Fitness Measures and Health Outcomes in Youth

DETAILS

274 pages | 6 x 9 | HARDBACK
ISBN 978-0-309-38550-3 | DOI 10.17226/13483

AUTHORS

Russell Pate, Maria Oria, and Laura Pillsbury, Editors; Committee on Fitness Measures and Health Outcomes in Youth; Food and Nutrition Board; Institute of Medicine

Fitness Measures and Health Outcomes in Youth

Committee on Fitness Measures and Health Outcomes in Youth

Russell Pate, Maria Oria, and Laura Pillsbury, *Editors*

Food and Nutrition Board

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Grant No. 61747 between the National Academy of Sciences and the Robert Wood Johnson Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-26284-2
International Standard Book Number-10: 0-309-26284-4

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Cover credit: Image designed by Casey Weeks.

Suggested citation: IOM (Institute of Medicine). 2012. *Fitness measures and health outcomes in youth*. Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON FITNESS MEASURES AND HEALTH OUTCOMES IN YOUTH

RUSSELL R. PATE (*Chair*), Professor, Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia
CAMERON BLIMKIE, Professor, Department of Kinesiology, McMaster University, Hamilton, Ontario
DARLA CASTELLI, Associate Professor, Department of Kinesiology and Health Education, University of Texas, Austin
CHARLES B. CORBIN, Professor Emeritus, Department of Exercise and Wellness, Arizona State University, Phoenix
STEPHEN R. DANIELS, Professor and Chair, Department of Pediatrics, University of Colorado School of Medicine, and Pediatrician-in-Chief, Children's Hospital Colorado, Aurora
HAROLD W. KOHL III, Professor, School of Public Health and Michael & Susan Dell Center for Healthy Living, University of Texas Health Science Center-Houston, and Department of Kinesiology and Health Education, University of Texas, Austin
ROBERT M. MALINA, Professor Emeritus, Department of Kinesiology and Health Education, University of Texas, Austin and Research Professor, Department of Kinesiology, Tarleton State University, Stephenville, Texas
JENNIFER SACHECK, Associate Professor, John Hancock Research Center on Physical Activity, Nutrition, and Obesity Prevention, Friedman School of Nutrition Science and Policy, Tufts University, Boston, Massachusetts
DAVID STODDEN, Associate Professor, Department of Health, Exercise, and Sport Sciences, Texas Tech University, Lubbock
MELICIA WHITT-GLOVER, President and CEO, Gramercy Research Group, LLC, Winston-Salem, North Carolina
WEIMO ZHU, Professor, Department of Kinesiology and Community Health, College of Applied Health Sciences, University of Illinois at Urbana-Champaign

Study Staff

MARIA ORIA, Study Director
LAURA PILLSBURY, Program Officer
ALLISON BERGER, Senior Program Assistant
ALICE VOROSMARTI, Research Associate
ANTON L. BANDY, Financial Officer

GERALDINE KENNEDO, Administrative Assistant
LINDA D. MEYERS, Director, Food and Nutrition Board

Consultants

MICHAEL W. BEETS, University of South Carolina, Columbia
RONA BRIERE, Briere Associates, Inc., Arlington, Virginia
WILLIAM H. DIETZ, Centers for Disease Control and Prevention, Atlanta, Georgia (retired)
JOAN M. DORN, Centers for Disease Control and Prevention, Atlanta, Georgia
JANET E. FULTON, Centers for Disease Control and Prevention, Atlanta, Georgia
SARAH M. LEE, Centers for Disease Control and Prevention, Atlanta, Georgia
MELINDA MILLARD-STAFFORD, Georgia Institute of Technology, Atlanta
JANE WARGO, President's Council on Fitness, Sports, and Nutrition, Rockville, Maryland

Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Virginia Rall Chomitz, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, Massachusetts
Wm. Cameron Chumlea, Lifespan Health Research Center, Boonshoft School of Medicine, Wright State University, Dayton, Ohio
Angela Diaz, Department of Pediatrics and Department of Preventative Medicine, Mount Sinai School of Medicine, New York, New York
Avery Faigenbaum, Department of Health and Exercise Science, The College of New Jersey, Ewing
B. Don Franks, Department of Kinesiology, School of Public health, University of Maryland, College Park
Maxine Hayes, State of Washington Department of Health, Tumwater

Steven Heymsfield, Pennington Biomedical Research Center, Baton Rouge, Louisiana
Duane Knudson, Department of Health & Human Performance, Texas State University, San Marcos
Matthew Mahar, Department of Kinesiology, The College of Health & Human Performance, East Carolina University, Greenville, North Carolina
J. Michael McGinnis, Institute of Medicine, Washington, DC
Thomas McKenzie, School of Exercise and Nutritional Sciences, San Diego State University, California
Mark Tremblay, Department of Pediatrics, University of Ottawa, Canada

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the report's conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by **Diane Birt,** Professor, Department of Food Science and Human Nutrition, Iowa State University, and **Elena O. Nightingale,** Scholar-in-Residence, Institute of Medicine. Appointed by the Institute of Medicine, both were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Preface

This report, initiated at the request of the Robert Wood Johnson Foundation, belongs to a series that is part of a broad effort of the Institute of Medicine (IOM) on preventing childhood obesity. Past reports in the series have focused on areas in which preventive interventions could make a difference in the struggle against obesity. This report is the first to look directly at the role of youth physical fitness in health. The committee that conducted this study was charged with recommending the best health-related measures of various components of fitness for inclusion in a national youth fitness survey and, secondarily, recommending test items for administration in educational settings. As an aid in accomplishing this task, the committee was provided with a scientific literature search conducted and managed by the Centers for Disease Control and Prevention.

Physical fitness test batteries for youth have been designed and administered widely in the United States since the mid–20th century. While the components of fitness measured and the specific test items included in protocols have varied considerably across test batteries and over time, youth fitness testing has become a well-established institution in school physical education programs. In addition, national surveys of youth fitness were conducted periodically between the late 1950s and the mid-1980s; the period from the mid-1980s to the present, however, has seen a hiatus in such surveys, perhaps as the result of an increased emphasis on surveys of physical activity behavior as distinct from fitness. With the increased prevalence of overweight and obesity in American youth and expanded knowledge of the impact of fitness on health outcomes, interest in the fitness status of contemporary children and adolescents has grown. Accordingly, this report is intended to provide guidance for the designers of a national survey of health-related fitness in American youth. Also, recognizing the importance of fitness testing in school physical education programs and in accordance with its statement of task, the committee recommends health-related fitness test items that are seen as both valid and feasible for administration in educational settings.

An important and sometimes vexing aspect of health surveys is the establishment of criteria for interpreting the survey findings. Accordingly, this report also includes guidelines for setting standards for performance on the various fitness test items included in the recommended battery. In doing its work, moreover, the committee encountered many gaps in the relevant scientific evidence; thus an important element of the report is a set of recommendations for future research on fitness testing in youth.

The committee comprised 11 experts with extensive knowledge in a range of areas related to fitness and physical activity, metabolic health, growth and maturation, body composition, and physical education, as well as the development of standards and validation of tests. In addition to its reviews of the literature and discussions in closed meetings, the committee benefited from rich discussions with other experts in fitness and youth during an open session. On behalf of the committee, I sincerely thank the participants and speakers who contributed to this open session, providing information critical to the completion of this report.

On behalf of the committee, I also would like to express my deep appreciation to the consultants who led the literature search, a main resource for the committee's recommendations. Michael W. Beets, William H. Dietz, Joan Dorn, Janet E. Fulton, Sarah M. Lee, Melinda L. Stafford-Millard, and Jane Wargo not only conducted the main literature search but also facilitated the committee's work by providing abstractions and summaries of the literature as requested. I would like to express my appreciation as well to Laura C. Leviton, senior adviser for evaluation at the Robert Wood Johnson Foundation, whose leadership has provided impetus for so many efforts on obesity prevention at the IOM, and to Tina J. Kauh, research and evaluation program officer at the Robert Wood Johnson Foundation.

I would also like to gratefully acknowledge the effort and skill the committee members brought to this study. Their backgrounds, experiences, and passion for the subject matter resulted in a report that will have a long life. Finally, I thank the project staff of the National Academies—Maria Oria, study director; Laura Pillsbury, program officer; and Allison Berger, senior program assistant—for their tireless dedication to the production of this report.

Robert R. Pate, *Chair*
Committee on Fitness Measures and Health Outcomes in Youth

Contents

Summary

Physical fitness is a state of being that reflects a person's ability to perform specific exercises or functions, and is related to present and future health outcomes. In the United States, serious efforts to assess the physical fitness of youth with a battery of tests began in the 19th century. These efforts intensified during times of war, focused primarily on improving athletic performance and military preparedness. Over time, the focus of such surveys shifted to assessing health rather than performance, reflecting growing concern about the current and future health of the nation's youth. While measures of performance-related fitness are designed to evaluate a person's capability to carry out certain physical tasks or activities, the focus of health-related fitness testing is on the concurrent or future health status of the subject under assessment.

The first U.S. national survey of youth fitness in 1958 was followed by surveys in 1965 and 1975 and then in 1985-1986. States and schools have continued to assess fitness in youth during the past two decades; after the 1985-1986 survey, however, there was no national-level assessment of youth fitness until 2012, as part of the National Health and Nutrition Examination Survey. Several factors may account for this gap, including the fact that interest and effort have been directed more toward understanding the role of physical activity in youth. This shift and the challenges inherent in associating fitness in youth with health have resulted in few advances in our understanding of the physiology and outcomes of fitness.

Assessment of fitness historically has encompassed such components as body composition, cardiorespiratory endurance, musculoskeletal strength and endurance, and flexibility. Examples of tests used historically in national

surveys and schools are the progressive shuttle run and mile run tests for cardiovascular endurance, measurement of body mass index (BMI) for assessment of body composition, the curl-up and push-up tests for muscular endurance, and the sit-and-reach tests for flexibility.

While the components of fitness measured and the specific test items included in protocols have varied considerably across test batteries and over time, youth fitness testing has become a well-established institution in school physical education programs during the past half-century. In a school context, tests are being used as institutional fitness assessment tools, as educational tools to teach youth and their families about the importance of physical fitness, and as communication tools to guide individuals on attainable goals for maintaining fitness and health. These efforts are costly in terms of not only human capacity and financial resources, but also the extensive training and organizational and communication skills required for their implementation. Most important, it is essential to use appropriate tests and understand the results in a health context to minimize misclassification and stigmatization of youth. Selection of the best tests is therefore a crucial process, and knowledge gaps in this area were an important motivator of the present study.

This study was undertaken in light of the past challenges encountered in identifying fitness tests related to health in youth, spurred by a renewed interest in fitness as one of the key tenets of health.

STUDY APPROACH

Given the gaps in knowledge noted above, the Institute of Medicine (IOM) convened an 11-member committee with expertise in fitness measures, body composition and maturity, physical activity, physical education, the development of cut-points (cutoff scores), motor development and skill, and modifiers of fitness to conduct this study. The committee was asked to assess the relationships between fitness tests and health outcomes in youth based on a review of the literature designed and conducted by the Centers for Disease Control and Prevention (CDC) (the committee's statement of task is shown in Box S-1). The CDC search criteria included longitudinal, experimental study designs in which fitness and health were measured in healthy[1] children aged 5-18 during 2000-2010. The CDC searches were conducted specifically for the fitness components cardiorespiratory endur-

[1]The criteria included overweight and obese youth, but excluded youth with various disabilities or congenital diseases. Since the primary task for this study was to identify fitness tests appropriate for a national youth fitness survey of the general youth population, the committee did not review additional literature specific to populations with disabilities, such as those with cognitive or physical impairments, activity limitations, or participation restrictions (as defined in Appendix B).

BOX S-1
Statement of Task

An ad hoc committee will recommend physical fitness test items for assessment of youth fitness components that are associated with health outcomes. The recommended items will be suitable for inclusion in a national survey of fitness in children and youth. The committee will make use of a systematic review of the literature conducted by the Centers for Disease Control and Prevention. In examining the review, the committee will evaluate the relationships between the fitness components and health outcomes (e.g., cardiovascular disease risk factors, musculoskeletal health, diabetes, obesity and others). Further, for selected fitness components the committee will examine the relationships between performance on specific test items and health outcomes.

In addition to the primary task above, the committee will answer the following questions:

1. For recommended test items for which there is evidence of an association with health, how should performance for the test items be interpreted? Should the interpretation be based on a cut-point approach? Are there alternative approaches to interpret performance?
2. If the association between a particular test and health outcomes reveals no obvious relationship to health, what strategy is most appropriate for identifying a criterion-referenced standard? In such a case, the committee may consider the use of norm-referenced standards.
3. How do demographic characteristics and overweight and obesity affect the tests scores and subsequent evaluations?
4. What additional research is needed to augment the evidence (or lack thereof) about the associations between fitness measures and health outcomes?

The committee will also study to what extent is change in performance on a fitness test item (e.g., handgrip strength or 1.5-mile walk/run) associated with change in health outcomes in youth who are apparently "healthy" but include both obese and nonobese. In addition, the committee will identify the strengths and weaknesses of fitness test items in regards to their practicality and as indicators of health outcomes in a school setting and, based on practicality, will provide recommendations for the most appropriate measures for each fitness component.

ance and musculoskeletal strength and endurance. The relationship of body composition measures to health is well established, so a systematic review of their relationship to health was not conducted. Although time and resources did not allow for a systematic review of the flexibility component, the committee evaluated the relationships between flexibility and health outcomes in studies from the CDC review that included a flexibility measure.

To guide its review and deliberations, the committee created a conceptual framework that depicts the potential relationships between physical fitness components and health as they are modified by various factors, such as demographic characteristics, maturity status, motor skills, and genetics. In addition to this conceptual framework, the committee developed the following set of general criteria for selecting tests to be implemented in the field:

- identification of a relationship between a fitness component and health markers;
- evaluation of the quality of the studies and the strength of the evidence for a relationship between fitness test items and health markers in youth, based on the CDC's systematic review;
- identification of health-related test items;
- evaluation of the integrity of test items (i.e., validity and reliability); and
- evaluation of the feasibility of implementing test items.

In its statement of task, the committee was asked not only to select test items for a national fitness survey but also to consider the practicality of their implementation. The committee reviewed only the evidence for field-based methods because, even if they are more prone to error than laboratory methods, they require less highly specialized training and are conducted with mobile equipment, adequate for assessing large samples of youth. The committee also recognized that national surveys and schools and other educational settings[2] raise different implementation issues. In addition, the conduct of fitness tests in schools may be driven by goals beyond health, such as educating about the importance of specific fitness components. Because of their role as educational tools, certain test items will be beneficial in a school fitness test battery even if their relationship to health cannot as yet be confirmed in youth. Therefore, the committee developed separate sets of recommendations for these two settings.

Implementing the best health-related fitness items entails important steps that relate to the interpretation and communication of the test results in a health context. Identifying one or more health outcomes that are

[2]Other educational settings include, for example, gymnasiums and fitness centers.

related to the test items of interest, then, is essential. Equally important is understanding the relationship between the test items and the associated health outcomes in quantitative terms so the results can be interpreted in a health context. For this purpose, a criterion-referenced cut-point—a test performance score below (or above) which a risk to health may exist—can be used. Ideally, criterion-referenced cut-points would be derived from population-based data on the relationship between a fitness test and a health outcome or marker in youth. As noted earlier, however, data on the relationship between fitness and health in youth are limited, mainly because of the difficulty of identifying such associations when health constructs in youth are not well defined. When data in youth populations are not available, alternative approaches can be followed to derive cut-points (interim cut-points). Box S-2 provides the committee's guidance on methods for selecting criterion-referenced cut-points and interim cut-points for health-related fitness testing in youth.

CONCLUSIONS AND RECOMMENDATIONS

The committee developed conclusions and recommendations regarding fitness measures for youth for each of four components of fitness: body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility. For each of these components, the committee identified test items, reviewed the evidence on these items, applied the general criteria for selection listed earlier, considered modifying factors, assessed the feasibility of implementation, and applied the guidance in Box S-2 for selecting cut-points. In general, the studies reviewed provided insufficient data with which to assess the influence of several potential modifiers—age, gender, race/ethnicity, body composition, maturation status, motor skill—on performance on tests of cardiorespiratory endurance, musculoskeletal fitness, and flexibility.

As noted earlier, the committee's recommendations are specific to the implementation of fitness measures either in a national youth fitness survey or in schools and other educational settings. The recommended tests for a national youth fitness survey represent valid, reliable, feasible, and safe tests for the assessment of health-related fitness in youth for population-level health-monitoring purposes. National survey fitness tests are intended to be implemented by skilled national administrators of such surveys (i.e., those familiar with the procedures for conducting large surveys and the protocols for fitness tests) in school settings. For schools, recommendations are made for fitness tests that are low in cost and equipment requirements such that they are practical for school-based implementation. Regardless of the setting, test administrators and those interpreting and communicating the results should receive appropriate training in conducting and interpret-

BOX S-2
Guidance for Developing Cut-Points

The committee determined that a criterion-referenced method should be employed in developing cut-points. That is, a test taker's performance should be compared against an absolute criterion that is related to health. The following are options, depending on the available evidence:

- When a confirmed concurrent relationship exists between health outcome measures and fitness tests in youth, criterion-referenced cut-points can be determined by using a data-mining procedure that establishes the statistical evidence for that relationship.
- When a confirmed concurrent relationship exists only in adults but not in youth, either a relative position or a panel-driven method can be used, whereby interim criterion-referenced cut-points in youth are derived from the percentile values (related to health outcomes) extrapolated from the adult population or by a panel of experts using cut-points for adults and other available information (e.g., growth curves and performance characteristics for different ages and genders).
- When no confirmed relationship exists in either youth or adults, a comparatively relative position method can be employed, whereby interim criterion-referenced cut-points are derived from the percentile values (related to health outcomes) extrapolated from a different test. When the percentile from another test is used, the two tests should be as comparable as possible in their nature (e.g., both require movement of the body) and in the dimension they measure (e.g., upper-body strength).

ing the tests to minimize measurement and classification errors and prevent adverse events. Finally, the committee offers recommendations for future research that would advance understanding of youth fitness measures and their association with health outcomes.

Conclusions About Components of Fitness

The committee's conclusions relate to the four components of fitness detailed above: body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility.

Body Composition

Body composition denotes the sum of the basic components that make up body weight, including fat, muscle, and bone content. The committee defined body composition operationally as a component of fitness, a health marker, and a modifier of fitness. Field-based measures of body composition relate to different dimensions. For example, skinfold is an indicator of subcutaneous fat, whereas waist circumference is an indicator of abdominal adiposity, and BMI measures body weight-for-height. These measures also vary in that they have been associated with different health markers; for example, skinfold measures are related to risk factors for cardiovascular disease and metabolic syndrome, waist circumference has been associated with cardiovascular disease, and BMI is related to risk of diabetes and hypertension. When implementing and interpreting measures of body composition, it is important to note that many factors, such as physical activity, calorie consumption, age, and maturation, influence body composition measures. The committee selected measures of body composition based on their relationship to health markers, their integrity, and their feasibility.

The committee concluded that the above three measures of body composition—skinfold, waist circumference, and BMI—are important to collect in a national youth fitness survey. Each is a proximal estimation of body fat and has increased standard of error over laboratory measures. Moreover, the measurement of body composition is multidimensional; no single measure is considered representative of all body composition tenets for youth of all morphologies.

In selecting measures of body composition, some feasibility factors must be considered: the availability of administrators with the highly specialized training required and the accessibility of appropriate space in which to conduct the test. The reliability of skinfold and waist circumference measurements depends on the skill of the test administrator; to avoid the introduction of errors in the measurements, specific and intense training is required. Training is also required to minimize concerns related to privacy in the administration of these measures. Also to ensure privacy, the appropriate space should be available for conducting the tests. Given the challenges associated with avoiding measurement errors, maintaining good reliability, and ensuring privacy in the administration of skinfold and waist circumference measurements, only BMI measurement is recommended for administration in schools.

Cardiorespiratory Endurance

Cardiorespiratory endurance is the ability to perform large-muscle, whole-body exercise at moderate to high intensity for an extended period of time. There is a well-established association between cardiorespiratory endurance and health outcomes in adults and health markers or risk factors in children—in particular, body weight, body composition, cardiometabolic risk factors, blood pressure, cognitive function, and pulmonary function. Although the fitness tests and protocols used vary substantially, the cardiorespiratory endurance tests associated most frequently and strongly with a positive change in health markers or risk factors are heart rate extrapolation tests (i.e., those that use a treadmill or cycle ergometer and measure cardiorespiratory endurance as maximal oxygen consumption [VO_2max]) and the progressive shuttle run. The health markers most frequently assessed are related to body weight or adiposity and cardiometabolic risk factors. The heart rate extrapolation and progressive shuttle run tests have high validity and reliability. In terms of feasibility, the progressive shuttle run is advantageous when time and financial constraints exist with respect to the necessary training and equipment. Treadmill and cycle ergometer tests are valid and reliable alternatives for a national survey in which space limitations are a concern, but extensive training is feasible. The validity and reliability of distance runs are more variable and in general lower than has been reported for the heart rate extrapolation and progressive shuttle run methods; however, these tests are appropriate for a school setting for practical reasons.

Musculoskeletal Fitness

Musculoskeletal fitness is a multidimensional construct that encompasses three related components: muscle strength (the ability of skeletal muscle to produce force under controlled conditions), muscle endurance (the ability of skeletal muscle to perform repeated contractions against a load), and muscle power (the peak force of a skeletal muscle multiplied by the velocity of the muscle contraction). Neither any of these components individually nor any single test can describe overall musculoskeletal fitness. Therefore, a number of tests that measure various dimensions of musculoskeletal fitness often are used in combination. As with other fitness components, a wide variety of tests, such as the curl-up, the push-up, the handgrip, and jumps, have been used to measure musculoskeletal fitness in the past.

The committee concluded that adequate experimental and prospective longitudinal evidence supports the relationship between the multidimensional construct of musculoskeletal fitness and health. Empirical evidence also is increasing for the importance of musculoskeletal fitness, especially muscle strength and power, to health outcomes in adults. There is, however, insufficient high-quality evidence to support a strong association between any single musculoskeletal fitness test item and health markers in youth. Based predominantly on evidence indicating a relationship to health outcomes in adults, the committee concluded that musculoskeletal fitness should be assessed in a national youth fitness survey.

Growing evidence supports use of the handgrip strength test and the standing long jump as health-related musculoskeletal fitness test items in youth. Studies reviewed show a relationship between performance on these tests and bone health and body composition. The handgrip strength test demonstrates moderate to strong validity when assessed against upper- and lower-body criterion muscle strength measures. The standing long jump, although not strictly a measure of muscle strength, demonstrates acceptable validity against lower- and upper-body criterion muscle strength measures and lower-body muscle power measures. The handgrip strength and standing long jump tests demonstrate strong and moderate reliability, respectively. The committee recommends that the handgrip strength and standing long jump tests be included in a national survey. While these tests should not be interpreted in a health context until their relationships with health outcomes have been established more firmly in youth, they can be included for their educational value. Other measures of muscular strength, such as the modified pull-up or push-up as an alternative for measuring upper-body musculoskeletal strength and power or the curl-up for measuring core strength and endurance, also can be used in schools.

Flexibility

Flexibility is the intrinsic property of body tissues, including muscle and connective tissue, that determines the range of motion achievable without injury at a joint or group of joints. Like musculoskeletal fitness, flexibility is specific; a person can have a good range of motion around a shoulder joint, for example, but lack range of motion in the hip. Such specificity precludes any relationship between a given measure of flexibility and any systemic health markers (e.g., back pain, risk of injury, posture problems). Moreover, clinical theory suggests that the complex interaction among multiple musculoskeletal components (e.g., flexibility, strength, endurance), rather than one component alone, is most likely to be associated with health markers or outcomes. Further, possible associations with health are complicated by the fact that risk may be higher for those with low or exceptionally high flexibility than for those in the middle ranges. Finally, although evidence suggests a link between flexibility and health among adults (e.g., low-back pain), such evidence is more difficult to establish in youth given that the commonly used health risk outcomes may take years to manifest.

The literature review did not reveal a relationship between any flexibility test and health in youth. In addition to the challenges mentioned above, this could be due to the study designs included in this review. Specifically, in contrast to studies on other fitness components, there was a lack of high-quality longitudinal and experimental studies measuring the association between flexibility and health markers in youth. For example, many studies did not include health markers hypothesized to be related to flexibility and typically did not include a control. Future efforts to study the relationship of flexibility to health will require a multivariate approach. Although no relationship to health has been shown, the sit-and-reach test is feasible to implement and has acceptable validity and reliability.

Recommendations for National Surveys

A substantial body of evidence supports the idea that specific tests measuring cardiorespiratory endurance and body composition are related to health markers in youth. The evidence for an association between musculoskeletal fitness and health markers in youth is less extensive. The committee concluded that insufficient evidence has thus far been accumulated to support recommending a health-related measure of flexibility for youth at this time.

The committee concluded that a criterion-referenced approach using cut-points associated with health markers is the ideal approach for interpreting scores. There is, however, insufficient evidence with which to develop age- and gender-specific criterion-referenced cut-points for all measures except for BMI. Until data are collected with which to establish criterion-referenced cut-points, age- and gender-specific interim cut-points corresponding to percentiles for adults on tests related to the same component or for youth on tests related to a different or the same component should be used.

RECOMMENDATION 8-1.[3] A national survey of health-related physical fitness in youth should include measures of cardiorespiratory endurance, body composition, and musculoskeletal fitness. The survey should include the following fitness test items: (1) measures of BMI, waist circumference, and skinfold thickness (triceps and subscapular sites) to assess body composition; (2) a progressive shuttle run, such as the 20-meter shuttle run (or a submaximal treadmill or cycle ergometer test if there are space limitations) to measure cardiorespiratory endurance; and (3) handgrip strength and standing long jump tests to measure musculoskeletal fitness.

RECOMMENDATION 8-2. Standard protocols for the administration of measures of youth fitness in national surveys should be developed

[3]The committee's recommendations are numbered according to the chapter of the main text in which they appear.

and implemented. The focus should be on maximizing the measures' reliability, validity, and safety. Trained personnel should be used for test administration and data collection.

RECOMMENDATION 8-3. Developers of fitness test batteries should use age- and sex-specific cut-points to determine which individuals are at risk of poor fitness-related health outcomes. Optimum cut-points should be based on criterion values when population-based evidence is available on the relationship between the level of performance on a fitness test and a health outcome or marker. In the absence of criterion values, interim population-based percentile values should be applied. These values might be derived from adults on tests for the same component or from youth on tests for a different or the same component. Specifically, the guidance of the committee should be applied as follows:

- **Body composition: For BMI, the CDC-established cut-points for underweight, overweight, and obesity evaluations should be used. Interim cut-points for skinfold and waist circumference measures could be derived from the CDC-established percentiles for BMI.**
- **Cardiorespiratory endurance: For measures of cardiorespiratory endurance, interim cut-points could be derived from the lowest performers (e.g., 20th percentile) on the cardiorespiratory endurance distribution curve.**
- **Musculoskeletal fitness: For musculoskeletal fitness tests, interim cut-points could be derived by borrowing the percentile (e.g., 20th percentile) from the cardiorespiratory endurance tests.**

Recommendations for Schools and Other Educational Settings

The preceding recommendations outline the optimum test items for measuring fitness in youth in national surveys. Conducting fitness tests in educational settings can yield further benefits, such as contributing to the body of evidence on the association between health-related fitness components and learning outcomes, improving individuals' fitness performance, and educating about the importance of physical fitness. The committee considered the strengths and weaknesses of the test items recommended for a national survey with regard to their practicality in schools and other educational settings.

School leaders and teachers should apply the following recommendation and select applicable test items in light of the contextual variables that characterize their schools, such as available equipment, space, and test administrators, as well as cost, as schools differ greatly on these variables. Factors related to culture and race/ethnicity, as well as how a test item aligns with the existing curriculum, should also be considered. Finally, perhaps the most important element of fitness testing in schools is the interpretation and dissemination of results. This element represents an opportunity to assist participants in preventing disease and understanding fitness, but can have detrimental effects on the individuals involved if not carried out appropriately. As mentioned above, training in the administration of protocols and the interpretation and communication of test results is essential.

