Measures Registry User Guide: Individual Physical Activity

NATIONAL COLLABORATIVE ON CHILDHOOD OBESITY RESEARCH

Suggested citation:

Welk, G. Morrow, J. Saint-Maurice, P. Measures Registry User Guide: Individual Physical Activity. Washington (DC):
National Collaborative on Childhood Obesity Research, January 2017. http://nccor.org/tools-mruserguides/wp-content/uploads/2017/NCCOR_MR_User_Guide_Individual_PA-FINAL.pdf

NCCOR
National Collaborative on Childhood Obesity Research

Measures Registry User Guide: Individual Physical Activity

Table of Contents

Acknowledgments ii

Section 1. Introduction 1
Overview of the Individual Physical Activity Measures Guide 2
Organization of this User Guide 3

Section 2. Behavioral Epidemiology Framework 6
Description of the Behavioral Epidemiology Framework 7

Section 3. Key Concepts for Understanding Individual Physical Activity 10
Definitions and Terminology 11
Physical Activity and Sedentary Behavior Guidelines 13
Uniqueness of Assessment in Children and Adolescents 14
Behavioral Patterns 14
Perceptions and Cognitions 15
Physiological and Metabolic Responses 15

Section 4. Measurement, Evaluation, and Statistics 18
Key Terms and Principles 19
Understanding Reliability and Validity 20
Distinctions and Definitions 20
Interpreting Reliability and Validity Statistics 21
Advanced Concepts in Measurement Research 24
Measurement Error 24
Statistical Power 24
Concepts of Equivalence 25

Section 5. Overview of Individual Physical Activity Assessment Tools 26
The Feasibility/Validity Continuum 27
Summary of Assessment Tools 29
Report-based Measures 29
Monitor-based Measures 30
Criterion Measures 32

Section 6. Selecting Measures 34
Population 35
Activity Outcomes 36
Research Type 36
Basic Research and Health Outcomes Research 36
Surveillance Research 37
Theory and Correlate Research/ Intervention Research 38
Resources 38
Decision Confirmation 39

Section 7. Case Studies 40
Case Study 1: Examining the Independent and Joint Associations of Physical Activity and Sedentary Behavior on Body Mass Index Among Middle and High School Students 42
Background 42
Considerations 42
Measure Selection 42
Case Study 2: Determining Compliance with Physical Activity Recommendations Across Different Grade Levels 43
Background 43
Considerations 43
Measure Selection 43
Case Study 3: Identifying Predisposing Factors for Active Commuting in Elementary School Children Who Live in Urban and Suburban Settings 44
Background 44
Considerations 44
Measure Selection 44
Case Study 4: Testing the Potential of a New Recess-Based Physical Activity Program Designed to Increase the Time Children Spend in MVPA During Recess 45
Background 45
Considerations 45
Measure Selection 45

Section 8. Supplemental Consideration for Monitor-based Assessments (New Technology and Data Processing Techniques) 46
Background on Accelerometry-based Activity Monitors 47
Considerations for Collecting and Processing Physical Activity Data 47
Considerations of Hip versus Wrist Placement 48
Handling Missing Physical Activity Data 48
Newer Monitoring Technologies and Methods 48
Considerations with Consumer-based Monitors 48
Distinctions between Raw and Count-based Accelerometer Data 49
Applications of Pattern Recognition Methods for Activity Classification 49

Section 9. Supplemental Considerations for Evaluating Sedentary Behavior 50
Background on Sedentary Behavior 51
Assessments of Sedentary Behavior 51
Quantifying Sedentary Behavior with Report-based Measures 51
Quantifying Sedentary Behavior with Monitor-based Measures 52

Section 10. Supplemental Considerations for Scaling and Scoring METs in Youth 54

Section 11. Conclusion 58

References 60

Acknowledgments

NCCOR gratefully recognizes The JPB Foundation for funding the development of the Measures Registry User Guides. NCCOR also extends its sincere thanks to the NCCOR Steering Committee, NCCOR Members, and the following individuals for their contributions to the guides.

Authors

Individual Diet
- Sharon Kirkpatrick, PhD, MHSc, RD
- Amanda Raffoul, MSc

Food Environment
- Leslie Lytle, PhD
- Allison Myers, PhD, MPH

Individual Physical Activity
- Gregory Welk, PhD
- James Morrow, PhD, FACSM, FNAK
- Pedro Saint-Maurice, PhD

Physical Activity Environment
- Jordan Carlson, PhD
- Kelsey Dean, MS, RD, LD, CCRP
- James Sallis, PhD

Reviewers

Food and Nutrition Expert Panel
- Alice Ammerman, DrPH, RD
- Carol Boushey, PhD, MPH, RD
- Karen Webb, PhD, MPH
- Gail Woodward-Lopez, MPH, RD

Physical Activity Expert Panel
- Genevieve Dunton, PhD, MPH
- Patty Freedson, PhD
- Brian Saelens, PhD

Measures Registry User Guide Project Team

NCCOR Members
- Rachel Ballard, MD, MPH
- David Berrigan, PhD, MPH (co-lead)
- Stephanie George, PhD, MPH
- Jill Reedy, PhD, MPH, RD (co-lead)

NCCOR Coordinating Center
- LaVerne Canady, MPA
- Adee Kennedy, MS, MPH
- Todd Phillips, MS
- Anne Rodgers
- Amanda Samuels, MS, MPH
- Namita Vaidya, MPH

We would also like to thank Seung Hee Lee-Kwan, PhD, MS, Latetia Freeman Moore, PhD, MSPH, Sohyun Park, PhD, Jesus Soares, ScD, MSc, and Tina Kauh, PhD, reviewing the final draft of the guides.

1

Introduction

SECTION 1

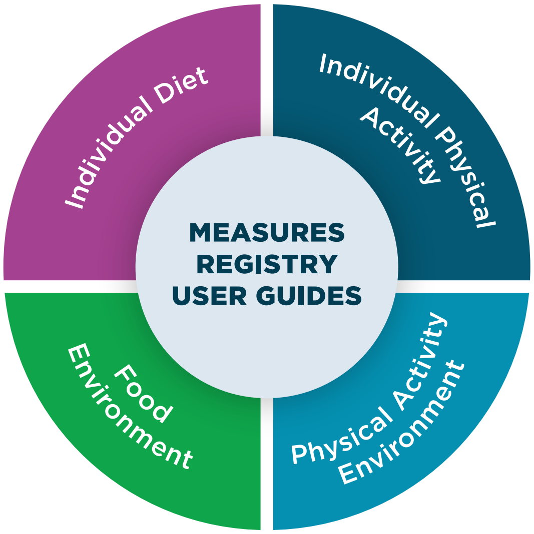
Introduction

Measurement is a fundamental component of all forms of research and it is certainly true for research on childhood obesity. A top priority for the National Collaborative on Childhood Obesity Research (NCCOR) is to encourage the consistent use of high-quality, comparable measures and research methods across childhood obesity prevention and research.

NCCOR's Measures Registry—a free, online repository of articles about measures—helps achieve this aim. It is widely recognized as a key resource that gives researchers and practitioners access to detailed information on measures in one easy-to-search location. The Registry's measures focus on four domains that can influence childhood obesity on a population level:

- Individual Diet
- Food Environment
- Individual Physical Activity
- Physical Activity Environment

Figure 1: NCCOR Measures Registry User Guides



Even with this resource, however, it can be challenging for users to choose the most appropriate measures for their work. To address this need, NCCOR began the Measures Registry User Guide project in 2015. Organized by the same four domains as the Measures Registry, the User Guides are designed to provide an overview of measurement, describe general principles of measurement selection, present case studies that walk users through the process of using the Measures Registry to select appropriate measures, and direct researchers and practitioners to additional resources and sources of useful information (Figure 1). The User Guide will help move the field forward by fostering more consistent use of measures, which will allow for standardization, meta-analyses, and synthesis.

Overview of the Individual Physical Activity Measures Registry User Guide

This Guide focuses on enhancing use of measures and tools in the Individual Physical Activity category. Accurate estimates of physical activity are essential for advancing research on the health benefits of physical activity; for understanding patterns and correlates that influence physical activity behavior; and for evaluating interventions designed to promote physical activity, improve health, or reduce obesity. Indicators of physical fitness (including body fatness) have generally shown stronger links with health indicators than with physical activity. However, this is due in part to the less precise methods available to assess physical activity. Physical activity directly improves fitness (and body composition) and consensus suggests that it improves health independently of both fitness and fatness. These findings clearly justify the emphasis on physical activity for advancing public health research focused on obesity and health.

Considerable attention has been given to improving physical activity assessment methods but progress has been hampered by limitations in the way that physical activity measures are used, scored, and interpreted. Many options are available for assessing physical activity, so it is important to appreciate and consider the relative advantages and disadvantages of the various measurement approaches. Decisions typically depend on the type of study or project being conducted as well as on the degree of precision needed for the assessment. However, consideration must also be given to the inherent challenges in collecting, processing, scoring, and interpreting physical activity data. Specific expertise may be needed to appropriately process and interpret data (particularly when using electronic monitoring devices). The various decisions and challenges involved can make it extremely difficult for researchers and practitioners to select, find, and use physical activity assessments effectively. Obtaining accurate assessments of physical activity is challenging in all populations but the issues are even further complicated when studying youth due to a variety of age and maturation effects.

A primary goal of the NCCOR Measures Registry is to move the field toward a more consistent use of common physical activity measures and research methods so that science (and practice) in childhood obesity prevention can progress in a systematic way. Consistent with this goal, the present Guide is designed to help users of the Measures Registry make informed decisions when selecting and using measurement tools to assess physical activity behaviors in youth. It is not intended to provide a comprehensive review of specific instruments, strategies, or assessment frameworks, as many excellent reviews have already been published.[1-4] Instead, it fills a different niche by emphasizing the measurement issues that should be considered when selecting and using physical activity measures in research and other evaluation contexts.

In its report Bridging the Evidence Gap in Obesity Prevention, the National Academy of Medicine (formerly known as the Institute of Medicine) established a decision-making framework called L.E.A.D. (*Locate Evidence, Evaluate it, Assemble it, and Inform Decisions*) to guide the progression of research on obesity prevention.[5] This framework has proved to be a useful strategy for any type of decision-making process. It was used to guide a prominent public health workshop on physical activity measures[6] and the concepts are also implicit in the vision and structure of the NCCOR Measures Registry. In this case, the NCCOR Measures Registry provides a definitive source to "*Locate evidence*" while this Guide provides frameworks to Evaluate options and to *Inform decisions* about the best ways to assess physical activity.

NCCOR: WORKING TOGETHER TO REVERSE CHILDHOOD OBESITY

NCCOR is a partnership of the four leading funders of childhood obesity research: The Centers for Disease Control and Prevention (CDC), the National Institutes of Health (NIH), the Robert Wood Foundation (RWJF), and the U.S. Department of Agriculture (USDA). These four leaders joined forces in 2008 to continually assess the needs in childhood obesity research, develop joint projects to address gaps and make strategic advancements, and work together to generate fresh and synergetic ideas to reduce childhood obesity. For more information about NCCOR, visit www.nccor.org.

Organization of this User Guide

The sections in this Guide build sequentially but can be reviewed independently depending on the needs or interests of the reader. Background information on measurement and evaluation principles is provided in the early sections along with coverage of physical activity terminology and calculations, as these provide the foundation for the Guide. Readers who have experience with physical activity assessment might consider jumping right to the set of Case Studies in Section 7 that summarize the factors that are most relevant for different types of research applications. Readers interested in deeper coverage on some topics can consult Section 8, Section 9, and Section 10, which provide supplemental information, although these sections only hint at the additional complexities and new methods being explored to further enhance methods.

In addition to this Introduction, this User Guide includes the following sections:

- Section 2. Behavioral Epidemiology Framework provides a framework to understand the unique needs of different types of studies and an introduction to the various categories of physical activity assessment options. It is important to understand these distinctions because appropriate decisions depend on the nature of the study goals, the target population, the relative need for precision, and the role that the physical activity measure plays in the analyses.

- Section 3. Key Concepts for Understanding Individual Physical Activity describes the complexities of quantifying physical activity and the inherent challenges involved in assessing a multi-dimensional and dynamic behavior rather than a more stable construct or trait. Assessments are challenging in all population segments but they are further compounded in research with children due to children's more sporadic activity behaviors, varying cognitive abilities, and different physiological responses as well as the inherent variability in growth and maturation.

- Section 4. Measurement, Evaluation, and Statistics provides an overview of measurement issues and key terminology, with a particular focus on the distinction between reliability and validity because these terms are reported for most of the tools in the Measures Registry. It is important to note that inclusion of a measure in the Measures Registry does not imply that it is recommended or that it has good psychometric properties. It is up to the researcher to evaluate the measurement properties for each tool and to determine the best measure for a specific application.

- Section 5. Overview of Individual Physical Activity Assessment Tools provides details about the strengths and limitations of different measurement tools by building on the psychometric properties and considerations outlined in Section 4. Methods are divided into criterion measures, monitor-based measures, and report-based measures and their relative utility is explained using a conceptual "*feasibility/validity*" continuum.

- Section 6. Selecting Measures compiles content from a number of sections and introduces a basic decision-making framework to facilitate the selection of appropriate assessment methods from the Measures Registry. Details of the typical measurement considerations involved in different types of research are summarized using the categories outlined in the behavioral epidemiology framework from Section 2.

- Section 7. Case Studies provides case studies to facilitate the application of the information in this Guide. Examples of different types of research are provided for the five distinct types of research described in Section 6. Different populations and outcomes are included in the scenarios to illustrate how the Measures Registry can be used to find details about specific tools.

- Section 8. Supplemental Considerations for Monitor-based Methods includes supplemental content that relates to continued advances in physical activity assessment research. Specific content is provided on new monitoring technologies and approaches with ecological momentary assessment, consumer-based monitors, and smartphone applications because these are dynamic growth areas in physical activity research.

- Section 9. Supplemental Considerations for Evaluating Sedentary Behavior includes supplemental material and content relevant to the evaluation of sedentary behavior. Many of the principles described in the Guide are relevant to the study of sedentary behavior, but this behavior also has unique considerations that must be taken into account.

- Section 10. Supplemental Considerations for Scaling and Scoring METs in Youth includes reference material and supplemental material for additional review. Specific consideration is given to implications for collecting and interpreting physical activity data on youth obtained from activity monitors and self-report instruments as they are the most commonly used tools. Readers are encouraged to review the references for additional information about some of the topics to obtain additional insights.

- Section 11. Conclusion.

- References

2

Behavioral Epidemiology Framework

Research advances occur over time through systematically using the scientific method and incorporating new ideas and approaches to answer progressively more complex questions. Medical and public health research is inherently driven by the need to develop and evaluate more effective methods to promote health and well-being in the population.

A classic definition of epidemiology (which underlies all public health research) is "*the study of the distribution and determinants of health-related states in the population, and the application of this study to the control of health problems.*" This definition captures all health-related states but specific "behavioral epidemiology frameworks" have been proposed to facilitate the progression[a] of research needed on specific health-related behaviors.[7] This Guide adapts a behavioral epidemiology framework developed for physical activity research.[8]

Description of the Behavioral Epidemiology Framework

The conceptual model in Figure 2 presents different types or categories of research needed to understand physical activity behavior and how to promote it more effectively.

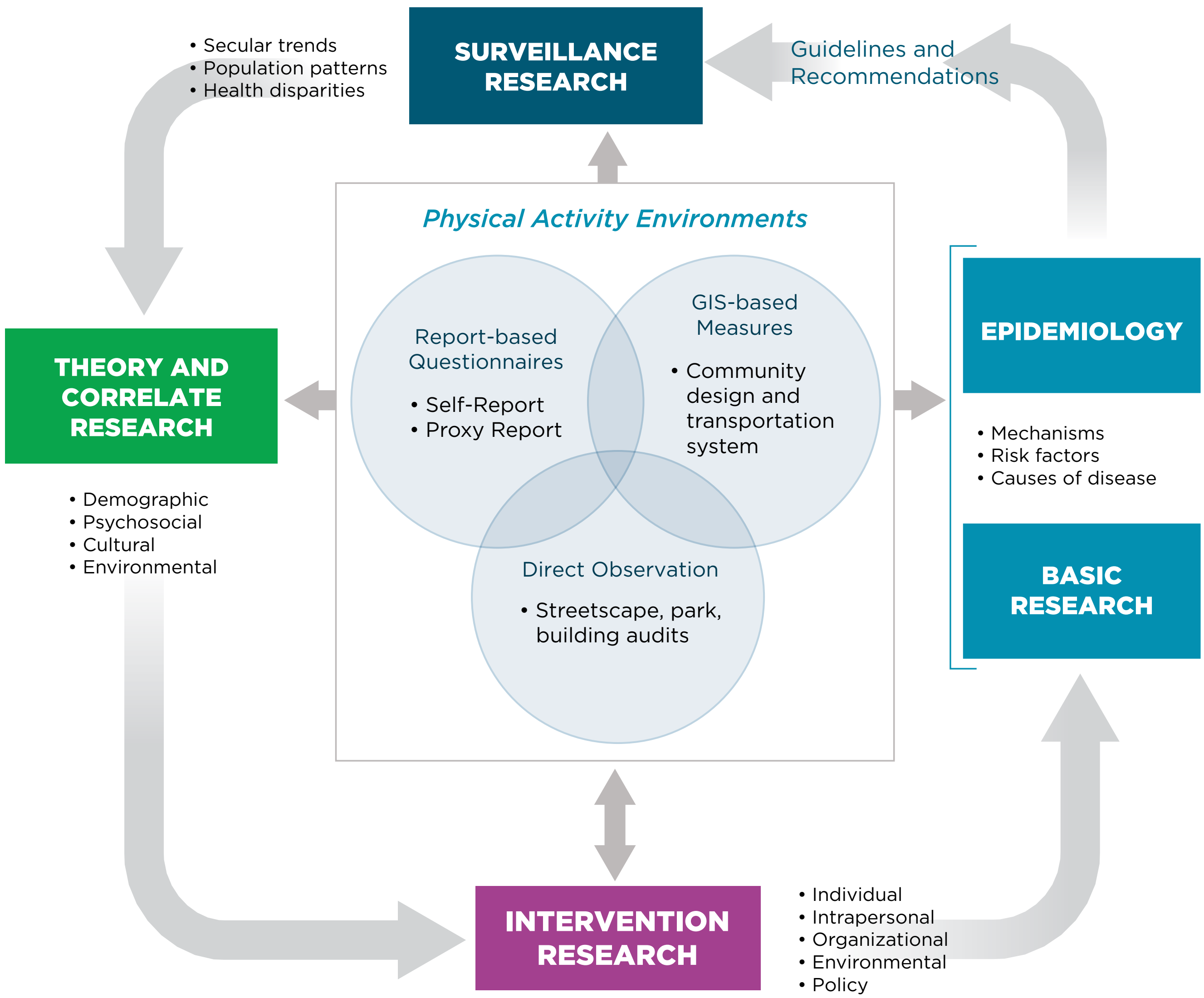
Five specific types of research are depicted around the perimeter of the figure, with each level building sequentially on the others to systematically advance behavioral research on physical activity. **Basic Research** provides the foundation for understanding physiological and biochemical mechanisms influencing disease risk and health. **Epidemiology** (health outcomes) then establishes specific associations with health indicators and facilitates the establishment of physical activity guidelines. **Surveillance Research** provides insights about secular trends with regard to physical activity as well as information about patterns, trends, and disparities in the population. **Theory and Correlates Research** seeks to understand the causes and correlates that influence the behavioral patterns and evaluates theoretical models designed to study and influence physical activity behavior. **Intervention Research** applies these insights to plan and evaluate intervention methods to influence physical activity and sedentary behavior in different settings and populations. The various types of research are linked to the center box labeled Physical Activity and Sedentary Behavior because these are the central behaviors of interest. The line between this center box and Intervention Research is a double-headed arrow because this is the only form of research designed specifically to promote change in these behaviors.

