33(6 Suppl):S370–8 ; discussion S419–20.
15. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity—a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovasc Prev Rehabil*. 2010;17(2):127–139.
16. Blair SN, Haskell W, Ho P, et al. Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. *Am J Epidemiol*. 1985;122(5):794–804.
17. Sallis JF, Haskell WL, Wood PD, et al. Physical activity assessment methodology in the five-city project. *Am J Epidemiol*. 1985;121:91–106.
18. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–1395.
19. Wolin KY, Heil DP, Askew S, Matthews CE, Bennett GG. Validation of the International Physical Activity Questionnaire-Short among Blacks. *J Phys Act Health*. 2008;5(5):746–760.
20. Jacobs DR, Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc*. 1993;25:81–91.
21. Kriska AM, Knowler WC, LaPorte RE, et al. Development of questionnaire to examine relationship of physical activity and diabetes in Pima Indians. *Diabetes Care*. 1990;13(4):401–411.
22. Bauman A, Ainsworth BE, 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.
23. Armstrong BK, White E, Saracci R. *Principles of exposure measurement in epidemiology*. Oxford: Oxford University Press; 1994.
24. Sowers MF, Crawford SL, Sternfeld B. A multicenter, multiethnic, community-based cohort study of women and the menopausal transition. San Diego, CA: Academic Press: Biology and Pathobiology; 2000.
25. Paffenbarger RS, Jr, Wing AL, Hyde RT. Physical activity as an index of heart attack risk in college alumni. *Am J Epidemiol*. 1978;108:161–175.
26. Baecke JAH, Burema J, Fritjers JER. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *Am J Clin Nutr*. 1982;36:936–942.
27. Crocker PR, Bailey DA, Faulkner RA, Kowalski KC, McGrath R. Measuring general levels of physical activity: preliminary evidence for the Physical Activity Questionnaire for Older Children. *Med Sci Sports Exerc*. 1997;29(10):1344–1349.
28. Kowalski KC, Crocker RE, Faulkner RA. Validation of the physical activity questionnaire for older children. *Pediatr Exerc Sci*. 1997;9:174–186.
29. Aaron DJ, Kriska AM, Dearwater SR, et al. The epidemiology of leisure physical activity in an adolescent population. *Med Sci Sports Exerc*. 1993;25(7):847–853.
30. Schatzkin A, Subar AF, Moore S, et al. Observational epidemiologic studies of nutrition and cancer: the next generation (with better observation). *Cancer Epidemiol Biomarkers Prev*. 2009;18(4):1026–1032.
31. Calabro MA, Welk GJ, Carriquiry AL, Nusser SM, Beyler NK, Mathews CE. Validation of a computerized 24-hour physical activity recall (24PAR) instrument with pattern-recognition activity monitors. *J Phys Act Health*. 2009;6(2):211–220.
32. Ainsworth BE, Richardson M, Jacobs DR, Jr, Leon AS. Gender differences in self-reported physical activity. *J Women Sport Activity*. 1993;23:1–16.
33. Ainsworth BE, Sternfeld B, Richardson MT, Jackson K. Evaluation of the Kaiser Physical Activity Survey in women. *Med Sci Sports Exerc*. 2000;32(7):1327–1338.
34. Sternfeld B, Ainsworth BE, Quesenberry CP. Physical activity patterns in a diverse population of women. *Prev Med*. 1999;28(3):313–323.
35. Matthews CE, Freedson PS, Hebert JR, et al. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. *Am J Epidemiol*. 2001;153(2):172–183.
36. Pivarnik JM, Reeves MJ, Rafferty AP. Seasonal variation in adult leisure-time physical activity. *Med Sci Sports Exerc*. 2003;35(6):1004–1008.
37. Rifas-Shiman SL, Gillman MW, Field AE, et al. Comparing physical activity questionnaires for youth: seasonal vs annual format. *Am J Prev Med*. 2001;20(4):282–285.
38. Shephard RJ. Absolute versus relative intensity of physical activity in a dose response context. *Med Sci Sports Exerc*. 2001;33:s400–s418.
39. Paffenbarger RS, Jr, Hyde RT, Wing AL. Physical activity, all-cause mortality, and longevity of college alumni. *N Engl J Med*. 1986;314:605–613.
40. Bernstein L, Patel AV, Ursin G, et al. Lifetime recreational exercise activity and breast cancer risk among black women and white women. *J Natl Cancer Inst*. 2005;97(22):1671–1679.
41. Ainsworth BE, Irwin ML, Addy CL, Whitt MC, Stolarczyk LM. Moderate physical activity patterns of minority women: The Cross Cultural Activity Participation Study. *J Womens Health Gend Based Med*. 1999;8(6):805–813.
42. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical activity questionnaire comprehension: lessons from cognitive interviews. *Med Sci Sports Exerc*. 2009;41(2):336–343.
43. 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;9:448–459.
44. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical activity scale for the elderly (PASE): evidence for validity. *J Clin Epidemiol*. 1999;57:643–651.
45. Weston AT, Petosa R, Pate RR. Validation of an instrument for measurement of physical activity in youth. *Med Sci Sports Exerc*. 1997;29(1):138–143.
46. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act*. 2008;5:56.
47. Pettee Gabriel K, McClain JJ, Lee CD, et al. Evaluation of physical activity measures used in middle-aged women. *Med Sci Sports Exerc*. 2009;41(7):1403–1412.

48. Jorstad-Stein EC, Hauer K, Becker C, et al. Suitability of physical activity questionnaires for older adults in fall-prevention trials: a systematic review. *J Aging Phys Act*. 2005;13(4):461–481.
49. Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. *J Appl Physiol*. 2008;105(3):977–987.
50. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. *Sports Med*. 2001;31(6):439–454.
51. Department of Public Health UoWA. Report to the physical activity taskforce evaluation and monitoring working group. 1-50. 2002.
52. 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):1282–1291.
53. Heesch KC, van Uffelen JG, Hill RL, Brown WJ. What do IPAQ questions mean to older adults? Lessons from cognitive interviews. *Int J Behav Nutr Phys Act*. 2010;7:35.
54. Irwin ML, Ainsworth BE, Conway JM. Estimation of energy expenditure from physical activity measures: determinants of accuracy. *Obes Res*. 2002;9:517–523.