RECOMMENDATION 9-1. Developers and administrators of fitness test batteries in schools and other educational settings should consider including the following test items:

- **standing height (measure of linear growth status) and weight (measure of body mass) to calculate BMI as an indicator of body composition;**
- **a progressive shuttle run, such as the 20-meter shuttle run, to measure cardiorespiratory endurance; and**
- **handgrip strength and standing long jump tests to measure upper- and lower-body musculoskeletal strength and power, respectively.**

Additional tests that have not yet been shown to be related to health but that are valid, reliable, and feasible may also be considered as supplemental educational tools. For cardiorespiratory endurance, alternatives to the shuttle run include distance and/or timed runs, such as the 9-minute or 1-mile run, while the modified pull-up and push-up are possible alternatives for measuring upper-body musculoskeletal strength. The curl-up may be considered in addition to the suggested musculoskeletal fitness tests for measuring core strength and endurance. Although the committee does not recommend a flexibility measure as a core component of a fitness test battery, administrators in schools and other educational settings may wish to include the sit-and-reach test or its alternatives (e.g., backsaver sit-and-reach) to measure flexibility. Experts who establish cut-points for interpreting performance on these fitness test items should follow the guidance provided earlier (Box S-2 and Recommendation 8-3).

Recommendations for Future Research

Altogether, the CDC's literature review revealed many gaps in understanding of the relationship between fitness measures and health in youth. Although the review revealed a number of associations between the two, many of the studies reviewed were not designed to assess the independent association between performance on a fitness test and a health outcome or marker. Moreover, while not included in the search strategy, studies predicting health outcomes in adulthood would be valuable in characterizing the importance of a health marker. For example, it remains to be determined whether changes in muscle strength and power during youth are predictive of health outcomes in later life.

The committee offers the following recommendations for designing and conducting research on some of the most pressing questions that must be answered if progress is to continue in selecting the best measures of fitness in youth. It should be noted that the committee is recommending research only for those test items that have been studied well enough to justify their inclusion here. At the same time, it is not the intent of the committee to eliminate from future consideration those test items that currently do not meet the level of evidence necessary for inclusion in a battery of tests.

RECOMMENDATION 10-1. Well-designed research studies aimed at advancing understanding of the associations between fitness components and health in youth should be undertaken. Researchers should ensure that the interventions studied are both specific and sufficient (i.e., appropriate dosage and duration) to induce a change in fitness. In addition, studies should be designed so that the effect of potential confounders (e.g., nutrition, physical activity, demographic variables, maturity status) and the potential for adverse events can be analyzed.

RECOMMENDATION 10-2. Longitudinal studies should be conducted to provide empirical evidence concerning how health markers related to fitness track from youth into adulthood.

RECOMMENDATION 10-3. Randomized controlled trials and longitudinal studies should be undertaken to understand the following issues regarding the relationships between (1) specific fitness tests and health, and (2) fitness components and health:

- **Studies should explore the relationship between body composition measures and physical fitness tests and the potential interactions among body composition, fitness, and health in youth.**
- **Studies should examine the relationship between changes in cardiorespiratory endurance as measured by field tests, including the shuttle run and timed and distance runs, and subsequent changes in health risk factors in youth beyond weight status and cardiometabolic risk factors. Examples include bone health and neurocognitive function and behavior.**

- Studies should address the relationship between specific musculoskeletal fitness test items and health markers in youth. Priority should be given to test items for which there is growing evidence, such as the handgrip strength or standing long jump test, or others that are promising. Since musculoskeletal fitness is a multivariate construct, the studies should be designed so that a variety of tests are conducted.
- Studies should investigate the relationship between specific flexibility test items (e.g., sit-and-reach and its modifications), either by themselves or in combination with musculoskeletal fitness test items, and potential health markers (e.g., back pain, posture, injury prevention). Such studies should include stretching interventions specifically designed to produce changes in joint-specific flexibility. Since flexibility is a multivariate construct, the studies should be designed so that a variety of tests are conducted. Researchers should investigate the development and validation of a general marker of musculoskeletal systemic flexibility and its relationship to health markers and risk factors.
- Studies should examine the potential effects of modifying factors (i.e., age, gender, race/ethnicity, body composition, maturity status, training status/practice, motor skill, socioeconomic factors) on fitness components and on the relationship between a change in a health-related fitness component and health markers in specific populations.

RECOMMENDATION 10-4. Developers of national surveys of health-related physical fitness in youth should consider the inclusion of measures of cardiometabolic health, bone health, and neurocognitive function. The collection of fitness and health data in the same individuals would allow investigators to further confirm whether direct relationships between specific test items and health markers and risk factors exist.

RECOMMENDATION 10-5. When an association between a fitness test and a health marker is confirmed, research should be conducted to establish and validate health-related cut-points for that test. For example, given the association of skinfold measures with health markers, large national studies should be conducted to establish health-related cut-points for skinfold measures in youth.

1

Introduction

Health is influenced by many factors, genetic, behavioral, and environmental, that are present prior to conception and continue throughout an individual's life span. Since childhood health predicts adult morbidity and mortality, it is beneficial to attain, sustain, and monitor health from childhood to adulthood. The United States and other countries have designed programs to measure or improve the health of the population in many different domains, including physical fitness, and at all ages. Key to setting national health agendas and priorities, as well as goals for individuals, is having goals for public health, metrics with which to determine health reliably and accurately in various areas, and an understanding of how close the population or individuals are to the established goals.

Examples of surveillance programs designed to measure health in various dimensions in the United States are the National Health Interview Survey (NHIS), the National Immunization Survey (NIS), the Behavioral Risk Factor Surveillance System (BRFSS), the National Health and Nutrition Examination Survey (NHANES), and the National Youth Physical Activity and Nutrition Study, to name a few. In addition to such national surveys, states may conduct their own surveys to track health status. Health programs can be established in the school environment or be part of the school curriculum (e.g., physical education classes), with the purpose of improving or evaluating health status among youth. While surveys of physical activity among youth have been carried out in recent years, however, national fitness surveys have not been conducted since the 1980s (see Chapter 2). The

NHANES includes components pertinent to physical fitness and a 2012 NHANES Youth Fitness Survey is currently under way.

As discussed in Chapter 2, there have been many efforts to identify fitness tests and standardize a battery of such tests for youth. To this day, however, an array of tests continues to be used, selected based on various historical circumstances and understandings of the science. This report represents an effort to provide an evidence-based approach to selecting field-based fitness measures for youth for inclusion in a national fitness survey. Recommendations for tests to be used in schools and other educational settings are provided as well.

BACKGROUND

To better understand the content of this report, it is important to distinguish between physical fitness and physical activity. *Physical fitness* has been defined as "a set of attributes that people have or achieve that relates to the ability to perform physical activity" (HHS, 1996, p. 21). The focus of this report is on the potential health-related components of physical fitness: body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility. *Physical activity*, on the other hand, is defined as "any body movement produced by muscle action that increases energy expenditure" (Castillo-Garzon et al., 2006, p. 213). There are many types of physical activities, such as exercise (physical activity with the purpose of improving fitness), sports, dance, and recreational activities. Box 1-1 and Appendix B provide the committee's operational definitions of physical fitness and other terms used throughout the report.

Fitness tests are conducted for several purposes for both individuals (e.g., goal setting, planning for improvement, preparing for specific tasks) and society at large (e.g., assessing current fitness status, tracking changes, research). The ultimate purpose, however, is to improve the health and physical performance of individuals, as well as the population as a whole. As noted above and described in detail later in this report, fitness surveys have been conducted in the United States at both the national and state levels. Similarly, other countries have developed fitness test batteries and conducted national surveys (see Chapter 2).

Early national fitness tests included items commonly described as skill-related fitness, as well as items focused on health-related fitness. Since the first national fitness test was developed in 1958, appropriate items for inclusion in fitness test batteries have been the subject of debate. The first national health-related physical fitness test was developed in 1980 (AAHPERD, 1980), and since then there has been increased emphasis on defining the relationship of fitness items to health. While measures of performance-related fitness are designed to evaluate a person's capability to carry out certain physical tasks

BOX 1-1
Terms Used in This Report

Body composition: the components that make up body weight, including fat, muscle, and bone content.

Cardiorespiratory endurance: the ability to perform large-muscle, whole-body exercise at moderate to high intensities for extended periods of time (also referred to as aerobic fitness or aerobic capacity) (Saltin, 1973).

Criterion-referenced standards (criterion measures): evaluation standards used to interpret physical fitness test scores and provide information about a participant's health status.

Cut-point (cutoff score): a test score that represents the minimum level of performance that must be achieved for a participant to be said to be at reduced risk or fit/healthy.

Flexibility: "the intrinsic property of body tissues that determines the range of motion achievable without injury at a joint or group of joints" (Holt et al., 1996, p. 172).

Musculoskeletal fitness: a theoretical construct reflecting the integrated function of an individual's muscle strength, endurance, and power to enable the performance of work against one's own body weight or an external resistance.

Physical activity: "any body movement produced by muscle action that increases energy expenditure" (Castillo-Garzon et al., 2006, p. 213).

Physical fitness: "a set of attributes that people have or achieve that relates to the ability to perform physical activity" (HHS, 1996, p. 21).

Reliability: the dependability of test scores, their freedom from error, and their reproducibility in repeated trials on the same individual.

Validity: the extent to which a test measures what it is designed to measure; the degree to which evidence supports the interpretation of test scores (Eignor, 2001).

or activities, the focus of health-related fitness testing is on concurrent or future health status. The measurement of health-related fitness in youth is the focus of this report. As more sophisticated research and statistical methods, computer technologies, and data management systems have emerged, the link between fitness tests and health has been more firmly established. Nevertheless, there is more to be done. This report is based on a systematic review of the literature designed to answer key questions concerning fitness and health in youth.

STATEMENT OF TASK

This study was undertaken to identify measures of fitness for which there is evidence of an association with health outcomes and to provide guidance for interpreting fitness test scores (e.g., setting health-related cutpoints for specific tests). The committee was asked to be attentive to the practicality of the recommended tests and to discuss considerations and pros and cons for these tests. The specific questions posed in the committee's statement of task are shown in Box 1-2.

METHODS

An 11-member committee was convened to answer the questions posed in the statement of task shown above. The committee members had extensive expertise in fitness and physical activity and were selected specifically for their knowledge of youth health issues, body composition and maturation, and motor coordination; methodologies for developing fitness measures related to health; physical education, physical activity, and fitness in schools; and national fitness surveys. Many committee members also are familiar with the various fitness test batteries that have been used throughout history and in different countries and that have responded to specific situations and purposes. Committee members are knowledgeable as well about the many factors (e.g., demographic characteristics) that interact with youth performance on tests for the various fitness components. Because the statement of task also requested that the committee be mindful of practical considerations when selecting fitness tests for use in the field, many of the committee members have practical experience with implementing fitness test batteries.

In addition to its members' extensive knowledge of fitness and health, the committee drew on other sources to inform its decisions about the selection of fitness test items. A major resource for inferring relevant associations between specific fitness test items and health markers in youth was a systematic review of the peer-reviewed scientific literature, designed and conducted by the Centers for Disease Control and Prevention (CDC) and encompassing the period 2000-2010. Further detail on the conduct of this review is presented in Chapter 3. For two fitness components—cardiorespiratory endurance and musculoskeletal fitness—the committee received the results of the review in the form of abstracted tables along with the full articles, and then selected the articles to review in depth based on its assessment of the quality of the research. Although articles on flexibility were not coded separately in the literature review, the committee reviewed several studies focused on the other fitness components that included a flexibility measure. A systematic review of the literature with respect to body

BOX 1-2
Statement of Task

An ad hoc committee will recommend physical fitness test items for assessment of youth fitness components that are associated with health outcomes. The recommended items will be suitable for inclusion in a national survey of fitness in children and youth. The committee will make use of a systematic review of the literature conducted by the Centers for Disease Control and Prevention. In examining the review, the committee will evaluate the relationships between the fitness components and health outcomes (e.g., cardiovascular disease risk factors, musculoskeletal health, diabetes, obesity and others). Further, for selected fitness components the committee will examine the relationships between performance on specific test items and health outcomes.

In addition to the primary task above, the committee will answer the following questions:

1. For recommended test items for which there is evidence of an association with health, how should performance for the test items be interpreted? Should the interpretation be based on a cut-point approach? Are there alternative approaches to interpret performance?
2. If the association between a particular test and health outcomes reveals no obvious relationship to health, what strategy is most appropriate for identifying a criterion-referenced standard? In such a case, the committee may consider the use of norm-referenced standards.
3. How do demographic characteristics and overweight and obesity affect the tests scores and subsequent evaluations?
4. What additional research is needed to augment the evidence (or lack thereof) about the associations between fitness measures and health outcomes?

The committee will also study to what extent is change in performance on a fitness test item (e.g., handgrip strength or 1.5-mile walk/run) associated with change in health outcomes in youth who are apparently "healthy" but include both obese and nonobese. In addition, the committee will identify the strengths and weaknesses of fitness test items in regards to their practicality and as indicators of health outcomes in a school setting and, based on practicality, will provide recommendations for the most appropriate measures for each fitness component.

composition also was not conducted because, even though this component is frequently included in fitness test batteries, its relationship to health is well known. Although the committee did not participate in the design of the literature review, members had ample opportunities to interact with the CDC in order to understand the nature of the review. The CDC literature review also did not include integrity and feasibility studies. The committee conducted further literature searches and reviews in other areas, for example, to assess the integrity of specific fitness tests or to complement the CDC's systematic review.

In addition, the committee drew on the work and experience of other organizations and countries to the extent that this information is available to the public. The committee also benefited from expert presentations during an open session on November 15-16, 2011; the agenda for this open session is in Appendix A. Presenters had extensive experience in the development of fitness test batteries and in the associations of fitness with metabolic risk factors and body composition. Other presenters had experience in implementing and interpreting results of a battery of fitness tests in the field, providing the committee with insight into feasibility considerations and challenges encountered at the time of test implementation.

The main purpose of this report is to identify fitness measures that are associated with health markers in youth and that are also practical in a field setting. To accomplish this purpose, the committee agreed on various concepts and on a general conceptual framework that guided its decisions. Before reviewing the literature, the committee decided on a stepwise process for identifying the best test items for each fitness component. As will be obvious from the description in Chapter 3, the literature review was designed to provide information about potential associations of fitness tests with health markers (or risk factors) and outcomes in youth as opposed to those that might be seen later in life. For that reason, the majority of health issues explored during the review were health markers (or risk factors) for a disease and not health outcomes per se, as most health conditions or diseases do not manifest until adulthood. As is clear from the discussion throughout the report, studies that follow youth into adulthood are infrequent. Since the 1980s, moreover, organizations and relevant government agencies have focused their efforts on the health benefits of physical activity among youth rather than on fitness, which was the focus prior to the 1980s. The lack of a recent focus on fitness has resulted in a less than ideal scientific literature base addressing questions of fitness and health. Nevertheless, the knowledge base has increased sufficiently to support the conduct of a national fitness survey. The focus on health in youth is a unique feature of this report and one that presented many challenges given the inadequate amount and nature of the relevant literature. However, this focus is in tune with current thinking that factors related to health in adults cannot necessarily be extrapolated to youth, and therefore, health markers in youth need to be defined and reviewed.

While the committee provides guidance for developing cut-points (cut-off scores) for and interpreting performance on fitness tests, it did not develop specific cut-points for the recommended test items. Rather, the committee suggests an ideal approach to establishing cut-points. Recognizing that all the data necessary to establish cut-points do not exist for all the recommended tests, the committee also provides alternative approaches for establishing interim cut-points when such data are unavailable. In addition, there are aspects of fitness testing that the committee did not address in depth, such as protocols for the recommended tests, specific training for test administrators, or the appropriateness of fitness components that were not included in the committee's statement of task. Finally, the studies reviewed were designed to collect evidence on the relationship between fitness tests and health in healthy youth. Studies on overweight and obese youth were included in the review; however, studies in special populations, such as athletes or people with disabilities[1] or congenital diseases, were not reviewed. Therefore, the committee's findings, conclusions, and recommendations do not target those special populations.

ORGANIZATION OF THE REPORT

This report is organized into chapters dedicated to background on measuring fitness in youth; the committee's methodology; and its findings, conclusions, and recommendations. Chapter 2 provides a historical perspective on the origins of youth fitness testing and the changes that have occurred over the years both in the tests and in their uses. This chapter includes a table describing fitness test batteries currently used around the world. Chapter 3 describes in detail the methodology used by the committee to identify test items, including the CDC's systematic review, which was the primary basis for the committee's conclusions and recommendations. Chapters 4, 5, 6, and 7 present the committee's rationale for recommending test items for the four fitness components, respectively—body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility—highlighting the findings of the scientific literature. As noted earlier in this chapter, the primary purpose of this report was to make recommendations for a national survey. A secondary purpose was to make recommendations for

[1] A disability is defined as any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being. For the purposes of this report, this term should be construed in the broadest sense, covering impairments (i.e., a problem in body function or structure), activity limitations (i.e., a difficulty encountered by an individual in executing a task or action), and participation restrictions (i.e., a problem experienced by an individual in involvement in life situations).

the use of fitness tests in schools and other educational settings.[2] Because tests vary based on their potential uses, separate chapters were prepared for each of these two uses of fitness test items. Chapter 8 presents the committee's recommendations for national surveys of youth fitness. Chapter 9 describes the importance of fitness in the context of education, details factors to consider when implementing fitness tests in schools and other educational settings, and presents the committee's recommendations for specific fitness tests for educational settings. Finally, Chapter 10 includes the committee's recommendations for future research.

REFERENCES

AAHPERD (American Alliance for Health, Physical Education, Recreation and Dance). 1980. *Health related physical fitness test manual*. Reston, VA: AAHPERD.

Castillo-Garzon, M. J., J. R. Ruiz, F. B. Ortega, and A. Gutierrez. 2006. Anti-aging therapy through fitness enhancement. *Clinical Interventions in Aging* 1(3):213-220.

Eignor, D. R. 2001. Standards for the development and use of tests: The standards for educational and psychological testing. *European Journal of Psychological Assessment* 17(3):157-163.

HHS (U.S. Department of Health and Human Services). 1996. *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: HHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

Holt, J., L. E. Holt, and T. W. Pelham. 1996. Flexibility redefined. In *Biomechanics in sports XIII*, edited by T. Bauer. Thunder Bay, Ontario: Lakehead University. Pp. 170-174.

Saltin, B. 1973. Oxygen transport by the circulatory system during exercise in man. In *Limiting factors of physical performance*, edited by J. Keul. Stuttgart, Germany: Thieme Publishers. Pp. 235-252.

[2]Other educational settings include, for example, gymnasiums and fitness centers.

2

Measuring Fitness in Youth

Fitness testing for youth emerged from the field of physical education, which has a long-standing history of fitness testing. Over the years, social and political circumstances have dictated the emphasis, progress, and use of fitness testing in the United States. In particular, an early emphasis on performance outcomes—particularly military performance—has given way to a focus on health outcomes as a result of concerns about the health of the nation's youth. While the components of fitness have remained virtually the same, moreover, the tests and protocols used to measure it have evolved as more data have accumulated on their validity and reliability and their relationship to desired outcomes. Although efforts have been ongoing to standardize and validate the constructs for fitness testing, the range of fitness test batteries currently in use, as detailed in this chapter, reveals that consensus on these issues remains elusive. The research needs identified in Chapter 10 therefore include a comprehensive reevaluation of the past and current approaches to fitness testing in youth.

This chapter begins with a brief early history of physical fitness testing. It then describes more recent historical events related specifically to measuring physical fitness among U.S. youth.[1] The final section includes a table that lists the various batteries of fitness tests currently in use worldwide.

[1]For more information about the history of youth fitness testing in the United States, the reader is referred to Corbin (2012), Mood et al. (2007), Morrow (2005), Morrow et al. (2009), and Plowman et al. (2006).

EARLY HISTORY OF PHYSICAL FITNESS TESTING[2]

Although organized youth fitness testing did not begin until the mid–20th century, the foundation of national youth fitness testing began to be established a century earlier. Park (1989) notes that early leaders in physical education, many of whom were medical doctors and YMCA leaders, focused the outcomes of instruction on anthropometric measurements. During the last half of the 19th century, national physical education organizations emerged (e.g., the American Association for the Advancement of Physical Education [AAAPE]), and the leaders who founded the organizations continued with a measurement focus (e.g., strength and lung capacity assessments). Dudley Sargent, one of the pioneers of physical testing, developed the vertical jump test that is still used today and is commonly referred to as the "Sargent jump." It is generally believed that Sargent thought of the vertical jump as a general measure of fitness and health. He published the books *Health, Strength and Power* (Sargent, 1904) and *Universal Test for Strength, Speed and Endurance of the Human Body* (Sargent, 1902).

By the late 1800s and early 1900s, the purpose of fitness testing had expanded beyond anthropometric measurements with the introduction of the concept of "physical efficiency," characterized as efficient functioning of body systems, such as the circulatory, respiratory, muscular, and nervous systems (Park, 1989). Fitness testing evolved from a focus on athletic performance to a focus on health in the early 1900s as researchers such as McCurdy and McKenzie studied blood pressure fatigue (McCurdy, 1901; McKenzie, 1913), and Storey studied pulse rate (Storey, 1903). Prior to World War I, tests of "motor ability" that included tests of jumping, climbing, lifting, vaulting, and running were popular. One prominent test, the Playground Association of America Athletic Badge Test, was introduced for boys in 1913 and girls in 1916. During and immediately after World War I, the focus on physical education and physical training in schools increased, with a shift toward fitness for war. Many physical educators led physical training programs for the military during the war era. The theme that many Americans were unfit was popular in the media.

After World War I, the Public Health Service and many different organizations focused attention on fitness tests and programs because of their potential link to preparedness for war. The Public Health Service booklet *Keeping Fit* emphasized many personal factors (e.g., willpower, courage, self-control) in addition to those related to health (USPHS, 1918). During the early 1900s, there was considerable debate about the importance

[2]The information included in this section is based on Park (1989).

of "efficiency" testing and what should be included in physical efficiency tests. Various test batteries were developed, including the Physical Fitness Index (PFI), developed by Rogers (several strength items) (Rogers, 1925), and a test of athletic power developed by McCloy (general motor ability and strength) (McCloy, 1934, 1941). Both Rogers and McCloy conducted research that provided a basis for the items selected for their tests. Other tests of the era were often developed by groups of professionals based on group consensus.

"Financial austerities" due to the Great Depression resulted in decreases in physical education and a reduced emphasis on physical fitness testing (Park, 1989). The interest in general physical fitness testing in schools that was common after World War I diminished, while interest in laboratory-based measures of fitness grew.

As was the case prior to, during, and immediately after World War I, World War II produced much military, governmental, and societal interest in fitness programs and fitness testing. While there was much fanfare and many proposals for action were made, most efforts with youth relied on volunteer leaders and local funding. A 1941 supplement to the *Research Quarterly* focused on physical fitness and fitness testing (Carpenter, 1941; Cureton and Larson, 1941; Larson, 1941; McCloy, 1941). Park (1989) indicates that the U.S. Department of Education, in cooperation with the Army, Navy, and Public Health Service, prepared a fitness booklet (*Physical Fitness through Physical Education for the Victory Corps*) in 1942. In addition, at the request of the U.S. Department of Education, the American Association for Health, Physical Education, and Recreation's (AAHPER's) section on women's athletics prepared a fitness test battery for high school girls. Fitness manuals were prepared for college students as well, and the armed services developed fitness programs of their own during the war. Also during the war, many conferences and committees focused on youth fitness. Park (1989, p. 11) notes that the "predominant interpretation given to the term physical fitness during World War II was the ability to sustain long, hard, muscular effort." The joint involvement of health, education, physical education, and military groups underscores the mixed purposes of physical fitness testing. Health was a concern, but so were general fitness and fitness for war.

NATIONAL YOUTH FITNESS TESTING: 1950 TO 1980

The physical fitness focus that was prominent during World War II gave way to a more generalized emphasis for youth during the postwar years. The popularity of college and professional sports led physical education programs to focus on athletic capabilities. The Korean War in the early

TABLE 2-1 Key Historical Events/Publications in Youth Fitness Testing in the United States, 1950-1979

Year	Historical Event/Publication
1954	Publication of the results of minimum muscular fitness and flexibility tests in schoolchildren (Kraus and Hirschland, 1953, 1954)
1956	By Executive Order #10673, President Eisenhower creates the President's Council on Youth Fitness (July 16)
1957-1958	American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) holds meetings on youth fitness
1958	American Association for Health, Physical Education and Recreation (AAHPER) Youth Fitness Test published (AAHPER, 1958)
1965	Update of AAHPER Youth Fitness Test published (AAHPER, 1965)
1966	President's Council on Physical Fitness creates the Presidential Physical Fitness Awards Program
1973	Texas Physical Fitness-Motor Ability Test released by the Governor's Commission on Physical Fitness (Coleman and Jackson, 1973)
1976	Update of AAHPERD Youth Fitness Test published (AAHPERD, 1976)

SOURCE: Adapted from Morrow et al., 2009.

1950s did bring some focus back to physical fitness, but it was research by Kraus and Hirschland (1953, 1954) that provided the impetus for the national youth physical fitness testing movement. Their reports indicated that children in the United States passed fewer fitness test items than children from European countries. For their research, Kraus and Hirschland used the Kraus-Weber test, a battery of six items testing minimum muscular fitness and flexibility originally developed as a measure of potential for back pain. Although this test was rudimentary by current standards, the results gained traction after being reported in the mainstream media (see for example, the article in *Sports Illustrated* titled "The Report That Shocked the President" [Boyle, 1955]). Published during the Cold War era, the results implying less fitness in American than in European youth raised major concern about the nation's military preparedness. Consequently, Kraus was granted an audience with then President Eisenhower to discuss the study results. After that meeting, Eisenhower established a cabinet-level President's Council on Youth Fitness (now the President's Council on Fitness, Sports, and Nutrition [PCFSN]). A chronology of these and other key events/publications relating to youth fitness, 1950 to 1979, is presented in Table 2-1.

TABLE 2-2 Changes in the Youth Fitness Test, 1958-1976

	1958	1965	1976
Shuttle run	×	×	×
50-yard dash	×	×	×
500-yard run/walk	×	×	×
Pull-up (boys)	×	×	×
Modified pull-up (girls)	×		
Flexed arm hang (girls)		×	×
Softball throw	×	×	
Long jump	×	×	×
Sit-up (straight-leg)	×	×	
Sit-up (flexed-leg, timed, arms behind head)			×

SOURCE: Adapted from Corbin and Pangrazi, 1992.

In 1957, the Council and a citizen's advisory group called on professional groups to improve efforts to promote youth fitness. Many different organizations, including the newly created American College of Sports Medicine and the American Medical Association, urged action. The AAHPER Research Council appointed a committee, chaired by Anna Espenschade of the University of California, that created the first youth physical fitness test battery (the Youth Fitness Test) for use in a large-scale national survey. The test included the items shown in Table 2-2. These test items included measures of strength and muscular endurance common in earlier fitness test batteries and a 600-yard run/walk believed at the time to be a measure of cardiovascular fitness; these measures often were considered to be health related. Also included, however, were items more related to physical education objectives and skill-related fitness, such as the softball throw and the 50-yard dash, reflecting in part societal concerns at the time regarding the athletic capabilities and military preparedness of youth (Morrow et al., 2009). The test items were administered in a national survey conducted by the University of Michigan (led by Paul Hunsicker), with funding from the U.S. Department of Education (AAHPER, 1958). As was the case with other testing in schools at the time (e.g., achievement tests), normative standards were developed and reported in the first test manual (AAHPER, 1958). AAHPER also designed awards (certificates and emblems) for students who met those standards (Park, 1989).