A prerequisite to advance research in these areas is to have accurate estimates of the underlying physical activity and sedentary behaviors. Therefore, assessment strategies have been placed in the center of the model. A novel adaptation in this adapted version of the framework is the depiction of three overlapping circles capturing the major categories of assessments:

- **Report-based measures** include various types of self- or proxy-report measures (e.g., questionnaires, diaries, logs).
- **Monitor-based measures** include various measures that directly or indirectly evaluate movement (e.g., accelerometer-based activity monitors, pedometers, multi-sensor monitors (e.g., devices that combine monitor-based measures such as accelerometer and

a Not to be confused with classic experimental (e.g., clinical trials) or non-experimental (i.e., cohort, case-control, cross-sectional, and ecologic studies) study designs.

Figure 2: Behavioral Epidemiology Framework



Welk, 2002. The figure was adapted to feature physical activity environment measures.[122]

heart rate), heart rate monitors, various smartphone apps (see Section 8 for more detail), and global positioning system [GPS] devices).

- **Criterion measures** include the doubly-labeled water technique, indirect calorimetry, and various direct observation measures that involve direct coding of behavior (e.g., time, intensity, type, location).

The independent but partially overlapping portions of the circles illustrate the fact that the measures provide somewhat different views of the physical activity construct.[9-10] The overlapping circles also reinforce the advantages of employing multiple measures to triangulate outcomes.[10]

The large space outside of the circles illustrates the fact that the available methods only capture estimates of the actual (true amount of) physical activity conducted. Although it could be argued that criterion measures explain more of the variance in the true physical activity (compared with report-based or monitor-based measures), the point is that all measures provide incomplete views of physical activity behavior. Thus, the methods represented in the circles should be viewed as surrogate measures intended to provide estimates of what actually occurs. Thus, all measures (even criterion measures) inherently contain errors of estimation, which is commonly called measurement error. It is important to take steps to minimize error during data collection, to control the amount of error introduced during processing, and to interpret all results in the light of overall measurement error.

The focus of the NCCOR Measures Registry is on field-based measures that are widely used in research applications. Other summaries have distinguished these types of measures as either *objective* or *subjective*[3] but this categorization infuses an inherent, but unintentional, bias. Large discrepancies in prevalence rates and levels of physical activity have been observed when objective and subjective measures are compared.[11-13] However, it is important not to assume that one category is better than the other. Monitor-based (e.g., activity monitors) measures are certainly more objective than report-based measures (e.g., diaries), but the key distinction is that they capture amounts of movement as opposed to a person's perception or recall of physical activity experiences. An unfit individual may perceive a certain activity as vigorous in intensity while a fit person may perceive the same activity as being light intensity. This example describes physical activity intensity expressed in relative terms (i.e., accounting for one's individual level of fitness) and highlights the need to differentiate between relative and absolute intensity of the physical activity performed. Report-based measures capture relative intensity (i.e., individual perception of intensity) while monitor-based measures capture the absolute intensity and volume of activity without considering perceived or actual effort. It is true that report-based measures contain a considerable degree of subjectivity that can contribute toward sources of error and bias, but different sources of measurement error also plague monitor-based measures. The *U.S. Physical Activity Guidelines*[14] emphasize the distinction between absolute and relative indicators, and this needs to be considered when evaluating physical activity outcomes. The labels of "monitor-based" and "report-based" measures reinforce the message that physical activity behavior can be monitored or reported.

The key point of the framework is that the accurate assessment of physical activity is a priority in all facets of research. In making decisions about the most appropriate instrument, it is first important to fully consider the distinctions between movement and behavior. These distinctions are more fully explained in the next section.

3

Key Concepts for Understanding Individual Physical Activity

This section describes concepts that are important for understanding the remaining sections of the Guide. The fundamental definitions and distinctions related to both physical activity and sedentary behavior are described first, followed by summaries of physical activity and sedentary behavior recommendations. Emphasis is then placed on the unique challenges of assessing physical activity and sedentary behavior in youth, as that population is NCCOR's focus. These sections provide the foundation for understanding how to assess behavior and movement and estimate energy expenditure in youth.

Definitions and Terminology

Physical activity can be quantified and interpreted in a variety of ways. Caspersen et al. previously described physical activity as "Any bodily movement produced by skeletal muscles that result in caloric expenditure."[15] This definition has been widely accepted but a more recent conception, developed through a consensus conference on physical activity research,[6,9] provides an operational definition to avoid subjectivity and facilitate assessment: "behavior that involves human movement, resulting in physiological attributes including increased energy expenditure and improved physical fitness."

A critical element in the new definition is the labeling of physical activity as a *behavior*. This captures the volitional nature of physical activity and the various physiologic, psychosocial, and environmental factors that influence it. For youth, the *movement* captured in this behavioral definition can be categorized as either structured (i.e., repetitive, organized activity, often led by an adult and performed in physical education class) or unstructured (i.e., play, unsupervised, activity performed during recess or school breaks). Activities also be operationally characterized according to their frequency (i.e., number of movements per day), duration (i.e., recorded minutes of actual movement), intensity (i.e., associated effort to perform the movement), and type (i.e., nature of movement as being, for example, aerobic or bone-strengthening related activities). The combination of frequency, duration, and intensity is often referred to as the dose or volume of physical activity and reflects the total amount of movement performed within a specific time period. Three other important distinctions with the definition are summarized below.

First: In this new conception of physical activity, a better understanding of the context and settings where physical activity behavior occurs (e.g., home, work) as well as the purpose (e.g., recreation, occupation) is needed. Although categories can vary, four broad domains that effectively capture behaviors for both adults and youth include: (1) leisure-time physical activity (i.e., recreation, play), (2) work- or school-related physical activity, (3) home or domestic physical activity, and (4) transportation physical activity (commuting from place to place). The term "exercise" is viewed as a subcategory of leisure-time physical activity that is more structured (e.g., steady state running) and performed with a well-defined purpose in mind (e.g., improving or maintaining physical fitness). The distinctions between physical activity and exercise are more relevant for adults, but participation in sports or structured activity programs or lessons by youth can be considered analogous to "exercise" because it is also structured and purposed. Adults often report physical activity for leisure or recreation but in youth this may be captured as simply “play” or unstructured activity.

Second: The new definition of physical activity stipulates that movement needs to be of sufficient magnitude to increase energy expenditure.

This definition helps to distinguish physical activity from non-volitional forms of movement (e.g., fidgeting) and focuses attention more on larger contributions to energy expenditure. Energy expenditure is typically expressed in units of kilojoules (kJ) or kilocalories (kcal), but it is also frequently expressed as multiples of resting energy expenditure known as Metabolic Equivalent Tasks (METs).[16] Resting energy expenditure is often estimated because it is challenging to measure, and the value of 3.5 ml/kg/min has been widely adopted as the oxygen consumption of a person at rest. Using standard conversions and additional assumptions, resting energy expenditure (i.e., 1 MET) has been equated to an energy cost of 1 kcal/kg/hour. Other procedures yield different estimates but the consistent adoption of MET values and methods from the Compendium of Physical Activities[17] has helped to standardize outcomes. Levels of physical activity are routinely calculated using established ranges (Rest is 1.0 to 1.4, Light physical activity [LPA] is 1.5 to 2.9, Moderate physical activity [MPA] is 3.0 to 5.9, Vigorous physical activity [VPA] is 6.0+). Most physical activity research has used a combined indicator that captures both moderate physical activity and vigorous physical activity (MVPA). However, research has increasingly emphasized the importance of understanding the allocation of time spent in different intensity classifications, as they each contribute directly to overall energy expenditure and health. Considerable attention has been given to time spent in sedentary behavior because it has been shown to be independent of time spent in MVPA. By default, the time spent in LPA also has implications because it falls between these two intensities.[18] More time spent in LPA can be beneficial if it corresponds with less time in sedentary behavior. However, time spent in LPA does not provide benefits that come from participation in MVPA. Distinctions of the main components and dimensions of physical activity and sedentary behavior are summarized in Figure 3, based on an established model.[9]

UNDERSTANDING ENERGY EXPENDITURE TERMS

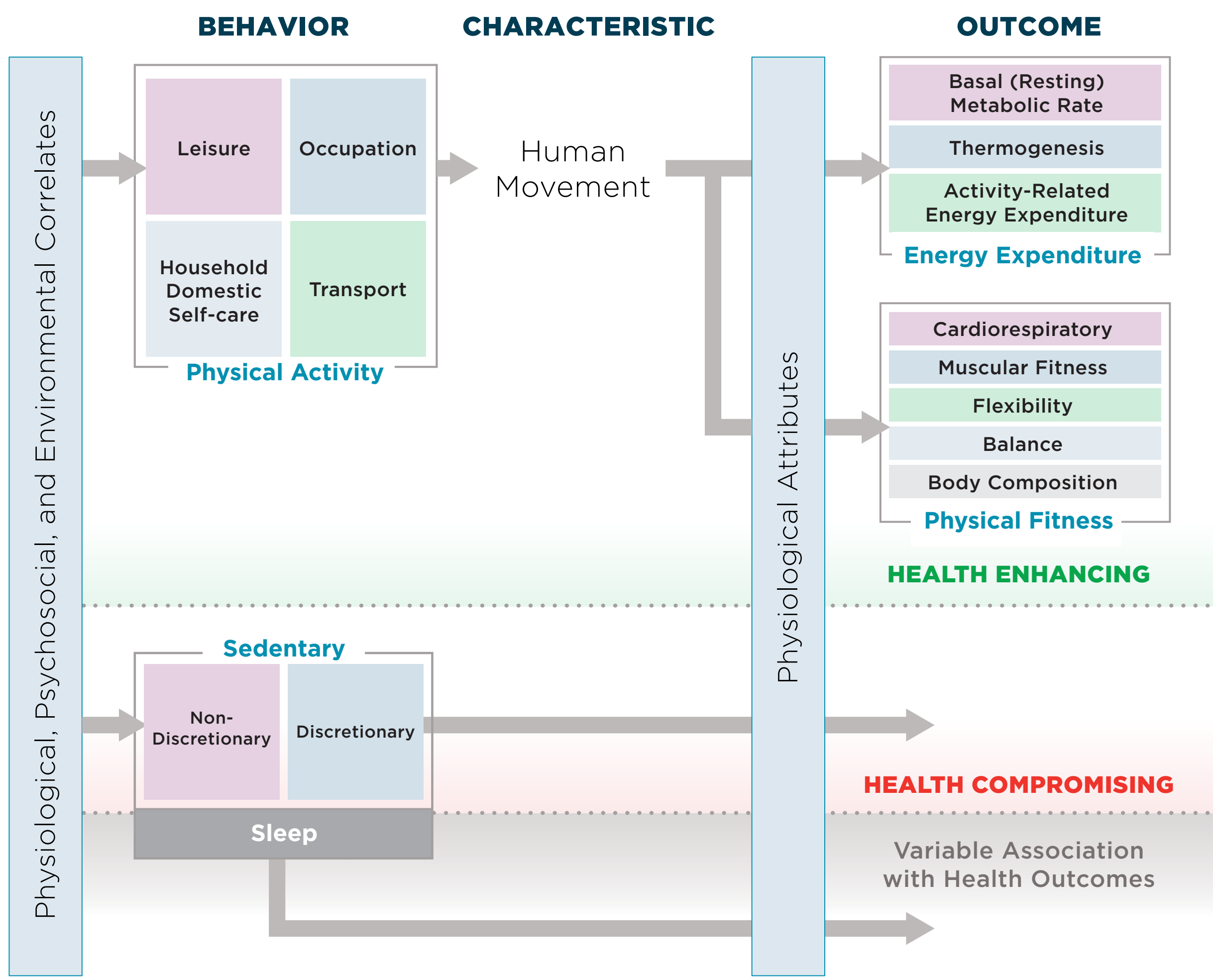
Total energy expenditure (TEE) is generally divided into three components: Resting energy expenditure (REE), the thermic effect of food (TEF), and the more volitional physical activity energy expenditure (PAEE). The REE value accounts for about 50% to 60% of total energy expenditure, but PAEE is usually of more relevance because it is the most variable component of TEE and is highly susceptible to change.

Third: The new definition of physical activity specifically references its contributions to improving dimensions of physical fitness. Physical fitness has generally been defined as “the ability to carry out daily tasks with vigor and alertness, without undue fatigue and with ample energy to enjoy leisure-time pursuits and meet unforeseen emergencies.”[15] It can be subdivided into performance-related fitness and health-related fitness but the latter is more relevant for the purpose of this Guide because the majority of physical activity research is focused on health-related outcomes. Another caveat with this definition is that participation in physical activity may not necessarily lead to predictable or measurable improvements in physical fitness. Improvements are influenced by baseline fitness, genetic predispositions and a number of other factors, so the emphasis should be placed on the potential of physical activity to improve dimensions of physical fitness.

Body composition is considered to be a dimension of health-related fitness and is obviously of particular relevance for research targeting childhood obesity. Based on the description above, physical activity has important implications for maintaining or improving body composition and can induce positive changes in body fat content and distribution.

Research and public health guidelines have distinguished physical activity and sedentary behavior as independent behavioral constructs and they also may have independent effects on health, although this is less established in youth.[19] No universally agreed-upon consensus has yet been achieved on defining sedentary behavior for both children

Figure 3: A Model of Sedentary and Physical Activity Behaviors



and adults, though concerted efforts have been made for adults. For instance, researchers in the Sedentary Behavior Research Network have come to agreement that sedentary behavior should be defined as "any waking behavior characterized by an energy expenditure ≤1.5 adult-METs while in a sitting or reclining posture."[20] The threshold of 1.5 adult-METs has been generally considered a cutpoint for identifying sedentary behavior in adults. However, different assumptions must be considered for children. Recommendations for addressing this issue have been included in Section 10.

Physical Activity and Sedentary Behavior Guidelines

Previous activity guidelines emphasized the total amount of MVPA that should be performed, and separate targets were provided for MPA and VPA. The current *U.S. Physical Activity Guidelines*[14] provide a more flexible model for tracking physical activity levels by focusing on the total volume of physical activity performed and the construct of "MET-Minutes." Recommendations call for individuals to obtain 500 MET-Minutes a week. However, the guidelines also emphasize the need for relative guidelines that take into account a person's individual level of fitness. Individuals are encouraged to perform 150 minutes of moderate-intensity physical activity a week but it can be accumulated in different ways. Consistent with the MET-Minute approach, vigorous minutes are multiplied by two to reflect the higher MET costs of VPA vs. MPA (6 METs vs. 3 METs) thereby allowing a person to meet the guideline with a combination of MPA and VPA. Because individuals vary in fitness level, it is important to acknowledge that MPA and VPA may be perceived very differently in the population. The distinction between relative and absolute intensities has important implications for the different physical activity measures. For example, report-based measures capture the perceptions of physical activity while monitor-based methods capture the movement that takes place. A fit person may report performing very little physical activity but the monitor may record considerable amounts. In contrast, an unfit person may have very little absolute movement in a day but it may be moderate in intensity. The examples are generally explained in the context of adult behavior but the same implications hold for quantifying youth physical activity, as youth who are unfit and have overweight may perceive their activity as of moderate or high intensity even though the monitor may record little absolute movement. These are simple examples, but the point is that the frequently observed discrepancies between measures may *not* be solely due to bias or recall problems, but rather to inherent differences in reported and measured data or how physical activity intensity is expressed. Thus, monitor-based and report-based measures capture different aspects of the same underlying construct of physical activity.

Guidelines for sedentary behavior have been harder to establish and are less consistently endorsed. Sedentary behavior is considered a construct that is independent of physical activity and that can carry different health implications. For this reason, the Canadian Society for Exercise Physiology in collaboration with the Healthy Active Living and Obesity Research Group developed the Canadian Sedentary Behavior Guidelines for Children and Youth. These were the first guidelines to specifically address recommendations for sedentary behavior in order to improve and maintain health. The guidelines suggest that children and youth should limit recreational screen time to a maximum of 2 hours per day and reinforce that lower amounts of screen time can offer additional health benefits.[21] Other national and international organizations, such as the Australian Department of Health, also have developed specific guidelines for children and youth while reinforcing the importance of avoiding long continuous periods of sitting time. The American Academy of Pediatrics (AAP) recommends that young children limit the amount of screen time per day to no more than 1 hour[22] and encourages pediatricians to work with children and families to promote a lifestyle with reduced sedentary behavior.[23] The focus in this Guide is on assessments of physical activity; additional details about issues with assessment of sedentary behavior are available in Section 9.

Uniqueness of Assessment in Children and Adolescents

Assessing physical activity is challenging for all populations but it is particularly difficult in children and adolescents. Children have unique behavioral patterns of physical activity, unique perceptions and cognitions related to physical activity, and distinct physiological and maturational responses and adaptations to physical activity. Most foundational work on assessing physical activity and energy expenditure has been derived in adults and the simple assumption has been that these also hold in youth. However it is clear that children are not just "little adults,"[24] so special considerations are needed to evaluate and study individual physical activity behavior in this segment of the population. Three specific considerations that must be taken into account for youth assessments are outlined below:

Behavioral Patterns

Children are known to engage in more sporadic and intermittent activity than adults, and this has important implications when trying to capture and assess physical activity with either report-based measures or monitor-based measures.[25] The patterns of physical activity also vary across childhood and throughout later stages of adolescence. For example, preschoolers go through phases of motor skill acquisition and refinement and demonstrate less refined and less efficient movement patterns than do older age groups. Elementary school children (i.e., ages 6 to 11 years) have increasingly efficient movements but highly sporadic and intermittent physical activity patterns due to the random nature of play. The transition to adolescence (i.e., ages 12 to 18 years) is typically characterized by drops in physical activity levels and a greater contribution of team sports toward total physical activity accumulated during the day. Youth do not commonly exhibit adult patterns of continuous physical activity despite the emergence of maturing physical and behavioral attributes.

The variability in movement patterns from childhood to adolescence imposes unique measurement constraints for both report-based and monitor-based measures. Preschoolers are unable to recall activity and standard energy expenditure conversions often do not account for lack of movement efficiency. Tools capturing reported physical activity and sedentary behavior are increasingly useful as youth move from elementary school to middle school (i.e., pre-adolescents). However, it is still inherently challenging to capture the sporadic movements and "play" that characterize youth activity. Activity patterns become more predictable in high school youth (i.e., adolescence) as participation in structured physical activity becomes more common. Structured physical activity is more easily reported on recall instruments and is also easier to detect and quantify with monitors so assessments become slightly easier. Nevertheless, youth participation in team sports and random forms of play are still more difficult to capture than the more structured activities common in adulthood (e.g., jogging).

Another independent challenge is capturing the representative nature of their behavior (i.e., assessment of habitual physical activity and sedentary behavior patterns). Monitor-based methods may assess behavior over multiple days or full weeks while report-based methods typically involve recall over time (e.g., previous 7 days) or estimates of "typical" behavior. Attention should be given to determining the time frame needed to obtain reliable indicators of actual behavior because it has important implications for research with youth. In addition to variability in overall behavior, it is important to consider inherent variability within a day (e.g., school-based physical activity vs. home-based physical activity), across days (e.g., days with physical education vs. days with no physical education), between days (e.g., school-day vs. non-school day), and across seasons (e.g., winter vs. summer activity patterns). Therefore, it is important to consider the definition of time frame as the period of time of interest and account for this variability when characterizing the physical activity behavior even though in most scenarios, if not all, the typical behavior is of most interest. Details on how to determine variability (i.e., reliability) of physical activity behavior are provided in Section 4.

Perceptions and Cognitions

Children's cognitions and perceptions (e.g., knowledge of physical activity) also must be considered when using report-based tools. When using these measures, ambiguous terms like "physical activity" and "moderate intensity" can generate confusion as children display a limited understanding of the concept of physical activity and have difficulties reporting the intensity of the activities in which they engage. These challenges become clear when children are asked to indicate how many bouts of moderate or vigorous physical activity they performed in the previous day or past week. A more prominent concern is related to the limited ability of children to provide details of past physical activity events with retrospective recall instruments (e.g., previous week, previous month). The recall of physical activity requires complex processes that can lead to inaccurate reports of this behavior in all populations, and particularly in youth.[26] The accuracy of reporting is highly dependent on the appropriate use and selection of episodic memories, which are associated with the child's capacity to remember a specific event within a particular place and moment in time. The memory of the episode is like re-experiencing it (e.g., when describing events such as school graduation, a person can clearly remember the setting and even feel the same sensations or emotions of that day). Episodic memory refers to experiences of daily living, such as eating breakfast or exercising and can be replaced by generic memories (i.e., memories of general events or patterns of events) that are used when individual memories or episodic memories are not available. For example, short-term or specific physical activity recall questionnaires (e.g., previous day, previous week, number of exercise bouts) are examples of instruments that refer to episodic memories.[27] This can be problematic considering that the "natural" intermittent patterns in youth behavior makes these events quite common and therefore harder to recall or report with sufficient level of detail.[26,28]

Physiological and Metabolic Responses

Another distinction with assessing physical activity and sedentary behavior in youth is that standard physiologic adaptations and relationships do not always hold when applied to youth. Perhaps the most critical distinction is the difference in metabolic cost of physical activity as a result of aging or growth. The amount of activity performed based on absolute intensity (e.g., use of MET values) assumes a standardized resting state of 3.5 ml/kg/min, a value established based on adult values of resting energy expenditure. Resting energy expenditure is primarily determined by body composition and more particularly by muscle mass, but other important predictors include age, sex, and body fat. These factors lead to error when using the standardized value of 3.5 ml/kg/min in adults but more significant errors (and systematic bias) when applied in youth populations.[24, 29-32] The error is introduced due to known differences in resting energy expenditure for youth. For example, the resting energy expenditure in a 13-year-old child can be approximated as 4.2 ml/kg/min or 1.2 adult METs.[33] If actual child resting energy expenditure values are used, light intensity would be described as activities eliciting up to 6.3 ml/kg/min (i.e., 1.5 times above their actual resting state) rather than up to 5.3 ml/kg/min (i.e., 1.5 METs x adult resting energy expenditure of 3.5 ml/kg/min). Failure to consider this difference leads to systematic over-estimation of children's physical activity intensity and a misclassification of performed activities.[b] Error is further compounded due to additional variability associated with differences in lean body mass in children classified as normal weight and children classified as overweight or obese.[c]

It is important to note that the concepts of METs were not intended to take into account inter-individual differences or effects of different body composition and fitness levels.[34] Detailed coverage of error due to estimation of METs is beyond the scope of the Guide but readers should consider the implications of these issues when processing and interpreting physical activity data. New methods to refine and standardize MET values for coding youth physical activity behaviors are described in the NCCOR Youth Compendium of Physical Activities website. Recommendations for standardizing youth MET outcomes are provided in Section 10.

b A recent evaluation confirmed that measured energy expenditure for sedentary activities in children tended to be closer to 2.0 METs instead of the adult threshold of 1.5 METs used to distinguish sedentary and light physical activity (see reference 35).

c Similar to the child vs. adult comparison, the differences in body composition and impact on REE will likely lead to systematic misclassifications of activity in children classified as having overweight or obesity (see reference 30).

4

Measurement, Evaluation, and Statistics

This section describes key concepts in research and statistical inference with special emphasis on assessing physical activity. The issues raised in this section are important regardless of the specific physical activity assessment protocol chosen. Key distinctions between the terms "measurement," "assessment," and "evaluation" are first explained, followed by principles of sampling and calibration that have specific implications for physical activity research. Specific attention will then be given to distinctions between reliability and validity so that viewers can effectively interpret tools summarized in the Measures Registry. A brief discussion of advanced concepts of measurement research concludes this section.

Key Terms and Principles

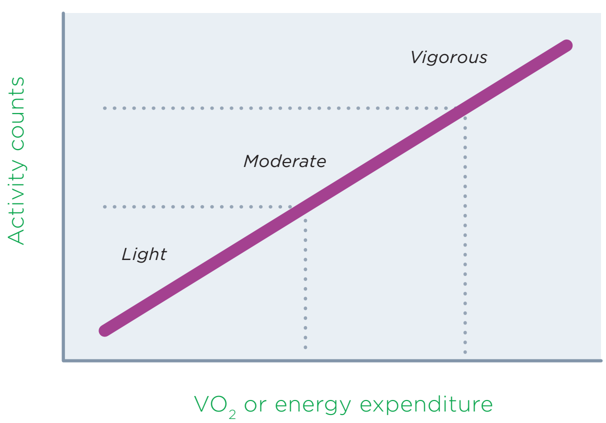
The terms "measurement," "assessment," and "evaluation" have important distinctions that need to be considered in physical activity research. The three items are often used interchangeably; however, they have very distinct meanings and interpretations. **Measurement** involves collecting specific information about an object or event and it typically results in the assignment of a number to that observation. **Assessment** is a broader term that refers to an appraisal or judgment about a situation or a scenario. Physical activity cannot be directly measured in the same way that height or weight are measured, for example, so the term "assessment" is generally preferred when referring to efforts to quantify the type or amount of physical activity that is performed. Available "measures" are used to assess physical activity with varying degrees of error and this must be considered when interpreting the information. **Evaluation** involves attributing a meaningful value to the information that is collected. As expected, evaluation denotes placing a "value" on the obtained measurement. The values can be compared to a reference population (i.e., are "norm-referenced") or to some type of standard (i.e., are "criterion-referenced").

The key point is that the three words each have different meaning and cannot be used interchangeably. In the context of physical activity, the *measurement* could be a set of responses obtained through a recall tool (e.g., the amount of time walking or playing tennis), the *assessment* refers to the procedures used to determine the estimate of time spent in physical activity, and the *evaluation* would determine whether the person met the established physical activity guidelines or not.

Another important distinction in measurement and evaluation is that of sample vs. population. It is certainly unrealistic to obtain information from every single member of a population, so a sample is typically used to reflect the population of interest. The distinction between sample and population is analogous to the inherent differences between a measure and an estimate. In essence, we are attempting to *measure* behaviors of a *population* with *estimates* obtained from a *sample* of individuals. We must infer and "speculate" what the estimate or sample would be equivalent to if we had a true measure from all of the individuals in the target population. This process is defined as inferential statistics and it consists of replication of the population parameters of unknown distributions by examining the distributions in a subset of individuals who were randomly selected and are part of the population of interest. Random sampling is a condition that is often not satisfied, so our inferences are typically based on convenience samples (i.e., individuals not randomly selected from a population). It is important that the sample drawn randomly represents the population to which one desires to generalize the results.

Another fundamental measurement consideration in physical activity research is that of calibration. As previously described, not all the dimensions of physical activity behavior can be directly measured, so assessment procedures typically necessitate a calibration process to obtain desired variables such as behavior type, VO_2 or energy expenditure. Monitor-based measures, for example, produce raw indicators of movement (e.g., activity counts) and these data may be analyzed directly as activity volume or translated to other units such as VO_2, energy expenditure, or minutes of MVPA).[35] The conceptual basis for calibrating activity monitors is depicted in Figure 4.

Figure 4: Conceptual Basis for Calibrating Activity Monitors



With a simple linear equation, algorithms can theoretically be developed to enable direct estimation of VO_2 or energy expenditure and categorization into LPA, MPA, or VPA. However, research has demonstrated that simple relationships are inadequate to capture the array of different activities performed under free-living conditions.[36] This is because the relationship between movement counts and energy expenditure varies greatly depending on the type of activity performed. The use of multiple equations or more complex pattern recognition approaches are now increasingly common for calibration purposes, but applying these methods requires additional expertise because the methods are not built directly into the software.

Some examples of more advanced methods will be introduced in later sections of this Guide (see Section 9) to facilitate additional exploration, but it is important to first have a basic understanding of the calibration process and associated statistics because these values are reported in papers highlighted in the Measures Registry. Common statistical indicators used to evaluate the resultant accuracy of calibrated physical activity measures include the test of mean differences (e.g., paired t-test), the Bland Altman procedure, and the Standard Error of Estimate (SEE). However, a detailed review of these terms is beyond the scope of the Guide.

Understanding Reliability and Validity

Regardless of the instrument or method used to assess physical activity behaviors or movement, users are often (and must be) concerned with the reliability (consistency) and validity (truthfulness) of the obtained measures. The distinctions of these two indicators are described along with statistics used to express them.

Distinctions and Definitions

Reliability refers to the consistency with which something is measured but it can be examined in several ways. For example, one might consider consistency of a response at a given point in time (e.g., *How consistently does a person respond to the same question if it is administered on two occasions?*). This would be analogous to having a person complete a physical activity questionnaire twice with a 30-minute interval in between. The comparison of the scores would reveal the extent to which the physical activity measure can provide similar and consistent information about activity levels. Reliability in this context is rarely assessed in physical activity assessment research because any short gap between two assessments will be confounded by memory. In other words, individuals are likely to remember what they answered and replicate their responses when asked to respond to the same questions. Alternatively, one can think of reliability across a longer period of time (e.g., *How consistent are individual physical activity behaviors performed across a full week?*) and avoid the short-term bias. This latter reliability, often referred to as stability reliability, provides information about the stability of the measure across a longer period of time. It is relatively easy to interpret the stability of a measure assessing a relatively stable (trait) characteristic, but it is challenging to evaluate and interpret in the context of physical activity assessment. The assumption of stability reliability in measuring physical activity is confounded by changes in physical activity patterns that occur from day to day, morning vs. afternoon, and weekday vs. weekend. As described in Section 3, children and youth have very particular movement patterns and, therefore, any measure is highly susceptible to low stability reliability indices. This has considerable implications because a low index of stability reliability is more likely to reflect variability in behavior rather than the properties of the assessment tool. Thus, it can prove difficult to separate out the reliability of the assessment tool from the reliability of the behavior.

Reliability is evaluated using interclass comparisons (based on the Pearson Product Moment correlation) or intraclass comparisons (based on analysis of variance). Using the interclass reliability is somewhat restrictive because it is limited to two points in time and does not take into account changes across time. It is important to note that the interclass reliability coefficient can be perfect even if the measures being compared are constantly changing. For example, if all participants increase their self-reported physical activity by about 30 minutes a day, the stability reliability for a full week would be very high. However, this example does not actually show stability (i.e., consistency) in the behavior but instead shows a systematic change in physical activity levels across the measurement period. The intraclass method is more robust and can examine consistency across multiple measures or over multiple days. The intraclass (alpha coefficient) reliability permits a more accurate estimate of the reliability. The intraclass reliability coefficient also is used to estimate the internal consistency reliability of a questionnaire or survey. The internal consistency reliability does not mean that the instrument is necessary reliable (consistent) across time. Rather, it means that the items on the instrument generally tap the same construct (i.e., measure the same outcome). This type of reliability is also very popular in social sciences but may have limited utility in the context of physical activity assessment. Again, consider the example of a questionnaire that asks about activity in different contexts (e.g., recess, physical education, after-school). Activity levels at each of these settings will vary and it is possible that a child would report low levels of physical activity at recess and after-school but indicate high amounts of physical activity at physical education. These various scores across different contexts (and items) would result in low internal consistency and indicate that the items do not assess the same construct when, in fact, they do. Depending on the measure being tested, the general use of reliability in physical activity research is more useful to assess variability in physical activity behavior and not the ability of the specific measure to provide consistent scores. For example, the agreement between two observers when coding observed physical activity behavior would be an important test of inter-rater reliability for the observation method and would indicate consistency of scores across testers or coders. Regardless of the type of reliability coefficient calculated, the range of possible reliabilities is 0 to 1.0, with higher values indicating greater consistency between the various measures. The Standard Error of Measurement (SEM) is often reported to reflect the degree to which a score might vary due to measurement error.

Validity refers to the *truthfulness* of the measure obtained. A measurement tool can result in reliable information but the data may not truthfully reflect the reported amount of physical activity behavior or movement. Validation of physical activity measures is typically accomplished with concurrent procedures in which a field or surrogate measure is compared with another more established or criterion measure. As shown in Figure 3, criterion measures are often used to calibrate monitor-based measures and these, in turn, are often used to calibrate report-based measures. From a validity perspective, the self-report physical activity measures are compared to this criterion to provide evidence of the truthfulness of the reported physical activity behaviors. These issues are discussed in Section 5. The Standard Error of Estimate (SEE) is often reported to reflect the degree to which an estimated value might vary due to measurement error.

Interpreting Reliability and Validity Statistics

Reading and interpreting the reliability and validity results in research reports can be difficult. Authors often make little reference to their procedures (other than to say, *"the results were reliable"* or *"the measure has been validated"*). Therefore, it is important to carefully review the procedures used to support reliability or validity of a specific measurement tool or process. The Measures Registry provides brief summaries of reliability and validity statistics but it is important to carefully review the actual study before determining whether it will have utility for a specific application. A number of different indicators are used to report reliability and validity; however, for the aforementioned reasons we will focus on the most popular indicators for validity. Table 1 provides a summary of common research statistics used to evaluate and report validity. Most would also be useful to determine reliability but the types of statistics will depend on whether the measures are continuous scores (e.g., steps per week) or categorical (e.g., achievement of physical activity guidelines).

When evaluating research findings on different physical activity methods, it is important to carefully consider the actual strength of the associations and not simply the "statistical significance." Research studies tend to focus on statistical significance to determine whether a value is reliable or valid, but significance is directly influenced by a number of factors, including sample size. To avoid over-interpreting findings, it is important to evaluate the absolute agreement and outcome measures rather than just the statistical significance. For example, focus should be on the magnitude of a Pearson Correlation Coefficient rather than the significance. Traditional interpretations characterize correlations below 0.4 as being low, 0.4 to 0.8 as moderate, and above 0.80 as being high. Validity indices of most report-based measures are below 0.4 (i.e., low), while values for monitor-based measures are at the high end of this scale (i.e., 0.6 to 0.8 or higher).

With validity statistics it is also important to keep in mind that the reported relationships are typically based on aggregated data from multiple people. This makes sense from a sampling perspective, but accurate group-level estimates of physical activity do not necessarily translate to accuracy for estimating individual physical activity levels. A useful indicator that captures the error for individual estimation is the Mean Absolute Percent Error (MAPE). As the name implies, the MAPE value reflects the mean absolute difference in outcomes and is computed by first calculating the absolute value of individual difference scores and then averaging them. This provides a more appropriate (and conservative) indicator of actual error for individual estimation because it captures the magnitude of

Table 1: Summary of Validity Indices Used in Physical Activity Research

CONTINUOUS VARIABLES	Pearson Product Moment	Measure A and Measure B are moderately and positively correlated (r = 0.50)
	Test of Mean Differences	Measure A and Measure B differ by a minimum amount and the difference is not statistically significant
	Bland Altman (Limits of Agreement)	The estimates of Measures A are within -452 and +560 kcal of that obtained by Measure B
	Standard Error of Estimate	Estimates obtained from Measure A are ±10.0 units apart from those of Measure B
CATEGORICAL VARIABLES	Proportion of Agreement	Measure A and Measure B agreed on 85% of their classifications of individuals meeting/not meeting PAG
	Kappa	The agreement among classifications of active individuals obtained from Measure B was good (kappa = 0.75)
	Sensitivity	Approximately 67% of the active individuals (Se = 67.0%) as indicated in Measure A were correctly identified by Measure B
	Specificity	Approximately 74% of inactive individuals (Sp = 74.0%) were correctly classified by Measure B

both overestimation and underestimation. Studies that report mean error do not capture the actual amount of error because overestimation and underestimation may average out.

Consider a very simple example of two children who have energy expenditure levels estimated with both a physical activity questionnaire (PAQ) and the doubly-labeled water (DLW) method. Assume that Child 1 had a DLW estimate of 2600 kcal/day and a PAQ estimate of 1100 kcal/day and that Child 2 had corresponding values of 900 kcal/day for DLW and 2000 kcal/day based on the PAQ. The average value from these two individuals is 1750 kcal/day for DLW and 1550 kcal/day for the PAQ. The group difference in the estimates is only 200 kcal/day (approximately 11% error) and may not be statistically significant. However, the group means mask the large individual error of 1500 kcal/day for Child 1 (i.e., underestimation) and 1100 kcal/day for Child 2 (i.e., overestimation). The absolute error for the individual estimates are 58% for Child 1 and 122% for Child 2 (leading to a computed mean value or MAPE of 90%).

Group means are typically evaluated in traditional validation studies, but it is important to not assume that this level of error would hold for individual estimates, as the example illustrates. A method may be reported to provide "valid" estimates for group-level comparisons, but error can be considerably higher when attempting to estimate values for individuals. Measures also can differ greatly at the individual level, yet it is possible (and common) for the measures to be moderately correlated (e.g., r = 0.40) and for group differences to be relatively small. Understanding the difference between individual- and group-level estimates is key to this distinction.[d]

A final point about interpreting validity statistics is to understand that significant associations do not imply equivalence or agreement. The scenario introduced above is further expanded to illustrate this. Envision that the estimates obtained from the DLW method are compared with two different physical activity questionnaires (PAQ 1 and PAQ 2) and that PAQ 2 provides estimates of energy expenditure that are twice as high as those obtained from PAQ 1 (see Table 2, which is adapted from Zaki et al.[37]) The average daily estimates of energy expenditure for DLW, PAQ 1, and PAQ 2 are 1712.8 kcal, 1733.5 kcal, and 3467 kcal, respectively. The PAQ 1 and PAQ 2 estimates are both strongly correlated with DLW (r = 0.88), perhaps indicating similar "agreement." However, the estimates from PAQ 1 differ from the DLW method by 20.7 kcal (at the group level) while estimates from PAQ 2 differ by 1754.2 kcal. The PAQ 2 assessment substantially overestimates energy expenditure even though it is highly correlated (r = 0.88) with the outcome measure. This example illustrates the limited value of correlations because it shows the inability to infer the presence or absence of systematic differences between two estimates. Measures that are correlated may be sufficient if the goal is to simply distinguish the most active individuals from the least active. However, in other situations, it may be important to be able to estimate the actual amount of physical activity performed or the overall energy expenditure. Therefore, care must be used when interpreting outcomes validity statistics from studies summarized in the Measures Registry or in other publications.

d Obtaining precise estimates of individual physical activity and energy expenditure is extremely challenging and is an unrealistic goal for most field-based measures. Therefore, it is important to recognize the limitations of the estimates, particularly when extrapolating or reporting estimates at the individual level. Parallel progress is being made with reducing error in monitor-based measures and developing and refining measurement error models for report-based measures.