During the late 1950s and early 1960s, the fitness movement continued. President Kennedy advocated for youth physical fitness in his article "The Soft American," published in *Sports Illustrated* (Kennedy, 1960). He convened a conference on youth fitness, and the President's Council subsequently prepared a booklet on the subject, commonly referred to as the "Blue Book," that included information about the seven-item Youth Fitness Test of 1958. This booklet was intended to emphasize the importance of having an active lifestyle and its role in establishing fitness and health. Kennedy wrote a second article in *Sports Illustrated* in 1962, entitled "The Vigor We Need" (Kennedy, 1962). By executive order, the name of the Council was changed to the President's Council on Physical Fitness (PCPF) to reflect interest in promoting fitness among people of all ages and ability levels. In 1965, a second survey was conducted using a modified version of the Youth Fitness Test (AAHPER, 1965). Changes in the test items used for the 1965 survey included the addition of a flexed arm hang test to replace the modified pull-up for girls (see Table 2-2). This change was made primarily to produce more reliable test scores. In 1966, the President's Council established the Presidential Physical Fitness Award Program, jointly administered by AAHPER and the PCPF, to acknowledge youth who met or exceeded the 85th percentile on all seven test items.

The third national survey using the Youth Fitness Test was published in 1976 (AAHPERD, 1976). As noted in Table 2-2, the softball throw was deleted, the sit-up was modified, and distance runs longer than 600 yards were included as options. The softball throw was deleted because it was considered to be a skill rather than a fitness-related item. The modification of sit-up testing was based on the idea that the bent-knee approach was less stressful on the back than the straight-leg approach. Finally, research indicating greater validity for longer runs and their association with aerobic capacity led to the inclusion of longer runs as optional items (Morrow et al., 2009).

During the 1960s and 1970s, evidence linking fitness and physical activity to good health accumulated. Correspondingly, interest grew in the development of youth fitness test batteries focused primarily on health-related physical fitness. The Texas Physical Fitness Motor Ability Test (Coleman and Jackson, 1973) included health-based test items, and evidence was included to support the test items selected.

NATIONAL YOUTH FITNESS TESTING: 1980 TO 1990

In the 1970s, several committees were appointed by the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) to study the Youth Fitness Test. Recommendations of these committees led to the development of a Health-Related Physical Fitness Test by AAHPERD in 1980. AAHPERD continued to maintain both the health-related test and the Youth Fitness Test. The Youth Fitness Test included an awards program administered by the renamed President's Council on Physical Fitness and Sports (PCPFS) and a newly created fitness report card created and administered by the Cooper Institute in Dallas, Texas. Table 2-3 lists these and other key events related to national youth fitness testing during 1980 to 1990.

In 1984, AAHPERD published a technical manual for the Health-Related Physical Fitness Test documenting the theoretical basis for the adopted test items and for replacements for the normative standards of the Youth Fitness Test (Morrow et al., 2009). Test items targeting cardiorespiratory fitness, musculoskeletal fitness, and body composition were included in the battery as fitness components related to health. Also in 1984, an ad hoc committee of AAHPERD recommended that the Health-Related Physical Fitness Test become the primary AAHPERD test and that the Youth Fitness Test be made a secondary test. However, that recommendation was not implemented, and in 1985 another AAHPERD committee was appointed (the Manual Task Force) to merge the two AAHPERD tests (see below). During this period, several national surveys were completed. In 1986, the School Population Fitness Survey was conducted by the then PCPFS (now PCFSN) using a revised version of the Youth Fitness Test. The revised battery is described later in this chapter. Of note are the removal of the 50-yard dash and the long jump and the addition of a V-sit test of flexibility.

Two additional youth fitness surveys were conducted in the mid-1980s. The National Children and Youth Fitness Study I, results of which were published in 1985 by the U.S. Department of Health and Human Services (McGinnis, 1985), focused on measuring the fitness of secondary school youth using health-related fitness test items developed specifically for the study. In 1986, the National Children and Youth Fitness Study II (Ross and Pate, 1987) was conducted to assess the fitness of elementary school youth using the health-related items from the National Children and Youth Fitness Study I.

AAHPERD's Manual Task Force "was charged with developing a single AAHPERD fitness test battery, establishing criterion-referenced standards, examining existing awards schemes, and writing the appropriate manual" (Plowman et al., 2006, p. S8). Before the task force could produce a document, however, the PCPFS initiated its fitness testing and awards program in 1986, based primarily on the 1985 version of the Youth Fitness Test and existing award schemes. Even after much discussion among relevant organizations (PCPFS, AAHPERD, Cooper Institute) regarding the establishment of a unified national fitness testing battery, the PCPFS continued with its test and awards program, named the President's Challenge Program (1987), while the Cooper Institute introduced a health-related fitness test and

TABLE 2-3 Key Historical Events/Publications in Youth Fitness Testing in the United States, 1980-1990

Year	Historical Event/Publication
1980	American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) *Health-Related Physical Fitness Test Manual* released (AAHPERD, 1980)
1982	Fitnessgram® pilot conducted in 30 Tulsa schools
1983	*A New Definition of Youth Fitness* published (Pate, 1983)
1983	Health-Related Physical Fitness Test user survey piloted (Safrit and Wood, 1983)
1984	AAHPERD's *Technical Manual*: *Health-Related Physical Fitness Test* released (AAHPERD, 1984)
1985	National Children and Youth Fitness Study I results published (McGinnis, 1985)
1985	AAHPERD's *Norms for College Students: Health Related Physical Fitness Test* published (Pate, 1985)
1986	Safrit and Wood (1986) report on tristate usage of the AAHPERD Health-Related Physical Fitness Test published, indicating many issues with adoption of the new test
1986	National School Population Fitness Survey results released (PCPFS, 1986)
1986	President's Challenge Program developed (PCPFS, 1987)
1986	*Fit Youth Today* (American Health Fitness Foundation, 1986) published; original test development begun under the Texas Governor's Commission on Physical Fitness
1987	National Children and Youth Fitness Study II results published (Ross and Pate, 1987)
1987	National Fitnessgram originally developed (Plowman et al., 2006)
1988	*Youth Fitness Testing: Validation, Planning, and Politics* published (Franks et al., 1988)
1988	AAHPERD's health-related fitness education program "Physical Best" published (McSwegin, 1989)
1988	Chrysler Amateur Athletic Union (AAU) Fitness Test (Chrysler Corporation and Amateur Athletic Union of the United States, 1992) initially distributed
1989	YMCA *Youth Fitness Test Manual* published (Franks, 1989)
1989	*The Case for Large-Scale Physical Fitness Testing in American Youth* published (Pate, 1989)
1989	*Physical Fitness Testing of Children: A 30-Year History of Misguided Efforts?* published (Seefeldt and Vogel, 1989)

SOURCE: Adapted from Morrow et al., 2009.

reporting program called Fitnessgram® (1988), and AAHPERD developed a health-based fitness testing and reporting program called Physical Best (1988). Table 2-4 shows the evolution of test items from the PCPFS/PCFSN and Fitnessgram batteries, including current and previously included items; the Physical Best battery is no longer in use. A comprehensive discussion of the events leading to the development of these test batteries is provided in Plowman et al. (2006). Issues that led these groups to devise different tests included the use of health versus motor fitness items, the use of health criteria versus normative standards, the inclusion of a body composition item, and the inclusion of award schemes.

In 1988, the Chrysler Amateur Athletic Union (AAU) Fitness Test was introduced. The YMCA *Youth Fitness Test Manual* was published the following year (Franks, 1989).

YOUTH FITNESS TESTING SINCE 1990

Despite the above efforts to develop a unified battery of fitness tests and the implementation of new tests, no new large-scale national fitness surveys have been conducted since the 1980s. In 1994, the Cooper Institute published *The Prudential Fitnessgram® Technical Reference Manual* (Morrow et al., 1994), which has been updated and published online (http://www.cooperinstitute.org/reference-guide). Fitnessgram uses health-based criterion references. Key events in youth fitness testing since 1990 are listed in Table 2-5.

In 1994, AAHPERD adopted Fitnessgram as its national fitness test. Physical Best, no longer a fitness test battery, became the AAHPERD fitness education program rather than a testing program. In 1996, the PCPFS introduced a new health-related fitness program using criterion-referenced health standards as opposed to normative standards, but it was subsequently discontinued. Items in the PCPFS battery (modified version of the Youth Fitness Test) introduced in 1986 are shown in Table 2-4. Over the years, the test battery has evolved to include mostly items considered to be health related (with the exception of the shuttle run). The PCFSN battery included in the President's Challenge Program still uses normative standards and offers awards based on those standards.

During the 1990s, calls for a public health basis for youth fitness testing received much attention (Sallis and McKenzie, 1991; Simons-Morton et al., 1988). Papers were published questioning the use of youth fitness tests and award schemes (Corbin et al., 1990; Keating, 2003; Rowland, 1995), and concerns about the proper use and misuse of tests were expressed (Corbin et al., 1990). Some research led to a call for the end of youth fitness testing as a result of findings implying the adverse effects of testing in academic settings and its ineffectiveness in promoting physical activity (Cale et al.,

TABLE 2-4 Evolution of President's Council on Physical Fitness and Sports (PCPFS)/President's Council on Physical Fitness, Sports, and Nutrition (PCFSN) and Fitnessgram® National Test Batteries

Test Item	PCPFS/PCFSN	Fitnessgram
600-yard run	1986	
Shuttle run (10 meters)	1986, current	
Mile run		1988, current alternative
Progressive Aerobic Cardiovascular Endurance Run (PACER) (20-meter shuttle)		1988, current alternative
PACER (15-meter shuttle)		Current alternative
Walk test		Current alternative
Mile, half-mile, quarter-mile run	Current based on age	
Pull-up	1986, current	1988, alternative
90-degree push-up		1988, current
Right-angle push-up, flexed arm hang	Current alternative	
Modified pull-up		1988, current alternative
Curl-up, feet held	1986, current	
Curl-up		1988, current
Partial curl-up	Current alternative	
Trunk lift		1988, current
Shoulder stretch		1988, current
V-sit reach	1986, current	
V sit-and-reach	Current	
Two-leg sit-and-reach	Current alternative	
Backsaver sit-and-reach		1988, current
Skinfold (body composition)		1988, current alternative
Body mass index (BMI)		1988, current alternative

NOTE: The year shown indicates when the test was first implemented. "Current" indicates items in the current version of the battery. "Alternative" means the item is an alternative for measuring the particular construct.

TABLE 2-5 Key Historical Events/Publications in Youth Fitness Testing in the United States, 1990-2012

Year	Historical Event/Publication
1992	Forum in the *Research Quarterly for Exercise and Sport* published, including a lead manuscript entitled "Are American Children and Youth Fit?" (Corbin and Pangrazi, 1992)
1994	*Physical Activity Guidelines for Adolescents: Consensus Statement* published (Sallis and Patrick, 1994)
1994	Fitnessgram® manual providing battery justification, description, and rationale released (Morrow et al., 1994)
1995	*Complete Guide to Youth Fitness Testing* published (Safrit, 1995)
1995	Rowland (1995) questions viability of youth fitness testing
1996	*Physical Activity and Health: A Report of the Surgeon General* published (HHS, 1996)
1998	*Physical Activity for Children: A Statement of Guidelines* published (NASPE, 1998)
2002	Keating and colleagues (2002) report on preservice teacher attitudes toward youth fitness tests published
2004	Keating and Silverman (2004) report on teacher use of youth fitness tests published
2004	National Association for Sport and Physical Education (NASPE) updates physical activity guidelines for children (NASPE, 2004)
2005	*Evidence Based Physical Activity for School-Age Youth* published (Strong et al., 2005)
2005	*Are American Children and Youth Fit?: It's Time We Learned* published (Morrow, 2005)
2007	Commentary on youth fitness testing published (Rowland, 2007)
2008	*Measurement in Physical Education and Exercise Science* special issue on *Youth Fitness Testing: A Positive Perspective* published (Liu, 2008)
2008	Physical activity guidelines for Americans released (HHS, 2008)
2012	National Health and Nutrition Examination Survey (NHANES) National Youth Fitness Survey launched

SOURCE: Adapted from Morrow et al., 2009.

2007; Rowland, 1995). Others, however, recognizing the potential of youth fitness to promote lifelong physical activity, emphasized the importance of continuing research aimed at improving the reliability and validity of pertinent test batteries and identifying strategies for preventing the misuse of such testing (Corbin, 2007; Corbin et al., 1995). Also emphasized was the role of professional and governmental agencies in developing fitness testing policies, including guidelines and standards for appropriate implementation and interpretation of results.

In light of the lack of a national youth fitness survey since the 1980s, several calls have been made for an updated survey. The National Children and Youth Fitness Studies (McGinnis, 1985; Ross and Pate, 1987) and the National School Population Fitness Survey (PCPFS, 1986) were the last national studies of youth fitness.

Several large-scale statewide surveys of youth physical fitness have been carried out (Morrow et al., 2010); more than a dozen states have conducted or are considering conducting fitness testing on a mandatory or large-scale voluntary basis (Morrow and Ede, 2009). Other large administrative units (e.g., New York City) also have implemented extensive youth fitness surveys in recent years. Some of these testing efforts have been on a routine basis, while others have not. For example, Texas has implemented mandatory health-related physical fitness testing for youth in grades 3 to 12 since 2007; California has been monitoring the fitness status of students in grades 5, 7, and 9; and New York City tests 600,000 students in grades K-12 annually (Morrow and Ede, 2009). Levels of funding to support such efforts have varied considerably.

In addition, the potential of fitness measures to reflect or predict health status has led to the inclusion of these measures in national health surveys such as the National Health and Nutrition Examination Survey (NHANES), which includes components pertinent to physical fitness, such as body composition, cardiovascular fitness, and physical activity (Morrow et al., 2009). Including fitness-related items in NHANES enables not only the longitudinal monitoring of fitness components but also the identification of associations between fitness status and health indices (Ortega et al., 2008; Suni et al., 1998). For example, epidemiological analyses with these data revealed the association of low cardiorespiratory endurance status with increased prevalence of the risk factors for cardiovascular disease in adolescents and adults (Carnethon et al., 2005) and with cases of obesity in adulthood (Wang et al., 2010). These data can also be used to establish cut-points (cutoff scores) for fitness tests, which are essential for interpreting test results and communicating them to individuals and families, as well as establishing individual performance goals related to health improvements.

Reflecting the initial impetus for developing fitness tests, all active duty U.S. military personnel are required to adhere to standards for body composition, physical fitness, and appearance for enlistment and retention (IOM, 1998). Recognizing the association of physical fitness with health and readiness for military tasks, the Department of Defense has mandated annual assessment of fitness components for service members (DoD, 2004) and also has implemented remedial programs for those who fail to meet the cut-points for physical fitness or readiness tests. The effectiveness of routine fitness tests for establishing a minimum level of fitness in military personnel has been widely acknowledged (IOM, 1998). Components assessed by these tests (which vary among the services) include cardiorespiratory endurance; muscular strength and endurance; whole-body flexibility; and parameters indicating balance, agility, and explosive power.

When implemented in schools, fitness testing also can serve the purpose of assessing and improving physical literacy among youth (Tremblay and Lloyd, 2010). Regular physical activity is widely considered part of a healthy lifestyle given its beneficial effects on various health outcomes and fitness levels (Bouchard and Shephard, 1994; Pate et al., 1995; Simons-Morton et al., 1988). Of note are the reciprocal effects of physical activity, fitness, and health. In other words, an increase in habitual physical activity can result in increased fitness, while health status affects one's fitness or ability to carry out physical activity (Bouchard and Shephard, 1994). Based on this relationship and considering fitness as a primary outcome of physical activity, fitness testing is regarded as an effective means of monitoring the status of physical activity in population-based studies, as well as promoting lifelong physical activity, which may eventually lead to improved health status (Corbin, 2007; Tremblay and Lloyd, 2010). Indeed, the rationale for programs described in this chapter is based on the influence of the dispositions and habits established during childhood on physical activity, fitness, and health status in adulthood (Malina, 2001; Morrow and Ede, 2009; Simons-Morton et al., 1988). At the same time, it is important to acknowledge the potential adverse effects of inappropriate uses of fitness testing (Ernst et al., 2006). The implementation of fitness measures for educational purposes is discussed in detail in Chapter 9.

CURRENT BATTERIES OF YOUTH FITNESS TESTS

A variety of tests have been designed to test physical fitness in youth in the United States and other countries by both governmental and nongovernmental organizations. Table 2-6 lists selected batteries of tests currently in use in the United States and other countries. The committee found that the evidence and criteria for selecting these tests are not always clear.

TABLE 2-6 International Batteries of Youth Fitness Tests

Battery Name (Country)	Age (yrs) of Tested Population	Tests	Component Assessed	References
Fitnessgram® (U.S.)	>5	• Progressive Aerobic Cardiovascular Endurance Run (PACER) test (recommended) or mile run or mile walk test (for secondary students) • Skinfold measurements (recommended) or body mass index (BMI) calculation • Cadence-based curl-up • Trunk lift • 90-degree push-up (recommended) or modified/traditional pull-up or flexed arm hang • Backsaver sit-and-reach or shoulder stretch	• Aerobic capacity, maximal oxygen consumption (VO_2max) • Body composition • Abdominal strength and endurance • Back extensor strength and flexibility • Upper-body strength and endurance • Hamstring or shoulder flexibility	Cooper Institute, 2010; Welk and Meredith, 2008
President's Council on Fitness, Sports, and Nutrition (PCFSN) (U.S.)	6-17	• Curl-up or partial curl-up • Shuttle run • Endurance run/walk • Pull-up or right angle push-up • V-sit reach or sit-and-reach	• Abdominal strength and endurance • Speed and agility • Cardiorespiratory endurance • Upper-body strength/endurance • Flexibility of lower back and hamstrings	PCFSN, 2010, 2011

Canadian Association for Health, Physical Education and Recreation (CAHPER) Fitness Performance Test Batteries (Canada)	7-17	• Speed sit-up • Standing broad jump (suggested take-off angle = 30-45 degrees) • Shuttle run (two 30-foot runs) • Flexed arm hang • 50-yard dash • 300-yard run	• Strength and endurance of the abdominal muscles • Explosive strength or power of the legs • Agility and speed • Endurance of the arm and shoulder muscles • Speed • Stamina	CAHPER, 1966; Cumming and Keynes, 1967; Thompson, 1975
The European Physical Fitness Test battery, (EUROFIT) (Europe)	6-18	• Flamingo balance test • Plate tapping • Sit-and-reach test (using 15 cm at the level of the feet) • Standing broad jump (Bosco test) • Handgrip test • Sit-up/curl-up • Bent arm hang • 10 × 5-meter shuttle run • Cycle ergometer or 20-meter endurance shuttle run	• Balance • Speed and coordination of limb movement • Flexibility • Explosive muscular (anaerobic) strength • Static muscular strength • Trunk strength • Muscular endurance and functional strength • Speed, agility, and coordination • Cardiorespiratory endurance	Council of Europe, 1983, 1988

continued

TABLE 2-6 CONTINUED

Battery Name (Country)	Age (yrs) of Tested Population	Tests	Component Assessed	References
Assessing Levels of Physical Activity (ALPHA) Health-Related Fitness Test Battery for Children and Adolescents (Europe)	13-17	• Handgrip strength • Standing long jump • 4 × 10-meter shuttle run • 20-meter shuttle run • BMI • Waist circumference • Skinfold thickness	• Isometric strength • Lower-body strength • Speed and agility • Cardiorespiratory fitness • Body composition • Body composition • Body composition	España-Romero et al., 2010; Ruiz et al., 2011
Singapore National Physical Fitness Award (NAPFA) Scheme (Singapore)	>12	• Sit-ups in 1 minute • Standing broad jump • Sit-and-reach • Pull-ups in 30 seconds (full pull-ups are performed by males aged >15; females and males aged ≤15 perform a modified inclined pull-up [an inclined flexed arm hang]) • 10 × 4-meter shuttle run • Walk-run test (run on firm and level surface over a distance of 2.4 km [1.5 mile] for secondary school students or 1.6 km [1 mile] for primary school students)	• Abdominal muscular endurance • Muscular power • Flexibility • Upper-body muscular endurance • Speed and agility • Muscular endurance and cardiovascular fitness	Ngee Ann Polytechnic, 2002; Schmidt, 1995

Nation-Wide Children and Youth Fitness Study (Taiwan)	7-18	• Bent-leg sit-up	• Abdominal muscular strength and endurance	Chen et al., 2002; Chiang et al., 1998
		• Standing long jump	• Explosive power	
		• Modified sit-and-reach	• Flexibility of lower back and upper thigh	
		• 800-meter (boys <13 and all girls) or 1600-meter (boys ≥13) run/walk; test not administered to children ≤8	• Cardiorespiratory endurance	
Physical Fitness and Athletic Ability Test (Japan)	6-9 10-17	• Side step	• Physical ability	Shingo and Takeo, 2002
		• Vertical jump	• Instantaneous power	
		• Back strength	• Muscle strength	
		• Grip strength	• Muscle strength	
		• Trunk extension	• Flexibility	
		• Standing flexion	• Flexibility	
		• Step test	• Endurance	
		• 50-meter run	• Athletic ability	
		• Long jump	• Athletic ability	
		• Ball throw (softball for ages 10-11; handball for older children)	• Athletic ability	
		• Pull-up (modified pull-up for children aged 10-11 and girls of all ages)	• Athletic ability	
		• Zigzag dribble (test implemented after 1966 and used only with children ≥12 years of age)	• Athletic ability	
		• Continuous going up foot over foot, using a low horizontal bar, in 10 seconds (test implemented after 1966 and used only with children ≥12 years of age)	• Athletic ability	

continued

TABLE 2-6 Continued

Battery Name (Country)	Age (yrs) of Tested Population	Tests	Component Assessed	References
		• Endurance run (used only with children ≥12 years of age; 1,500-meter for boys; 1,000-meter for girls)	• Athletic ability	
Australian Fitness Education Award (AFEA)[a] (Australia	9-18	• Multistage fitness test (also known as 20-meter shuttle run, Beep test, or PACER)	• Cardiorespiratory endurance	ACHPER, 1996
		• 1.6-km (1-mile) run/walk	• Cardiorespiratory endurance	
		• Curl-up	• Muscular endurance (and strength)	
		• Basketball throw	• Muscular strength (and endurance)	
		• Sit-and-reach	• Muscle and joint flexibility	
		• Shoulder stretch	• Muscle and joint flexibility	
Physical Fitness Score (Poland)	7-19	• 50-meter dash	• Speed (short-distance)	Pilicz et al., 2005
		• Standing broad jump	• Explosive power	
		• Long run (fixed distance or fixed time period)	• Cardiorespiratory endurance	
		• Handgrip	• Handgrip strength	
		• Relative strength (pull-up or arm hang)	• Muscular strength	
		• Shuttle run	• Speed, agility, coordination	
		• Sit-up	• Abdominal muscular strength (and endurance)	
		• Bend trunk	• Flexibility	

Canadian Assessment of Physical Literacy (CAPL) Test (Ontario, Canada)	Initial focus on children in grades 4-6; later development for other grades and ages	• PACER test (from Fitnessgram) • Partial curl-up • Push-up • Grip strength • Sit-and-reach • Arm flexibility	• Cardiorespiratory endurance • Muscular strength and endurance • Muscular strength and endurance • Muscular strength and endurance • Flexibility • Flexibility	Lloyd and Tremblay, 2011; Lloyd et al., 2010; Tremblay and Lloyd, 2010
YMCA Youth Fitness Test Manual (United States)	Ages 6-17	• 1-mile run • Tricep and calf skinfold • Sit-and-reach • Curl-up • Modified pull-up	• Cardiorespiratory endurance • Relative leanness • Flexibility, back health • Muscular strength and endurance • Muscular strength and endurance	Franks, 1989
Physical Fitness Tests in Nordic Armed Forces (Denmark)	>20	• Distance run • Shuttle run • Lunge • Dip • Pull-up • Dead-lift • Plank • March with loads and obstacles	• Aerobic fitness • Aerobic fitness • Muscular strength • Muscular strength • Muscular strength • Muscular strength • Muscular strength • Function	Malmberg, 2011

continued

TABLE 2-6 Continued

Battery Name (Country)	Age (yrs) of Tested Population	Tests	Component Assessed	References
Physical Fitness Tests in Nordic Armed Forces (Finland)	>20	• Timed run • Ergometer • Walk • Standing long jump • Sit-up • Push-up • BMI • Waist circumference • March	• Aerobic fitness • Aerobic fitness • Aerobic fitness • Muscular strength • Muscular strength • Muscular strength • Body composition • Body composition • Function	Malmberg, 2011
Physical Fitness Tests in Nordic Armed Forces (Norway)	>20	• Distance run • Swim • Cross-country ski • Bicycle • Pull-up • Sit-up • Push-up	• Aerobic fitness • Aerobic fitness • Aerobic fitness • Muscular strength • Muscular strength • Muscular strength • Muscular strength	Malmberg, 2011
Physical Fitness Tests in Nordic Armed Forces (Sweden)	>20	• Shuttle run • Swim • Push-up • Sit-up • Vertical jump • Back suspension • Arm suspension	• Aerobic fitness • Aerobic fitness • Muscular strength • Muscular strength • Muscular strength • Muscular strength • Muscular strength	Malmberg, 2011

NOTE: This table is not an exhaustive listing of international tests. China and South Korea, for example, conduct fitness tests regularly, but relevant publications are not available in English translations. All tests listed are field tests currently in use. In addition to field tests, laboratory tests, questionnaires such as the International Fitness Scale, and other assessment methods may be used.

[a] Available at http://www.achper.org.au/bookshop/achper-resources/afea-kit (accessed August 31, 2012).

REFERENCES

AAHPER (American Association for Health, Physical Education, and Recreation). 1958. *AAHPER youth fitness test manual*. Washington, DC: AAHPER.

AAHPER. 1965. *AAHPER youth fitness test manual, revised edition*. Washington, DC: AAHPER.

AAHPERD (American Alliance for Health, Physical Education, Recreation and Dance). 1976. *AAHPERD youth fitness test manual, revised 1976 edition*. Reston, VA: AAHPERD.

AAHPERD. 1980. *Health related physical fitness test manual*. Reston, VA: AAHPERD.

AAHPERD. 1984. *Technical manual: Health related physical fitness*. Reston, VA: AAHPERD.

ACHPER (Australian Council for Health, Physical Education and Recreation). 1996. *Australian fitness education award. User's manual and curriculum ideas*. Adelaide, Australia: ACHPER.

American Health Fitness Foundation. 1986. *Fit youth today: Educating, conditioning and testing students for health fitness*. Austin, TX: American Health and Fitness Foundation.

Bouchard, C., and R. J. Shephard. 1994. Physical activity, fitness and health: The model and key concepts: International proceedings and consensus statement. In *Physical activity, fitness and health*, edited by C. Bouchard, R. J. Shephard, and T. Stephens. Champaign, IL: Human Kinetics Publishers. Pp. 11-22.

Boyle, R. H. 1955 (August 15). The report that shocked the president. *Sports Illustrated Vault*, http://sportsillustrated.cnn.com/vault/article/magazine/MAG1130060/index.htm (accessed August 8, 2012).

CAHPER (Canadian Association for Health, Physical Education, and Recreation). 1966. *The CAHPER fitness-performance test manual: For boys and girls 7 to 17 years of age*. Ottawa: CAHPER.

Cale, L., J. Harris, and M. H. Chen. 2007. More than 10 years after "the horse is dead...": Surely it must be time to "dismount"?! *Pediatric Exercise Science* 19:115-131.

Carnethon, M. R., M. Gulati, and P. Greenland. 2005. Prevalence and cardiovascular disease correlates of low cardiorespiratory fitness in adolescents and adults. *Journal of the American Medical Association* 294(23):2981-2988.

Carpenter, A. 1941. An analysis of the relationships of the factors of velocity, strength, and dead weight to athletic performance. *Research Quarterly* 12:34-39.

Chen, W., C. C. Lin, C. T. Peng, C. I. Li, H. C. Wu, J. Chiang, J. Y. Wu, and P. C. Huang. 2002. Approaching healthy body mass index norms for children and adolescents from health-related physical fitness. *Obesity Reviews* 3(3):225-232.

Chiang, J., H. C. Wu, and D. S. Shih. 1998. *The 1997 nation-wide children and youth fitness study* [in Chinese]. Taipei, Taiwan: Ministry of Education.

Chrysler Corporation, and Amateur Athletic Union of the United States. 1992. *1992-93 testing packet: Chrysler fund-AAU physical fitness program*. Bloomington, IN: Amateur Athletic Union.

Coleman, A. E., and A. S. Jackson. 1973. *Texas physical fitness motor ability test*. Austin, TX: Governor's Commission on Physical Fitness.