Table 2: Example of Agreement in Estimates of Energy Expenditure (kcal/day) Using the Pearson Correlation

CHILD	DLW	PAQ 1	PAQ 2 (DOUBLE PAQ 1)
1	1350	1100	2200
2	1708	1650	3300
3	2340	2700	5400
4	1200	1340	2680
5	1090	980	1960
6	1670	1920	3840
7	2460	2060	4120
8	1320	1580	3160
9	2000	2125	4250
10	1990	1880	3760
Average EE	**1712.8**	**1733.5**	**3467**
Pearson Correlation		**0.88**	**0.88**

Source: Adapted from Zaki et al.[37]

Advanced Concepts in Measurement Research

Regardless of reliability and validity, it is important to remember that data collected from all measurement tools are *estimations* of the behavior or movement. Measurement errors cannot be avoided, but researchers must distinguish between systematic and random errors as well as the sources of error in research and evaluation. In considering effects of error, researchers should consider the relative importance of obtaining estimates of individual behavior or whether group level estimates are sufficient. As previously illustrated, data reported at the individual level will generally have more error than aggregate summary values that reflect the means of a sample of individuals. In many situations, group-level estimates are all that are needed, but it is important to minimize and control sources of error when possible.

Measurement Error

A key challenge in research is to minimize not only error in the measure itself (i.e., closely related to validity as described above), but to control for error in the measurement procedure. Overall, error can be attributed to multiple sources and divided into specification error, measurement error, and processing error.

Specification error occurs when the measurement method does not match the actual concept being studied. This is hard to avoid in physical activity research because the specific "construct" investigators are trying to capture often cannot be directly observed. For example, researchers are frequently interested in evaluating "typical" or "habitual" physical activity, but existing methods may capture only a few isolated days or a week and therefore do not directly capture the construct.

Measurement error refers to error that arises in obtaining a response or a measure during data collection. This occurs in research because physical activity is inherently difficult to assess and because our methods have important limitations. Examples include limitations of technology in capturing movement, mistakes in recording data, respondent errors in reporting or recalling data, and deliberate or subconscious bias in the reported data. Measurement error is generally assumed to be more of a problem in report-based measures due to their inherent subjectivity. However, equally challenging forms of measurement error complicate assessments with various monitor-based methods.

Processing error must be considered when manipulating raw data to produce outcome or summary variables. It occurs because it requires the researcher to summarize, aggregate, and categorize the data, and these steps necessitate a number of assumptions. Examples include assigning METs to process data or using algorithms or prediction equations to convert movement data into outcomes.

Being cognizant of the types of errors can help in designing more effective studies and in appropriately interpreting outcomes.

Statistical Power

Statistical power refers to the concept of identifying relationships between variables when such a relationships truly exist. The essence of inferential statistics is to investigate the relationship between variables. Statistical power is a key concern in intervention studies where the goal may be to determine whether a particular program or treatment increased MVPA in children. The ability to detect an association (i.e., have statistical power) depends on a number of factors, but the accuracy (i.e., validity) of the measurement instrument is one of the most important. With inaccurate measures it is possible for a researcher to incorrectly conclude that there is no relationship between the intervention and the intended outcome simply because the MVPA outcome is not accurately assessed. For example, think of an outcome (e.g., cardiovascular disease) and an exposure variable (e.g., participation in physical activity) that are moderately correlated. The sample size required to find this association with a 95% statistical power might increase by two- or three-fold depending on the accuracy of the measure selected to assess physical activity. This illustrates how the accuracy of the measure can have important implications for study design by affecting statistical power. Technically, the lack of accuracy of a measure and inability to detect a true association is referred to as an error of the second type (i.e., a Type II error). A Type II error can be made for a variety of reasons but a main one is a lack of a reliable measure that reduces the statistical power to identify a relationship. A detailed description of the interaction between statistical power and validity is beyond the scope of this Guide, but researchers and practitioners should be aware that error, as defined based on the previous descriptions, of validity reduces power and makes it difficult to detect associations when they actually exist.

Concepts of Equivalence

A common need in measurement research is to evaluate the equivalence of different measures. This is important for all calibration applications discussed previously and is typically expressed as a measure of group-level agreement. In many cases, a more practical field-based measure is evaluated to determine whether it agrees with a more expensive and accurate lab-based measure. For example, monitor-based instruments are often calibrated to estimate energy expenditure, so it is important to evaluate whether the procedure can produce group-level estimates that are statistically equivalent to a criterion measure obtained from indirect calorimetry or DLW. Measurement researchers often evaluate agreement between measures with the graphical Bland Altman method[37-38] but this method does not provide a way to empirically (or statistically) evaluate agreement. Standard statistical tests (e.g., ANOVAs or t-tests) are often used to compare measures but these are designed to test differences and are fundamentally flawed for evaluating agreement. Because the significance is influenced by sample size, studies with larger and more robust samples are more likely to detect significant differences than studies with fewer participants, regardless of the size of the difference. This issue relates to the concept of statistical power described above and has led many studies to erroneously conclude that two measures are equivalent when a non-significant difference is detected (or that two measures are not equivalent when a difference is significant). This is analogous to an intervention effect that is likely to be significant if the sample being assessed is considerably larger (i.e., great degree of statistical power). Therefore, working with larger groups is often desirable. In the context of physical activity assessment, a large sample will have the opposite effect and improve the likelihood to detect small differences between two measures even though these might not be relevant (i.e., not meaningful; not clinically important). This would artificially minimize the chances that the measure being tested is not deemed similar to the criterion assessment of physical activity. Alternative methods of "equivalence testing" avoid the inherent limitations of traditional statistical tests of differences and are more appropriate for evaluating agreement.[e,39] This approach has not been widely adopted so it is possible for studies to erroneously conclude that measures agree or that they are valid. The main message is that it is important to understand the limitations of traditional statistical methods when evaluating agreement between alternative measures.

e According to classic hypothesis testing for differences, the null hypothesis is that the differences between the two measures are not equal to "0" (i.e., differences are "non-random"). This is a rather unrealistic scenario, as no measure will result in identical values when compared to another. Equivalence testing flips the null hypothesis and this enables zones of equivalence to be established a priori and to be tested statistically. Details about equivalence testing methods are beyond the scope of this Guide, but the main point is to understand that equivalence testing is a more appropriate way to examine agreement than are standard tests of differences. Considering this issue when reviewing instruments and outcomes can help to facilitate interpretations.

5

Overview of Individual Physical Activity Assessment Tools

A number of techniques and tools are available to assess physical activity in children and youth, each with advantages and disadvantages. This section provides a brief description of physical activity measures while establishing an important relation between feasibility and validity. Many excellent review papers have already summarized the utility of various assessment tools[1,4,8,40,41] so this section emphasizes the considerations that are important when comparing tools described in the NCCOR Measures Registry.

The Feasibility/Validity Continuum

A common method for comparing measures is to determine their validity in relation to their feasibility. This relationship is illustrated in Figure 5A.[f] Measures that are highly valid are often too expensive or cumbersome to use in large-scale research applications. Measures that are more feasible to use may have lower validity. For simplicity, the measures are aggregated into three main categories (report-based measures, monitor-based measures, and criterion measures). The feasibility/validity continuum has been developed and discussed for measures of energy expenditure. However, there is a growing realization that energy expenditure alone does not fully capture other important aspect of physical activity, such as strength, flexibility, or other behaviors.[42] Bowles et al. have discussed this issue in the context of measurement challenges related to distinguishing between behavior, activity, and motion. They describe "motion," which can be captured by devices such as accelerometers; "activity," which represents a class of actions such as cleaning house or playing soccer; and "behavior," which is a specific action embedded in the activity, such as standing in the goalie box or wiping the windows with a paper towel.[6] More work is needed to explore the feasibility/validity continuum for measurements of behavior and activity, but clearly it will differ from the continuum for estimates of energy expenditure. For example, DLW has essentially zero validity for measuring specific activities, as it measures aggregate energy expenditure over one or more days. Figure 5B illustrates the feasibility/validity continuum for measures of behavior.[9] Note the distinct differences in the criterion measures and the importance of physical activity recall and time use methods.

Report-based measures include self-report tools and diaries that are designed to provide subjective information about physical activity levels and the context of physical activity behaviors. These measures tend to be the most feasible to use (due to both lower administrative and processing costs) but also tend to have lower validity. The choice of measurement approach for a given application depends to a great extent on the relative importance of feasibility and validity, but a number of other factors also come into play when selecting a measure.

Monitor-based measures include various devices designed to objectively quantify movement, such as accelerometers, GPS units, and pedometers or devices to measure the intensity and duration of physical activity such as heart rate monitors. These measures have a good balance between

f The original depiction of this relationship is in a book edited by Dr. Tom Rowland (Rowland TW. Aerobic fitness. In T. W. Rowland, ed. Developmental Exercise Physiology. 2nd edition. Champaign, IL: Human Kinetics, 2005. P. 89-108.). The concepts have been adapted and used in different ways in this Guide to characterize the relationships among the measures and the distinctions among the three main classes of assessments.

Figure 5A: Physical Activity Assessment Tools and Their Relative Positions on the Feasibility/Validity Continuum

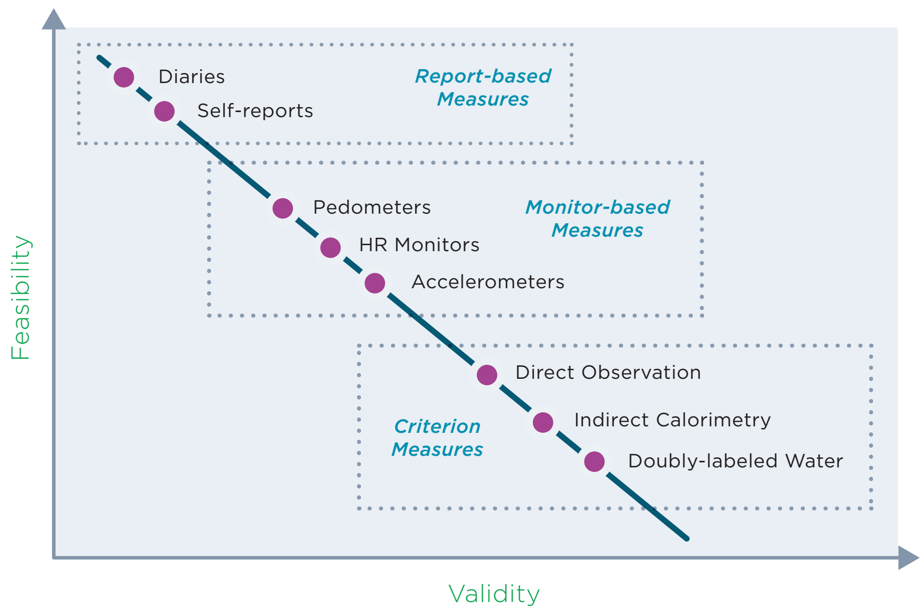
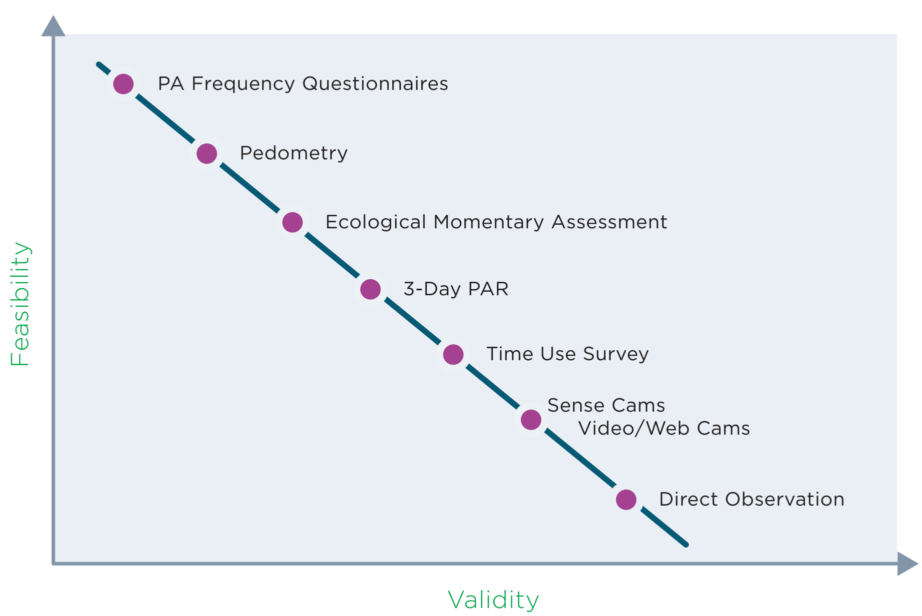


Figure 5B: Feasibility/Validity Continuum for Physical Activity Behavior[g]



g Figure 5B has been developed by David Berrigan and Richard Troiano for the purposes of this User Guide.

feasibility and validity, making them attractive for a number of research and evaluation applications.

Criterion measures include the DLW method, indirect calorimetry, and direct observation. These measures provide criterion estimates of energy expenditure and movement and are typically used for validation studies, smaller applications, or in lab-based study designs where precise indicators are needed.

The above assessments apply to efforts aimed at measuring energy expenditure from physical activity. This is likely to be important for many health studies because of the links between level of energy expenditure and diverse health and energy balance-related benefits of physical activity. However, studies focused on the social, strength, or flexibility benefits of physical activity may require measures selection from a distinct feasibility/validity continuum. More work is needed to better categorize measures as focusing on one or more of these three aspects of physical activity: behavior, activity, and motion. In any case, a critical step in measures selection is careful thinking about which aspect of physical activity needs to be measured.

As previously described in Section 3 and Section 4, measures obtained from participants are often raw movement or raw reports of behaviors performed. To enable the raw measures to be of value for physical activity research, they are typically calibrated against other more valid "criterion" measures. Thus, a key goal of calibration is to minimize error in the estimates of more feasible measures and to make physical activity values more interpretable. The most precise estimate of energy expenditure is known to be DLW. However, this is a very expensive method and not practical for calibration applications. Therefore, researchers have more commonly used other measures of energy expenditure such as indirect calorimetry. Monitor-based measures, for example, are commonly calibrated against indirect calorimetry systems to establish the relationship between movement and energy expenditure or exercise intensity. This process typically involves having individuals perform a series of different activities while being simultaneously assessed with both a monitor-based measure and indirect calorimetry (criterion measure). Direct observation is not as commonly used for calibration, but it is often used for validation purposes to test the classification of observed behaviors.

Although most work has focused on calibration of activity monitors, estimates obtained from self-report measures also need to be calibrated. Most, if not all, report-based measures are designed to capture free-living activity. Therefore, calibration of these measures requires a criterion measure that is also designed to capture activity in a variety of contexts. Indirect calorimetry is not well-suited to calibrate these measures, so monitor-based measures are typically the best option for calibrating report-based measures. The criterion measure in this case still has considerable error, so calibration of report-based measures will only be as accurate as the underlying criterion measure to which it was related. New methods and approaches offer promise for improving calibration methods and these are briefly introduced in Section 8.

A detailed explanation of calibration research is beyond the scope of this Guide, but it is important for researchers to have a conceptual understanding of the calibration process because it provides the foundation for how measures are used to "assess" physical activity behavior. The description of each tool will be presented in relation to its role in calibration and with reference to the inherent balance between feasibility and validity. The criterion measures will be introduced first, followed by monitor-based measures and report-based measures.

Summary of Assessment Tools

The following sections describe the major types of assessment tools used to measure individual physical activity.

Report-based Measures

Report-based measures described in this Guide include various self-report surveys, such as physical activity questionnaires and diaries, that capture a participant's perception and interpretation of physical activity behavior. These tools also can be defined as subjective measures because they rely on the person's ability to interpret and recall physical activity and are generally categorized by mode of administration: self-administered or interview (most are self-administered). Recall-based self-reports may prompt users to recall time periods ranging from 1 day to 3 months. The time needed to complete the questionnaires may range from 1 to 20 minutes, with most of the self-report measures requiring less than 10 minutes to be completed. Diaries have very different properties and characteristics because they generally require that a person records the activity performed throughout the day or right after it occurs. The level of detail varies with the instrument and a number of logging tools are now available for real time tracking with cell phone applications (see Section 8). Regardless of form, the information collected from self-reports and diaries is often used to convert to measures of energy expenditure (e.g., kcals) and both can provide information on frequency, intensity, duration, and type of physical activity and also include context of physical activity (e.g., inside vs. outside). The frequency and duration of the activities reported can be useful in determining compliance with physical activity guidelines and in computing volume of physical activity performed during a pre-defined time window.

Four major categories of self-report assessments are records or logs,[h] recall questionnaire, quantitative history,[i] and global self-report.[j] Recall questionnaires tend to be used more often than other types of self-report assessment, and examples for youth include the Previous Day Physical Activity Recall (1d-PAR), the 3-day Physical Activity Recall (3d-PAR), and the Youth Activity Profile (YAP). A major limitation of recall questionnaires is the level of subjectivity involved in the estimates provided and challenges imposed on participants as a result of the recall process (this is particularly true in youth, as described in Section 3). Another challenge is the limited utility of questionnaires for directly estimating time in physical activity or in computing total daily energy expenditure values (other than referring to standardized estimates of activities energy cost). Therefore, individual error tends to be very high. Individual error also often compromises the ability of a self-report to capture changes in physical activity that might occur over time or as a result of treatment effect. However, with calibration, it is possible to model the error from these tools and produce group estimates of physical activity that mirror estimates obtained from more accurate measures, such as activity monitors. Recall questionnaires are easy to administer and inexpensive and are the only field measure that can capture both the type and the context of physical activity. These particular attributes can offset the limitations described above. More detail on self-reports is available elsewhere.[10,30,33,43-48]

h Records or logs of physical activity can sometimes include diaries depending on the definition but are often placed into a separate category of self-reports (just as we did in this Guide). These include recording the frequency and/or duration of activities as they occur and providing comprehensive characterization of physical activity patterns. However, these are likely the least feasible method within self-reports as they place a great burden on individuals being assessed.

i Quantitative histories are typically long questionnaires (e.g., 50 items) that are designed to assess lifetime or long-term (e.g., over the previous year) physical activity patterns. These can provide a comprehensive characterization of physical activity and capture important dimensions such as duration and frequency. However, they are likely to have a considerable amount of error when compared to other categories of self-reports.

j These are very brief questionnaires (typically composed by a single item) that are designed to assess general physical activity levels and are often used to determine whether individuals meet or not a specific physical activity threshold, such as recommended guidelines. These types of self-reports provide limited characterization of physical activity levels as they do not ask about type, context, or patterns of physical activity.

Monitor-based Measures

Activity monitors: Activity monitors provide a good balance between accuracy and feasibility and therefore are widely used in contemporary physical activity research applications. They have historically been worn on the waist but recent trends have shifted to wrist-worn monitors.[k] Most devices use internal accelerometers to obtain an objective indicator of the amount of movement being performed. The resulting value has typically been called an “activity count,” which is a dimensionless value that is difficult to interpret because it has no real physical or physiological meaning. Activity counts have been calibrated to output meaningful outcomes such as energy expenditure and METs. These calibration equations (in addition to activity counts) typically use information, such as body weight, age, and sex, to predict energy expenditure and these estimates are often categorized into sedentary, light-, or moderate- and vigorous-intensity physical activity to determine the time spent in the different activities (e.g., percent time in moderate- and vigorous-intensity physical activity). The calibration process enables the monitors to evaluate compliance with physical activity guidelines, and overall, these tools can capture all domains of physical activity except for the “type” of activity being performed.[l]

Popular activity monitors include the Actigraph and the GENEactiv, but a number of devices are available and each has different features and capabilities. The price of monitors can range from $200 to $400 each and it may be necessary to use customized software from the manufacturer to process the data (at an additional cost). A few limitations to consider (other than the high cost) include the burden placed on participants, who often have

k The majority of work on accelerometry-based monitors has been conducted using waist/hip worn devices. However, investigators have moved toward using wrist-worn monitors. This transition has been fueled by the progression in consumer-based monitors as well as by evidence that compliance is enhanced when participants are asked to wear monitors on the wrist (more like a watch).

l New pattern recognition approaches have shown promise in detecting underlying movement patterns and classifying type of activities performed, but accurate detection of the diverse range of activities performed under free-living conditions remains elusive.

to use the monitor for long periods of time (e.g., 7 days or more). Additionally, monitors can place some burden on researchers and practitioners when extracting or processing the data. Another key limitation is that waist-worn monitors are not well-suited for capturing cycling, non-ambulatory movements, weight-bearing activities, or upper body activities (wrist-worn monitors can overcome some of these limitations). Despite these limitations, the objective nature of monitoring devices offers many advantages for field-based research, including the ability to detect the magnitude and temporal characteristics of the movement. Monitor-based tools continue to evolve along with methods for processing and calibrating the data from these devices. However, it is important to note that advances in technology and methods often come at the expense of feasibility. Recent trends in monitor technology are briefly described in Section 8 but readers are referred to other scientific papers for more detailed reviews.[42,49-53]

Heart rate monitors: Heart rate monitors are no longer as popular as they were in the early studies of physical activity behavior, but they are common in exercise training applications and for absolute determinations of exercise intensity. These measures capture the physiological response to movement in terms of heart rate (usually expressed in beats per minute) and reflect the level of stress imposed on the cardiorespiratory system. Early heart rate monitors relied on chest straps but contemporary monitors can measure heart rate through an optical sensor built into a wristwatch (i.e., wrist conductivity). These measures track the number of beats per minute (bpm), which is (assumed to be) linearly related to oxygen consumption. This relation serves as the foundation for existent calibrated heart rate measures developed to provide estimates of energy expenditure in kcal/day or kJ and that can be used to discriminate between different activity intensities.[m] Heart rate thresholds also have been used to determine the time or percent of time that individuals spend above pre-defined intensity levels (e.g., 140 bpm indicating moderate intensity). Heart rate monitors can assess frequency, intensity, and duration of physical activity, but like activity monitors, they provide no information about type or context of physical activity.

A variety of heart rate monitors are available on the market and the cost is typically $100 or higher depending on additional features and design. The key limitations of early heart rate monitors included the lack of accuracy to discriminate intensity at the lower spectrum of activities and the susceptibility of heart rate to factors other than movement (e.g., ambient temperature, level of hydration, anxiety). Newer heart rate monitors (placed on the wrist) rely on optical sensors to measure physical activity, but these alone have been deemed inaccurate during more vigorous exercise (e.g., running at a speed >6.0 mph). For these reasons, heart rate monitors are more commonly used in combination with other measures (e.g., activity monitors) or for controlled laboratory-based studies. Heart rate monitors are particularly useful for monitoring activity associated with non-ambulatory activities, such as cycling or swimming, and for evaluating individual responses to physical activity. A unique advantage is that heart rate monitors objectively capture the relative level of stress. The thresholds used to determine relative intensity are usually determined based on individual calibrated heart rate values that take into account resting or maximal heart rate (i.e., expressed as percent of maximal heart rate or percent of heart rate reserve).[54-58]

Pedometers: Pedometers are objective monitoring devices designed solely to quantify the number of steps performed as an indicator of movement. Pedometers have evolved substantially, but early models of pedometers used a horizontal lever-arm or piezo-electric mechanism that captured vertical accelerations of the hip. Many of the most recent models now use accelerometers to detect the number of steps. Pedometers track vertical hip movements that are recorded as steps and the data usually can be stored for 24 hours or several days depending on the pedometer. Pedometers can capture the frequency of movement (i.e., number of steps) but also are able to produce estimates of the distance covered (i.e., number of steps X individual stride length). The ability to predict energy expenditure is limited but some device-specific algorithms have been developed for this purpose. The number of steps accumulated is usually expressed per day, and recommended values for youth can range between 10,000 to 15,000 steps/day depending on age and sex. These recommendations often lack a criterion outcome to define sufficient steps for health. However, some of the guidelines have been validated to differentiate between youth of different body weight status while others have been generated to reflect an equivalent of 60 minutes per day of MVPA.

m Calibration equations are generally based on the assumption that heart rate is linearly related to energy expenditure. This assumption is particularly true for moderate- to vigorous-intensity activities. However, the assumption might not hold across sedentary and light-intensity activities.

Many brands of pedometers are available ranging in price from $10 to $200 each depending on the features and memory capability. The key limitations of pedometers include the inability to capture non-ambulatory activities (e.g., cycling), and the level of inaccuracy when predicting energy expenditure. However, the key advantage is the ease of use and the reliability and validity for estimating steps accumulated (at walking speeds) during the day. Steps can provide a good indicator of overall physical activity patterns among youth because a substantial portion of their physical activity derives from lower-body, locomotor movement. Pedometers also are particularly useful as a motivational tool and therefore are widely used in physical activity promotion studies. More detail is available elsewhere.[50,59-64]

Criterion Measures

Doubly-labeled water (DLW): Doubly-labeled water is the most accurate measure of total energy expenditure and allows an activity's energy expenditure to be determined if estimates for the thermic effect of food and resting energy expenditure are available. Total energy expenditure as measured by DLW is determined by evaluating the metabolic breakdown of two stable isotopes (deuterium [^{2}H] and oxygen-18 [^{18}O]) over time. The isotope-labeled water is administered orally using standardized doses depending on the individual's total body water. The evaluation requires between 7 and 21 days with the traces of the isotopes obtained through sequential urine samples. The rate of depletion of the isotopes is used to estimate total carbon dioxide produced over time and ultimately a calculation of total energy expenditure. This technique is extremely expensive and requires advanced expertise to handle both the measurement protocol and data processing, and estimates are limited to total energy expenditure. Thus, this tool is not able to assess other dimensions of physical activity, such as intensity, duration, frequency, and type. Despite these limitations, DLW provides the most accurate measure of total energy expenditure and is particularly useful for measurement protocols aimed at providing a summary measure of overall free-living energy expenditure.

Calorimetry: Calorimetry is a method based on the measurement of heat released due to the chemical processes occurring when metabolizing different body substrates (e.g., carbohydrates, fat, or protein). The resulting breakdown of energy associated with these chemical processes can be inferred by determining the amount of heat released from the body, and using either direct or indirect calorimetry. Direct calorimetry involves the direct measurement of body heat released to the air and requires the use of a room calorimeter (also known as heat chamber) so it is not commonly used. Indirect calorimetry is a widely accepted and more practical alternative tool for the measurement of energy expenditure. It provides an estimate of heat produced based on the relation between oxygen consumed and carbon dioxide produced, typically referred to as the respiratory exchange ratio. The method relies on the assumption that 1 liter of consumed oxygen is equivalent to known amounts of kcal depending on the substrate being metabolized. For simplicity, it is often assumed a respiratory exchange ratio of 1.0 and a caloric equivalent of 5.0 kcal per liter of oxygen consumed. Measuring energy expenditure using indirect calorimetry is commonly performed using laboratory oxygen and carbon dioxide gas analyzers, or a portable gas exchange/analysis system.[n] This method is commonly used as a criterion measure to establish relationships between movement and estimates of energy expenditure from monitor-based tools.[65-66]

Direct observation: Direct observation is considered to be a gold standard method of physical activity assessment because behavior is directly observed. Observation typically involves the choice of a participant to observe (because it is not possible to observe all participants at the same time), when to watch (because it is not practical to try to observe continuously for extended periods), and how to record the behavior (record every single behavior once it occurs or record if the behavior lasts for a pre-defined amount of time). Technical considerations when using direct observation include: (1) the definition of physical activity behaviors to be recorded, and (2) the selection of the most appropriate behavior recording technique. Additional considerations include the selection of the observation pacing method and the choice of software to record and analyze the data. The behaviors of interest also should be carefully defined and organized into classes of mutually exclusive behaviors. Examples of behavior classes observed include various postures (i.e., lying down, sitting, standing, and walking) or activity intensities (i.e., sedentary, light, moderate, and vigorous). With observation, it is also possible to determine time spent in a specific posture and assign an intensity category to the posture being coded. This method can be of great value when understanding behavior because

n The portable systems use the same principles as metabolic carts but require participants to wear a backpack-type harness that holds two light-weight sensors (O_2 and CO_2) and transmission modules secured to the body that enable estimates of oxygen consumption and carbon dioxide production to be sent to and displayed on a laptop computer.

environmental factors (i.e., the context of the behavior or movement) also can be assessed. Depending on the observation method, it is possible to accurately classify the type, intensity, duration, frequency, and context of activities performed.

Some examples of direct observation instruments include the System for Observing Fitness Instruction (SOFIT), Behaviors of Eating and Activity for Children's Health: Evaluation System (BEACHES), System for Observing Play and Leisure Activity in Youth (SOPLAY), and System for Observing Play and Active Recreation in Communities (SOPARC). The use of standardized procedures during observation provides good objectivity, but observation of behaviors always involves some degree of subjectivity and can impose a high experiment burden in terms of cost and time. Software is available to facilitate recordings and tracking of data but the method requires time, expertise, and practice. Overall, the direct observation method is considered to be an appropriate criterion measure of physical activity if conducted using standardized procedures and trained observers. Often, it is the only way to directly understand the context of behavior, but advances in video-based methods and ecological momentary analyses provide alternative views of behavior.[o] Readers interested in direct observation are encouraged to refer to definitive technical guides for the methods and tools.[67-68]

o Ecological momentary assessment techniques provide a way to capture context of behavior. Using text messages and smartphone prompts, it is possible to capture information about the type, intensity, purpose, or setting of activity. Ecological momentary assessment offers many advantages for physical activity research but it has a number of logistical and assessment challenges. (See references 69-71).

6

Selecting Measures

This section challenges readers to think specifically about key features of a research study design and understand how those will influence the choice of a measure. The variety of physical activity measures available can be overwhelming and, depending on the features of the project, researchers and professionals must carefully consider the advantages and disadvantages of each measure. To help select a measure, investigators should consider the features of their project.

This Guide emphasizes four features: (1) the population being assessed (i.e., children, adults, older adults), (2) the activity outcome (e.g., leisure-time physical activity or school-based physical activity; physical activity volume or physical activity frequency), (3) the research type (e.g., intervention study vs. surveillance), and (4) the resources available (e.g., existing personnel in the project, timeline for data collection, expertise required to handle the data, immediate feedback required).[p] These four key considerations (*Population, Activity outcome, Research Type*, and *Resources*) are abbreviated with the acronym PARR, and each is described below. The description of the PARR uses the specific example of report- vs. monitor-based measures because these two measures are the most commonly used in studies of physical activity.

Population

The nature of the population is perhaps the most important consideration in selecting an instrument and that is why it is included as a key filter in the NCCOR Measures Registry. The selection of a specific age group (preschool, elementary, adolescent) will immediately facilitate the filtering choices available for review. Tools are often used across all ages but need to be adjusted to meet the needs of the population of interest. As described in Section 3, children and youth have unique physical activity patterns, cognitive skills, and metabolic responses to physical activity that have to be considered when deciding on what measures to use.

The impact of age on measure selection is well illustrated when choosing between report- vs. monitor-based measures. For example, report-based measures are typically ineffective for preschool children because they require cognitive skills that are still immature at this stage of life. Proxy measures that rely on parents to either assist in the completion of the surveys or report their perceptions about their child's physical activity levels can provide an alternative report-based approach for this population. Although this is a viable method, parental perception about children's physical activity may be biased, which can add substantial error to these estimates. The utility of report-based measures increases with age due to more advanced cognitive abilities to recall and report past events. Alternatively, monitor-based measures of physical activity avoid issues with subjectivity but researchers and practitioners need to carefully consider other limitations of these approaches for children. Pedometers, for example, are easy to use but children are more likely (than adolescents) to accidentally reset the device, which deletes the data. A more technical limitation relates to the inability of pedometers to account for differences in leg length, which can complicate age-related comparisons when interpreting overall amounts of physical activity. Activity monitors are easy to use but may offer practical limitations and may not provide the same type of feedback (thereby influencing compliance). An important

p This list of considerations is aligned with the steps proposed by Strath et al. (see reference 3) and colleagues, but differs in some respects. An additional consideration of "Population" was added because issues with assessments vary greatly by age and other demographic factors. Several of the steps proposed by Strath and colleagues also were combined to facilitate interpretation.

technical limitation includes the challenges in handling the large amounts of data and the implications of different decisions involved in data collection. There is a consensus, for example, that the use of a 60-second epoch will lead to substantial underestimation of activity levels in preschoolers and young children due to the sporadic nature of their physical activity patterns.[q,25] Readers are referred to Section 10, which provides supplemental information on youth.

Activity Outcomes

As described in Section 3, physical activity data must be operationalized so it can be scored and interpreted. Within the broad scope of the definition of physical activity introduced in this section, researchers and practitioners should identify what activity outcome is of most interest by selecting both the domain (e.g., school-based physical activity) and dimension (e.g., intensity) of physical activity being studied. The choice of domain will require that researchers and practitioners select a measure that can capture the appropriate physical activity contexts (e.g., during commuting, recess, physical education, or after-school depending on whether it is leisure, school-related, or other). Most tools are flexible enough to capture this information depending on the data collection protocol used.[r] Therefore, this section focuses on issues related to the dimension of physical activity while providing examples of the implications of using either report- or monitor-based measures of physical activity.

In studies related to obesity, the total volume of physical activity (or overall energy expenditure) is usually of interest because it allows the researcher or practitioner to examine energy balance. This indicator results from the combination of frequency and intensity and is highly comparable across different studies or projects. If the volume of physical activity needs to be determined, then the measure selected needs to capture both frequency and intensity of the activity performed over the time period defined by the researcher or practitioners. For example, report-based measures could be a good choice if items ask about the activities performed, duration, and frequency of participation. From the type of activities performed, it is possible to infer the intensity of the activity, and convert to an associated energy expenditure equivalent (i.e., METs). The duration and frequency can then be multiplied to obtain total energy expenditure or physical activity reported by the child (e.g., MET-minutes per week, Kcal/day). The challenge with this approach is that it often relies on absolute estimates of energy expenditure, which have some important implications (these have been described in Section 3). Alternatively, monitor-based measures, such as activity monitors, can store recorded movement over several days and weeks and can capture both intensity and frequency of events. Validated calibration equations can be used to convert recorded movement into estimates of energy expenditure.[72]

Research Type

The research type inherently dictates the relative needs with regard to feasibility versus validity. The feasibility portion of the continuum described in Section 5 is influenced by the funding available for the study or project (i.e., money available to buy either expensive or affordable tools), the necessary sample size (e.g., number of measures needed), and the level of burden placed on the child as a result of the assessment (i.e., time and effort required to comply with the assessment). These factors need to be weighed in relation to the accuracy of the measure. The general considerations with each type of behavioral epidemiology research (see Figure 2) are summarized below. Although the behavioral epidemiology framework includes five categories, they are summarized in three major divisions here.

Basic Research and Health Outcomes Research

Measures of physical activity typically serve as independent variables in basic research and health outcomes research. To elucidate mechanisms and understand health impacts, it is common to compare changes before or after training adaptations take place or between active or inactive groups. Measures used in these types of research often favor precision or accuracy over feasibility. Many designs are possible and the need for precision varies based on the application and outcomes of interest. Some lab-based studies may rely on criterion measures of indirect

q Assume that a certain activitiy monitor uses a cutpoint for MVPA of 2000 counts per minute. A child who is playing a tag game can accumulate 1200 counts in 30 seconds and then remain sitting or in a standing position for the other 30 seconds and therefore accumulate 0 counts during the remaining fraction of the minute. The aggregated counts for this entire minute (1200) would be less than the threshold (e.g. < 2000) and indicate that the child was not active during that minute even though half of the time was spent running. If the epoch was 30 seconds, the counts would exceed an adjusted threshold (e.g., 1000) and the same period would be categorized as active. The use of 1-minute epochs essentially "ignores" these shorter bouts of activity, resulting in underestimations of activity levels in children.

r Existing self-report measures tend to overlook important physical activity domains, such as activity associated with transportation, while monitor-based measures cannot provide direct information about context. In both cases, the measurement protocol can be adapted to capture this information by either including additional items in a self-report (report-based measures) or by obtaining detailed schedule information (e.g., school time) and extract raw data during the period of interest.

calorimetry but field-based studies or longitudinal studies (with larger samples) may necessitate simpler and more feasible assessment options. The choice between report- and monitor-based measures can be difficult, as both sets of tools have unique advantages for this type of research. Report-based measures can facilitate the assessment of large samples, which would improve the representativeness of the population being studied (thereby improving external validity) while monitor-based measures have higher accuracy (thereby improving internal validity). The decision may depend on the sensitivity needed to capture associations with the specific outcome of interest.

For example, the majority of scientific evidence on the health benefits of physical activity has been accumulated predominantly with report-based measures. The increased availability of monitor-based measures has facilitated the inclusion of these tools in this type of research. However, report-based measures are still predominantly used. This is because in many types of epidemiology studies, the main need is to simply classify individuals into general levels of physical activity participation. Report-based measures such as self-reports have proven to be useful to rank individuals according to their activity level and therefore provide sufficient accuracy to categorize individuals based on their level of physical activity (e.g., quintiles).

This type of stratification is sufficient for some applications, but research in this area also is aimed at determining the dose-response between physical activity and outcomes of interest, and this necessitates more precision (particularly if the goal is to establish clinically meaningful thresholds). More precise estimates of physical activity may be needed if researchers or practitioners are attempting to determine the dose of physical activity necessary to achieve health benefits. The use of activity monitors would be strongly recommended in these situations even though some procedures would allow the improvement of the estimates obtained from self-reports.[73] Obtaining precise estimates of physical activity at the individual level is still challenging with activity monitors, but the error is substantially lower when compared to self-reports.[s] The need for precision is also greater in situations where the association between physical activity and the health outcome is subtler or harder to detect. In these situations, the more precise the measure, the greater the likelihood that it will be able to capture the associations and possible effect of physical activity on the outcome of interest. The diverse range of applications and designs make it difficult to generalize about the most appropriate measures.

Surveillance Research

The goal in many surveillance applications is to evaluate levels of physical activity in the population, so physical activity measures most typically serve as dependent variables. Measures used in studies or projects of this nature tend to emphasize feasibility over validity due to the greater emphasis on sampling and external validity. Report-based measures have historically been more common in these types of studies but monitor-based measures are now widely used in large-scale surveillance applications. The key need is to capture population-level estimates so emphasis is on ensuring that the measures have adequate group-level measurement properties (see Section 4). Unique challenges related to this design involve selecting a tool that is feasible for large samples; that has measurement properties not affected by the population being assessed (e.g., equally valid for youth ages 8 to 18 years, or equally valid for youth of different countries); that is sensitive enough to capture sex, seasonal, or age-related group differences in physical activity; or that can capture either changes over time or differences between subgroups (e.g., boys vs. girls, children vs. adolescents).