Cooper Institute. 2010. *Fitnessgram & Activitygram test administration manual (updated 4th edition)*. Champaign, IL: Human Kinetics Publishers.

Corbin, C. B. 2007. Commentary to "more than 10 years after the horse is dead...." *Pediatric Exercise Science* 19:123-125.

Corbin, C. B. 2012. C. H. Mccloy lecture: Fifty years of advancements in fitness and activity research. *Research Quarterly for Exercise and Sport* 83(1):1-11.

Corbin, C. B., and R. P. Pangrazi. 1992. Are American children and youth fit? *Research Quarterly for Exercise and Sport* 63(2):96-106.

Corbin, C. B., P. Y. Lovejoy, P. Steingard, and R. Emerson. 1990. Fitness awards: Do they accomplish their intended objectives? *American Journal of Health Promotion* 4:345-351.

Corbin, C. B., R. P. Pangrazi, and G. J. Welk. 1995. A response to "the horse is dead; let's dismount." *Pediatric Exercise Science* 7:347-351.

Council of Europe. 1983. *Testing physical fitness: Eurofit experimental battery—provisional handbook*. http://www.bitworks-engineering.co.uk/linked/eurofit%20provisional%20handbook%20leger%20beep%20test%201983.pdf (accessed June 28, 2011).

Council of Europe. 1988. *Eurofit: Handbook for the Eurofit tests of physical fitness*. Rome, Italy: Council of Europe, Committee for the Development of Sport.

Cumming, G. R., and R. Keynes. 1967. A fitness performance test for school children and its correlation with physical working capacity and maximal oxygen uptake. *Canadian Medical Association Journal* 96(18):1262-1269.

Cureton, T. K., and L. A. Larson. 1941. Strength as an approach to physical fitness. *Research Quarterly* 12:391-406.

DoD (Department of Defense). 2004. *DoD physical fitness and body fat program*. Directive No. 1308.1. Washington, DC: DoD.

Ernst, M. P., C. B. Corbin, A. Beighle, and R. P. Pangrazi. 2006. Appropriate and inappropriate uses of Fitnessgram®: A commentary. *Journal of Physical Activity and Health* 3(Suppl. 2):S90-S100.

España-Romero, V., E. G National children and youth fitness study. Artero, D. Jimenez-Pavon, M. Cuenca-Garcia, F. B. Ortega, J. Castro-Piñero, M. Sjostrom, M. J. Castillo-Garzon, and J. R. Ruiz. 2010. Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA study. *International Journal of Sports Medicine* 31(7):490-497.

Franks, B. D. 1989. *YMCA youth fitness test manual*. Champaign, IL: Human Kinetics Publishers.

Franks, B. D., J. R. Morrow, and S. A. Plowman. 1988. Youth fitness testing: Validation, planning, and politics. *Quest* 40(3):187-199.

HHS (U.S. Department of Health and Human Services). 1996. *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: HHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

HHS. 2008. *Physical activity guidelines for Americans*. Washington, DC: HHS.

IOM (Institute of Medicine). 1998. *Assessing readiness in military women: The relationship of body, composition, nutrition, and health*. Washington, DC: National Academy Press.

Keating, X. D. 2003. The current often implemented fitness tests in physical in physical education programs: Problems and future directions. *Quest* 55:141-160.

Keating, X. D., and S. Silverman. 2004. Teachers' use of fitness tests in school-based physical education programs. *Measurement in Physical Education and Exercise Science* 8(3): 145-165.

Keating, X. F. D., S. Silverman, and P. H. Kulinna. 2002. Preservice physical education teacher attitudes toward fitness tests and the factors influencing their attitudes. *Journal of Teaching in Physical Education* 21(2):193-207.

Kennedy, J. F. 1960. The soft American. *Sports Illustrated* 13(26):14-23.

Kennedy, J. F. 1962. The vigor we need. *Sports Illustrated* 17:12-15.

Kraus, H., and R. P. Hirschland. 1953. Muscular fitness and health. *Journal of Physical Education and Recreation* 24(10):17-19.

Kraus, H., and R. P. Hirschland. 1954. Minimum muscular fitness tests in school children. *Research Quarterly* 25:178-188.

Larson, A. 1941. A factor analysis of motor ability variables and tests with tests for college men. *Research Quarterly* 12:499-517.

Liu, Y. 2008. Youth fitness testing: If the "horse" is not dead, what should we do? *Measurement in Physical Education and Exercise Science* 12(3):123-125.

Lloyd, M., and M. S. Tremblay. 2011. Introducing the Canadian assessment of physical literacy. In *Children and exercise XXV: Proceedings of the 25th pediatric work physiology meeting*, edited by G. Baquet and S. Berthoin. London: Routledge Taylor and Francis Group. Pp. 239-243.

Lloyd, M., R. Colley, and M. S. Tremblay. 2010. Advancing the debate on 'fitness testing' for children: Perhaps we're riding the wrong animal. *Pediatric Exercise Science* 22(2): 176-182.

Malina, R. M. 2001. Physical activity and fitness: Pathways from childhood to adulthood. *American Journal of Human Biology* 13(2):162-172.

Malmberg, J., ed. 2011. *Physical fitness tests in the Nordic armed forces: A description of basic test protocols*. http://www.nih.no/upload/Forsvarets%20institutt/Nordic%20Test%20 Protocoll.pdf (accessed August 31, 2012).

McCloy, C. H. 1934. The measurement of general motor capacity and general motor ability. *Research Quarterly* 5(Suppl. 1):46-61.

McCloy, C. H. 1941. The factor analysis as a research technique. *Research Quarterly* 12:22-23.

McCurdy, J. H. 1901. The effect of maximum muscular effort on blood-pressure. *American Journal of Physiology—Legacy Content* 5(2):95-103.

McGinnis, J. M. 1985. National Children and Youth Fitness Study. *Journal of Physical Education, Recreation and Dance* 56(1):44-90.

McKenzie, R. T. 1913. The influence of exercise on the heart. *American Journal of Medical Sciences* 145:69-75.

McSwegin, P. 1989. *Physical best: The AAHPERD guide to physical fitness, education, and assessment*. Reston, VA: AAHPERD.

Mood, D. P., A. W. Jackson, and J. R. Morrow, Jr. 2007. Measurement of physical fitness and physical activity: Fifty National children and youth fitness study years of change. *Measurement in Physical Education and Exercise Science* 11(4):217-227.

Morrow, J. R., Jr. 2005. Are American children and youth fit? It's time we learned. *Research Quarterly for Exercise and Sport* 76(4):377-388.

Morrow, J. R., Jr., and A. Ede. 2009. Statewide physical fitness testing: A big waist or a big waste? *Research Quarterly for Exercise and Sport* 80(4):696-701.

Morrow, J. R., H. B. Falls, and H. W. Kohl. 1994. *The Prudential Fitnessgram® technical reference manual*. Dallas, TX: Cooper Institute for Aerobics Research.

Morrow, J. R., Jr., W. Zhu, B. D. Franks, M. D. Meredith, and C. Spain. 2009. 1958-2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise and Sport* 80(1):1-11.

Morrow, J. R., Jr., S. B. Martin, G. J. Welk, W. Zhu, and M. D. Meredith. 2010. Overview of the Texas Youth Fitness Study. *Research Quarterly for Exercise and Sport* 81(Suppl. 2):S1-S5.

NASPE (National Association for Sport and Physical Education). 1998. *Physical activity for children: A statement of guidelines*. Reston, VA: NASPE.

NASPE. 2004. *Physical activity for children: A statement of guidelines for children ages 5-12*. 2nd ed. Reston, VA: NASPE.

Ngee Ann Polytechnic. 2002. *National Physical Fitness Assessment (NAPFA) test*. http://www.np.edu.sg/sdar/napfa/Pages/default.aspx (accessed July 11, 2011).

Ortega, F. B., J. R. Ruiz, M. J. Castillo, and M. Sjöström. 2008. Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity* 32(1):1-11.

Park, R. J. 1989. Measurement of physical fitness: A historical perspective. In *Office of Disease Prevention and Health Promotion Monograph Series*. Washington, DC: HHS, Public Health Service.

Pate, R. R. 1983. A new definition of youth fitness. *The Physician and Sportsmedicine* 11: 77-83.

Pate, R. R. 1985. *Norms for college students: Health related physical fitness test*. Reston, VA: AAHPERD.

Pate, R. R. 1989. The case for large-scale physical fitness testing in American youth. *Pediatric Exercise Science* 1(4):290-294.

Pate, R. R., M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Heath, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, Jr., K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore. 1995. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of the American Medical Association* 273(5):402-407.

PCFSN (President's Council on Fitness, Sports, and Nutrition). 2010. *The president's challenge: Physical activity & fitness awards program 2010-2011*. https://www.presidentschallenge.org/tools-resources/docs/PresChal_booklet_10-11.pdf (accessed July 14, 2011).

PCFSN. 2011. *The president's challenge: Physical fitness test*. https://www.presidentschallenge.org/challenge/physical/index.shtml (accessed July 14, 2011).

PCPFS (President's Council on Physical Fitness and Sports). 1986. *National school population fitness survey.* Ann Arbor, MI: University of Michigan.

PCPFS. 1987. *The presidential physical fitness award program, featuring the president's challenge. Instructor's guide*. Washington, DC: HHS.

Pilicz, S., R. Przeweda, J. Dobosz, and S. Nowacka-Dobosz. 2005. *Physical fitness score tables of Polish youth*. Warsaw, Poland: Akademia Wychowania Fizycznego.

Plowman, S. A., C. L. Sterling, C. B. Corbin, M. D. Meredith, G. J. Welk, and J. Morrow, Jr. 2006. The history of Fitnessgram. *Journal of Physical Activity & Health* 3(Suppl. 2):S5-S20.

Rogers, F. R. 1925. *Physical capacity tests in the administration of physical education* (Vol. 173). New York: Teacher College, Columbia University.

Ross, J. G., and R. R. Pate. 1987. The National Children and Youth Fitness Study II. *Journal of Physical Education, Recreation and Dance* 58(9):50-96.

Rowland, T. W. 1995. The horse is dead: Let's dismount. *Pediatric Exercise Science* 7(2): 117-120.

Rowland, T. W. 2007. Fitness testing in the schools: Once more around the track. *Pediatric Exercise Science* 19(2):113-114.

Ruiz, J. R., J. Castro-Piñero, V. España-Romero, E. G. Artero, F. B. Ortega, M. M. Cuenca, D. Jimenez-Pavon, P. Chillon, M. J. Girela-Rejon, J. Mora, A. Gutierrez, J. Suni, M. Sjostrom, and M. J. Castillo. 2011. Field-based fitness assessment in young people: The ALPHA health-related fitness test battery for children and adolescents. *British Journal of Sports Medicine* 45(6):518-524.

Safrit, M. J. 1995. *Complete guide to youth fitness testing*. Champaign, IL: Human Kinetics Publishers.

Safrit, M. J., and T. M. Wood. 1983. Health-related fitness test opinionnaire: A pilot survey. *Research Quarterly for Exercise and Sport* 54(2):204-207.

Safrit, M. J., and T. M. Wood. 1986. The health-related physical fitness test: A tri-state survey of users and non-users. *Research Quarterly for Exercise and Sport* 57(1):27-32.

Sallis, J. F., and T. L. McKenzie. 1991. Physical education's role in public health. *Research Quarterly for Exercise and Sport* 62:124-137.

Sallis, J. F., and K. Patrick. 1994. Physical activity guidelines for adolescents: Consensus statement. *Pediatric Exercise Science* 6(4):302-314.

Sargent, D. A. 1902. *Universal test for strength, speed and endurance of the human body*. Cambridge, MA: Powell Press.

Sargent, D. A. 1904. *Health, strength and power*. New York and Boston, MA: H. M. Caldwell Co.

Schmidt, G. J. 1995. Muscular endurance and flexibility components of the Singapore national physical fitness award. *Australian Journal of Science and Medicine in Sport* 27(4): 88-94.

Seefeldt, V., and P. Vogel. 1989. Physical fitness testing of children: A 30-year history of misguided efforts? *Pediatric Exercise Science* 1(4):295-302.

Shingo, N., and M. Takeo. 2002. The educational experiments of school health promotion for the youth in Japan: Analysis of the "sport test" over the past 34 years. *Health Promotion International* 17(2):147-160.

Simons-Morton, B. G., G. S. Parcel, N. M. O'Hara, S. N. Blair, and R. R. Pate. 1988. Health-related physical fitness in childhood: Status and recommendations. *Annual Review of Public Health* 9:403-425.

Storey, T. A. 1903. The influence of fatigue upon the speed of voluntary contraction of human muscle. *American Journal of Physiology* 8:355-375.

Strong, W. B., R. M. Malina, C. J. R. Blimkie, S. R. Daniels, R. K. Dishman, B. Gutin, A. C. Hergenroeder, A. Must, P. A. Nixon, J. M. Pivarnik, T. Rowland, S. Trost, and F. Trudeau. 2005. Evidence based physical activity for school-age youth. *Journal of Pediatrics* 146(6):732-737.

Suni, J. H., P. Oja, S. I. Miilunpalo, M. E. Pasanen, I. M. Vuori, and K. Bös. 1998. Health-related fitness test battery for adults: Associations with perceived health, mobility, and back function and symptoms. *Archives of Physical Medicine and Rehabilitation* 79(5): 559-569.

Thompson, P. 1975. *North York fall norms for boys and girls age 12-14 for: CAHPER tests; measures of aerobic fitness; peak flow; muscle strength; percent body fat*. Willowdale, Ontario: North York Board of Education.

Tremblay, M., and M. Lloyd. 2010. Physical literacy measurement: The missing piece. *Physical and Health Education Journal* 76(1):26-30.

USPHS (U.S. Public Health Service). 1918. *Keeping fit*. Washington, DC: USPHS.

Wang, C. Y., W. L. Haskell, S. W. Farrell, M. J. Lamonte, S. N. Blair, L. R. Curtin, J. P. Hughes, and V. L. Burt. 2010. Cardiorespiratory fitness levels among US adults 20-49 years of age: Findings from the 1999-2004 National Health and Nutrition Examination Survey. *American Journal of Epidemiology* 171(4):426-435.

Welk, G. J., and M. D. Meredith, eds. 2008. *Fitnessgram/Activitygram reference guide*. Dallas, TX: Cooper Institute.

3

Methodology for Selection and Interpretation of Health-Related Fitness Measures in Youth

This chapter describes the methodology followed by the committee in reaching conclusions and making recommendations on the most appropriate health-related physical fitness test items for youth. Before reviewing the scientific literature, the committee developed a conceptual framework to illustrate its thinking on the theoretical associations among the various components of fitness, their modifiers, and relevant health markers. The chapter begins by describing this framework. It then describes the committee's approach to the selection of test items for each of the four fitness components—body composition, cardiorespiratory endurance, musculoskeletal fitness, and flexibility. Included is a description of the literature review and the set of criteria that guided the selection process. The next section examines potential modifiers of fitness or of the associations between fitness and health, examples of which are included in the committee's conceptual framework. Just as the extent of the evidence on the association of each fitness component and test item with health markers varies, so, too, does the evidence for the effect of potential modifying factors. In general, there is more evidence on the importance of gender and age, while less is known about the effect of developmental maturity, motor skill, and practice. Similarly, there is a dearth of information about the influence of some demographic factors, such as ethnicity and race or socioeconomic status, on performance on fitness tests and its interaction with health markers. The final section of the chapter presents the committee's guidance for establishing cut-points (cutoff scores) for use in interpreting the results of youth fitness tests. Interpretation of test results is one of the most crucial aspects of such testing because it serves as a way of communicating with participants, health and school officials, and parents about their risk of negative health outcomes based on test performances.

CONCEPTUAL FRAMEWORK

To illustrate the overall challenge of its task and create a model for physical fitness measures that are most clearly associated with health outcomes in youth, the committee developed a conceptual framework (Figure 3-1). This framework guided the committee's analysis of research findings. Figure 3-1 depicts the potential relationships between physical fitness components—which can be measured by a variety of fitness test items represented by the smaller embedded boxes—and markers of health.

As illustrated in Figure 3-1, these relationships can be affected by both modifying factors and risk factors. As defined by the committee, modifying factors are those that can independently affect an individual's level of fitness. They include both factors that are measurable in the field (e.g., gender, race, ethnicity, maturity) and those that are not (e.g., heredity, practice level, skill level). Likewise, health outcomes are modified by certain risk factors that characterize an individual (e.g., low HDL cholesterol is a risk factor for cardiovascular disease). In the case of youth, health outcomes (i.e., diseases or conditions) are defined in terms of health markers or risk factors since youth are unlikely to experience a disease or condition (e.g., heart disease) as a result of their fitness level. The potential health outcomes that result from a specified level of performance on a fitness test are depicted within five categories: four categories of (positive or negative) markers of health-related outcomes (i.e., cardiovascular/respiratory health, metabolic health and obesity, mental and cognitive health, and musculoskeletal health) and a category that includes adverse events. Note that in this report, the terms *health marker* and *health risk factor* are used in a broad sense and interchangeably to refer to indicators of health outcomes.

The committee included body composition as a component of fitness, even though perspectives on this categorization vary. Body composition is also considered a modifier of performance on fitness tests and a health marker. Thus, it appears in all three categories of variables in the framework—fitness components, modifying factors, and health markers—and is highlighted in a different color from that of the other fitness components because of this unique nature.

The next section describes the approach used by the committee to select the best youth fitness test items, considering (1) the strength of their association with health markers in youth, (2) their integrity (validity and reliability), and (3) the relative feasibility of their administration in the field.

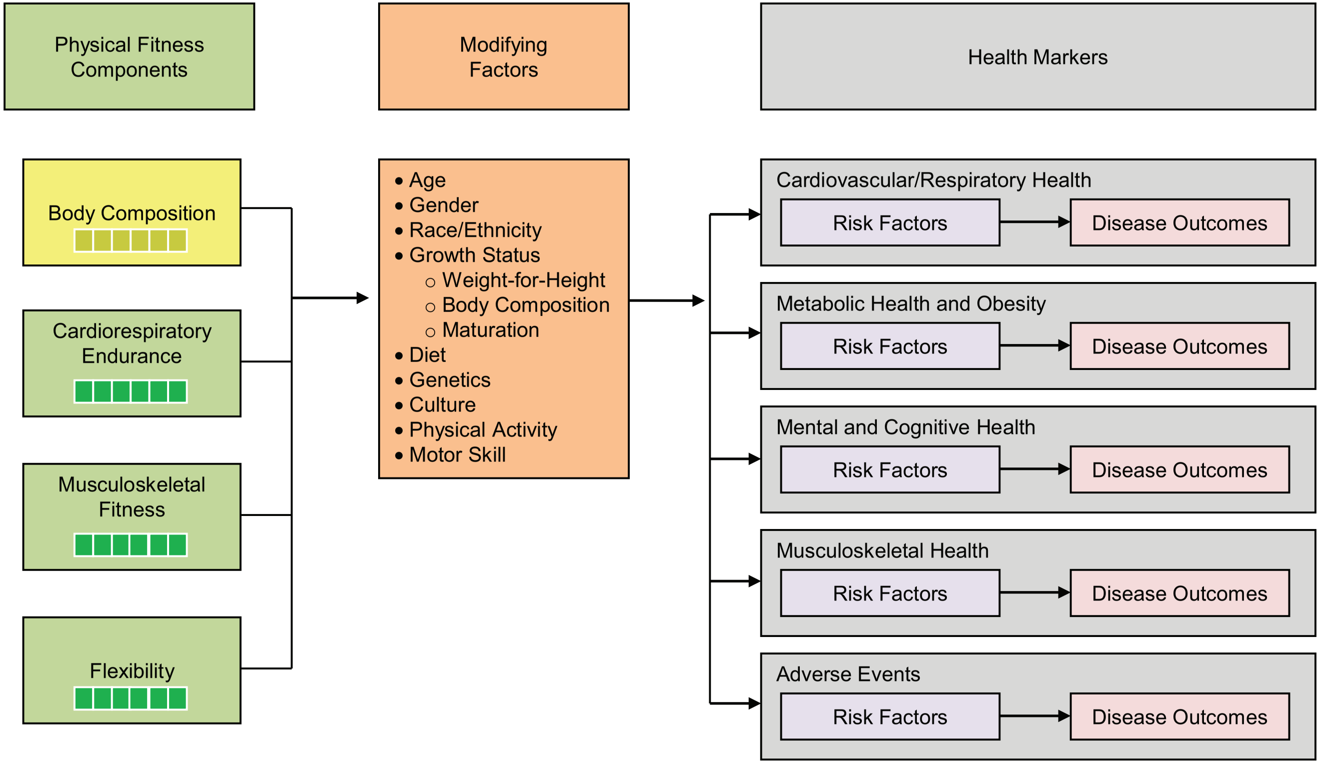

FIGURE 3-1 Physical fitness measures and health outcomes: Conceptual framework.
NOTE: The variety of fitness tests that can measure a particular fitness component is represented by the embedded boxes under each component.

SELECTION OF APPROPRIATE FITNESS TEST ITEMS

Review of the Literature

The committee used various resources to collect scientific data to inform its selection of fitness test items. A main source of information for the committee was a systematic review of the literature conducted by the Centers for Disease Control and Prevention (CDC). The CDC search criteria and vocabulary are described in Box 3-1. The committee supplemented this systematic search with selected publications based on the members' knowledge of the scientific literature.

For the purposes of its review, the CDC defined *health* as a "human condition with physical, social, and psychological dimensions, each characterized as a continuum with positive (i.e., the absence of disease, along with a capacity to enjoy life and withstand its challenges) and negative (i.e., illness and premature death) aspects." The CDC defined *health-related fitness* as the fitness components that have an association with health-related outcomes and are typically identified as aerobic fitness (i.e., cardiorespiratory endurance), muscular strength, muscular endurance, body composition, flexibility, and balance. Only the literature on cardiorespiratory endurance, muscle strength, and muscle endurance was examined systematically, further selected based on the inclusion/exclusion criteria presented in Box 3-1, and abstracted (only experimental and longitudinal studies were abstracted). The CDC considered that the relationship between body composition and health outcomes is well established and therefore conducted a systematic review of it only as a health outcome. Although the CDC performed a systematic search for flexibility, the articles on this component were not abstracted because of limited time and resources. When studies addressing cardiorespiratory endurance or musculoskeletal strength or endurance also included flexibility as a fitness component, however, the CDC abstracted such information. A breakdown of the total number of studies that satisfied the CDC search criteria in Box 3-1 and were abstracted is as follows:

- Cardiorespiratory endurance: 47 experimental, 29 quasi-experimental, 35 experimental (no control), 24 longitudinal
- Musculoskeletal strength: 23 experimental, 12 quasi-experimental, 22 experimental (no control), 6 longitudinal
- Musculoskeletal endurance: 12 experimental, 6 quasi-experimental, 15 experimental (no control), 5 longitudinal
- Flexibility: 7 experimental, 5 quasi-experimental, 9 experimental (no control), 4 longitudinal

Identification and Selection of Test Items

The committee followed the criteria listed in Box 3-2 in a stepwise fashion to select test items for the various components of fitness. Although the search for studies on tests measuring musculoskeletal strength and endurance was conducted separately by the CDC, the committee discusses those tests in Chapter 6 since they all measure dimensions of the same component, musculoskeletal fitness. As Box 3-2 shows, the committee applied five broad criteria in filtering and selecting the best test items for each fitness component: (1) the test item has been described and is currently being utilized; (2) the quality of the research of individual studies showing the relationship between a test item and a health marker is high; (3) based on all the evidence, there is an association between performance on the test item and one or more health markers; (4) the test item has adequate integrity (validity and reliability); and (5) administering the test item in the field is feasible.

The selection of high-quality studies for cardiovascular endurance, musculoskeletal fitness, and flexibility test items was based on the following criteria: (1) study design (e.g., randomized controlled trials versus longitudinal studies), (2) representativeness of the population (e.g., age range), (3) freedom from bias, (4) sample size, (5) validity of health markers, (6) adequacy of description of the intervention, (7) relationship between performance on the test item and one or more health markers, (8) statistical rigor, and (9) adjustment for confounders. Limitations of the scientific literature with regard to these criteria are described in the chapters on the fitness components that were assessed for their relationship to health in youth (Chapters 5, 6, and 7).

Attempting to find associations between performance on fitness tests and health in youth entails important limitations that can help in understanding the committee's approach to reviewing the evidence. In addition to challenges inherent in using field-based (as opposed to laboratory-based) fitness tests, two important challenges arise in analyzing associations between fitness performance and health in youth. First, health constructs in youth are not as well defined as they are in adults; for example, there are questions about whether elevated blood pressure in youth is directly associated with a poor health outcome. Second, diseases that typically are related to low levels of fitness in adults are found with low frequency in youth; therefore, finding an association between performance on fitness tests and health in youth is not highly probable, particularly in studies with small sample sizes. Consequently, studies that investigate the association between performance on a fitness test and health in youth often use health markers rather than health outcomes as health variables. Finding health markers with good ability to predict a future negative health outcome is in itself difficult. Although biological significance and strength of association are typical

BOX 3-1
CDC Literature Search: Criteria and Vocabulary

Literature Databases Searched

- CINAHL
- Sport Discus
- Embase
- Medline
- PsychINFO
- Pubmed
- Web of Knowledge
- Web of Science
- ProQuest
- PsyInfo

Inclusion/Exclusion Criteria

- Peer-reviewed original research
- At least one fitness measure and at least one health measure
- No minimum sample size
- All study designs
- English language
- Population aged 5-18, healthy, obese and nonobese, sedentary and athletic
- Publication date: January 2000-December 2010
- Exclusions: congenital diseases, disabilities (e.g., cerebral palsy, cystic fibrosis, heart abnormalities, motor deficits)

Search Vocabulary for Fitness Measures

Aerobic capacity: aerobic capacity, cardiorespiratory fitness, cardiovascular health, exercise tolerance, endurance capacity, maximum oxygen uptake, oxygen consumption, endurance training, aerobic exercise

Flexibility: range of motion (articular), joint flexibility, range of motion, joint range of motion, motor skills, motor fitness, sit-and-reach, fitnessgram, fitness gram, muscle stretching exercises, hamstring flexibility, low-back flexibility

criteria applied in reviewing scientific evidence, the above challenges led the committee to consider as evidence any significant positive association reported in a high-quality study. The strength of the association between

Muscular endurance: exercise test, fitnessgram, fitness gram, bicep curl, curl-up, pull-up, sit-up, muscle endurance, muscular endurance, physical exertion

Muscular strength: climb, stairs, fitness centers, strength training, resistance training, weight lifting, gymnastics, tree climbing, rope climbing, rock climbing, calisthenics, muscle strength

Search Vocabulary for Health Outcomes

Body composition/obesity/diabetes: obesity, overweight, adiposity, body weight, body mass index, abdominal fat, adipose tissue, glucose metabolism disorders

Cancer: neoplasms

Cardiovascular/pulmonary: cardiovascular diseases, cardiovascular system, cerebrovascular disorders, cardiac function, respiratory tract diseases, chronic obstructive pulmonary disease, lung diseases (obstructive), respiratory function tests, lung function, asthma, cardiovascular risk, bronchoconstriction

Cognitive/neurological: mental health, anxiety, depression, sleep disorders, cognition, memory, attention, dementia, stress (psychological), anger, self concept, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, attention deficit disorder with hyperactivity, pain, gait, gait disorders (neurologic)

Skeletal/muscular/connective: osteoporosis, calcification (physiologic), muscular atrophy, fat-free mass, arthritis, bone and bones, bone density, bone mineral density, scoliosis, rhabdomyolysis

Miscellaneous: risk, injury risk, functional health

a test item and a health marker was categorized as sufficient when most high-quality studies showed a significant association between the test results and a specific health outcome or marker.

BOX 3-2
Stepwise Application of Criteria for Selection of Fitness Test Items

Phase 1: Identification of fitness test items for consideration

- Step 1. Review fitness test batteries currently used (those listed in Table 2-6 in Chapter 2 and other sources).
- Step 2. Identify appropriate test items for each fitness component.
- Step 3. If appropriate and feasible, add test items identified through sources other than the CDC literature search.

Phase 2: Summary of relationship of fitness components to health outcomes

- Step 1. Review and summarize the literature that establishes a link between each fitness component and health outcomes.