Report-based measures (primarily self-reports) are the most common tool for surveillance studies, but it can be challenging to find a survey that can fulfill critical measurement requirements. For example, well-designed self-report tools can provide reasonable estimates for some groups of children, but the instrument may have differential properties (e.g., reliability and validity) for various ages or in youth from different backgrounds and may lack the ability to capture physical activity changes over time. For example, the development of the International Physical Activity Questionnaire was specifically aimed at standardizing a self-report tool so that it could be used in different countries to provide a common metric. Although report-based measures are easier to use and less expensive, monitor-based measures offer advantages for standardization because they capture only the movement performed. The use of monitor-based measures is now common in both large population surveys and smaller studies in which the focus is on comparing levels of physical activity in different groups or segments of the population.

s The use of measurement error models has helped to refine the precision of some self-report measures. This has been shown to strengthen the associations between physical activity and health outcomes such as obesity or diabetes by 30% to 50% (see Reference 88).

Theory and Correlate Research/ Intervention Research

Theory and correlate research often focuses on identifying factors (i.e., correlates) that may explain differences in physical activity levels in the population or in testing theories that may explain physical activity behavior. Intervention research then seeks to use insights to plan and evaluate strategies designed to promote physical activity in the population. Designs in these realms of behavioral research vary widely and physical activity can serve as either an independent variable or a dependent variable. Studies, for example, may compare a battery of psychosocial predictors for active or inactive individuals (physical activity as an independent variable) or use a battery of correlates to explain physical activity behavior (physical activity as a dependent variable). Studies can be set up to compare health outcomes in groups with different levels of physical activity (physical activity as an independent variable) or to quantify actual differences in physical activity outcomes (physical activity as a dependent variable). Measures used in theory and correlate research tend to be at the middle to high left end of the feasibility vs. validity continuum, while measures for intervention designs are often at the middle and lower right end of the continuum because precise estimates are often needed to detect any small differences between intervention groups. The distinctions in the design can have important implications for the need for precision and the type of measure that would work best.

Overall, report-based measures (self-report measures in particular) are still very popular in these study designs, because they can not only provide reasonable estimates of group-level physical activity but also add contextual information that is usually of interest depending on the intervention being conducted. However, one key need in interventions is the ability to detect changes in physical activity as a result of the intervention or to capture differences in physical activity among people with different health conditions. These two require a level of precision that is often not characteristic of self-report tools. This level of precision requires measures that can provide more accurate estimates at the individual level, and this need can vary depending on the expected impact of a physical activity intervention on activity levels or association between physical activity and a health outcome. For example, in situations where the intervention is expected to increase physical activity by a small but still meaningful amount (e.g., increase physical activity during recess), researchers or practitioners might need a precise measure that is sensitive to these changes in physical activity. Assuming recess is likely to last for 15 to 20 minutes, the measure must be able to capture changes of 10 to 15 minutes of activity or less during the recess period. It would be unrealistic to expect that a self-report can capture such effects. Monitor-based measures would be better suited for this purpose. However, depending on the monitor-based measure, important drawbacks may need to be considered. This is particularly true for pedometers, which are known to also serve as a motivational tool because they can provide immediate and interpretable feedback. Their use is strongly recommended in studies examining motivation. However, their use in interventions that are aimed at manipulating other factors (e.g., inducing changes in the environment) can add bias when determining the treatment (e.g., physical activity program) effect or changes in physical activity levels.

Resources

The selection of a measure also will need to take into consideration the timeline for the project or study, available or planned human resources allocated for the project or study, and the need for immediate feedback. The timeline can often dictate the schedule and timing of data collection and this, in turn, can dictate the most practical assessment strategy. The ratio between the sample size and the timeline can give a good indication of what measure property needs to be prioritized. The human resources relate to the availability of human capital to collect data or the expertise required to handle the data processing tasks, while immediate feedback involves having estimates of physical activity available to participants as they participate in the measurement protocol.

If the sample size is high or the timeline is short, a simpler and more practical assessment, such as a self-report tool, may be warranted. However, if the sample size is small or the timeline long it becomes more feasible to use a monitor-based approach. This factor plays an important role in the measurement protocol and is heavily influenced by the availability of staff to collect and process the data. Having more staff will allow for more intense data collection protocols and can increase the likelihood that more people be assessed per unit of time. The availability of staff is critical for both data collection (e.g., setting up devices or assisting with completion of surveys) and the data processing steps. In general, report-based measures are better suited for short timelines and large sample sizes, while monitor-based measures might work better for a larger window of time for data collection and small to medium size samples. Again, this relation is further compounded by the availability of research staff. For example, if the size of the research team is limited, it becomes more challenging to initialize and distribute a large number of activity monitors or to download and process data from multiple monitors. In this case, it may be necessary to reduce the sample size or to collect data over a longer span of time to ensure that the data on sufficient number of participants can be collected. With monitor-based methods, the availability of monitors also can become a rate-limiting factor and may dictate the rate at which data can be collected. Therefore, it is important to carefully evaluate both the timeline and the human resources available to determine how many assessments can be conducted per day or week. This will directly determine the feasibility of a given assessment approach.

Two other important factors that are often overlooked are the expertise required to handle the data collected and the need for immediate feedback. Overall, data processing protocols for report-based measures often require less technical work, while monitor-based measures are highly susceptible to processing decisions and require a greater level of expertise. For example, the staff working with report-based tools may only need to calculate energy expenditure or physical activity variables while staff collecting monitor-based tools may need to be familiar with software and data processing methods. With regard to feedback, most report- and monitor-based measures require some level of processing before data can be interpreted and feedback provided, but there are some exceptions. For example, web-based questionnaires can automate processing and offer immediate feedback. Similarly, pedometers and some consumer-based devices can provide immediate feedback to participants (i.e., number of accumulated steps). This type of feedback can be an advantage for intervention studies focused on changing behavior but can be problematic for studies attempting to capture "typical" behavior. There is considerable interest in new lines of consumer activity monitors but more work is needed to understand the measurement properties of the various models available (see Section 9). Thus, the need for immediate feedback can narrow the list of measures suitable for a particular project, but some of the challenges can be overcome depending on the human resources allocated for the project. A larger research team will permit staff members to be allocated to data processing once data are collected and can therefore provide feedback once the protocol is completed. This approach does not replace the need for immediate feedback but instead allows for feedback at the end of participation in the project.

Decision Confirmation

A final step, once a set of possible measures is selected, involves filtering among the available report-based or monitor-based measures, models, or versions to determine which is most appropriate considering their measurement properties when applied to the population of interest. The NCCOR Measures Registry helps to summarize the documented evidence regarding the different tools to facilitate this review. For example, if the assessments are to be conducted in adolescents, then researchers or professionals can filter the existing measures for adolescents ages 12 to 18 years, and then select the type of measure preferred. Once the measure is selected, it is possible to access a variety of studies that summarize the properties or measurement characteristics of the instrument. This information can help determine whether the measure is indeed appropriate for the design of the research study or project. It is important to note that the Measures Registry only summarizes the information. It is up to the researcher to carefully review the findings and to determine whether the results generalize or apply to their situation.

Before formal adoption and use, it is essential to test the selected measure under real-world conditions through pilot testing or a formalized feasibility study. Pilot testing involves replicating the design of the project or study but in a small fraction of the population of interest. Pilot testing can help determine whether the physical activity assessment protocol is appropriate. More importantly, such a step allows the researcher to test the different steps associated with the measurement protocol: preparing for data collection, collecting data, and processing and handling the data generated from the collection phase. This is particularly important if the user is not familiar with the tool selected. Some considerations for data processing and data management are summarized in Section 10. However, it is first important to provide practical applications of how to select a measure.

7

Case Studies

SECTION 7

This section provides a series of case studies or scenarios that show how to apply the insights from the Guide to identify and select assessment tools through the NCCOR Measures Registry. As described in other sections, the measurement considerations for physical activity related studies vary greatly depending on the nature of the research question and the relative need for accuracy and precision, as well as a number of other considerations.

CASE STUDY 1 EXAMINING THE INDEPENDENT AND JOINT ASSOCIATIONS OF PHYSICAL ACTIVITY AND SEDENTARY BEHAVIOR ON BODY MASS INDEX AMONG MIDDLE AND HIGH SCHOOL STUDENTS

Background

The increased prevalence of childhood obesity among developed countries can be partially attributed to the increased availability of emerging technologies (e.g., smartphones) and lack of sufficient physical activity environments or opportunities. However, little is known about the independent contribution of sedentary behaviors and physical activity to obesity.

To address this, a school board planned a project to examine the independent and joint associations of physical activity and sedentary behavior on body mass index (BMI) among middle and high school students. The project team envisions categorizing youth based on compliance with existing guidelines for both physical activity (i.e., >60 min/day of moderate-to-vigorous intensity physical activity [MVPA]/day) and sedentary behavior (i.e., <2h/day).

The study is to be conducted with all students at one school during the spring semester (i.e., a 3- to 4-month period).

Considerations

Given the nature of this study, the project team decides the design requires a measure appropriate for adolescents or middle school students. The measure also must be able to capture time spent in sedentary time and MVPA per week or compliance with the guidelines.

The team decides that it needs a measure that can provide precise estimates at the individual level. However, considering the large sample and the need to categorize individuals in either sedentary vs. non-sedentary and active vs. not-active groups, the team feels it should be satisfied with a measure that is accurate for groups of individuals.

This project involves assessing a large sample of adolescents. However, the study has some flexibility because data can be collected across an extended period of time (over a full semester). A key challenge in the design is the need to obtain measures of both physical activity and sedentary behavior outcomes. Rough categorical estimates from a large sample could be sufficient to detect associations, but more objective data from activity monitors could increase the accuracy of the estimates. Regardless of the measure chosen, expertise is needed to process and interpret the physical activity data in the large sample; immediate feedback is not required.

Measure Selection

With respect to the population, the lowest age range would be 12 years so any of the measures in the Measures Registry would be suitable.

Indirect calorimetry, direct observation, activity monitors, heart rate monitors, pedometers, self-reports, and diaries can all provide an estimate of compliance with physical activity guidelines (i.e., 60 min of MVPA/day). This same list of measures can be used to infer about sedentary time with the exception of pedometers, but direct observation, self-reports, and diaries are the only direct measures of sedentary behavior.

Indirect calorimetry and direct observation are not feasible for population studies, and activity monitors are the next most accurate in the possible list of measures. The few monitors that can provide accurate estimates of time spent in sedentary behavior are not accurate for MVPA. The team therefore needs two different measures and opts for a self-report to measure sedentary behavior.

Data are collected using activity monitors and self-reports (i.e., several participants at a time). The sample size is large but so is the timeline (i.e., 3 to 4 months). Therefore, the ratio of sample size to timeline is favorable.

A midsize to large project team is needed to assist in data collection. Staff with expertise in processing and interpreting accelerometer data also are needed. The team does not need to provide feedback to participants so this factor will not affect its decision.

Activity monitors and self-reports are a good choice for this study. The next step involves selecting a particular activity monitor model and choosing a self-report from the Measures Registry. Pilot testing is recommended to review the protocol and obtain training in handling activity monitor data.

CASE STUDY 2 DETERMINING COMPLIANCE WITH PHYSICAL ACTIVITY RECOMMENDATIONS ACROSS DIFFERENT GRADE LEVELS

Background

Schools are considered the ideal setting to educate youth about physical activity and healthy lifestyles. National recommendations indicate that youth should accumulate at least 30 minutes/day of MVPA in the school-setting (150 minutes per week). However, it is difficult to determine compliance with these guidelines because of a lack of documented records. A team managing a large national research network plans a study to determine compliance with physical activity recommendations across different grade levels (i.e., elementary, middle, and high school). They are interested in determining what percentage of youth meet the public health goal of 60 minutes of MVPA per day and whether children are getting at least half of this activity at school. As a school-based project, they also want to ensure that the assessment provides an educational experience for both the students and the teachers.

Considerations

This is a large project that will require a measure that is equally appropriate for children and adolescents. The measure needs to capture the frequency, duration, and intensity of physical activity and address most of the physical activity domains, particularly physical activity occurring at school.

The team recognizes that it needs a measure that is highly feasible but decides it can sacrifice individual-level accuracy so long as it has a measure that is relatively accurate at the group level. Feasibility is a priority in surveillance research (and in this particular project) because the team needs to rely on teachers to perform the assessments, while ensuring that data can be collected within a similar period of time (e.g., spring semester) to allow for direct comparisons between schools.

The team needs to include staff responsible for coordinating all the contacts with schools and facilitate the assessments. It also needs a measure that can provide immediate feedback because the annual assessment was intentionally designed to educate students about the importance of physical activity and how to self-monitor their activity behaviors.

Measure Selection

All the measures are suitable for children and adolescents but the team needs to make a careful choice if they decide to use report-based measures because these can present challenges when administered to young children.

Indirect calorimetry, direct observation, activity monitors, heart rate monitors, self-reports, and diaries can all provide an estimate of time spent in MVPA. However, only self-reports and diaries can capture the context of physical activity and specifically partition daily activity accumulated at school and out-of-school. Activity at school can be estimated with activity monitors or heart rate measures if additional information is collected (e.g., school schedule).

Self-reports can easily be shared with schools throughout the country for assessment and be sent along with assessment instructions for teachers. A web-based self-report would be a particularly good option, while diaries can be less accurate and be difficult for young children to complete.

The scale of the project creates some challenges for resources and time. Self-reports allow for the collection of large amounts of data within a short time period so this is an advantage. This particular study will not require great human power to administer the self-report because the data will be collected by school staff (e.g., physical education teachers). Self-reports also can provide immediate feedback and therefore have unique educational value.

A wide variety of self-reports is available for children and adolescents, but they can be narrowed down once the Measures Registry filters are applied to include context-related physical activity and those that have demonstrated validity across different age groups (ages 8 to 18 years). A web-based version would be preferred and applying this filter would narrow the available self-reports even further. Pilot testing would require testing the web-based tool to determine, for example, how scores are saved (e.g., how the server saves the data) and how feedback is provided.

CASE STUDY 3 IDENTIFYING PREDISPOSING FACTORS FOR ACTIVE COMMUTING IN ELEMENTARY SCHOOL CHILDREN WHO LIVE IN URBAN AND SUBURBAN SETTINGS

Background

Promoting walking and active transportation is an important public health strategy for the whole population, but specific efforts have been made to increase the percentage of children who walk or bike to school. To make it easier to promote this behavior, it is important to better understand factors that influence adoption of this strategy and the barriers that must be overcome to promote it on a large scale.

A project team therefore plans an evaluation to identify predisposing factors for active commuting in elementary school children who live in urban and suburban settings. Although some children may bike to school, the focus was on understanding walking behavior and barriers to walking. In addition to walking behaviors, the team also plans to collect information on parents' perceptions of neighborhood safety and benefits of physical activity, along with detailed mapping of distance between home and school, and availability of sidewalks and crosswalks.

Considerations

This project needs to use measures that capture activity primarily in elementary school children. The indicator of active commuting is defined as walking behaviors and, therefore, the team's measure needs to include an indicator of number of steps accumulated per week.

The team is particularly interested in exploring the associations between walking and predisposing family and environmental factors for active commuting. A reasonable degree of accuracy is needed to ensure that the measures are sensitive enough to capture differences in walking due to the hypothesized factors.

A team of individuals will be needed to help collect the data and to evaluate the school and community factors that may influence active transportation.

Measure Selection

All measures can be adjusted for possible use, but pedometers are a logical choice for this purpose. Considering the age of the sample, it would be difficult to get accurate estimates from report-based measures.

Direct observation, activity monitors, pedometers, self-reports, and diaries can provide a measure of steps. Report-based measures would be particularly useful to obtain information about the context, or parents' perceptions of neighborhood safety and benefits of physical activity. However, this case study focuses on selecting a measure of physical activity and not related barriers.

A high degree of precision is not needed, but the nature of the study necessitates a reasonably accurate estimate of activity from a large sample of youth. Pedometers are viewed as the most valid measure for steps and as being very feasible (i.e., affordable) when compared to activity monitors. It also would be challenging to estimate the number of steps (or alternatively, distance covered) using direct observation or report-based measures such as self-reports or diaries. Report-based measures could be a good alternative if project staff are interested only in determining the mode of transportation/commuting (e.g., walking vs. car).

The team can divide data collection across several weeks. However, if it chooses to use pedometers, it will need to expand the project team in order to have sufficient personnel to distribute the pedometers and record the number of steps on each day of the week once participants arrive at school. It still will be necessary to ask participants to reset their pedometer (i.e., number of steps) at the beginning of each day before going to school.

The team wants to omit immediate feedback so that its associations are not confounded by motivation as a result of self-monitoring. This can be accomplished with sealed monitors.

The project team chooses to use pedometers but is aware of the implications and adjustments that it will need to include in its physical activity measurement protocol. The team now selects the most appropriate pedometer model by consulting the Measures Registry and testing its properties in a pilot test.

CASE STUDY 4 TESTING THE POTENTIAL OF A NEW RECESS-BASED PHYSICAL ACTIVITY PROGRAM DESIGNED TO INCREASE THE TIME CHILDREN SPEND IN MVPA DURING RECESS

Background

Public health agencies have advocated for coordinated efforts to promote physical activity across the school day. Activity breaks during classroom time are encouraged along with enhancement of opportunities and programming during recess. Evidence-based strategies to promote physical activity during recess are needed because it has proven difficult to create engaging opportunities that promote physical activity in youth without constraining natural free play. An intervention study is planned to test the potential of a new recess-based physical activity program designed to increase the time children spend in MVPA during recess. Schools in a Midwestern state of the United States are randomly assigned to either a 1-year intervention or control group. The design calls for physical activity to be assessed at baseline, 6 months (post-intervention), and 1-year follow-up.

Considerations

The intervention requires a measure that is appropriate for children. The project team also needs a measure that can capture the frequency, duration, and intensity of physical activity to quantify the volume of MVPA.

The measure needs to be relatively accurate at the individual level to capture small variations in physical activity associated with the intervention (i.e., assuming physical activity is likely to increase by 15 to 20 minutes per week if the intervention is successful), but be cost effective because the team needs to buy several devices so that it can assess physical activity within a short amount of time (i.e., assuming all the schools need to be assessed within a 3- to 4-week period).

Additionally, the project team needs a diverse group of people who are able to collect a variety of data within a short time and who have the expertise to handle physical activity output. The team does not need to have immediate feedback so the more detailed data processing can be done after all data are collected.

Measure Selection

Most tools are suitable for young children, but report-based measures would likely not be a good option because either recall or recording activity at these younger ages is very challenging.

Indirect calorimetry, direct observation, activity monitors, heart rate monitors, self-reports, and diaries can provide an indicator of volume or MVPA. Only direct observation, self-reports, and diaries can be used to obtain information on context (e.g., recess). Monitor-based measures, such as activity monitors or heart rate monitors, also can be used if the measurement protocol is adjusted and additional information is collected to define the context where activity occurs (i.e., time periods when recess takes place and recess environment).