Phase 3: Evaluation of relationship of fitness test items to health markers in youth

- Step 1. Review technical reports from currently used fitness tests and consider literature referenced in those reports, especially as it relates to health markers and the integrity of the fitness test items.
- Step 2. Review the CDC literature (and additional sources).
- Step 3. Evaluate the quality of the research of each study; select publications based on the quality of the research.
- Step 4. Based on all the evidence, identify health-related test items that qualify for further review in Phases 4 and 5.

Phase 4: Evaluation of integrity of test items (i.e., validity and reliability)

- Step 1. Review the relevant literature on validity and reliability.
- Step 2. Identify the best test items (based on their association with health markers as identified in Phase 3 and on their integrity).

The committee deliberated at length on the issue of considering studies in adults to draw conclusions about the associations between fitness in youth and health outcomes. The committee also discussed the inclusion of studies exploring the relationship between meeting a level of fitness early in life and specific health outcomes as an adult. The committee concluded

Phase 5. Evaluation of administrative feasibility of test items

- Step 1. Review the relevant literature.
- Step 2. Complete a scorecard for the test items resulting from Phase 4 using the following list of questions:
  - Can the item be conducted in a timely and efficient manner?
  - Does the item impose an acceptable preparation burden on participants?
  - Does the item impose an acceptable preparation burden on administrators?
  - Is the item relatively free of motivational or self-esteem influence?
  - Is the test item free of interpretation misuse?
  - Can it be administered with acceptable privacy?
  - Can it be administered with minimal equipment and space?
  - Is (interpretation of) performance on the item independent of reading comprehension by the participant, socioeconomic status, and age bias?
  - Is performance on the item independent of familiarity with the item?
  - Is performance on the item relatively independent of prior practice?
- Step 3. Assess the test items using the scorecard, considering whether they will be implemented in a national survey or in schools or other educational settings.

Phase 6. Formulation of conclusions and recommendations on appropriate fitness test items for youth.

that the assumptions entailed in extrapolating studies in adults to youth are sometimes uncertain, and therefore decided not to include such studies as the main or sole evidence of an association of a test item with health in youth but only as supportive evidence. In addition, because the literature search was limited to health in youth and prospective studies did not cover the length of time until youth reach adulthood, studies of the long-term consequences of fitness (i.e., whether reaching a specific fitness level as a youth would result in health consequences as an adult) were considered only as secondary evidence in identifying fitness test items that are associated with health in youth. However, the committee understands the importance of assessing how well a health marker tracks into adulthood and identifying appropriate health markers in youth and makes recommendations to this effect.

The approach used to select appropriate test items for body composition differed. As noted earlier, the relationships between body composition, particularly percent body fat, and health outcomes are well established in both youth and adults. For this reason, the CDC did not provide the committee with a systematic review of literature demonstrating the relationship between test items for body composition and health outcomes. Therefore, to identify appropriate test items for body composition, the committee selected field-based items that were valid, reliable, and feasible for either a national survey or an educational setting.

CONSIDERATION OF MODIFYING FACTORS

Gender and age often are analyzed as confounders or effect modifiers in research studies, and gender- and age-specific cut-points have been set for fitness test items. There are, however, other factors that may modify performance on fitness test items or the association between an item and health but are not routinely included in surveys of youth fitness or in study designs.

For some modifiers (e.g., gender, age, race/ethnicity, body composition, maturation status, motor skill), evidence for their influence on specific fitness test items varies in quality (see, for example, Chapter 4 on the influence of body composition on fitness tests). In the studies reviewed, insufficient data were available with which to assess the influence of several potential modifiers—gender, age, race/ethnicity, body composition, maturation status, motor skill—on performance on cardiorespiratory endurance, musculoskeletal fitness, and flexibility tests. When these modifiers were considered in the designs of the studies reviewed, the committee comments on the results in the respective chapters (Chapters 5-7).

The committee recommends including measures of some of these factors in surveys and study designs to gain a better understanding of their influence. As more research is undertaken and survey data are collected, it may be appropriate to establish cut-points based on selected modifiers, but at this time the committee recommends that only age- and gender-based cut-points be established (see Chapter 8). Age- and gender-based cut-points, however, need to be interpreted in the context of other potential modifiers (e.g., body composition, demographic characteristics). For that reason, the committee highlights the importance of training those who will be interpreting and communicating results (see Chapters 8 and 9). An important aspect of this training is learning to be cognizant of the influence of modifying factors.

The following is a summary of what is known about the potential effects of maturation status, motor skill, and demographic factors (race/ethnicity, socioeconomic factors) on fitness performance. Although body composition is also categorized as a modifying factor in the committee's framework, a summary of what is known about the influence of selected elements of body composition on fitness performance is included in Chapter 4 and is not repeated here. The committee's recommendations for including some of these factors in national surveys and research study designs can be found in Chapters 8 and 10, respectively.

Maturation Status

The literature review revealed extensive documentation of changes in fitness with age per se (Branta et al., 1984; Carron and Bailey, 1974; Ellis et al., 1975; Espenschade, 1960; Froberg and Lammert, 1996; Haubenstricker and Seefeldt, 1986; Keogh, 1965; Malina and Roche, 1983; Malina et al., 2004; Mizuno et al., 1973; Ostyn et al., 1980; Simons et al., 1990) and with individual differences in maturity status within a given age group (Espenschade, 1940; Jones, 1949; Little et al., 1997; Malina et al., 2004). This effect may be due in part to differentials in the timing of maturation. There are two indicators of this timing: age at menarche and age at peak height velocity (a measure of the maximum rate of growth in stature during a growth spurt) (Malina et al., 2004). Several measures of physical fitness have their own spurts, which have been documented more in boys than in girls (Beunen and Malina, 1988; Beunen et al., 1988; Carron and Bailey, 1974; Heras Yague and de la Fuente, 1998; Kemper and Verschuur, 1985; Mirwald and Bailey, 1986). Variation in fitness among youth has been classified as late, on time (average), or early in skeletal age, age at menarche, and timing of peak height velocity (Jones, 1949; Lefevre et al., 1990; Little et al., 1997; Malina et al., 2004). Differences in maturity are more marked for boys than for girls, although relevant data are not as extensive for girls (Malina et al., 2004). While evidence suggests a relationship between maturity stages per se and performance on fitness tests, some questions remain to be answered—in contrast to the influence of body size and/or body composition on fitness, which is affected independently by individual differences in maturation (Malina et al., 2004). The following four methods can be used to measure maturity.

Skeletal Age

Skeletal age indicates the level of skeletal maturity of the hand-wrist attained at a given point in time and can be measured from infancy through adolescence. Three measurement methods are commonly used: the Greulich-Pyle and Fels methods, based on American children, and the Tanner-Whitehouse method, based on British children. Measurement of skeletal age requires a small radiation dose and expertise in assessing films (Beunen et al., 2006; Malina, 2011; Malina et al., 2004).

Secondary Sex Characteristics

Secondary sex characteristics include pubic hair and genitalia in boys, and breasts and pubic hair in girls. Privacy and cultural issues arise with this measure, although self-assessments increasingly are being used. Use of this measure also is limited to the pubertal period. The criteria described by Tanner (1962) are most commonly used, whereby secondary sex characteristics are rated on five discrete-point scales (stages from prepubertal to mature status) superimposed upon a continuous process of sexual maturation. It is important to recognize that youth should not be grouped by developmental stage across chronological ages (Beunen et al., 2006; Malina et al., 2004). Moreover, there are differences among white, African American, and Mexican American youth, with African Americans beginning pubertal maturation in advance of their Mexican American and white counterparts (Chumlea et al., 2003; Sun et al., 2002, 2005).

Age at menarche and menarcheal status are an indicator of maturity in girls. As with pubertal stages, girls should be grouped by menarcheal status within each chronological age year (Malina et al., 2004).

Somatic Maturation

Somatic maturation is an after-the-fact indicator. It is defined as the age at the maximum rate of growth in height during the adolescent spurt (peak height velocity) and is an indicator of maturity timing. Estimation of age at peak height velocity requires longitudinal data spanning at least 5-6 years around the spurt (Beunen and Malina, 1988; Beunen et al., 2006; Malina et al., 2004).

Noninvasive Indicators

Two indicators of maturity status considered noninvasive have recently been used: (1) percentage of predicted mature (adult) height attained at a given age (Roche et al., 1983), and (2) maturity offset or predicted time before or after peak height velocity (Mirwald et al., 2002). The former is an indicator of status, while the latter is an indicator of timing.

The most accurate prediction equations for mature height require the age, height, and weight of the child and midparent height (i.e., the average of the parents' height). The prediction equations for maturity offset require age, height, weight, sitting height, and estimated leg length (height minus sitting height, technically subischial length) (Mirwald et al., 2002). Limitations of this measure are that it requires an additional measurement and a flat sitting surface and the fact that ethnic variation is a potential confounder (Hamill et al., 1973; Malina et al., 1974, 1986, 1987; Martorell et al., 1988). It should also be noted that leg length has its adolescent growth spurt before sitting height (Malina et al., 2004).

Although the above measures have been used to assess various populations of athletes (Cumming et al., 2006; Malina et al., 2005; Nurmi-Lawton et al., 2004; Sherar et al., 2007), they have not been applied and validated in large samples of youth (Malina et al., 2012).

Motor Skill

The association between motor skill (i.e., motor coordination and control) development and health-related fitness performance and health outcomes has not been thoroughly investigated. However, a growing body of cross-sectional and longitudinal evidence demonstrates positive relationships between motor skill competence levels and multiple aspects of health-related fitness in youth. Given the lack of experimental data, the literature does not provide adequate support for a recommendation to include a motor skill measure in a national youth fitness test battery. Further research is needed to examine the relationships between the development of motor skill and health-related fitness performance and health outcomes.

Motor Skill and Health-Related Fitness Performance

Children do not develop motor skill through maturation alone, but also through context-specific engagement in physical activity (Logan et al., 2011; Stodden et al., 2008). Successful technical completion of multijoint motor skill and fitness tests is promoted through practice and experience, and also is linked to a child's maturation status. Therefore, it is logical to consider the potential influence of an individual's motor skill status on the performance of tasks involving coordinated movements. Without intervention or formal instruction, such as physical education, youth with lower levels of motor coordination and control (i.e., low motor skill) are more likely to exhibit decreased performance on physical fitness tests (Cantell et al., 2008; Castelli and Erwin, 2007; Castelli and Valley, 2007; Erwin et al., 2007; Haga, 2009; Matvienko and Ahrabi-Fard, 2010; O'Beirne et al., 1994; Okely et al., 2001; Schott et al., 2007). However, the use of multiple types of motor skill assessments (i.e., process- and product-oriented assessments) (Fisher et al., 2005; McKenzie et al., 2002; Okely et al., 2001; Stodden et al., 2008) makes it difficult to compare relationships across studies and developmental time. Additionally, issues related to a lack of developmental validity, the sensitivity and skill-level discrimination capabilities of various assessments, and a lack of consensus on how motor skill versus health-related fitness is defined need to be addressed in future research (Fisher et al., 2005; McKenzie et al., 2002; Okely et al., 2001; Stodden et al., 2008).

Recently published data show highly variable correlation strengths (r = 0.00-0.74) between individual or composite product- or process-oriented motor skill assessments and individual health-related fitness measures in children, adolescents, and young adults, including both males and females (Anliker et al., 2011; Barnett et al., 2008a,b; Castelli and Valley, 2007; Castro-Piñero et al., 2010; Hands et al., 2009; Matvienko and Ahrabi-Fard, 2010; Okely et al., 2001; Tveter and Holm, 2010). Using multivariate regression, explained variance in either individual or composite measures of fitness by multiple individual or composite motor skill assessments has ranged from 0 to 79 percent (Barnett et al., 2008a,b; Hands, 2008; Stodden et al., 2009). With the exception of one quasi-experimental study (Matvienko and Ahrabi-Fard, 2010) and three longitudinal studies (Barnett et al., 2008a,b; Hands et al., 2009), these data were derived from cross-sectional study designs. Sample sizes varied from 230 to 2,026 in all studies except those of Matvienko and Ahrabi-Fard (2010) and Hands and colleagues (2009), which included only 90 and 19 subjects, respectively. It is important to note that the strength of relationships generally increases across age in both males and females. Relationships between various motor skills and flexibility generally are weaker (r = 0.01-0.25) than is the case for other aspects of health-related fitness performance.

Results of recent cross-sectional and longitudinal research examining associations between motor skill competence levels and body weight status (i.e., percent body fat and body mass index [BMI]), as either a component of fitness or a health outcome, indicate that motor skill competence is inversely correlated with body composition. As demonstrated with other aspects of health-related fitness, the strength of associations between motor skill competence and body weight status varies, but generally increases over time (r = 0.05-0.73) (Burgi et al., 2011; D'Hondt et al., 2009, 2011, 2012; Hands and Larkin, 2006; Lopes et al., 2011, 2012; Martins et al., 2010; Stodden et al., 2009). As body weight status may influence both motor skill and health-related fitness performance, it is difficult to identify a causal pathway for these relationships. Independent of a causal pathway, however, these cross-sectional and longitudinal data indicate that youth with low motor skill competence have a higher risk of unhealthy trajectories not only for weight gain but also for physical activity and multiple aspects of health-related physical fitness (Barnett et al., 2009; Haga, 2009; Stodden et al., 2009). These data indirectly support the hypothesis of Stodden and colleagues (2008) that the development of motor skills may promote improvements in body weight status, physical activity, and health-related physical fitness through the dynamic and reciprocal relationships that occur among these variables across childhood. As mentioned previously, there is a need for long-term experimental studies to better understand the impact of motor skill development on body weight status and various aspects of health-related fitness.

It has been suggested that associations between cardiorespiratory endurance and fundamental motor skills are indirectly related to developmental trajectories of motor skill development, are reciprocal in nature (Barnett et al., 2008b, 2011; D'Hondt et al., 2011; Stodden et al., 2008), and may be sustained over time from childhood (Barnett et al., 2008b; Burgi et al., 2011; D'Hondt et al., 2011; Hands and Larkin, 2006; Martins et al., 2010). Associations between motor skill (i.e., motor control and coordination) performance and muscular strength and endurance performance are linked to direct mechanisms involving many aspects of neuromuscular adaptation (Enoka, 2002; Ratamess, 2008). This link supports the notion that motor skill development influences these variables (Myer et al., 2011; Stodden et al., 2008).

Although most of the data reported above were derived from correlational or prospective longitudinal studies, the increasing relationship strength trajectories between motor skill and fitness levels across ages suggest the need for additional research on the relationship trajectories between motor skill development and health outcomes. The committee could identify only a few studies examining the relationship between motor skill and any health outcomes.

Motor Skill and Health Outcomes

A small body of research (cross-sectional and longitudinal studies) indicates that low motor skill competence is associated with poor bone health in youth (Anliker et al., 2011; Vicente-Rodriguez et al., 2004, 2008). High-loading activities such as jumping and hopping, which integrate skill as well as strength and power attributes, have demonstrated strong relationships ($r = 0.65$-0.81) with site-specific lower-extremity bone mineral density in studies with 323 and 28 adolescents, respectively (Anliker et al., 2011; Vicente-Rodriguez et al., 2004). However, these relationships also can be attributed to lean mass (Anliker et al., 2011; Vicente-Rodríguez et al., 2008). Weaker associations have been demonstrated with other skill-related tests (e.g., running, agility), as well as activities demanding specific sports skills (Vicente-Rodriguez et al., 2004). Overall, the proposed synergistic relationships and mechanisms involving skill development and muscle strength, power, and endurance make it difficult to delineate the contributions of skill, strength, and power to bone health (see Chapter 6 for a discussion of the relationship between musculoskeletal fitness and bone mineral density).

Even fewer studies have examined the relationship between any aspect of motor skill and cardiovascular or metabolic health outcomes in youth. In a cross-sectional study (N = 149), blood lipid profiles were weakly associated with motor skill (Cantell et al., 2008). One longitudinal study (N = 1,192) addressing cardiovascular health (Monyeki et al., 2008) showed no significant relationships between motor skill levels and blood pressure (systolic and diastolic).

In longitudinal and cross-sectional studies, relationships between aspects of motor skill competence and mental health outcomes generally range from weak to moderate (r = 0.10-0.68), depending on the nature of the behavior or disorder. Many of these studies, however, involved participants with mental or associated cognitive, motor, emotional, or behavioral developmental disorders (Emck et al., 2011; Piek et al., 2007, 2008, 2010; Skinner and Piek, 2001). Thus, these data may not be representative for normal populations of children.

Influence of Amount of Practice Time on Fitness Testing Performance

Evidence for the effects of practice on performance on specific health-related physical fitness tests (i.e., 1-mile run, 20-meter shuttle run, curl-ups) is lacking. The relevant literature on the relationship between motor learning and development and general skill learning indicates that adequate learning and completion of a fitness test depend on many factors, including experience, instruction, feedback, cognitive capabilities, motivation, and the complexity of the test (Farpour-Lambert and Blimkie, 2008; Raudsepp and Pall, 2006). In addition, factors associated with age, level of coordination and control, and body composition may affect the short-term capability to learn and perform a test (D'Hondt et al., 2011; Hands and Larkin, 2006; Lloyd et al., 2003).

Race/Ethnicity and Socioeconomic Factors

The associations among race/ethnicity, gender, socioeconomic status, and health outcomes are well established (Baker et al., 2006; CDC, 2011; Floyd et al., 2009; Miech et al., 2006; National Center for Health Statistics, 2011; Whitt-Glover et al., 2009). The known presence of health disparities by race/ethnicity and socioeconomic status highlights the potential importance of understanding the effects of these factors on fitness performance. Race/ethnicity and socioeconomic status have been identified as potential modifiers of test performance and interpretation, but their role is unknown and should be further investigated and included in surveys. For example, although the committee found no evidence to support the influence of socioeconomic status on test selection, delivery, or interpretation, one could hypothesize possible limitations due to lack of equipment if a school had a suboptimal built environment (e.g., having no outdoor track would limit the options for cardiorespiratory fitness tests). Measurement of race/ethnicity and socioeconomic status in research studies and surveys has become increasingly common as (1) social demography has expanded as a scholarly enterprise, (2) the U.S. population has become more diverse, and (3) evidence of the scientific relevance of demographic characteristics to human development and health outcomes has become more widely known (Entwisle and Astone, 1994).

The committee's review included some studies that describe the population in terms of ethnic and racial background; often, however, no statistical analysis of the effects of these factors was carried out. One exception is the case of body composition, where the differences associated with race/ethnicity and socioeconomic status are well established. Previous youth fitness surveys in the United States have failed to consider these factors (see Malina, 2007).

Previous studies have typically used self-report, parent questionnaires, information taken from existing records, or assignment by a field worker to assess race/ethnicity among youth (Entwisle and Astone, 1994). A more appropriate method of assessing race/ethnicity in surveys and research would be the data collection standard for race and ethnicity developed by the Office of Management and Budget (OMB, 1997), which federal programs have been required to follow since 2003.

Previous studies have used proxy measures to collect information on socioeconomic status from youth themselves, such as the number of books or cars the family owns, parental education or occupation, and the number of rooms in the household.[1] However, this method is potentially problematic and could have different implications for different parts of the country (e.g., children living in areas where public transportation is easily available and preferred, such as New York City, may not live in households with cars, and increased access to electronic books may mean children are not actually aware of the number of books the family owns). A better way to

[1]Available at http://www2.lse.ac.uk/media@lse/research/EUKidsOnline/BestPracticeGuide/FAQ26.aspx (accessed August 27, 2012).

collect information on socioeconomic status would be a questionnaire to parents (Entwisle and Astone, 1994; Merola, 2005) at the time they are asked to provide consent for their children to participate in fitness testing. If collecting this information directly from parents is not feasible, it may be estimated from official school-level statistics, such as percentage of students eligible for free or reduced-price lunch (Merola, 2005).

Disability

The information in this report is driven by the evidence for healthy study populations and directed to the general population. However, the committee recognizes the significance of physical fitness for the health of youth with various disabilities (as defined in Appendix B), especially since they are likely to be less active than their nondisabled peers (Physical Activity Guidelines Advisory Committee, 2008). Other reviews, such as the report of the Physical Activity Guidelines Advisory Committee (2008), specifically examine the relationship between health outcomes and physical activity in people with disabilities (see also Seaman, 1999). More information on including youth with disabilities in school-based fitness testing is presented in Chapter 9.

ESTABLISHMENT OF CUT-POINTS FOR HEALTH-RELATED YOUTH FITNESS TESTS

As noted earlier, interpretation of the results of health-related fitness testing is one of the most important aspects of such testing, making the setting of cut-points essential. Cut-points serve as a way of discerning between individuals and populations that may be at risk of poor health outcomes based on performance on a fitness test and those that are not. For example, a 10-year-old boy with a BMI above 22 would be considered at increased risk of poor health. Cut-points thus are tools for communicating with participants, families, health and school officials, and the public in general about fitness status and setting goals for improvement.

Many challenges are entailed in setting and interpreting these cut-points (or standards). An important challenge, alluded to earlier, is the fact that most health outcomes can take years to develop, and there is no concurrent relationship between fitness test performance and any actual health outcomes in youth. In light of this challenge, the committee adopted the term *health marker* (or health risk factor) instead of *health outcome*. This was done with the assumption that those health markers are associated with a risk of a future health outcome. Another challenge is that, although the fitness and health variables are often continuous (i.e., a continuous improvement in fitness would be related to a continuous improvement in health), for communication purposes, evaluation standards frequently are established as a discrete variable (e.g., meets or does not meet the standard). The results can still be presented as part of a continuous distribution to convey the concept that health is a continuous variable, and any improvement in fitness will likely be associated with a lower risk of negative health outcomes.

This section describes two evaluation approaches commonly used in setting health-related youth fitness standards—norm- and criterion-referenced (Safrit and Wood, 1989). It then presents the committee's guidance on methods for establishing standards for health-related youth fitness testing based on the available evidence.

Norm-Referenced Versus Criterion-Referenced Evaluation

Norm-Referenced Setting of Cut-Points and Evaluation of Fitness

In the norm-referenced evaluation approach, a test taker's performance is compared with that of his/her peers, often by gender and age. For example, students might be categorized depending on whether they score below, equal to, or higher than the 85th percentile of their peers. When the interest is in performance (e.g., strength of an individual's upper body), norm-referenced evaluation is appropriate. Computing and deriving norms (e.g., percentiles) is relatively simple as long as data from a nationally representative sample exist and can be updated. The following are four important limitations associated with this norm-referenced evaluation approach to setting cut-points:

- *Time dependence*—Population distributions, and therefore norm-referenced values, change with time. Updating these values is costly in terms of both time and human resources, but is necessary to avoid misleading results.
- *Population dependence*—The interpretation depends on the fitness of the reference population. If a population is abnormally healthy or unhealthy, the comparisons with an individual fitness level (e.g., better than average) will not be meaningful.
- *Discouraging unfit youth*—The use of a norm-referenced approach tends to reward youth who are already fit while potentially discouraging those who are not fit because they know their chances of achieving the standard are low.
- *Favoring genetically talented or punishing disadvantaged youth*—Standards based on the norm-referenced approach often are set at the high end of a population and thus may favor genetically talented or punish disadvantaged youth.

Criterion-Referenced Setting of Cut-Points and Evaluation of Fitness

Most of the limitations listed above are overcome with a criterion-referenced evaluation approach, whereby a test taker's performance is compared with an absolute criterion related to whether a child meets a minimal necessary physical fitness level. In contrast with the norm-referenced approach, because the criterion is defined independently, it is not impacted by changes in a population that occur over time or in the level of fitness of a specific population. Limitations related to genetic differences or the potential for discouraging unfit participants also can be minimized with this approach.

Many methods have been developed for setting performance standards (Cizek, 2001; Livingston and Zieky, 1982), but for criterion-referenced evaluation of health-related fitness, the health outcome–centered method (Zhu et al., 2011) has predominated. Basically, this method involves linking health-related fitness performance with a particular (set of) health outcome measure(s). Specifically, this approach identifies a level of test performance that discriminates, with acceptable specificity and sensitivity, between participants who have and do not have a defined health characteristic. An example is identifying a level of performance on a measure of cardiorespiratory endurance that discriminates between groups of youth who have or do not have an at-risk score for metabolic syndrome.

Steps completed before developing cut-points for the health-outcome centered method include determining the components of health-related fitness (e.g., cardiorespiratory endurance) and selecting valid, reliable criterion measures and field tests and health outcomes or markers. Field tests are selected because even though criterion measures (criterion-referenced standards) are the most accurate measure of a construct, they often are more expensive and time-consuming and require sophisticated equipment. Field tests are more practical, less costly, and less time-consuming for mass testing. It is important, however, to determine the validity and reliability of field tests by deriving the predictive relationship and determining its consistency with the selected criterion measure.

The selection of health outcomes can be based on the expected relationship between field tests and health markers or outcomes. Health is a construct, so there are many possible health outcome measures, such as mortality, a single risk factor (e.g., blood pressure), or a group of risk factors (e.g., metabolic syndrome). Because no specific measure is considered superior, it is advisable to use multiple outcome measures to validate the test results when possible. Those selecting cut-points also need to identify health outcome measures using existing standards (e.g., a systolic blood pressure level of ≥103 mmHg for a 5-year-old child 104 cm in height) and make adjustments for specific populations if needed.

Guidance for Establishing Cut-Points for Youth Fitness Tests

As mentioned above, setting cut-points for evaluating results of health-related fitness tests entails a number of challenges, one of which is the lack of appropriate data. Ideally, cut-points are established from data on performance on a specific fitness test and health outcomes in a broad-based youth population. More often, however, these data are not available; instead, there may be enough data on an association with health in the adult population but only growing evidence from studies in small samples of the youth population. In yet a different scenario, only growing evidence for an association between a specific test and health exists, with no data coming from broad populations. Until data in broad youth populations are generated so cut-points can be derived, cut-points should be referred to as *interim criterion-referenced cut-points* (or *interim cut-points*). Various approaches can be used to select these interim criterion-referenced cut-points, depending on the available data. The committee's guidance on these approaches is presented in this section. Often a number of field tests are used simultaneously to measure the same fitness component. The reader is referred to Zhu et al. (2010) and Jackson (1989) for information on setting cut-points for multiple tests of a single component.

Several considerations apply in interpreting the results of fitness tests. For health-related fitness testing in youth, the key interest is not only whether a test taker is "fit enough" to be free of potential health risks but also whether the test taker is "fit enough for the future." In addition, because the key outcome of interest of the criterion-referenced approach to evaluating test results is classification (e.g., being at risk of a health outcome versus not being at risk), the accuracy of the classification is key. Further, regardless of how well the related cut-point is established, it will be possible to misclassify individuals. There are two kinds of misclassification: (1) when a fit test taker is misclassified as unfit and (2) when an unfit test taker is misclassified as fit. The committee considered the first of these to be more problematic because it would result in a greater likelihood of recommending an exercise intervention to youth who do not need it, thereby depleting already limited resources that should be used for youth who need them the most. To minimize the effects of misclassification, cut-points need to be validated or cross-validated using additional measures and samples.

Finally, whether cut-points should be established differently for various subpopulations must be examined and determined empirically. As discussed earlier, while age and gender often have been taken into consideration in setting cut-points, many other factors, such as race/ethnicity, maturation status, and disability, have not been considered.

Once cut-points have been established for a specific test and age/gender group, they should be used in interpreting test results and communicating with participants, families, school and health officials, and the general public. Doing so will minimize the confusion that might arise from communicating in terms of the percentiles used to derive the cut-points. For example, the CDC has used the 95th percentile from a previous decade to derive cut-points for obesity in children, yet more than 15 percent of youth currently exceed that 95th percentile. These are challenging issues to communicate. In cases where percentiles may allow for a clearer presentation of the results than cut-points, as with BMI, the year of data collection should be reported with the percentile. In this connection, researchers developing percentile data with which to derive cut-points should also report the time of data collection.