Indirect calorimetry is not feasible for this study but the team strongly considers direct observation (the next and most accurate measure of physical activity in the continuum) because it would allow them to combine both MVPA and context-related information, such as the availability and use of recess equipment. Activity monitors and heart rate monitors also could be an option and could provide more interpretable outcomes when compared to direct observation. The choice of report-based tools would not be ideal given that the tools have considerable error at the individual level (particularly in young children) and it would be very challenging to capture any intervention effects.

Working with several schools will create some challenges and require that the team collect large amounts of data in a short window of time. This will require a comprehensive project team to collect the data and undergo additional training to ensure that observation data are collected in a standardized way. The team does not need immediate feedback so this should not affect its decision on measures.

The team chooses to use both direct observation and activity monitors concurrently to provide a comprehensive and interpretable measure of physical activity. It now navigates through the Measures Registry to decide which direct observation and activity monitor to use. The team then follows up with staff training followed by pilot testing to ensure that direct observation data are collected with accuracy.

8

Supplemental Considerations for Monitor-based Assessments (New Technology and Data Processing Techniques)

The rapid development and evaluation of new monitoring techniques and technologies makes it challenging for researchers and practitioners to determine the best way to collect, process, and interpret data with these methods. Recent advances in data processing methodologies available for activity monitors also may necessitate advanced data management and analytic skills, which can further complicate assessment decisions. This supplemental section introduces some of the complexities and refers to other sources for additional information.

Background on Accelerometry-based Activity Monitors

Considerations for Collecting and Processing Physical Activity Data

Accelerometry-based activity monitors are relatively easy to use, but a number of decisions must be made when collecting, processing, and interpreting data. One of the first decisions is the epoch or sampling timeframe for the assessments. Most monitors allow for the selection of different time interval windows (e.g., 1-second epochs, 5-second epochs, 30-second epochs) and the choice can have implications for the outcomes. Researchers routinely use a 60-second epoch because a minute is a reasonable unit for comparison and evaluation. However, the intermittent nature of children's activity patterns requires that lower epochs be used to capture more sporadic activity patterns (i.e., 5 seconds).

A number of other decisions must then be made when processing activity monitor data. Some of the more critical decisions include selecting thresholds or equations to interpret or scale the data, selecting a method to determine non-wear time (e.g., 60 minutes vs. 20 minutes of consecutive zeros), defining minimum wear time to consider a day to be representative (e.g., 60% vs. 80% of total day), identifying spurious data (e.g., ≥20,000 counts vs. ≥16,000 counts in Actigraph data), and selecting the number of valid days needed to characterize "habitual physical activity" (e.g., 4 days vs. 7 days; week vs. weekend). The availability of product-specific software can facilitate this process and allow the user to create customized settings for the data. For example, Actigraph data can be processed using Actilife data analysis software platform. The software can easily allow the user to set specific non-wear time algorithms, activity cutpoints, and a minimum wear requirement, among others. The software also allows data from the monitors to be converted to a variety of formats (e.g., csv, dat) and analyzed using various statistical packages (e.g., SPSS, SAS, R).

The different decisions regarding the data generated from activity monitors require that researchers clearly identify the data reduction procedures used when using activity monitor data. Depending on the decisions about data reduction protocols, researchers and practitioners can expect significant differences in wear time, activity counts per minute, average activity per day (in counts/day), average MVPA levels (in minutes/day), and average MVPA bouts per day. The number of participants meeting physical activity guidelines also will differ depending on the different data reduction protocols. The more conservative the data reduction protocol is, the lower the number of participants with valid data, the lower the number of minutes of inactivity, and the higher the number of minutes spent in light and

MVPA.[74] Many papers in the literature have alluded to this problem, and there have been calls for standards to help facilitate comparisons among studies.[74-76] Some examples of excellent detailed procedures used for data reduction protocols include the National Health and Nutritional Examination Survey (NHANES),[12] the International Children's Accelerometry Database (ICAD),[77] the Canadian Health Measures Survey,[78] and the ENERGY-project.[79]

Considerations of Hip versus Wrist Placement

The majority of work on accelerometry-based monitors has been conducted using devices worn at the waist or hip. However, investigators have begun to transition toward the use of wrist-worn monitors. This transition has been fueled by the progression in consumer-based monitors as well as by evidence that compliance is enhanced when participants are asked to wear monitors on the wrist (more like a watch). The wrist placement may offer some advantages, but it is important to note that equations and methods developed for hip-worn devices cannot be directly applied to data collected with wrist-worn monitors. Acceleration at the wrist is generally higher than that at the hip and therefore requires new calibration studies to determine physical activity intensity cutpoints so that wrist data can be interpretable. For example, only a few wrist cutpoints have been proposed for youth, and the evidence supporting them is still limited. The uncertainties about the utility of existing cutpoints for the wrist limit the ability to make direct comparisons with previous research when monitors were worn at the hip. It is important to carefully consider the relative advantages and disadvantages of monitor position when planning a study. Even so, several large epidemiological studies including, NHANES,[t] have elected to use the wrist position; this will likely drive additional development and innovation. The reports in the Measures Registry may or may not specify the location used in the various validation studies because this distinction is a relatively new consideration with the use of monitor-based approaches.

t The National Health Examination Survey (NHANES) is a combined surveillance program led by the National Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC) that tracks health indicators in the U.S. population. NHANES has collected data every 2 years since 1999 and uses a complex, multistage, probability sampling design of all ages to adequately characterize the U.S. population.

Handling Missing Physical Activity Data

One of the challenges associated with physical activity measurement protocols using direct measures is the burden placed on the children being assessed. Complete, full-week measures of activity using accelerometers would ideally require that youth use an activity monitor device for 24 hours a day, during 7 consecutive days. Instead, youth often forget to replace the monitor after showering or sleeping, or choose not to wear the accelerometer during some periods of the day. It is important to account for these periods when the activity monitors are not worn.[80-81] One strategy is to remove these bouts of non-wear time from further analysis. However, after removing these periods, it is important to check whether the remaining recorded data can still provide a representative picture of activity levels during the week being assessed or whether the children who were ultimately excluded from the data differ from those who were compliant (e.g., less active). It is not surprising to find non-compliance rates of 30% of the total sample, meaning that 1 out of every 3 children does not comply with the physical activity measurement protocol.[82-84] Previous research has shown that youth characteristics, such as BMI, age, and screen time, can predict non-compliance.[82-83] However, the more important distinction is whether data are missing at random or systematic patterns exist. This influences whether missing data can be imputed or not. Readers interested in this topic should consult with statisticians or the literature for guidance.

Newer Monitoring Technologies and Methods

Considerations with Consumer-based Monitors

The consumer marketplace has been flooded with an array of activity monitors designed to enhance self-monitoring and behavior change, and these features also have led to interest among researchers and health professionals. Products have been released with little or no evidence of reliability and validity, but researchers have started to identify potential strengths and limitations of the various devices. Evidence suggests that the accuracy of some consumer monitors may be comparable to findings from other, established research monitors. However, it cannot be assumed that all monitors have similar utility. A key distinction is the relative utility of step count estimates from devices. Many products and smartphone apps can provide estimates of steps, but the accuracy is questionable in many devices when directly compared with pedometer counts or manual step counts. Consumer monitors may be fashionable and trendy, but they may have limitations when used in research applications. It is up to the researcher to ensure that the selected device has sufficient reliability and validity for the desired application. Efforts are underway to establish benchmarks or standards for accuracy within the wearable monitor industry, which will facilitate comparisons in the future.

Distinctions between Raw and Count-based Accelerometer Data

A major challenge with the use of accelerometry-based activity monitors has been the lack of standardization about the processing and filtering of the raw accelerometer data. The use of different filtering and processing methods by various manufacturers has prevented movement "counts" from one monitor from being compared to those of another monitor. As a result, support for tracking and processing "raw" accelerometer data has increased. In theory, this would enable standardization of output in terms of real acceleration units (i.e., g values) and promote standardization of methods using open source processing techniques. This transition has some advantages but it also has dramatically complicated the data processing methods. The sheer volume of data is one challenge. If data are collected on a minute-by-minute basis, researchers must process 1,440 lines of data per day of assessment. However, this number grows to 8.64 million lines when processing raw data at 100 Hz (100 samples per second). New, open-source macro processing methods are being released to facilitate the processing, but additional expertise and time are needed to process these types of files. The new methods offer considerable promise for standardization in the future, but they present challenges for curent researchers and practitioners interested in using them in studies or projects.

Applications of Pattern Recognition Methods for Activity Classification

The availability of raw data at a low resolution, as described previously, creates a variety of opportunities to enhance the accuracy of activity monitors. Traditional cutpoints based on counts are still widely used, although the field is evolving to more advanced methods that can make use of large amounts of data to predict activity type or posture and use this information to estimate energy expenditure. The accuracy of single prediction equations to estimate energy expenditure is influenced by the activity being performed because the relation between energy expenditure and movement counts varies from activity to activity. Machine learning is a popular method in computer sciences but just recently has been used in physical activity research. This method involves selecting and extracting features from movement signals obtained from wearable sensors such as an accelerometer. Examples of actual features from movement signals include for wrist algorithms (from g's) the mean of vector magnitude (vm or mvm), the standard deviation of the vector magnitude (sdvm), the percentage of the power of the vector magnitude that is in 0.6–2.5 Hz range, the mean angle of acceleration relative to vertical on the device (mangle), the SD of the angle of acceleration relative to the device, and many others. The variability in the raw acceleration signal is then used to detect patterns and create activity classification schemes using advanced methods such as Hidden Markov models or Random Forests models. The resultant models are known to be able to differentiate between a set of postures (e.g., sitting vs. standing vs. walking) and also determine absolute activity intensities. Therefore, they can create implicit prediction equations based on the activity type or posture detected. This method can overcome the limitation associated with existing calibration equations. However, this method is still in the early phases of development. More information is available elsewhere.[85]

9

Supplemental Considerations for Evaluating Sedentary Behavior

The health and behavioral consequences of sedentary behavior has generated considerable public health interest. Research is still in its infancy, but consensus has emerged that sedentary behavior is distinct from physical activity behavior in youth. Low levels of physical activity cannot be inferred to reflect high sedentary behavior, and high levels of physical activity cannot be assumed to reflect low sedentary behavior.

Assessment strategies are also inherently different, and challenges in assessing sedentary activities have confounded efforts to better understand this behavior in youth. The Measures Registry is well positioned to facilitate selection of effective sedentary behavior measures to guide future research. This section will provide brief summaries of issues associated with the evaluation of sedentary behavior using both monitor- and report-based measures. More detail is available elsewhere.[86-87]

Background on Sedentary Behavior

Early concerns about sedentary behavior in youth stemmed from studies showing that excess television (TV) viewing was a likely contributor to the epidemic of childhood obesity. Further, interest was sparked by studies in adults showing that sedentary behavior may influence health risks independent of physical activity behavior.[88] Research to date, however, has not supported the independence of sedentary behavior as a health risk in youth when physical activity is taken into account. The health implications of sedentary behavior in youth warrant further evaluation, but there is no doubt that it is an important behavioral target for intervention and a priority for family, school, and community programming.

Assessments of Sedentary Behavior

The focus of this Guide has been on evaluating physical activity behavior, but evaluating sedentary behavior has unique considerations. A complicating factor in sedentary behavior research is the ever-changing nature of technology in society. Common forms of physical activity have remained relatively consistent over the years, but this is not the case with sedentary behavior. Public health recommendations for sedentary behavior have focused on excess TV viewing and computer games,[89] but it is likely that youth now spend more time on their smartphones or on handheld or tablet devices than watching or using TV. The blurring of technology makes it difficult to characterize sedentary behavior but another challenge is that time spent in sedentary behavior may include desirable behaviors, such as doing homework, reading, or playing music. Thus, it is important to distinguish discretionary or recreational sedentary behavior from required or desired forms of sedentary behavior. Device-based measures that provide objective information on movement are being used to assess sedentary behavior, but certain caveats and assumptions must be considered when interpreting the data. Research has demonstrated differences in the nature and patterns of findings depending on how sedentary behavior is assessed, so it is important to understand the advantages and disadvantages of each approach.

Quantifying Sedentary Behavior with Report-based Measures

An advantage of report-based measures is that they can provide information about the type and context of sedentary behavior, but a disadvantage is the difficulty in quantifying sedentary behavior due to recall bias. Guidelines have been proposed to assist researchers in choosing appropriate self-report instruments for evaluating sedentary behavior.[2] Comprehensive systematic reviews have also provided specific insights on the validity and reliability of self-report measures for youth.[87,90] The reviews have demonstrated clear limitations with standard time-use methods. However, few self-report tools have been developed specifically to assess youth sedentary behaviors.

Quantifying Sedentary Behavior with Monitor-based Measures

Monitor-based measures are designed to evaluate movement so it is somewhat paradoxical to operationalize “lack of movement” as sedentary behavior, especially considering that physical activity and sedentary behavior are thought to be independent of each other. A key advantage for sedentary behavior work is that monitor-based measures can provide detailed information about breaks and bouts of sedentary behavior. However, they cannot distinguish specific types of sedentary behavior or provide contextual information about sedentary behavior. The activPAL is a unique example of a monitor-based measure that can differentiate sitting from other postures with a high degree of precision and therefore adds great value for measuring sedentary behavior. However, the majority of work has used the Actigraph monitor and support is strong for using a threshold of 100 counts/minute for detecting sedentary behavior at the hip.[u,91] Even though this threshold was not empirically derived, the assumption is that the accumulation of little or no movement over a certain time interval could only occur if a person was sedentary. Thresholds do not hold for other monitors, and recent evidence demonstrates that unique thresholds are needed for monitors worn at the wrist.[91] This makes conceptual sense because arms and hands can be performing small movements even when a person is sedentary.

u It is important to acknowledge that debate continues about the most appropriate threshold for hip-worn monitors but studies have used the value of 100 and this threshold has been well supported. However, it does not hold for other monitors or other positions.

10

Supplemental Considerations for Scaling and Scoring METs in Youth

This section provides supplemental information about some key measurement challenges and issues that come about when using METs to standardize activity intensities in youth physical activity assessments. METs are often used to impute energy expenditure values in self-reports or/and to classify activity intensities using calibration equations developed for activity monitors. However, the use of METs assumes standard resting energy expenditure rates and therefore requires a basic understanding of changes that occur in boys and girls as they transition from childhood into adolescence and adulthood.

Children undergo systematic changes in body composition as a result of growth and maturation and these changes are particularly relevant when considering the use of MET values in youth. In general, both boys and girls (though more so in girls) experience gains in fat mass during pre-adolescent ages, but patterns change after puberty. Boys experience increases in muscle mass across adolescence and reductions in fat mass, while girls' fat mass continues to increase throughout adolescence. The marked events in body composition are strongly related with the onset of maturity, which also varies between boys (usually at an average age of 14.0 years) and girls (usually at an average age of 12.0 years), and within sex, with some children maturing earlier or later than others. These patterns are relevant considering that muscle mass is a strong predictor of resting energy expenditure and that resting energy expenditure is often expressed as ml or kcal per kg of body weight, which reflects a composite of the partitions of body fat, muscle mass, and other tissues. The MET as it stands assumes a linear and positive relation between body weight and energy demand, which would be defensible if most changes in body weight were to occur as a result of muscle mass alone (i.e., metabolically active tissue). However, as changes in body weight occur and the fractions of muscle mass and body fat are altered as a result of growth, the MET assumption of linearity between body weight and energy expenditure is violated. The main cause of this violation is changes in body fat that can account for a significant portion of the changes in body weight (such as in girls during adolescence). These changes explain the differences in resting energy expenditure when comparing children with adults, children of different body weight, children of different ages, and children of different sexes.

The variability in body composition during development has important implications for activity intensity classifications based on METs and requires that cutpoints be adjusted for differences in resting energy expenditure that occur during growth. The MET was not intended to capture these differences, but interestingly, it has become a popular energy expenditure metric when establishing cutpoints or converting activity behaviors into estimates of energy expenditure or classifications of intensity. The simplicity of METs and relative ease for comparison across different subgroups of the population might well justify its use. However, researchers and practitioners are now more aware of the limitations of this metric and how it might affect activity classification and measurement of physical activity in general. By definition, METs assume resting energy expenditure as being 3.5 ml/kg/min (based on adult values and therefore often referred to as adult-METs), and multiples of this value (e.g., 3 METs) are used to distinguish light-intensity from moderate-intensity activity. The systematic change in resting energy expenditure during growth in youth implies that moderate activity might be better characterized by 4.0 METs and not 3.0 if resting energy expenditure is higher than 3.5 ml/kg/min (e.g., 4.2 to 6.0 ml/kg/min as seen in previous research). Adjusted MET values for sedentary, light, and moderate-to-vigorous intensity are directly affected by changes in resting energy expenditure and result in MET values that are higher in younger ages but decrease as youth get older, reaching adult values at about the age of 18 years. Figure 6 is an illustration of the discrepancy between adjusted and unadjusted MET values. The dashed lines indicate the traditional thresholds of 1.0, 1.5, and 4.0 adult METs (using 3.5 ml/kg/min), respectively, used to classify youth activity intensity, while the solid lines of same color indicate these same thresholds when accounting for maturation and differences in resting energy expenditure as growth occurs.

The proposed values were extrapolated based on average resting energy expenditure values published by Harrell and colleagues[30]. The resulting, adjusted resting energy expenditure values in METs were then multiplied by 1.5, 3.0, and 6.0 to obtain energy expenditure thresholds for light, moderate, and vigorous activity, respectively. Based on the published values from Harrell and colleagues, the resting energy expenditure in children (ages 8 to 12 years) can be on average 5.92 ml/kg/min, which is equivalent to 1.7 METs (substantially higher than the commonly, 1 MET value). When multiplied by 3.0, this value results in a threshold of 5.1 METs that should be used to classify activities of moderate intensity. Again, the 5.1 METs is substantially higher than the traditional threshold of 3.0 METs, which implies that if the latter is used, time spent in moderate physical activity is likely to be overestimated in children. Table 3 provides adjusted values for boys and girls ages 8 to 19+ years that can be used to overcome this misclassification and inappropriate assumption of the 1 MET value for resting energy expenditure when assessing activity levels in youth.