Establishing Cut-Points When the Relationship Between the Test of Interest and a Health Marker in Youth Is Known

In the ideal situation, when there is a concurrent relationship between a health outcome and a fitness test, the cut-points for the test are determined using a data mining procedure[2] to establish statistical evidence for the relationship. While not common, this kind of concurrent relationship does exist and has been used for setting cut-points. For example, based on the concurrent relationship between body composition and a set of health outcome measures (total cholesterol, serum lipoprotein ratio, and blood pressure) (Williams et al., 1992), a set of cut-points was derived for evaluating body composition (Going et al., 2008). Similar applications have been reported for setting cut-points for cardiorespiratory endurance (Lobelo et al., 2009) and waist circumference in Chinese school-aged children (Liu et al., 2010) and for body composition and cardiorespiratory endurance tests (Going et al., 2011; Welk et al., 2011).

Establishing Cut-Points When a Concurrent Relationship Has Not Been Confirmed in Youth

Even if a concurrent relationship between a health outcome measure and a putative health-related fitness test has been well established in adults and cut-points exist for that population, such a relationship often has not been confirmed in youth. Because a negative health outcome (e.g., low-back pain, cardiovascular diseases) may take years to develop, children's health

[2]Data mining involves varying cut-points, computing agreement-related statistics with the classification of health outcome measures each time, and determining the cut-points according to optimal statistical results (e.g., highest P- and kappa-coefficients, specificity index and relative risk statistics, findings and illustration of receiver operating characteristic [ROC] curve analysis). If the cut-points are set across a large range of groups, the data from these groups usually are smoothed before the data mining procedure is applied.

outcome measures may be in a normal range even if they are not fit. Methods of setting cut-points based on evidence in adults assume that a fit child will likely become a fit adult. While there is some evidence to support this assumption for certain fitness components (see, e.g., Beunen et al., 1992; Campbell et al., 2001; Malina, 1996, 2001), more research, especially longitudinal studies, is needed to confirm this assumption.

When a relationship can be confirmed only in adults, there are two methods for estimating the cut-points for health-related fitness tests in youth—the relative position and the panel-driven methods. With the *relative position method,* the percentile of adults considered to be at risk based on their performance on a fitness test is taken as the fitness standard in youth. For example, the lowest 20th percentile for performance on a cardiorespiratory endurance test item could be selected based on the demonstration by Blair and colleagues (1989) that morbidity and mortality are disproportionately elevated in the lowest quintile for performance on a maximal treadmill test in adults. In the *panel-driven method*, a panel of experts uses the cut-points from adults and all available information (e.g., growth curves and performance characteristics for different ages and genders) to derive the cut-points for youth. For example, the criterion maximal oxygen uptake (VO_2max) value in youth could be determined in various ways, ranging from expert opinion to extrapolation from associations between VO_2max and health outcomes in adults. The panel-driven method was used to set the Fitnessgram® standards for cardiorespiratory endurance test items (Cureton, 1994; Cureton and Warren, 1990).

Establishing Cut-Points When the Relationship Between a Fitness Test and Health Outcomes Is Not Confirmed in Youth or Adults

While the importance of some fitness components to health has been suggested, the relationship between specific fitness test items and health outcomes may not be confirmed. For example, while the validity and reliability of commonly used tests have generally been well established, evidence for the importance of muscular strength for health in adults is still growing and may be equivocal for some tests, and for youth remains largely unconfirmed. Until these relationships are confirmed, an alternative approach for setting cut-points is to use the *comparatively relative position method*, in which a percentile established for another measure is borrowed. If the percentile from another test is borrowed, the two tests should be as comparable as possible in their nature (e.g., both require movement of the body) and in the dimension they measure (e.g., upper-body strength). For example, if the cut-points for tests of the cardiorespiratory endurance component derived through a criterion-referenced evaluation procedure were set at about the 20th percentile, the cut-points for tests of the musculoskeletal fitness component would also be set temporarily by being derived from the 20th percentile. Until health-related cut-points were developed specifically for the test of interest, these interim cut-points might be used.

REFERENCES

Anliker, E., R. Rawer, U. Boutellier, and M. Toigo. 2011. Maximum ground reaction force in relation to tibial bone mass in children and adults. *Medicine and Science in Sports and Exercise* 43(11):2102-2109.

Baker, E. A., M. Schootman, E. Barnidge, and C. Kelly. 2006. The role of race and poverty in access to foods that enable individuals to adhere to dietary guidelines. *Preventing Chronic Disease* 3(3):A76.

Barnett, L. M., P. J. Morgan, E. van Beurden, and J. R. Beard. 2008a. Perceived sports competence mediates the relationship between childhood motor skill proficiency and adolescent physical activity and fitness: A longitudinal assessment. *International Journal of Behavioral Nutrition and Physical Activity* 5.

Barnett, L. M., E. Van Beurden, P. J. Morgan, L. O. Brooks, and J. R. Beard. 2008b. Does childhood motor skill proficiency predict adolescent fitness? *Medicine and Science in Sports and Exercise* 40(12):2137-2144.

Barnett, L. M., E. V. Beurden, P. J. Morgan, L. O. Brooks, and J. R. Beard. 2009. Childhood motor skill proficiency as a predictor of adolescent physical activity. *Journal of Adolescent Health* 44(3):252-259.

Barnett, L. M., P. J. Morgan, E. Van Beurden, K. Ball, and D. R. Lubans. 2011. A reverse pathway? Actual and perceived skill proficiency and physical activity. *Medicine and Science in Sports and Exercise* 43(5):898-904.

Beunen, G., and R. M. Malina. 1988. Growth and physical performance relative to the timing of the adolescent spurt. *Exercise and Sport Sciences Reviews* 16:503-540.

Beunen, G., R. M. Malina, M. Van't Hof, J. Simons, M. Ostyn, R. Renson, and D. Van Gerven. 1988. *Adolescent growth and motor performance: A longitudinal study of Belgian boys*. Champaign, IL: Human Kinetics Publishers.

Beunen, G. P., R. M. Malina, R. Renson, J. Simons, M. Ostyn, and J. Lefevre. 1992. Physical activity and growth, maturation and performance: A longitudinal study. *Medicine and Science in Sports and Exercise* 24(5):576-585.

Beunen, G. P., A. D. Rogol, and R. M. Malina. 2006. Indicators of biological maturation and secular changes in biological maturation. *Food and Nutrition Bulletin* 27(4, Suppl. Growth Standard):S244-S256.

Blair, S. N., H. W. Kohl, III, R. S. Paffenbarger, Jr., D. G. Clark, K. H. Cooper, and L. W. Gibbons. 1989. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *Journal of the American Medical Association* 262(17):2395-2401.

Branta, C., J. Haubenstricker, and V. Seefeldt. 1984. Age changes in motor skills during childhood and adolescence. *Exercise and Sport Sciences Reviews* 12:467-520.

Burgi, F., U. Meyer, U. Granacher, C. Schindler, P. Marques-Vidal, S. Kriemler, and J. J. Puder. 2011. Relationship of physical activity with motor skills, aerobic fitness and body fat in preschool children: A cross-sectional and longitudinal study (Ballabeina). *International Journal of Obesity* 35(7):937-944.

Campbell, P. T., P. T. Katzmarzyk, R. M. Malina, D. C. Rao, L. Pérusse, and C. Bouchard. 2001. Prediction of physical activity and physical work capacity (PWC150) in young adulthood from childhood and adolescence with consideration of parental measures. *American Journal of Human Biology* 13(2):190-196.

Cantell, M., S. G. Crawford, and P. K. Doyle-Baker. 2008. Physical fitness and health indices in children, adolescents and adults with high or low motor competence. *Human Movement Science* 27(2):344-362.

Carron, A., and D. A. Bailey. 1974. Strength development in boys from 10 through 16 years. *Monographs of the Society for Research in Child Development* 157.

Castelli, D. M., and H. E. Erwin. 2007. A comparison of personal attributes and experiences among physically active and inactive children. *Journal of Teaching in Physical Education* 26(4):375-389.

Castelli, D. M., and J. A. Valley. 2007. The relationship of physical fitness and motor competence to physical activity. *Journal of Teaching in Physical Education* 26:358-374.

Castro-Piñero, J., F. B. Ortega, E. G. Artero, M. J. Girela-Rejon, J. Mora, M. Sjostrom, and J. R. Ruiz. 2010. Assessing muscular strength in youth: Usefulness of standing long jump as a general index of muscular fitness. *Journal of Strength and Conditioning Research* 24(7):1810-1817.

CDC (Centers for Disease Control and Prevention). 2011. CDC grand rounds: Childhood obesity in the United States. 2011. *Morbidity and Mortality Weekly Report* 60(2):42-46.

Chumlea, W. C., C. M. Schubert, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and S. S. Sun. 2003. Age at menarche and racial comparisons in US girls. *Pediatrics* 111(1):110-113.

Cizek, G. J. 2001. *Setting performance standards: Concepts, methods, and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cumming, S. P., R. A. Battista, S. Martyn, M. E. Ewing, and R. M. Malina. 2006. Estimated maturity status and perceptions of adult autonomy support in youth soccer players. *Journal of Sports Sciences* 24(10):1039-1046.

Cureton, K. J. 1994. Aerobic capacity. In *The Prudential Fitnessgram technical reference manual,* edited by J. R. Morrow, H. B. Falls, and H. W. Kohl. Dallas, TX: The Cooper Institute for Aerobics. Pp. 33-55.

Cureton, K. J., and G. L. Warren. 1990. Criterion-referenced standards for youth health-related fitness tests: A tutorial. *Research Quarterly for Exercise and Sport* 61(1):7-19.

D'Hondt, E., B. Deforche, I. De Bourdeaudhuij, and M. Lenoir. 2009. Relationship between motor skill and body mass index in 5- to 10-year-old children. *Adapted Physical Activity Quarterly* 26(1):21-37.

D'Hondt, E., B. Deforche, R. Vaeyens, B. Vandorpe, J. Vandendriessche, J. Pion, R. Philippaerts, I. de Bourdeaudhuij, and M. Lenoir. 2011. Gross motor coordination in relation to weight status and age in 5- to 12-year-old boys and girls: A cross-sectional study. *International Journal of Pediatric Obesity* 6(2-2):e556-e564.

D'Hondt, E., B. Deforche, I. Gentier, I. De Bourdeaudhuij, R. P. Vaeyens, R., and M. Lenoir. 2012. A longitudinal analysis of gross motor coordination in overweight and obese children versus normal-weight peers. *International Journal of Obesity* [Epub ahead of print].

Ellis, J. D., A. V. Carron, and D. A. Bailey. 1975. Physical performance in boys from 10 through 16 years. *Human Biology* 47(3):263-281.

Emck, C., R. J. Bosscher, P. C. Van Wieringen, T. Doreleijers, and P. J. Beek. 2011. Gross motor performance and physical fitness in children with psychiatric disorders. *Developmental Medicine and Child Neurology* 53(2):150-155.

Enoka, R. M. 2002. *Neuromechanics of human movement*. 3rd ed. Champaign, IL: Human Kinetics Publishers.

Entwisle, D. R., and N. M. Astone. 1994. Some practical guidelines for measuring youth's race/ethnicity and socioeconomic status. *Child Development* 65:1521-1540.

Erwin, H. E., A. M. Woods, M. K. Woods, and D. M. Castelli. 2007. Children's environmental access in relation to motor competence, physical activity, and fitness. *Journal of Teaching in Physical Education* 26(4):404-415.

Espenschade, A. 1940. Motor performance in adolescence, including the study of relationships with measures of physical growth and maturity. *Monographs of the Society for Research in Child Development* 24.

Espenschade, A. 1960. Motor development. In *Science and medicine of exercise and sports*, edited by W. R. Johnson. New York: Harper. Pp. 419-439.

Farpour-Lambert, N. J., and C. J. R. Blimkie. 2008. Muscle strength. In *Paediatric exercise science and medicine*. 2nd ed., edited by N. Armstrong and W. Van Mechelen. Oxford, United Kingdom: Oxford University Press. Pp. 37-53.

Fisher, A., J. J. Reilly, L. A. Kelly, C. Montgomery, A. Williamson, J. Y. Paton, and S. Grant. 2005. Fundamental movement skills and habitual physical activity in young children. *Medicine and Science in Sports and Exercise* 37(4):684-688.

Floyd, M. F., W. C. Taylor, and M. Whitt-Glover. 2009. Measurement of park and recreation environments that support physical activity in low-income communities of color: Highlights of challenges and recommendations. *American Journal of Preventive Medicine* 36(Suppl. 4):S156-S160.

Froberg, K., and O. Lammert. 1996. Development of muscle strength during childhood. In *The Encyclopaedia of sports medicine VI. The child and adolescent athlete*. Oxford, United Kingdom: Blackwell Scientific Publications. Pp. 25-41.

Going, S. B., T. G. Lohman, and H. B. Falls. 2008. Body composition assessment. In *Fitnessgram/Activitygram reference guide,* edited by G. J. Welk and M. D. Meredith. Dallas, TX: Cooper Institute. Pp. 121-128.

Going, S. B., T. G. Lohman, E. C. Cussler, D. P. Williams, J. A. Morrison, and P. S. Horn. 2011. Percent body fat and chronic disease risk factors in U.S. children and youth. *American Journal of Preventive Medicine* 41(4, Suppl. 2):S77-S86.

Haga, M. 2009. Physical fitness in children with high motor competence is different from that in children with low motor competence. *Physical Therapy* 89(10):1089-1097.

Hamill, P. V. V., F. E. Johnston, and S. Lemeshow. 1973. Body weight, stature, and sitting height: White and negro youths 12-17 years. *Vital and Health Statistics* 11:126.

Hands, B. 2008. Changes in motor skill and fitness measures among children with high and low motor competence: A five-year longitudinal study. *Journal of Science and Medicine in Sport* 11(2):155-162.

Hands, B. P., and D. Larkin. 2006. Physical fitness differences in children with and without motor learning difficulties. *European Journal of Special Needs Education* 21(4):447-456.

Hands, B., D. Larkin, H. Parker, L. Straker, and M. Perry. 2009. The relationship among physical activity, motor competence and health-related fitness in 14-year-old adolescents. *Scandinavian Journal of Medicine and Science in Sports* 19(5):655-663.

Haubenstricker, J., and V. Seefeldt. 1986. Acquisition of motor skills during childhood. In *Physical activity and well-being*, edited by V. Seefeldt. Reston, VA: American Alliance for Health, Physical Education, Recreation and Dance. Pp. 41-102.

Heras Yague, P., and J. M. de la Fuente. 1998. Changes in height and motor performance relative to peak height velocity: A mixed-longitudinal study of Spanish boys and girls. *American Journal of Human Biology* 10:647-660.

Jackson, A. S. 1989. Application of regression analysis to exercise science. In *Measurement concepts in physical education and exercise science,* edited by M. J. Safrit and T. M. Wood. Champaign, IL: Human Kinetics Publishers. Pp. 181-205.

Jones, H. E. 1949. *Motor performance and growth. A developmental study of static dynamometric strength*. Berkeley, CA: University of California Press.

Kemper, H. C. G., and R. Verschuur. 1985. Motor performance fitness tests. In *Growth, health and fitness of teenagers*, edited by H. C. G. Kemper. Basel, Switzerland: Karger Publishers. Pp. 96-106.

Keogh, J. 1965. *Motor performance of elementary school children*. Los Angeles, CA: Department of Physical Education, University of California.

Lefevre, J., G. Beunen, G. Steens, A. Claessens, and R. Renson. 1990. Motor-performance during adolescence and age 30 as related to age at peak height velocity. *Annals of Human Biology* 17(5):423-435.

Little, N. G., J. A. P. Day, and L. Steinke. 1997. Relationship of physical performance to maturation in perimenarcheal girls. *American Journal of Human Biology* 9:163-171.

Liu, A., A. Hills, X. Hu, Y. Li, L. Du, Y. Xu, N. Byrne, and G. Ma. 2010. Waist circumference cut-off values for the prediction of cardiovascular risk factors clustering in Chinese school-aged children: A cross-sectional study. *BMC Public Health* 10(1):82.

Livingston, S. A., and M. J. Zieky. 1982. *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.

Lloyd, L. K., P. A. Bishop, J. L. Walker, K. R. Sharp, and M. T. Richardson. 2003. The influence of body size and composition on Fitnessgram® test performance and the adjustment of Fitnessgram® test scores for skinfold thickness in youth. *Measurement in Physical Education and Exercise Science* 7(4):205-226.

Lobelo, F., R. R. Pate, M. Dowda, A. D. Liese, and J. R. Ruiz. 2009. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Medicine and Science in Sports and Exercise* 41(6):1222-1229.

Logan, S. W., L. E. Robinson, A. E. Wilson, and W. A. Lucas. 2011. Getting the fundamentals of movement: A meta-analysis of the effectiveness of motor skill interventions in children. *Child: Care, Health and Development* 38(3):305-315.

Lopes, V. P., J. A. R. Maia, L. P. Rodrigues, and R. Malina. 2011. Motor coordination, physical activity and fitness as predictors of longitudinal change in adiposity during childhood. *European Journal of Sport Science* 12(4):384-391.

Lopes, V. P., D. F. Stodden, M. M. Bianchi, J. A. R. Maia, and L. P. Rodrigues. 2012. Correlation between BMI and motor coordination in children. *Journal of Science and Medicine in Sport* 15(1):38-43.

Malina, R. M. 1996. Tracking of physical activity and physical fitness across the lifespan. *Research Quarterly for Exercise and Sport* 67(Suppl. 3):S48-S57.

Malina, R. M. 2001. Physical activity and fitness: Pathways from childhood to adulthood. *American Journal of Human Biology* 13(2):162-172.

Malina, R. M. 2007. Physical fitness of children and adolescents in the United States: Status and secular change. *Medicine and Sport Science* 50:67-90.

Malina, R. M. 2011. Skeletal age and age verification in youth sport. *Sports Medicine* 41(11):925-947.

Malina, R. M., and A. F. Roche. 1983. *Manual of physical status and performance in childhood. Vol. 2 Physical performance*. New York: Plenum Press.

Malina, R. M., P. V. Hamill, and S. Lemeshow. 1974. Body dimensions and proportions, white and negro children 6-11 years, United States. *Vital and Health Statistics* 11(143): 1-66.

Malina, R. M., R. Martorell, and F. Mendoza. 1986. Growth status of Mexican American children and youths: Historical trends and contemporary issues. *Yearbook of Physical Anthropology* 29:45-79.

Malina, R. M., K. H. Brown, and A. N. Zavaleta. 1987. Relative lower extremity length in Mexican American and in American black and white youth. *American Journal of Physical Anthropology* 72(1):89-94.

Malina, R. M., C. Bouchard, and O. Bar-Or. 2004. *Growth, maturation, and physical activity*. 2nd ed. Champaign, IL: Human Kinetics Publishers.

Malina, R. M., S. P. Cumming, P. J. Morano, M. Barron, and S. J. Miller. 2005. Maturity status of youth football players: A noninvasive estimate. *Medicine and Science in Sports and Exercise* 37(6):1044-1052.

Malina, R. M., M. J. Coelho E Silva, A. J. Figueiredo, C. Carling, and G. P. Beunen. 2012. Interrelationships among invasive and non-invasive indicators of biological maturation in adolescent male soccer players. *Journal of Sports Sciences* 1-13.

Martins, D., J. Maia, A. Seabra, R. Garganta, V. Lopes, P. Katzmarzyk, and G. Beunen. 2010. Correlates of changes in BMI of children from the Azores islands. *International Journal of Obesity* 34(10):1487-1493.

Martorell, R., R. M. Malina, R. O. Castillo, F. S. Mendoza, and I. G. Pawson. 1988. Body proportions in three ethnic groups: Children and youths 2-17 years in NHANES II and HHANES. *Human Biology* 60(2):205-222.

Matvienko, O., and I. Ahrabi-Fard. 2010. The effects of a 4-week after-school program on motor skills and fitness of kindergarten and first-grade students. *American Journal of Health Promotion* 24(5):299-303.

McKenzie, T. L., J. F. Sallis, S. L. Broyles, M. M. Zive, P. R. Nader, C. C. Berry, and J. J. Brennan. 2002. Childhood movement skills: Predictors of physical activity in Anglo American and Mexican American adolescents? *Research Quarterly for Exercise and Sport* 73(3):238-244.

Merola, S. S. 2005. *The problem of measuring SES on educational assessments*. Paper presented at American Sociological Association, Philadelphia, PA, August 16.

Miech, R. A., S. K. Kumanyika, N. Stettler, B. G. Link, J. C. Phelan, and V. W. Chang. 2006. Trends in the association of poverty with overweight among US adolescents, 1971-2004. *Journal of the American Medical Association* 295(20):2385-2393.

Mirwald, R. L., and D. A. Bailey. 1986. *Maximal aerobic power*. London, Ontario: Sports Dynamics.

Mirwald, R. L., A. D. Baxter-Jones, D. A. Bailey, and G. P. Beunen. 2002. An assessment of maturity from anthropometric measurements. *Medicine and Science in Sports and Exercise* 34(4):689-694.

Mizuno, T., S. Ebashi, and K. Yamaji. 1973. Longitudinal study on the physique and physical performance of Japanese adolescents (12-18 years, boys and girls). *Bulletin of the Faculty of Education of the University of Tokyo* 13:219-235.

Monyeki, K. D., H. C. G. Kemper, and P. J. Makgae. 2008. Relationship between fat patterns, physical fitness and blood pressure of rural South African children: Ellisras Longitudinal Growth and Health Study. *Journal of Human Hypertension* 22(5):311-319.

Myer, G. D., A. D. Faigenbaum, K. R. Ford, T. M. Best, M. F. Bergeron, and T. E. Hewett. 2011. When to initiate integrative neuromuscular training to reduce sports-related injuries and enhance health in youth? *Current Sports Medicine Reports* 10(3):155-166.

National Center for Health Statistics. 2011. *Health, United States, 2010: With special feature on death and dying*. Hyattsville, MD: National Center for Health Statistics.

Nurmi-Lawton, J. A., A. D. Baxter-Jones, R. L. Mirwald, J. A. Bishop, P. Taylor, C. Cooper, and S. A. New. 2004. Evidence of sustained skeletal benefits from impact-loading exercise in young females: A 3-year longitudinal study. *Journal of Bone and Mineral Research* 19(2):314-322.

O'Beirne, C., D. Larkin, and T. Cable. 1994. Coordination problems and anaerobic performance in children. *Adapted Physical Activity Quarterly* 11(2):141-149.

Okely, A. D., J. W. Patterson, and M. Booth. 2001. Relationship of cardiorespiratory endurance to fundamental movement skill proficiency among adolescents. *Pediatric Exercise Science* 13(4):380-391.

OMB (Office of Management and Budget). 1997. *Revisions to the standards for the classification of federal data on race and ethnicity.* Directive No. 15. Washington, DC: Office of Management and Budget.

Ostyn, M., J. Simons, G. Beunen, R. Renson, and D. Van Gerven. 1980. *Somatic and motor development of Belgian secondary school boys*. Leuven, Belgium: Catholic University of Leuven Press.

Physical Activity Guidelines Advisory Committee. 2008. *Physical Activity Guidelines Advisory Committee report*. Washington, DC: U.S. Department of Health and Human Services.

Piek, J. P., D. Rigoli, J. G. Pearsall-Jones, N. C. Martin, D. A. Hay, K. S. Bennett, and F. Levy. 2007. Depressive symptomatology in child and adolescent twins with attention-deficit hyperactivity disorder and/or developmental coordination disorder. *Twin Research and Human Genetics* 10(04):587-596.

Piek, J. P., G. S. Bradbury, S. C. Elsley, and L. Tate. 2008. Motor coordination and social-emotional behaviour in preschool-aged children. *International Journal of Disability, Development and Education* 55(2):143-151.

Piek, J. P., N. C. Barrett, L. M. Smith, D. Rigoli, and N. Gasson. 2010. Do motor skills in infancy and early childhood predict anxious and depressive symptomatology at school age? *Human Movement Science* 29(5):777-786.

Ratamess, N. A. 2008. Adaptations to anaerobic training programs. In *Essentials of strength training and conditioning*. 4th ed., edited by T. R. Baechle and R. W. Earle. Champaign, IL: Human Kinetics Publishers. Pp. 94-99.

Raudsepp, L., and P. Pall. 2006. The relationship between fundamental motor skills and outside-school physical activity of elementary school children. *Pediatric Exercise Science* 18(4):426-435.

Roche, A. F., F. Tyleshevski, and E. Rogers. 1983. Non-invasive measurement of physical maturity in children. *Research Quarterly for Exercise and Sport* 54:364-371.

Safrit, M. J., and T. M. Wood. 1989. Criterion-referenced measurement: Validity. In *Measurement concepts in physical education and exercise science*. 1st ed. Champaign, IL: Human Kinetics Publishers.

Schott, N., V. Alof, D. Hultsch, and D. Meermann. 2007. Physical fitness in children with developmental coordination disorder. *Research Quarterly for Exercise and Sport* 78(5):438-450.

Seaman, J. A. 1999. Physical activity and fitness for persons with disabilities. *PCPFS Research Digests*, http://www.fitness.gov/publications/digests/digest_mar1999.html (accessed July 18, 2012).

Sherar, L. B., A. D. Baxter-Jones, R. A. Faulkner, and K. W. Russell. 2007. Do physical maturity and birth date predict talent in male youth ice hockey players? *Journal of Sports Sciences* 25(8):879-886.

Simons, J., G. P. Beunen, R. Renson, A. L. M. Claessens, B. Vanreusel, and J. Lefevre, eds. 1990. *Growth and fitness of Flemish girls: The Leuven Growth Study*. Champaign, IL: Human Kinetics Publishers.

Skinner, R. A., and J. P. Piek. 2001. Psychosocial implications of poor motor coordination in children and adolescents. *Human Movement Science* 20(1-2):73-94.

Stodden, D., J. D. Goodway, S. Langendorfer, M. A. Roberton, M. E. Rudisill, G. Garcia, and L. E. Garcia. 2008. A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest* 60:290-306.

Stodden, D., S. Langendorfer, and M. A. Roberton. 2009. The association between motor skill competence and physical fitness in young adults. *Research Quarterly for Exercise and Sport* 80(2):223-229.

Sun, S. S., C. M. Schubert, W. C. Chumlea, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and A. S. Ryan. 2002. National estimates of the timing of sexual maturation and racial differences among US children. *Pediatrics* 110(5):911-919.

Sun, S. S., C. M. Schubert, R. Liang, A. F. Roche, H. E. Kulin, P. A. Lee, J. H. Himes, and W. C. Chumlea. 2005. Is sexual maturity occurring earlier among U.S. children? *Journal of Adolescent Health* 37(5):345-355.

Tanner, J. M. 1962. *Growth at adolescence*. 2nd ed. Oxford, United Kingdom: Blackwell Scientific Publications.

Tveter, A. T., and I. Holm. 2010. Influence of thigh muscle strength and balance on hop length in one-legged hopping in children aged 7–12 years. *Gait and Posture* 32(2):259-262.

Vicente-Rodríguez, G., I. Ara, J. Perez-Gomez, J. A. Serrano-Sanchez, C. Dorado, and J. A. Calbet. 2004. High femoral bone mineral density accretion in prepubertal soccer players. *Medicine and Science in Sports and Exercise* 36(10):1789-1795.

Vicente-Rodríguez, G., A. Urzanqui, M. I. Mesana, F. B. Ortega, J. R. Ruiz, J. Ezquerra, J. A. Casajús, G. Blay, V. A. Blay, M. Gonzalez-Gross, and L. A. Moreno. 2008. Physical fitness effect on bone mass is mediated by the independent association between lean mass and bone mass through adolescence: A cross-sectional study. *Journal of Bone and Mineral Metabolism* 26(3):288-294.

Welk, G. J., K. R. Laurson, J. C. Eisenmann, and K. J. Cureton. 2011. Development of youth aerobic-capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine* 41(4, Suppl. 2):S111-S116.

Whitt-Glover, M. C., W. C. Taylor, M. F. Floyd, M. M. Yore, A. K. Yancey, and C. E. Matthews. 2009. Disparities in physical activity and sedentary behaviors among U.S. children and adolescents: Prevalence, correlates, and intervention implications. *Journal of Public Health Policy* 30(Suppl. 1):S309-S334.

Williams, D. P., S. B. Going, T. G. Lohman, D. W. Harsha, S. R. Srinivasan, L. S. Webber, and G. S. Berenson. 1992. Body fatness and risk for elevated blood pressure, total cholesterol, and serum lipoprotein ratios in children and adolescents. *American Journal of Public Health* 82(3):358-363.

Zhu, W., S. A. Plowman, and Y. Park. 2010. A primer-test centered equating method for setting cut-off scores. *Research Quarterly for Exercise and Sport* 81(4):400-409.

Zhu, W., M. T. Mahar, G. J. Welk, S. B. Going, and K. J. Cureton. 2011. Approaches for development of criterion-referenced standards in health-related youth fitness tests. *American Journal of Preventive Medicine* 41(4):S68-S76.

4

Health-Related Fitness Measures for Youth: Body Composition

KEY MESSAGES

Body composition is a physiologic characteristic that affects an individual's ability to carry out daily tasks with vigor. Although body composition is not a demonstrative action like other health-related fitness components, the committee has operationally defined it as a component of fitness, a health marker, and a modifier of fitness for the purposes of this report. Both body weight (mass) and body fat (absolute fatness and relative fat distribution) are elements of body composition that have implications for health and fitness. It is important to measure weight and height in national youth fitness surveys to derive body mass index (BMI), an indicator of weight-for-height; waist circumference, an indicator of abdominal adiposity; and skinfolds, an indicator of subcutaneous adipose tissue. These three recommended field indicators of body composition for a national youth fitness survey uniquely measure different elements, and each can be linked to health markers and outcomes in both youth and adults. For example,

- A high BMI is related to the risk of type 2 diabetes and hypertension.
- Waist circumference is linked to risk factors for cardiovascular disease, type 2 diabetes, and all-cause mortality.
- Elevated skinfold thicknesses and proportionally more subcutaneous fat on the trunk are associated with an elevated risk for cardiovascular disease and metabolic syndrome.

Only standing height and weight should be measured in school and other educational settings to calculate BMI given such concerns as measurement errors and privacy.

Two approaches to interpreting the results of the above three measures are recommended to determine whether individuals or populations are at risk of poor health outcomes. For BMI, the cut-points (cutoff scores) based on the 2000 Centers for Disease Control and Prevention (CDC) growth charts and percentiles should be applied for underweight, overweight, and obesity evaluations. Interim cut-points for waist circumference and skinfold measures should be set at levels analogous to those currently being applied by the CDC for BMI. This approach should be used until evidence becomes available to support establishing waist circumference and skinfold cut-points by associating those measures with cardiometabolic risk factors.

Body weight (mass) and body fat distribution are elements of body composition that have implications for health and fitness. No element on its own adequately and comprehensively describes an individual's body composition, and each element has been linked with various health markers and outcomes in youth.

Measures of body composition have been used in the past as a component of fitness test batteries (see Table 2-6 in Chapter 2). The background paper for the Second International Consensus Symposium on Physical Activity, Fitness and Health in 1992 offered an outline of "components and factors of health-related fitness" (Bouchard and Shephard, 1994; Bouchard et al., 2007) in which body composition was included as a morphological component of health-related fitness. In a review of existing fitness tests, 10 of 15 physical fitness test batteries for children and adolescents included body composition as a component of health-related fitness (Artero et al., 2011; Castro-Piñero et al., 2010), but the supporting evidence for their inclusion was quite variable.

Body composition differs from the other fitness components reviewed in this report at various levels. First, there are different perspectives on whether body composition should be considered a component of fitness. The committee considered body composition to be a physiologic characteristic that affects an individual's ability to carry out daily tasks with vigor and to be influenced by physical activity behavior. Second, body composition influences performance on many fitness tests and itself is also an indicator of health. The committee thus defined body composition operationally as a component of fitness, a marker of health, and a modifier of fitness, for the purposes of this report. Finally, the relationships between body composition, in particular percent body fat, and health outcomes are well established in both youth and adults. Thus, the committee did not collect evidence on the relationship between any body composition field measures and health outcomes. The committee identified appropriate measures of body composition by selecting field-based items that were valid, reliable, and feasible for implementation in either a national survey or a school or other educational setting.

This chapter provides an overview of the existing measures of body composition and presents the committee's conclusions about the best measures of body composition based on their relationship to health in youth, as well as their integrity and feasibility. The committee's full recommendations for measuring body composition in a national fitness survey and in schools and other educational settings can be found in Chapters 8 and 9, respectively.

MEASUREMENT OF BODY COMPOSITION

Body weight is a gross measure of the mass of the body. The partitioning and quantification of mass into its basic elements has been a major focus of study historically and has accelerated with the refinement of models (Wang et al., 1992, 2005) and the development of technology (Ackland et al., 2012; Heymsfield et al., 2005; Roche et al., 1996). A variety of models and methods—developed largely in adults—have been used to partition body mass into several elements: fat-free mass, fat mass, total body water, fat-free dry mass, and bone mineral. As assessment techniques have become refined, models have evolved from those including the traditional two elements (body mass = fat-free mass + fat mass) to those including three, four, or five elements (Wang et al., 1992, 2005), with fat mass—or adipose tissue as it is labeled in anatomical models—being basic to all models.

Body composition can be approached at several levels: atomic, molecular, cellular, tissue, and whole body (Wang et al., 1992, 1995). The technology for measuring specific elements of body mass at each level and factors influencing body composition have been summarized (Heymsfield et al., 2005; Roche et al., 1996). While no single criterion measure of body composition is universally accepted as the gold standard (Ackland et al., 2012), laboratory measures (e.g., underwater weight, body potassium, total body water, and dual-energy X-ray absorptiometry [DXA]) are considered the most accurate techniques. DXA provides measures of bone mineral and of fat and lean tissues. The other three laboratory measures have major limitations. Underwater weighing is used to estimate body density, which is then converted to percentage body fat; total body potassium and total body water provide measures of fat-free mass. Quite often, the laboratory measures are used together (especially underwater weighing, total body water, and a measure of bone mineral) to provide an estimate of body composition, depending on the model selected (see, e.g., Gutin et al., 1996). Several additional laboratory techniques, as well as field measurements available for estimating body composition in youth in various settings (e.g., research, education/practice, clinical), have been reviewed extensively by others (Heymsfield et al., 2005; Roche et al., 1996).

Critical evaluation of body composition methodology at each level of analysis (Wang et al., 1992, 1995) is beyond the scope of this report. Further, while laboratory methods—such as DXA, hydrostatic weighing, ultrasound, densitometry, and air displacement plethysmography (e.g., Bod Pod®)—have many advantages (such as more specific measurements and reduced measurement variation, measurement of the whole body, minimal subject burden, and relative ease of administration), they typically require highly specific training and special and expensive equipment. Some of the equipment also lacks the mobility that may be necessary to access large samples of youth.

Overall, the measurement of body composition is dependent on the question being addressed, the information necessary, and the application of the assessment protocols (Ackland et al., 2012). For example, techniques used for collecting data to track the health status of a given population epidemiologically will likely differ from those used to collect data to achieve advances in sports performance. Also, many laboratory methods are not feasible for use in the field because of limitations cited earlier. Direct evaluation of body composition (i.e., DXA measure of adiposity) is not the same as indirect and associated proximity measures of body composition. Body mass index (BMI) is used for classification of weight status (underweight, normal, overweight, obese), although it does not accurately predict percent body fat (Moreno et al., 2006). Accordingly, the selection of body composition measures depends on what element is of interest, which in turn depends on the question(s) being asked. Given the desirability of a comprehensive understanding of an individual's body composition, as well as considerations of feasibility in practice settings such as schools, the committee focused its review on low-cost field-based measures of body composition.

Laboratory measures, such as DXA, may be appropriate for some national surveys, like the National Health and Nutrition Examination Survey (NHANES), for which the equipment is transported in a trailer and only a small sample of youth is studied. For large national surveys in which youth are tested nationwide in a school setting, however, field-based measures are more suitable than laboratory measures. Field-based measures include anthropometry (skinfolds, weight, height [weight-for-height in the form of BMI], waist circumference) and bioelectric impedance analysis

(BIA). These measures were selected based on their relationship to health markers, their integrity and reliability, and their previous use.

- **Skinfolds** provide an indication of subcutaneous fat at specifically defined measurement sites. Skinfolds can also be used to predict percent body fat. They can be expressed as a sum of skinfolds (overall subcutaneous fat) and as a ratio of trunk to extremity skinfolds (relative subcutaneous fat distribution).
- **BMI (kg/m^2)** is an indicator of weight-for-height. It is used internationally in public health and nutrition surveys to monitor weight status, specifically overweight and obesity. At the extremes of heaviness, BMI is probably a reasonable indicator of fatness in the general population.
- **Waist circumference** increasingly is used as an indicator of central or abdominal adiposity rather than percent body fat, which can vary greatly among individuals with a similar BMI. Located in the abdominal region, abdominal fat is composed of three elements: visceral, retroperitoneal, and subcutaneous.
- **BIA** provides a measure of resistance or impedance; some BIA systems are calibrated directly to fat-free mass (or fat). Some other systems provide a measure of total body water that is then transformed to an estimate of fat-free mass. The equation for converting resistance to total body weight usually includes height. Algorithms used in estimating body composition with BIA with commercially available units are considered proprietary information, which is a major shortcoming. Moreover, it has been suggested that equations provided by manufacturers may not be suitable for estimating body composition in youth of different racial/ethnic backgrounds (Haroun et al., 2010). Given its shortcomings and the abundant evidence on the effectiveness of other field-based measures, the committee did not explore BIA further. However, if shortcomings due to proprietary equations were resolved and the algorithms accounted for changes in the composition of fat-free mass during childhood and adolescence, for sex differences, and for racial/ethnic variations in body composition, BIA might be considered a useful measure of body composition in youth, particularly given its ease of administration.

LITERATURE REVIEW PROCESS

As noted earlier, the fact that body composition is a measure of health is well established. Also well established is that percent body fat is related to health outcomes and that there are various tests with which to measure percent body fat, subcutaneous fat, or abdominal adiposity. Therefore, the Centers for Disease Control and Prevention's (CDC's) literature review, described in Chapter 3, did not include body composition as a fitness component; the CDC review, however, included articles in which body composition appeared as a modifier of fitness or as a health marker or both. For this reason, the committee considered each such article for inclusion in its review, and this chapter includes findings from selected studies from the CDC review, as well as others, in which body composition was considered as a fitness component, a health marker, or a modifier of physical fitness. A body of literature from obesogenic intervention research was also reviewed. In addition, the committee reviewed validity and reliability data specific to field tests that measure various aspects of body composition. Further, articles on specific topics related to body composition (e.g., growth and maturation, race/ethnicity, and body composition and health) were identified and chosen for inclusion in this chapter.

BODY COMPOSITION IN YOUTH

In addition to age, energy intake, and other factors, individual differences in biological maturation, gender, and race/ethnicity affect elements of body composition such as fat mass and fat-free mass, subcutaneous fat, and fat distribution. The following discussion is based on trends described in Malina (1996, 2005) and Malina et al. (2004).

Variation in Body Composition with Age, Gender, and Maturation Status

Fat-free mass, fatness, and relative fat distribution in late childhood and adolescence (approximate school ages) vary with age, between genders, and among individuals of contrasting maturation status. Age- and gender-related changes are discussed in the subsections that follow. Variation associated with maturation status is then briefly considered.

Fat-Free Mass

The principles underlying several methods for estimating body composition warrant special consideration when applied to growing and maturing youth. It is important to determine when mature (adult) levels of the primary elements of fat-free mass are reached. This relates to the concept of chemical maturity, defined by Moulton (1923, p. 80): "The point at which the concentration of water, proteins, and salts [minerals] becomes comparatively constant in the fat-free cell is named the point of chemical maturity of the cell." Chemical maturity of fat-free mass is not attained until after the adolescent growth spurt, probably about ages 16-18 in girls and 18-20 in boys (Lohman, 1981, 1986; Malina, 2005; Wells et al., 2010). When adult values of its primary components are reached, fat-free mass is chemically mature.

On average, fat-free mass has a growth pattern like that of stature and body mass. Differences over time until chemical maturity is reached reflect a larger fat-free mass, specifically bone mineral content and skeletal muscle mass in males. That is, the relative contribution of water to body mass decreases while the relative contributions of solids—protein, mineral, and fat—increase during approximately the first two decades of life, which are dominated by the biological activities of growth and maturation. Sex differences are apparent during the adolescent growth spurt. On average, fat-free mass in males contains relatively less water and more protein and mineral compared with that in females from childhood into young adulthood (Malina et al., 2004). Density of fat-free mass is also greater in males, which reflects primarily sex differences in skeletal muscle mass and bone mineral. These are only average trends, and it should be noted that there are variations among individuals and with biological maturation (status and timing).

Although efforts continue to derive more accurate estimates of the chemical composition of fat-free mass, three points should be noted: (1) the composition of fat-free mass changes during growth and maturation, (2) variation among individuals is considerable, and (3) chemical maturity is not attained until late adolescence or young adulthood. Ideally, equations and constants used to estimate body composition should be adjusted for the chemical immaturity of the fat-free mass in growing and maturing individuals.

Fat Mass

Fat mass increases more rapidly in girls than in boys during childhood and continues to increase through adolescence in girls (Malina, 2000). Fat mass appears to reach a plateau, or to change only slightly, near the time of the adolescent spurt in boys (around age 13-15) (Malina, 2000). Fat mass as a percentage of body mass (percent body fat) increases gradually during childhood, and sex differences are small. Percent body fat increases through adolescence in girls; it increases into early adolescence in boys and then declines. The decline during male adolescence is a function of the adolescent spurt in fat-free mass, more specifically muscle mass. Sex differences in body composition are negligible in childhood, and are established during the adolescent spurt and sexual maturation (Malina, 2000). Although estimated fat mass is similar in male and female adolescents, females have greater percent body fat.

Weight-for-Height

BMI, as a measure of weight-for-height, declines from infancy through early childhood and reaches its lowest point at about age 5-6. It then increases linearly with age through childhood and adolescence and into adulthood. Sex differences in BMI are small during childhood, rise during adolescence, and persist into adulthood (Malina et al., 2004). The rise in BMI after it reaches a nadir at age 5-6 has been labeled the "adiposity rebound" (Rolland-Cachera et al., 1984). It has been suggested that individuals who have an early adiposity rebound have an increased probability of being overweight in late adolescence and young adulthood (Rolland-Cachera et al., 1984). The concept of adiposity rebound needs further study. In the context of body composition, there is a need to document specific changes in body composition during the rebound. Does it reflect an increase in fat mass or an increase in fat-free mass? Some evidence suggests the latter (Katzmarzyk et al., 2012).

Subcutaneous Fat

A skinfold thickness is a double fold of skin and underlying soft tissues, primarily adipose tissue. Two commonly used skinfolds are the triceps and subscapular. The former increases with age from childhood through adolescence in females, whereas it decreases during the adolescent spurt in males (Malina, 2000). On the other hand, the latter increases from childhood through adolescence in both sexes. As a result, the adolescent sex difference in the triceps skinfold is marked compared with the relatively small difference in the subscapular skinfold.

Any number of skinfolds can be and have been measured. Changes in individual skinfolds are variable during growth and specifically relative to the timing of peak height velocity, more so in boys than in girls (Malina et al., 1999). Such variation may influence age-associated trends.

Relative Fat Distribution

Fat distribution refers to regional variation in the accumulation of adipose tissue in the body. With advances in technology (computed tomography [CT] scans, magnetic resonance imaging [MRI]), attention shifted to abdominal fatness, specifically visceral versus subcutaneous. With widespread availability of DXA, trunk and extremity distribution of adipose tissue also has received more attention (Malina, 1996, 2005). Ratios of skinfolds measured on the trunk to those measured on the extremities are commonly used to estimate relative subcutaneous fat distribution. Ratios of trunk to extremity skinfold thicknesses increase gradually through childhood in both sexes, and there is no sex difference in the ratios. Subsequently, ratios tend to be rather stable in females but to increase in males through adolescence. Males accumulate proportionally more subcutaneous fat on the trunk than the extremities, while females accumulate relatively similar amounts on both the trunk and extremities (Malina, 1996; Malina and Bouchard, 1988). Ratios of trunk to extremity adiposity based on DXA show similar trends (He et al., 2004; Taylor et al., 2010). Ratios of abdominal visceral and subcutaneous adiposity show small differences with age and sex from childhood into early adolescence in normal-weight youth, but males have proportionally more visceral adiposity in later adolescence. A similar trend associated with age and sex is not clearly apparent in overweight/obese youth (Katzmarzyk et al., 2012).

Maturity-Associated Variation

Children and adolescents advanced in maturity status compared with their chronological-age peers tend to be fatter and to have proportionally more subcutaneous fat on the trunk (Malina and Bouchard, 1988; Malina et al., 2004). The maturity-associated trend also continues into young adulthood (Beunen et al., 1994a; Kindblom et al., 2006; Labayen et al., 2009; Sandhu et al., 2006). Samples in studies using DXA, CT scans, and MRI generally involve several age groups, so that it is difficult to clearly specify maturity differences in each gender. Stage of puberty (clinically and/or self-assessed) is often described, but not systematically analyzed. When it is described, youth typically are grouped by pubertal stage or stages so that several ages are represented within a group (Katzmarzyk et al., 2012).

Variation in Body Composition with Race/Ethnicity

The pattern of gender- and maturity-related differences is similar in all racial/ethnic groups. African Americans have greater total bone mineral content during childhood, adolescence, and adulthood than whites. Comparisons between Mexican Americans and whites show small differences in fat-free mass and bone mineral content, although Mexican Americans tend to have greater adiposity. Data for skinfolds indicate proportionally more subcutaneous adipose tissue on the trunk in African Americans, Mexican Americans, and Asian Americans compared with whites. In contrast, data on the distribution of visceral and subcutaneous adiposity overlap among ethnic groups. Unfortunately, the available data often combine multiple age groups and in some instances combine males and females and/or youth of different ethnic groups (Katzmarzyk et al., 2012; Malina, 2005).

BODY COMPOSITION, FITNESS, AND HEALTH IN YOUTH

This section considers body composition as both a modifier of physical fitness and a health marker.

Body Composition as a Modifier of Physical Fitness

Body composition is one of many factors that influence performance on laboratory- and field-based tests of physical fitness. Fat-free mass and its major tissue component, skeletal muscle mass (force-generating tissue of the body), obviously are important to performance on many tests. Absolute fat-free mass is significant in tests requiring the projection of objects (e.g., overhead medicine ball throw) and a variety of strength tests. Fat-free mass also is highly correlated with height in children and adolescents. Fat mass and percent body fat are more variable, but excess fatness (absolute and relative) tends to exert a negative influence on performances on fitness tests that require movement or projection of the body through space (i.e., runs and jumps) as opposed to projection of objects, as well as on endurance tests on cycle ergometers (Boileau and Lohman, 1977; Houtkooper and Going, 1994; Malina, 1975, 1992).

Two studies of national samples of Belgian youth considered relationships between the sum of skinfolds (four in boys, five in girls) and a variety of fitness tests (Beunen et al., 1983; Malina et al., 1995). Among males aged 12-20, partial correlations (controlling for height and body mass) and several relevant fitness test items (static arm pull strength, sit-and-reach, vertical jump, left lifts, flexed arm hang, agility) were negative and low to moderate, –0.13 to –0.40. Corresponding partial correlations in girls aged 7-17 ranged from –0.08 to –0.42 (Physical Working Capacity-170 [PWC-170], step test recovery, arm pull strength, sit-and-reach, sit-ups, leg lifts, flexed arm hang, standing long jump, vertical jump agility). Comparison of the fattest and leanest 5 percent of participants highlighted the negative influence of excessive subcutaneous fatness for all fitness test items except sit-and-reach, arm pull strength, and PWC-170. Differences in flexibility were negligible. The fattest were absolutely stronger (boys and girls) and generated more watts (girls only), reflecting their larger body size. The fattest youth of both sexes also were advanced in skeletal maturation compared with their leanest peers of the same age groups (Beunen et al., 1982, 1994b).

Trends were generally similar in more recent samples of normal-weight and obese youth. For example, sum of skinfolds was inversely related to schoolchildren's performance on the progressive aerobic cardiovascular endurance run (PACER), push-up, and curl-up tests ($r = –0.30$ to $–0.49$) (Lloyd et al., 2003). Low levels of cardiorespiratory endurance were associated with percent body fat in African American and white adolescents

(Gutin et al., 2005); with visceral and subcutaneous abdominal fat and waist circumference in African American and white youth aged 8-17, controlling for age, sex, pubertal status, and BMI (Lee and Arslanian, 2007); with percent body fat, percent abdominal fat, and waist circumference in 8-year-old boys and girls (Stigman et al., 2009); and with BMI, skinfolds, and predicted percent body fat in obese children of both sexes aged 6-13 (Nassis et al., 2005).

Studies evaluating the relationship between BMI and fitness tests generally have focused on the negative influence of obesity (Chatrath et al., 2002; Deforche et al., 2003; Kim et al., 2005; Mota et al., 2006; Stratton et al., 2007) or BMI as a covariate of aerobic fitness (Beets and Pitetti, 2004; Beets et al., 2005). Youth aged 5-14 classified as "underfit" (based on pass-fail scores on five fitness tests) were at greater risk of obesity (Kim et al., 2005). Correlations between BMI and indicators of fitness tended to be linear in well- and undernourished children aged 6-14 (Malina et al., 1998), but were curvilinear in youth aged 12-15 (Bovet et al., 2007) and young adults (Sekulic et al., 2005; Welon et al., 1988). A recent study with a representative sample of Brazilian youth showed that, after adjusting for potential confounding factors, weight and BMI were negatively correlated with performance on the long jump, curl-up, pull-up, 9-minute run, 20-meter run, and 4-meter shuttle run (Dumith et al., 2012).

Data on variations in fitness across the broad spectrum of BMI within relatively narrow age groups are limited. Relationships between BMI and fitness varied among fitness test items in four age groups—9-10, 11-12, 13-15, and 16-18—in a national sample of Taiwanese youth (Huang and Malina, 2010). Correlations were low to moderate and did not vary among age groups or between sexes. They were highest for a distance run/walk (800/1,600 meters, 0.17 to 0.34) and lowest for the sit-and-reach (0.04 to 0.12). For sit-ups and the standing long jump, however, correlations were higher for boys than for girls ($r = 0.15$ to 0.23 versus 0.10 to 0.14 and 0.22 to 0.27 versus 0.12 to 0.17, respectively). In a national sample of Taiwanese youth (Huang and Malina, 2010), sex-specific regressions of fitness items on BMI, using a nonlinear quadratic model, indicated differential effects for individual tests, which varied with age and sex. Relationships for the sit-and-reach were similar and slightly curvilinear in girls aged 9-10 and boys aged 9-12, but were parabolic among girls aged 11-18 and boys aged 13-18. Peaks of the parabola were sharper in adolescent boys than in adolescent girls. Youth of both sexes aged 13-18 with either higher or especially lower BMIs had the poorest flexibility. Relationships for sit-ups were similar in girls and boys aged 9-12. Above ~20 kg/m^2 in girls and ~18 kg/m^2 in boys, sit-ups declined with increasing BMI. The decline was initially slight but accelerated with increasing BMI, more so in girls. The relationship between BMI and performance on sit-ups was parabolic in youth aged 13-18, more so in boys. Performance on the standing long jump (muscle power/explosive strength) declined linearly with increasing BMI in boys aged 9-10. The relationship was curvilinear in other age groups of boys and all age groups of girls, and was especially parabolic in boys aged 13-18. Times to complete the 800-meter run/walk increased (indicative of poorer performance) linearly with an increase in BMI in girls aged 9-10 and boys aged 9-12. Times varied little in older girls with BMI of 10-20 kg/m^2 but became progressively higher (indicative of poorer performance) with increasing BMI. Among older boys, times on the 1,600-meter run/walk declined (improved performance) with an increase in BMI from 10 to ~20 kg/m^2, and subsequently increased linearly (poorer performance) with increasing BMI.

Body Composition and Health

Excessive adiposity and especially abdominal obesity have been associated with risk factors for cardiovascular disease and metabolic syndrome in youth (e.g., Reed et al., 2008; Tjonna et al., 2009; Williams et al., 1992). The Bogalusa Heart Study, a longitudinal study investigating the risk factors for cardiovascular disease since 1972, provides many insights into the relationship between different measures of body composition and cardiovascular health in a biracial population (African American and white). The study found that by age 10, 60 percent of children who were overweight had at least one biomarker or risk factor for cardiovascular disease (Freedman et al., 1999a). Likewise, a dose-response relationship between increasing body weight and the presence of two or more cardiovascular disease risk factors—specifically prehypertension/hypertension, low high-density lipoprotein (HDL) cholesterol, and prediabetes/diabetes—was found in the NHANES 1999-2008 data set (May et al., 2012). More than one cardiovascular disease risk factor (besides body weight) was seen in 49 percent of overweight and 60 percent of obese adolescents (May et al., 2012). Other studies have confirmed these findings, suggesting that elevated weight status is associated with increased risk for cardiovascular disease (Chang et al., 2008; Jekal et al., 2009; Ribeiro et al., 2004), particularly with low cardiorespiratory fitness in children (Carnethon et al., 2005; Lobelo et al., 2010). Markers of metabolic syndrome (a cluster of precursory risk factors) also are associated with elevated childhood weight status (Rizzo et al., 2007). For example, obese youth have an increased risk for prediabetes compared with nonobese youth (Li et al., 2009). Among overweight children, those with prediabetes had 4 percent lower bone mineral content than those without prediabetes (Pollock et al., 2010).

Higher levels of body fatness predicted from skinfolds have been linked to increased cardiovascular disease risk in youth (Going et al., 2011). Estimated fatness levels above 20 percent in boys and 30 percent in girls were associated with risk factors for cardiovascular disease and metabolic syndrome, especially elevated C-reactive protein and insulin levels (Going et al., 2011).

The location of adipose tissue stores also influences cardiometabolic risk. Compared with more generalized obesity (measured as BMI), abdominal obesity was more strongly associated with risk of myocardial infarction, stroke, and premature death in adult men (Larsson et al., 1984). Waist circumference has been associated with cardiovascular risk factors such as insulin levels (Bassali et al., 2010) and may be a better predictor of cardiovascular disease than BMI alone (Savva et al., 2000; WHO, 2011). A review of the adult literature, however, found that neither waist circumference nor BMI had superior discriminatory capability in identifying cardiovascular disease risk (Huxley et al., 2010), thus demonstrating the value of measuring both. Abdominal adipose tissue also was associated with higher blood pressure, type 2 diabetes, and dyslipidemia in adults, but the relationship remains unclear in children (Daniels et al., 1999; Freedman et al., 1999b, 2012; Goran and Gower, 1998). Nevertheless, relationships between markers of disease risk and body composition were noted in school-aged youth in the Bogalusa Heart Study, specifically among African American girls aged 5-17, in whom a 20-cm increase in waist circumference was associated with a decrease in concentration of HDL cholesterol and increases in triacylglycerol and insulin (Freedman et al., 1999b). The association with the risk factors was stronger for waist circumference than for skinfold thickness (subscapular and triceps measurements). In overweight children, neither BMI nor the sum of skinfolds was a good predictor of risk factors for cardiovascular disease; a measure of fat distribution (waist-to-height ratio) was a better predictor of cardiovascular risk factors (low-density lipoprotein [LDL] cholesterol, HDL cholesterol, fasting insulin, and systolic and diastolic blood pressure) (Freedman et al., 2009).

Evidence for the contribution of measures of body composition to cardiovascular health also was noted in the European Youth Heart Study. Among children aged 9 and 15 from Denmark, Estonia, and Portugal, waist circumference and the sum of skinfolds were associated with clustered cardiovascular disease risk, determined by a composite score of systolic blood pressure, triglycerides, insulin resistance (using the homeostasic model assessment-insulin resistance [HOMA-IR] level), and ratio of total cholesterol to HDL (Andersen et al., 2008).