These new MET thresholds can be easily matched with available, published energy costs of physical activities in youth, and researchers and practitioners can alter their interpretation of the intensity of the activity performed or reported using the thresholds described above. The Youth Compendium of Physical Activity provides the energy cost using different metrics of the MET values (with both unadjusted and adjusted values for youth resting energy expenditure). Therefore, the above thresholds are appropriate only for energy cost when expressed in adult-METs (using 3.5 ml/kg/min). For example, according to the corrected thresholds, a reported activity for a boy age 10

Figure 6: Discrepancies Between Adjusted and Unadjusted MET Values

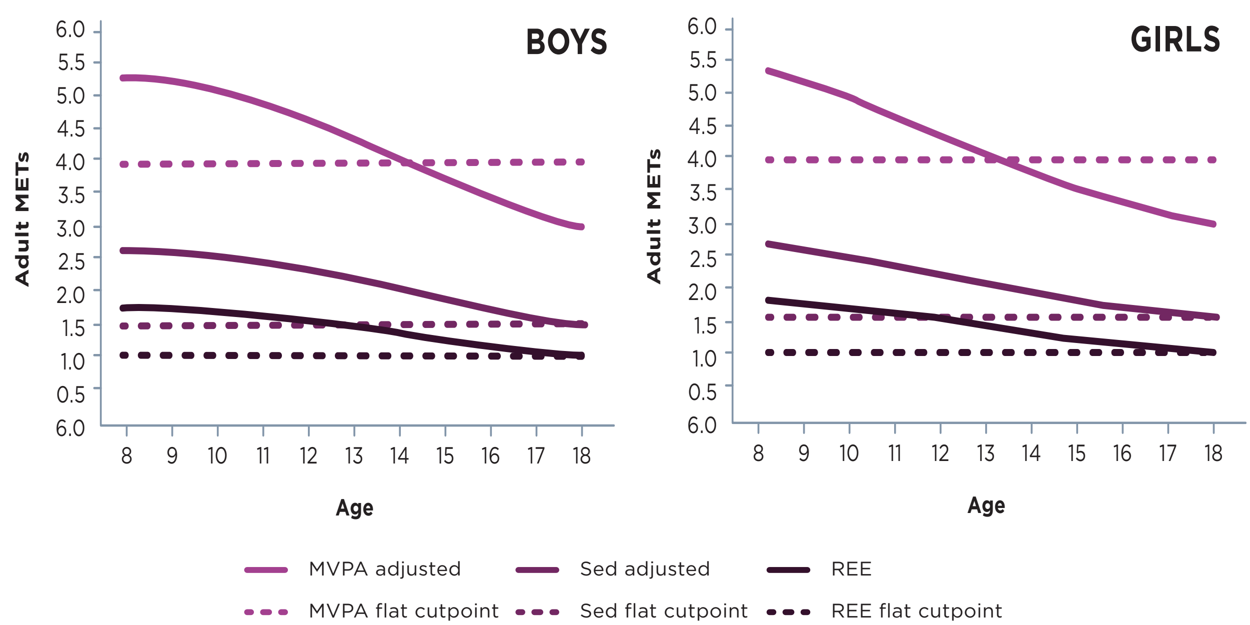


Table 3. Extrapolated Resting Energy Expenditure (REE) and Associated Activity Intensity Thresholds for Children and Adolescents Using Adult-MET Values

	REE*	SEDENTARY†	MVPA§
BOYS			
8 yr	1.8	2.6	5.3
9 yr	1.7	2.6	5.2
10 yr	1.7	2.5	5.1
11 yr	1.6	2.4	4.9
12 yr	1.5	2.3	4.6
13 yr	1.4	2.2	4.3
14 yr	1.3	2.0	4.0
15 yr	1.2	1.9	3.7
16 yr	1.2	1.7	3.5
17 yr	1.1	1.6	3.2
18 yr	1.0	1.5	3.0
19+yr	1.0	1.5	3.0
GIRLS			
8 yr	1.8	2.7	5.4
9 yr	1.7	2.6	5.1
10 yr	1.6	2.4	4.9
11 yr	1.5	2.3	4.6
12 yr	1.4	2.2	4.3
13 yr	1.4	2.0	4.1
14 yr	1.3	1.9	3.8
15 yr	1.2	1.8	3.5
16 yr	1.1	1.7	3.3
17 yr	1.0	1.6	3.1
18 yr	1.0	1.5	3.0
19+yr	1.0	1.5	3.0

* REE - Resting Energy Expenditure in adult-METs (1 MET = 3.5 ml/kg/min); Extrapolated from Harrell et al (2005) Table 4 REE values

† Sedentary threshold; Computed as a 1.5 multiple of extrapolated REE

§ MVPA – Moderate-to-vigorous physical activity threshold; Computed as a 3.0 multiple of extrapolated REE

Reprinted from Reference 61

years that is estimated at a value of 4.4 adult-METs values would be interpreted as a light intensity activity (because it is <5.1 METs as shown in Table 3) and not as being moderate as it would be if based on the traditional and unadjusted cutpoints (i.e., 3.0 METs). Alternatively, crude data obtained from activity monitors also calibrated against adult-MET values and the output activity intensity would need similar adjustments.

11

Conclusion

Accurate estimates of physical activity are essential for advancing research on the health benefits of physical activity; for understanding patterns and correlates that influence physical activity behavior; and for evaluating interventions designed to promote physical activity, improve health, or reduce obesity. Considerable attention has been given to improving physical activity assessment methods, and research continues to further improve these methods. Many options are available for assessing physical activity, so it is important to appreciate and consider the relative advantages and disadvantages of the various measurement approaches. The goal of this User Guide is to provide an overview of the field of individual physical activity, offer general guidance about selecting measures to suit each user's needs, and make the Measures Registry a more user-friendly and valuable resource. Our most important recommendation is to think carefully about the aim of your study or evaluation project and to select a physical activity outcome measure that addresses this aim as directly as possible. This guidance should be applicable even as more measures are added to the Registry. We hope this User Guide encourages greater use of individual physical activity measures in research and practice.

References

1. Ainsworth B, Cahalin L, Buman M et al. The current state of physical activity assessment tools. Prog Cardiovasc Dis. 2015;57(4):387-95.

2. Sternfeld B, Goldman-Rosas L. A systematic approach to selecting an appropriate measure of self-reported physical activity or sedentary behavior. J Phys Act Health. 2012;9(Suppl 1):S19-28.

3. Strath SJ, Kaminsky LA, Ainsworth BE et al. Guide to the assessment of physical activity: Clinical and research applications: A scientific statement from the American Heart Association. Circulation. 2013;128(20):2259-79.

4. Warren JM, Ekelund U, Besson H et al. 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):127-39.

5. Institute of Medicine. Bridging the evidence gap in obesity prevention: A framework to inform decisions making. Kumanyika SK, Parker L, Sim LJ, editors. Washington (DC): National Academies Press, 2010. Available at http://www.nap.edu/catalog/12847/bridging-the-evidence-gap-in-obesity-prevention-a-framework-to. Accessed November 1, 2016.

6. Bowles HR. Measurement of active and sedentary behaviors: Closing the gaps in self-report methods. J Phys Act Health. 2012;9(Suppl 1):S1-4.

7. Sallis JF, Owen N, Fotheringham MJ. Behavioral epidemiology: A systematic framework to classify phases of research on health promotion and disease prevention. Ann Behav Med. 2000;22(4):294-8.

8. Welk G. *Physical Activity Assessment for Health-Related Research*. Welk G, ed. Champaign (IL): Human Kinetics; 2002.

9. Pettee Gabriel KK, Morrow JR, Jr., Woolsey AL. Framework for physical activity as a complex and multidimensional behavior. J Phys Act Health. 2012;9(Suppl 1):S11-8.

10. Troiano RP, Pettee Gabriel KK, Welk GJ et al. Reported physical activity and sedentary behavior: Why do you ask? J Phys Act Health. 2012;9(Suppl 1):S68-75.

11. LeBlanc AG, Janssen I. Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. Pediatr Exerc Sci. 2010;22(4):523-34.

12. Troiano RP, Berrigan D, Dodd KW et al. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008;40(1):181-8.

13. Tucker JM, Welk G, Nusser SM et al. Estimating minutes of physical activity from the previous day physical activity recall: validation of a prediction equation. J Phys Act Health. 2011;8(1):71-8.

14. Physical Activity Guidelines Advisory Committee (PAGAC). *Physical Activity Guidelines Advisory Committee Report, 2008.* Washington DC; 2008.

15. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. Public Health Rep. 1985;100(2):126-31.

16. Howley ET. Type of activity: Resistance, aerobic and leisure versus occupational physical activity. Med Sci Sports Exerc. 2001;33(6 Suppl):S364-9; discussion S419-20.

17. Ainsworth BE, Haskell WL, Herrmann SD et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc. 2011;43(8):1575-81.

18. Levine J, Melanson EL, Westerterp KR et al. Measurement of the components of nonexercise activity thermogenesis. Am J Physiol Endocrinol Metab. 2001;281(4):E670-5.

19. Welk GJ, Kim Y. Sedentary Behavior in Youth. In: Zhu W, Owens N., eds. *Sedentary Behavior and Health: Concepts, Assessment & Intervention.* Champaign (IL): Human Kinetics; 2016.

20. Sedentary Behaviour Research Network. Letter to the editor: standardized use of the terms “sedentary” and “sedentary behaviours.” Appl Physiol Nutr Metab. 2012;37(3):540-2.

21. Tremblay MS, Leblanc AG, Janssen I et al. Canadian sedentary behaviour guidelines for children and youth. Appl Physiol Nutr Metab. 2011;36(1):59-64; 5-71.

22. American Academy of Pediatrics Council on Communications and Media. Media and young minds. Pediatrics. 2016;138(5):320162591.

23. Daniels SR, Hassink SG. The Role of the pediatrician in primary prevention of obesity. Pediatrics. 2015;136(1):e275-e292.

24. Saint-Maurice PF, Kim Y, Welk GJ et al. Kids are not little adults: What MET threshold captures sedentary behavior in children? Eur J Appl Physiol. 2016;116(1):29-38.

25. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport. 2000;71(2 Suppl):S59-73.

26. Baranowski T. Validity and reliability of self report measures of physical activity: An information-processing perspective. Res Q Exerc Sport. 1988;59:314-27.

27. Matthews C. Use of self-report instruments to assess physical activity. In: Welk GJ, ed. Physical Activity Assessments for Health-Related Research. Champaign, IL: Human Kinetics; 2002. p. 107-23.

28. Kohl III H, Fulton, JE, Caspersen CJ. Assessment of physical activity among children and adolescents: A review and synthesis. Prev Med 2000;31:S54-S76.

29. Byrne NM, Hills AP, Hunter GR et al. Metabolic equivalent: One size does not fit all. J Appl Physiol (1985). 2005;99(3):1112-9.

30. Harrell JS, McMurray RG, Baggett CD et al. Energy costs of physical activities in children and adolescents. Med Sci Sports Exerc. 2005;37(2):329-36.

31. McMurray RG, Soares J, Caspersen CJ et al. Examining variations of resting metabolic rate of adults: A public health perspective. Med Sci Sports Exerc. 2014;46(7): 1352-8.

32. Molnar D, Schutz Y. The effect of obesity, age, puberty and gender on resting metabolic rate in children and adolescents. Eur J Pediatr. 1997;156(5):376-81.

33. Ridley K, Olds TS. Assigning energy costs to activities in children: A review and synthesis. Med Sci Sports Exerc. 2008;40(8):1439-46.

34. Ainsworth BE, Haskell WL, Leon AS et al. Compendium of physical activities: Classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993;25(1):71-80.

35. Carroll R, Ruppert D, Stefanski L et al. *Regression Calibration. Measurement Error in Nonlinear Models: A Modern Perspective, 2nd ed*. Oxford (UK):Taylor and Francis Group; 2006. p. 65-95.

36. Welk GJ. Principles of design and analyses for the calibration of accelerometry-based activity monitors. Med Sci Sports Exerc. 2005;37(11 Suppl):S501-11.

37. Zaki R, Bulgiba A, Ismail R et al. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: A systematic review. PLoS One. 2012;7(5):e37908.

38. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

39. Robinson AP, Froese RE. Model validation using equivalence tests. Ecological Modelling. 2004; 176:349-58.

40. Corder K, Ekelund U, Steele RM et al. Assessment of physical activity in youth. J Appl Physiol. 2008; 105(3):977-87.

41. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. Sports Med. 2001;31(6):439-54.

42. Troiano RP, McClain JJ, Brychta RJ et al. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014;48(13):1019-23.

43. Biddle SJ, Gorely T, Pearson N et al. An assessment of self-reported physical activity instruments in young people for population surveillance: Project ALPHA. Int J Behav Nutr Phys Act. 2011;8:1.

44. Chinapaw MJ, Mokkink LB, van Poppel MN et al. Physical activity questionnaires for youth: A systematic review of measurement properties. Sports Med. 2010;40(7):539-63.

45. Corder K, van Sluijs EM, Wright A et al. Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? Am J Clin Nutr. 2009;89(3):862-70.

46. Haskell WL. Physical activity by self-report: A brief history and future issues. J Phys Act Health. 2012;9(Suppl 1):S5-10.

47. Saint-Maurice P. *Validation and Calibration of Self-Report Methods: The Youth Activity Profile.* Ames (IA): Iowa State University; 2013.

48. Saint-Maurice PF, Welk GJ. Web-based assessments of physical activity in youth: Considerations for design and scale calibration. J Med Internet Res. 2014;16(12):e269.

49. Cain KL, Sallis JF, Conway TL et al. Using accelerometers in youth physical activity studies: A review of methods. J Phys Act Health. 2013;10(3):437-50.

50. Corder K, Brage S, Ekelund U. Accelerometers and pedometers: Methodology and clinical application. Curr Opin Clin Nutr Metab Care. 2007;10(5):597-603.

51. De Vries SI, Van Hirtum HW, Bakker I et al. Validity and reproducibility of motion sensors in youth: A systematic update. Med Sci Sports Exerc. 2009;41(4):818-27.

52. John D, Freedson P. ActiGraph and Actical physical activity monitors: A peek under the hood. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S86-9.

53. Pedisic Z, Bauman A. Accelerometer-based measures in physical activity surveillance: Current practices and issues. Br J Sports Med. 2015;49(4):219-23.

54. Engström E, Ottosson E, Wohlfart B et al. Comparison of heart rate measured by Polar RS400 and ECG, validity and repeatability. Advances in Physiotherapy. 2012;14(3):115-22.

55. Lee CM, Gorelick M. Validity of the Smarthealth watch to measure heart rate. Meas Phys Educ Exerc Sci 2011;15(1):18-25

56. Porto LG, Junqueira LF, Jr. Comparison of time-domain short-term heart interval variability analysis using a wrist-worn heart rate monitor and the conventional electrocardiogram. Pacing Clin Electrophysiol. 2009;32(1):43-51.

57. Rowlands AV, Eston RG, Ingledew DK. Measurement of physical activity in children with particular reference to the use of heart rate and pedometry. Sports Med. 1997;24(4):258-72.

58. Terbizan DJ, Dolezal BA, Albano C. Validity of seven commercially available heart rate monitors. Meas Phys Educ Exerc Sci 2002;6(4):243-47.

59. Basset DR, Dinesh J. Use of pedometers and accelerometers in clinical populations: Validity and reliability issues. Physical Therapy Reviews. 2010;15(3):135-42.

60. Laurson KR, Eisenmann JC, Welk GJ, et al. Evaluation of youth pedometer-determined physical activity guidelines using receiver operator characteristic curves. Prev Med. 2008;46(5):419-24.

61. Laurson KR, Welk GJ, Eisenmann JC. Estimating physical activity in children: Impact of pedometer wear time and metric. J Phys Act Health. 2015;12(1):124-31.

62. Schneider PL, Crouter S, Bassett DR. Pedometer measures of free-living physical activity: Comparison of 13 models. Med Sci Sports Exerc. 2004;36(2):331-5.

63. Tudor-Locke C, Craig CL, Beets MW et al. How many steps/day are enough for children and adolescents? Int J Behav Nutr Phys Act. 2011;8:78.

64. Tudor-Locke C, McClain JJ, Hart TL et al. Pedometry methods for assessing free-living youth. Res Q Exerc Sport. 2009;80(2):175-84.

65. Bassett DR, Jr., Howley ET, Thompson DL et al. Validity of inspiratory and expiratory methods of measuring gas exchange with a computerized system. J Appl Physiol (1985). 2001;91(1):218-24.

66. Bassett DR, Jr., Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc. 2012;44(1 Suppl 1):S32-8.

67. McKenzie TL. 2009 C. H. McCloy Lecture. Seeing is believing: Observing physical activity and its contexts. Res Q Exerc Sport. 2010;81(2):113-22.

68. McKenzie TL, van der Mars H. Top 10 research questions related to assessing physical activity and its contexts using systematic observation. Res Q Exerc Sport. 2015;86(1):13-29.

69. Dunton GF, Liao Y, Intille SS et al. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring). 2011;19(6):1205-12.

70. Dunton GF, Whalen CK, Jamner LD et al. Mapping the social and physical contexts of physical activity across adolescence using ecological momentary assessment. Ann Behav Med. 2007;34(2):144-53.

71. Shiffman S, Stone AA, Hufford MR Ecological momentary assessment. Annu Rev Clin Psychol. 2008;4:1-32.

72. Trost SG, Loprinzi PD, Moore R et al. Comparison of accelerometer cut points for predicting activity intensity in youth. Med Sci Sports Exerc. 2011;43(7):1360-8.

73. Arem H, Keadle SK, Matthews CE. Invited commentary: Meta-physical activity and the search for the truth. Am J Epidemiol. 2015;181(9):656-8.

74. Masse LC, Fuemmeler BF, Anderson CB et al. Accelerometer data reduction: A comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc. 2005;37(11 Suppl):S544-54.

75. Cliff DP, Reilly JJ, Okely AD. Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0-5 years. J Sci Med Sport. 2009;12(5):557-67.

76. Esliger D, Copeland JL, Barnes JD, et al. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. J Phys Act Health. 2005;3:366-83.

77. Sherar LB, Griew P, Esliger DW et al. International Children's Accelerometry Database (ICAD): Design and methods. BMC Public Health. 2011;11(1):485.

78. Colley R, Gorber SC, Tremblay MS. Quality control and data reduction procedures for accelerometry-derived measures of physical activity. Health Rep. 2010;21(1): 63-9.

79. Yildirim M, Verloigne M, de Bourdeaudhuij I et al. Study protocol of physical activity and sedentary behaviour measurement among schoolchildren by accelerometry—Cross-sectional survey as part of the ENERGY-project. BMC Public Health. 2011;11:182.

80. De Meester F, De Bourdeaudhuij I, Deforche B et al. Measuring physical activity using accelerometry in 13-15-year-old adolescents: The importance of including non-wear activities. Public Health Nutr. 2011;14(12):2124-33

81. Ottevaere C, Huybrechts I, De Meester F et al. The use of accelerometry in adolescents and its implementation with non-wear time activity diaries in free-living conditions. J Sports Sci. 2011;29(1):103-13.

82. Loprinzi PD, Smit E, Cardinal BJ et al. Valid and invalid accelerometry data among children and adolescents: comparison across demographic, behavioral, and biological variables. Am J Health Promot. 2014;28(3):155-8.

83. Toftager M, Kristensen PL, Oliver M et al. Accelerometer data reduction in adolescents: Effects on sample retention and bias. Int J Behav Nutr Phys Act. 2013;10:140.

84. Zhuang J, Chen P, Wang C et al. Characteristics of missing physical activity data in children and youth. Res Q Exerc Sport. 2013;84 Suppl 2:S41-7.

85. Mannini A, Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors (Basel). 2010;10(2):1154-75.

86. Atkin AJ, Gorely T, Clemes SA et al. Methods of Measurement in epidemiology: Sedentary behaviour. Int J Epidemiol. 2012;41(5):1460-71.

87. Lubans DR, Hesketh K, Cliff DP et al. A systematic review of the validity and reliability of sedentary behaviour measures used with children and adolescents. Obes Rev. 2011;12(10):781-99

88. Healy GN, Dunstan DW, Salmon J et al. Television time and continuous metabolic risk in physically active adults. Med Sci Sports Exerc. 2008;40(4):639-45.

89. American Academy of Pediatrics. Committee on Public Education. Children, adolescents, and television. Pediatrics. 2001;107;423.

90. Foley L, Maddison R, Olds T et al Self-report use-of-time tools for the assessment of physical activity and sedentary behaviour in young people: Systematic review. Obes Rev. 2012;13(8):711-22.

91. Kim Y, Lee JM, Peters BP et al. Examination of different accelerometer cut-points for assessing sedentary behaviors in children. PLoS One. 2014;9(4):e90630.

JANUARY 2017

National Collaborative on Childhood Obesity Research