The relationship between weight status and risk for cardiovascular disease tends to track from childhood and adolescence into adulthood (Freedman et al., 2001; Herman et al., 2009). For example, changes in BMI and childhood blood pressure were found to be strongly correlated with adult blood pressure in both sexes (Lauer and Clarke, 1989). Based on logistic models developed with BMI data from youth aged 3-20 and adults aged 30-39 in the Fels Longitudinal Study, childhood and adult obesity are related, and the risk becomes stronger in adolescence (Guo et al., 2002). Another analysis of data from the Bogalusa study showed that the sum of skinfold thicknesses (triceps and subscapular measurements) and BMI z-score in childhood were the main contributors to higher levels of high-sensitivity C-reactive protein in adults, a risk factor for cardiovascular disease. The effect of skinfold thicknesses was greater in African Americans than in whites and in girls than in boys (Toprak et al., 2011). A recent systematic review of the tracking of obesity and its association with metabolic risk in adulthood, however, suggests that weight status and cardiometabolic risk factors should perhaps be considered independently of each other, given the limitations of current research designs (Lloyd et al., 2012). Finally, analysis of data from the Bogalusa study shows that BMI and subscapular skinfold measurements are positively correlated with the risk of becoming a diabetic adult (Nguyen et al., 2008).

In summary, indicators of body composition, specifically adiposity, are determinants of health. Prevention of the accumulation of excess adiposity and specifically overweight/obesity among youth holds potential for yielding immediate and long-term health benefits. Health implications of fat-free mass in youth apparently have not been systematically addressed.

VALIDITY AND RELIABILITY OF SELECTED BODY COMPOSITION MEASURES

By definition, field-based measurements such as skinfolds, BMI, and waist circumference are, to a large extent, indirect estimates of body composition. The advantages of field measures include minimal subject burden, adequate reliability in the hands of trained technicians (see more on quality control in Annex 4-1), and relatively rapid data acquisition for a large number of subjects. Limitations include reduced accuracy, high variability, and lack of broad applicability in all populations. The validity of field-based measurements is usually assessed by comparison with laboratory measures (e.g., underwater weight, body potassium, total body water, and DXA). Although evidence for the validity of field-based measures of body composition is variable (see below), their associations with markers of health risk justifies including them in a survey of youth fitness.

It is important to note that the accuracy of anthropometric measures is strongly dependent upon the experience and training of technicians in implementing the measurements. Care in interpreting the data also is necessary in working with youth. Anthropometric measures (BMI, percent body fat, skinfolds, circumferences) and other indicators of body composition (fat-free mass based on densitometry, total body water and potassium concentration, lean and fat tissue, bone mineral via DXA) change with age and are influenced by gender, race/ethnicity, and biological maturation status (Daniels et al., 1997; Malina, 1996, 2005; Malina et al., 2004). Elements of body composition, especially bone mineral content, also may be influenced by regular physical activity (Strong et al., 2005).

Body Mass Index (BMI)

BMI is an indicator of weight-for-height. As discussed earlier, in contrast to height and weight, which increase with age during childhood, BMI declines from infancy through early childhood and reaches its lowest point at about age 5-6.

BMI is reasonably well correlated with fat mass and percent body fat in heterogeneous samples of youth, but has limitations (Goran et al., 1995); it also is related to fat-free mass. Among youth aged 8-18 in the Fels Longitudinal Study, age-specific correlations between BMI and components of body composition ranged from 0.37 to 0.78 for percent body fat, 0.67 to 0.90 for fat mass, and 0.39 to 0.72 for fat-free mass in girls, and from 0.64 to 0.85 for percent body fat, 0.83 to 0.94 for fat mass, and 0.25 to 0.78 for fat-free mass in boys (Maynard et al., 2001). When chronological age was statistically controlled in five samples of boys and girls aged 8-18, correlations for BMI were a bit lower: percent body fat, 0.28 to 0.61; fat mass, 0.46 to 0.81; and fat-free mass, 0.27 to 0.64 (Malina and Katzmarzyk, 1999). Correlations for fat mass and fat-free mass were similar in four of the five samples, but those for BMI and percent body fat were variable. In a nationally representative sample of American children aged 2-19 in NHANES III, BMI was better than other anthropometric indicators (Rohrer index and weight-for-height) in predicting underweight and overweight when percent body fat or total fat mass based on DXA was the criterion measure (Mei et al., 2002).

Nevertheless, youth with the same BMI can differ considerably in fat mass and percent body fat, so care is essential when interpreting BMI as an indicator of fatness in youth. BMI is, more appropriately, an indicator of heaviness and, indirectly, of adiposity; at the extremes of heaviness, BMI is probably a reasonable indicator of fatness in general population surveys, but its limitations must be recognized (Pietrobelli et al., 1998).

Limited evidence supports higher intra- and interobserver reliability for BMI and waist circumference than for skinfold thicknesses (Artero et al., 2011). Beyond the debate about what the measurement of BMI actually represents (body composition, body fat, body weight, etc.), the association between BMI and health markers justifies its use among school-aged children as a means of tracking health status.

Waist Circumference

Waist circumference is an emergent measure of body composition. Its use as a dimension of body composition is justified for various reasons. First, it is an indicator of abdominal fat as opposed to waist-to-hip circumference ratio, which is an indicator of fat distribution (Despres et al., 1989). Second, criterion measures that relate to health have already been established in certain populations of youth (Liu et al., 2010). Further, other measures are more challenging to administer, such as measuring hip circumference to determine waist-to-hip ratio (WHO, 2011). Additionally, other waist measures have insufficient data to support their consideration and have not been found to be a better predictor of health risk than waist circumference (Huxley et al., 2010). Waist circumference is strongly associated with intra-abdominal (visceral) adipose tissue (r = 0.84) and subcutaneous abdominal adipose tissue (r = 0.93) in prepupertal children (Goran and Gower, 1998) and with trunk fat (r = 0.92) in children and youth aged 3-19 (Taylor et al., 2000). On the other hand, it has been suggested that waist circumference has no advantage over BMI for diagnosing high fat mass in youth aged 9-10 (Reilly et al., 2010).

According to a measurement protocol for adolescents, intra- and interobserver technical errors of measurement for waist circumference have been calculated at 1.31 cm and 1.56 cm, respectively (Malina et al., 1973). As mentioned above, a review of reliability found higher intra- and interobserver reliability for BMI and waist circumference than for skinfold thicknesses (Artero et al., 2011). Mueller and Malina (1987) report high intra- and interobserver reliabilities for waist circumferences of 0.97 and 0.96, respectively, based on data from the Health Examination Survey for youth aged 12-17. Technical errors have also been reported in national surveys (ODPHP, 1985, 1987).

Unlike other field measures, waist circumference may not be a good indicator of percent body fat or fatness in youth. However, it is an indicator of abdominal adiposity (Lee et al., 2011), which provides information about a different dimension of body composition that is linked to health risks.

Sum of Skinfolds

Skinfolds are considered valid and reliable estimates of subcutaneous fat and predictors of percent body fat, assuming they are measured by trained individuals. Criterion-related validity for the sum of skinfold (triceps and calf) measurements ranged from r = 0.70 to 0.90 compared with hydrostatic weighing (Boileau et al., 1984). Specifically, reliability coefficients for sum of skinfolds vary by pubertal status in girls (Gutin et al., 1996). Others have reported acceptable interobserver reliability coefficients of 0.89 to 0.98 for children aged 11-14 (Safrit, 1990). The technical error of measurement for subscapular skinfold varies from 0.88 to 1.16 mm for intraobserver error and from 0.88 to 1.53 mm for interobserver error (Harrison et al., 1988). For triceps skinfold, the technical error of measurement ranges from 0.4 to 0.8 mm for intraobserver error and from 0.8 to 1.89 mm for interobserver error (Harrison et al., 1988).

The triceps and subscapular skinfolds are the most widely used in growth studies, and national reference data were developed by using the samples included for BMI in the CDC 2000 growth charts (Addo and Himes, 2010). Skinfolds can be and have been measured on any number of bodily sites. A key is standard definition and location of the sites and proper marking of the sites prior to application of the skinfold calipers. As noted earlier, ratios of skinfolds measured on the trunk to those measured on the extremities are commonly used to estimate relative subcutaneous fat distribution, which has been related to chronic disease risk factors in youth.

ADMINISTRATIVE FEASIBILITY

When selecting fitness test items, an important criterion is the feasibility and practicality of the measures. The committee evaluated the feasibility and practicality of body composition measures assuming that they would be implemented by trained personnel as recommended in this report.

The measurements recommended for inclusion in a youth fitness test battery are height, weight, waist circumference, and triceps and subscapular skinfolds. All of these measurements can be taken reasonably quickly. The selected measurements, however, are not free of potential motivational or self-esteem influences; self-esteem may be affected by the interpretation of results for estimated body composition. For this reason and to protect privacy, waist circumference and subscapular skinfold thickness should not be assessed in group settings. It is assumed that appropriate space (e.g., a nurse's office) would be available to ensure the privacy of the measurement process since measurement of skinfolds and waist circumference requires exposure of the trunk to allow the test administrator to access the subscapular area on the back and the waist. This setting also would minimize the potential for embarrassment when two test administrators are needed in the room (see below).

Equipment needed to measure body composition using the tests recommended above includes a stadiometer, a scale, skinfold calipers, and a tape measure. The NHANES measurement techniques are presented in Annex 4-1 as an example of commonly used methodology for indicators of body composition. As noted there, these anthropometric measurements, while not difficult, are highly error-prone. To avoid error, only high-quality equipment should be used, and test administrators should have the necessary technical training. This training includes proper positioning for measuring height (standard erect posture with the head and eyes in the Frankfurt horizontal plane), procedures for stepping on and off the scale (for example, some children may require assistance, and children must be kept from jumping on the scale), positioning for measuring waist circumference (feet together), identification of the correct level for measuring waist circumference, and identification of the correct sites for measuring the triceps and subscapular skinfolds. The level for measuring waist circumference and the sites for each skinfold measurement should be marked on the skin.

Height and weight typically are measured without shoes and in light indoor clothing (e.g., shorts and a t-shirt); the subscapular site is easily accessed by raising the back of the t-shirt. Waist circumference is measured from the side (measurements taken face-to-face are generally invasive). Two technicians may be needed to measure waist circumference in some overweight and obese youth. This should not be a problem as a separate individual (who is well versed in the measurement protocols) should serve as recorder for the measurements. Other duties of the recorder include observation of the position of the subject (e.g., young children often slouch after being placed in the standard erect posture), proper identification of skinfold and waist circumference sites; and in measurement of waist circumference, checking to ensure that the tape is horizontal or is not pulled too tightly, resulting in major skin and soft tissue compression. The lack of a recorder will slow down the measurement process and contribute to potential error in transcribing measurements.

GUIDANCE FOR INTERPRETATION OF TEST RESULTS

The committee acknowledges that there are multiple approaches to establishing cut-points (cutoff scores) for estimates of body composition depending on the purpose and on the available data. In general, the committee considered the following two approaches:

- *Direct associations with health-related biomarkers*—This method involves examining associations between BMI, waist circumference, or sum of skinfold scores and cardiometabolic risk factors in youth. Ideally, as discussed in Chapter 3, data necessary to establish those associations will exist from broad populations of youth.
- *Indirect associations with health-related biomarkers using adult cut-points or data from other body composition measures*—When the necessary data in youth are not available, associations can be examined in adult data, and cut-points established for adults can be projected to the corresponding percentile in children, if appropriate. When data needed to establish associations between a specific

test and health do not exist in youth or adults or when a cut-point exists in adults but extrapolation to youth is not appropriate (i.e., waist circumference, sum of skinfolds), the percentiles from another fitness measure (such as the 85th and 95th percentiles used for BMI) can be used temporarily to derive interim cut-points. As research progresses, cut-points based on the measure's relationship to health in youth should be developed.

The committee concludes that these two approaches are appropriate for interpretation of body composition measurements administered in the context of a national youth fitness survey. For these approaches, a cut-point is determined specifically for each body composition test recommended. Obtaining information for the different indicators of body composition in this manner allows for a more complete and accurate description of an individual's body composition. As a result, the interpretation of the tests in terms of health risks is expanded and possibly more accurate than if only one test is administered.

A third approach, which involves transforming the raw data to a measure of percent body fat by using prediction equations, has been used for interpreting body composition test items. This approach makes it possible to compare results from more than one measurement with a selected standard for percent body fat. This approach may be appropriate when test administrators must select one measure of body composition from multiple alternatives, such as when a battery of tests is applied in schools and other educational settings.

Body Mass Index

Cut-points for BMI have been calculated by age and gender from percentiles developed using the CDC growth charts based on data from large national surveys. The CDC growth charts are based primarily on data from the National Health Examination Survey (NHES) and NHANES from 1963 to 1994 (NHES II and III and NHANES I, II, and III). Data on body weight from NHANES III for subjects ≥6 years of age were not used so as to avoid the influence of an increase in body weight from previous years.[1] Details are described in the CDC growth charts (Kuczmarski et al., 2000, 2002). Sample sizes were sufficiently large in the national surveys, which

[1]Questions arose over which population was appropriate for establishing such percentiles related to health given concern for the increasing prevalence of obesity between NHANES II (1976-1980) and NHANES III (1988-1994) (Kuczmarski et al., 2002; Ryan et al., 1999). Developing percentiles for weight using elevated values from NHANES III would have raised the percentiles and thus resulted in a false sense of having a satisfactory weight, specifically relative to stature.

TABLE 4-1 Centers for Disease Control and Prevention (CDC) Reference Values for Body Mass Index (BMI)

Percentile Ranking	Weight Status
Less than 5th percentile	Underweight
5th percentile to less than 85th percentile	Healthy weight
85th percentile to less than 95th percentile	Overweight
Equal to or greater than 95th percentile	Obese

were combined to produce the charts. Percentiles were derived for specific age groups by sex and were subsequently smoothed. Recommendations for the 85th percentile (P) to identify overweight, initially labeled as at risk of overweight ($P85 \leq BMI < P95$), and the 95th percentile to identify obesity, initially labeled overweight ($BMI \geq P95$), were based on the findings of an expert committee (Barlow and Dietz, 1998; Himes and Dietz, 1994). The recommendations have remained unchanged with the exception of overweight being used in place of risk of overweight and obesity in place of overweight (Barlow and Expert Committee, 2007). The CDC developed cut-points for underweight youth based on the 5th percentile as recommended by the World Health Organization's Expert Committee on Physical Status (WHO, 1995).

A national survey of youth fitness in the United States should use the CDC cut-points for weight status (Table 4-1). An alternative set of cut-points is that developed by the International Obesity Task Force (Cole et al., 2000), which are widely used internationally.[2]

Skinfolds

The same U.S. national data used to develop growth charts for body weight and BMI were used to develop reference curves for the triceps and subscapular skinfolds for youth through age 19 (Addo and Himes, 2010). The committee recommends using the established percentiles for BMI to derive interim cut-points until more studies are conducted to determine health-related cut-points in youth. The interim cut-points could be verified using the corresponding percentiles for the concurrent relationship to health in adults.

[2]Data used to develop the criteria for children and adolescents were based on six nationally representative cross-sectional samples from Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States. In establishing the cut-points for children and adolescents, curves were mathematically fit to the pooled BMI data from the six studies so that they passed through the adult criteria for overweight (BMI of 25.0 kg/m^2) and obesity (30 kg/m^2) at age 18. Cut-points recommended by the World Health Organization (1995) were used for adults: overweight, BMI 25.0 to 29.9 kg/m^2 and obesity, BMI ≥30.0 kg/m^2.

Waist Circumference

The development of cut-points for waist circumference is complicated by methodological variation, that is, different levels at which the measurement is taken in various studies. Although the recommended level is midway between the iliac crest and the lowest rib (WHO, 2008), NHANES reference data (1999-2002 and 2003-2006) were based on measures taken at the uppermost lateral border of the iliac crest (McDowell et al., 2005, 2008). Corresponding reference values for Canadian youth were derived at the narrowest waist (Katzmarzyk et al., 2004). A report describing suggested percentiles for British youth indicates that waist circumference was measured midway between the tenth rib and the iliac crest, but later in the paper, the authors indicate it was measured at the "natural waist" (McCarthy et al., 2001). Wang and colleagues (2003) found that waist circumference estimates taken at different levels on adults are not comparable, especially among females. A systematic review of the adult literature, however, found that differences in the level of measurement did not have a considerable influence on the relationship between waist circumference and health outcomes (Ross et al., 2008).

Criterion-referenced cut-points for waist circumference have been established in adults (WHO, 2011). Although recommendations for cut-points have been developed for different samples, generally reflecting the 90th percentile by age and gender, standardized cut-points for youth have not yet been established. The committee recommends using the established percentiles for BMI to derive interim cut-points until more studies are conducted to determine health-related cut-points in youth. The interim cut-points could be verified using the corresponding percentiles for the concurrent relationship to health in adults.

CONCLUSIONS

Body composition can influence performance on some physical fitness tests and is also a health-related risk factor associated with physical fitness. The committee operationally defined body composition as a component of fitness, a marker of health, and a modifier of fitness for the purposes of this report. Given its well-known central role in both fitness and health, body composition should be included in a survey of youth fitness and measured across the life span.

The committee's recommendations with respect to body composition are premised on the committee's intention that the test administrators will have the necessary knowledge and training in test protocols and interpretation of results. The committee recommends inclusion of the following anthropometric measurements in a youth fitness test battery: (1) height and weight for the derivation of BMI, (2) waist circumference, and (3) triceps and subscapular skinfold thicknesses. Height also serves as an indicator of linear growth status. (See Annex 4-1 for common examples of measurement techniques.)

The committee concluded that the above three measures of body composition are important to collect in a national youth fitness survey for several reasons. First, each measure is a proximal estimation of body fat and has increased standard of error of over laboratory measures. Also, there is consensus that the measurement of body composition is multidimensional (Bouchard et al., 1994). Second, no single measure is considered the gold standard and representative of all body composition tenets for children of all morphologies: BMI is a marker of obesity, waist circumference is a marker of abdominal adiposity, and skinfold thicknesses are a measure of subcutaneous fat. The measures recommended have acceptable validity and reliability.

To interpret the findings of body composition testing and determine whether individuals or populations are at risk of negative health outcomes, the committee recommends employing two approaches. For BMI, the CDC's current established cut-points for underweight, overweight, and obesity should be applied. Interim cut-points for the waist circumference and skinfold measures should be set at levels that are analogous to those currently being applied by the CDC for BMI. This approach should be used until the necessary evidence becomes available to support establishing waist circumference and skinfold cut-points by associating those measures with cardiometabolic risk factors. The committee's full recommendation for including body composition in a national youth fitness survey is presented in Chapter 8.

When body composition is measured in schools and other educational settings, important concerns arise related to the measurement of waist circumference and skinfolds. Therefore, the committee recommends that only BMI be used in these settings. A full description of considerations and the committee's recommendation for schools and other educational settings is included in Chapter 9.

ANNEX 4-1
MEASUREMENT OF BODY COMPOSITION QUALITY CONTROL AND TECHNIQUES

Body composition measurements should be taken by trained individuals using established techniques. Error—the discrepancy between a measured value and its true quantity—is inherent in anthropometry. It can be random[3] or systematic.[4]

Replicate measurements of the same subject are used to estimate variability or error. Replicates on the same individual are taken independently by the same technician after a period of time has elapsed (intraobserver) or are taken on the same individual by two different technicians (interobserver). Replicate measurements provide an estimate of imprecision. The technical error of measurement is a widely used measure of replicability (Malina et al., 1973; Mueller and Martorell, 1988). Technical errors are reported in the units of the specific measurement. Intra- and interobserver technical errors for a variety of dimensions in national surveys and several more local studies are summarized by Malina (1995).

Accuracy is another aspect of the measurement process. It refers to how closely measurements taken by one or several technicians approximate the "true" measurement. Accuracy ordinarily is assessed by comparing measurements taken by technicians with those obtained by a well-trained or "criterion" anthropometrist (the reference). Note, however, that well-trained, expert anthropometrists do make errors.

The height, weight, waist circumference, and triceps and subscapular skinfold measurement techniques described below are provided as examples from the commonly used *NHANES Anthropometry Procedures Manual*.[5]

Stature and Weight

Stature, or standing height, is the linear distance from the floor or standing surface to the top (vertex) of the skull. It is measured to the nearest millimeter with the subject in standard erect posture, without shoes.

[3]Random error is associated with variation within and among individuals in measurement technique, problems with the measuring instruments (e.g., variation in or even lack of calibration or random variation in manufacture), and errors in recording. Random error is nondirectional, i.e., above or below the true dimension. Random errors tend to cancel each other out in large-scale surveys and ordinarily are not a major concern.

[4]Systematic error results from the tendency of a technician or a measuring instrument to consistently under- or overmeasure a dimension. Such error is directional and introduces bias. Measurement variability also is associated with the individual (e.g., normal variation in physiology, temperament, cooperativeness, and stranger anxiety).

[5]Available at http://www.cdc.gov/nchs/data/nhanes/nhanes_07_08/manual_an.pdf (accessed May 10, 2012).

Body weight is a measure of body mass. It is measured to the nearest 100 grams (depending on the type of scale) with the individual attired in ordinary, light, indoor clothing without shoes (e.g., gym shorts and a t-shirt). It assumed that the scales would be calibrated regularly for a national survey. Height and weight are used to derive body mass index (BMI) (kg/m^2).

Waist Circumference

The protocol for waist circumference calls for measuring just above the uppermost lateral border of the right illium after normal expiration. The level should be marked on the skin. When the tape is applied, it should make contact with the skin without indenting it. The measurement should not be made over clothing. Two individuals may be needed, especially for some overweight and obese individuals.

Skinfolds

A skinfold thickness is a double fold of skin and underlying soft tissue at a specific site. Skinfolds are measured to the nearest 0.5 mm (some calipers measure to the nearest millimeter, while others measure to the nearest 0.2 mm). Three measurements usually are taken for each skinfold (some protocols recommend two).

Triceps skinfold is measured on the back of the arm (over the triceps muscle) at the level midway between the lateral border of the acromial process of the scapula (acromion) and the inferior border of the olecranon process of the ulna. With the arm flexed to 90 degrees at the elbow, the acromion is marked. A measuring tape is placed on the acromion (zero marker) and run down the lateral side of the upper arm. The distance midway between the acromion and the olecranon is marked and extended to the back of the arm. The skinfold is measured with the arm hanging relaxed (loosely) at the side by grasping a vertical fold about 1 cm above the mark, with the caliper being placed at the level of the mark.

Subscapular skinfold is measured 1 cm below the tip (inferior angle) of the scapula. The measurement site should be marked on the skin. The skinfold should follow the natural anatomical (cleavage) lines of the skin. It is not a vertical fold like that taken over the triceps.

REFERENCES

Ackland, T. R., T. G. Lohman, J. Sundgot-Borgen, R. J. Maughan, N. L. Meyer, A. D. Stewart, and W. Muller. 2012. Current status of body composition assessment in sport review and position statement on behalf of the ad hoc research working group on body composition health and performance, under the auspices of the I.O.C. Medical Commission. *Sports Medicine* 42(3):227-249.

Addo, O. Y., and J. H. Himes. 2010. Reference curves for triceps and subscapular skinfold thicknesses in US children and adolescents. *American Journal of Clinical Nutrition* 91(3):635-642.

Andersen, L. B., L. B. Sardinha, K. Froberg, C. J. Riddoch, A. S. Page, and S. A. Anderssen. 2008. Fitness, fatness and clustering of cardiovascular risk factors in children from Denmark, Estonia and Portugal: The European Youth Heart Study. *International Journal of Pediatric Obesity* 3(Suppl. 1):58-66.

Artero, E. G., V. España-Romero, J. Castro-Piñero, F. B. Ortega, J. Suni, M. J. Castillo-Garzon, and J. R. Ruiz. 2011. Reliability of field-based fitness tests in youth. *International Journal of Sports Medicine* 32(3):159-169.

Barlow, S. E., and W. H. Dietz. 1998. Obesity evaluation and treatment: Expert Committee recommendations. The Maternal and Child Health Bureau, Health Resources and Services Administration and the Department of Health and Human Services. *Pediatrics* 102(3):E29.

Barlow, S. E., and Expert Committee. 2007. Expert Committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: Summary report. *Pediatrics* 120(Suppl. 4):S164-S192.

Bassali, R., J. L. Waller, B. Gower, J. Allison, and C. L. Davis. 2010. Utility of waist circumference percentile for risk evaluation in obese children. *International Journal of Pediatric Obesity* 5(1):97-101.

Beets, M. W., and K. H. Pitetti. 2004. One-mile run/walk and body mass index of an ethnically diverse sample of youth. *Medicine and Science in Sports and Exercise* 36(10): 1796-1803.

Beets, M. W., K. H. Pitetti, and B. J. Cardinal. 2005. Progressive aerobic cardiovascular endurance run and body mass index among an ethnically diverse sample of 10-15-year-olds. *Research Quarterly for Exercise and Sport* 76(4):389-397.

Beunen, G., R. M. Malina, M. Ostyn, R. Renson, J. Simons, and D. Van Gerven. 1982. Fatness and skeletal maturity of Belgian boys 12 through 17 years of age. *American Journal of Physical Anthropology* 59(4):387-392.

Beunen, G., R. M. Malina, M. Ostyn, R. Renson, J. Simons, and D. Van Gerven. 1983. Fatness, growth and motor fitness of Belgian boys 12 through 20 years of age. *Human Biology* 55(3):599-613.

Beunen, G., R. M. Malina, J. Lefevre, A. L. Claessens, R. Renson, J. Simons, H. Maes, B. Vanreusel, and R. Lysens. 1994a. Size, fatness and relative fat distribution of males of contrasting maturity status during adolescence and as adults. *International Journal of Obesity and Related Metabolic Disorders* 18(10):670-678.

Beunen, G. P., R. M. Malina, J. A. Lefevre, A. L. Claessens, R. Renson, and B. Vanreusel. 1994b. Adiposity and biological maturity in girls 6-16 years of age. *International Journal of Obesity and Related Metabolic Disorders* 18(8):542-546.

Boileau, R. A., and T. G. Lohman. 1977. Measurement of human physique and its effect on physical performance. *Orthopedic Clinics of North America* 8(3):563-581.

Boileau, R. A., T. G. Lohman, M. H. Slaughter, T. E. Ball, S. B. Going, and M. K. Hendrix. 1984. Hydration of the fat-free body in children during maturation. *Human Biology* 56(4):651-666.

Bouchard, C., and R. J. Shephard. 1994. Physical activity, fitness and health: The model and key concepts. In *Physical activity, fitness and health: International proceedings and consensus statement*, edited by C. Bouchard, R. J. Shephard, and T. Stephens. Champaign, IL: Human Kinetics Publishers. Pp. 11-22.

Bouchard, C., S. N. Blair, and W. L. Haskell. 2007. Why study physical activity and health? In *Physical activity and health*, edited by C. Bouchard, S. N. Blair, and W. L. Haskell. Champaign, IL: Human Kinetics Publishers. Pp. 3-19.

Bovet, P., R. Auguste, and H. Burdette. 2007. Strong inverse association between physical fitness and overweight in adolescents: A large school-based survey. *International Journal of Behavioral Nutrition and Physical Activity* 4:24-31.

Carnethon, M. R., M. Gulati, and P. Greenland. 2005. Prevalence and cardiovascular disease correlates of low cardiorespiratory fitness in adolescents and adults. *Journal of the American Medical Association* 294(23):2981-2988.

Castro-Piñero, J., E. G. Artero, V. España-Romero, F. B. Ortega, M. Sjostrom, J. Suni, and J. R. Ruiz. 2010. Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* 44(13):934-943.

Chang, C., W. Liu, X. Zhao, S. Li, and C. Yu. 2008. Effect of supervised exercise intervention on metabolic risk factors and physical fitness in Chinese obese children in early puberty. *Obesity Reviews* 9(Suppl. 1):135-141.

Chatrath, R., R. Shenoy, M. Serratto, and D. G. Thoele. 2002. Physical fitness of urban American children. *Pediatric Cardiology* 23(6):608-612.

Cole, T. J., M. C. Bellizzi, K. M. Flegal, and W. H. Dietz. 2000. Establishing a standard definition for child overweight and obesity worldwide: International survey. *British Medical Journal* 320(7244):1240-1243.

Daniels, S. R., P. R. Khoury, and J. A. Morrison. 1997. The utility of body mass index as a measure of body fatness in children and adolescents: Differences by race and gender. *Pediatrics* 99(6):804-807.

Daniels, S. R., J. A. Morrison, D. L. Sprecher, P. Khoury, and T. R. Kimball. 1999. Association of body fat distribution and cardiovascular risk factors in children and adolescents. *Circulation* 99(4):541-545.

Deforche, B., J. Lefevre, I. De Bourdeaudhuij, A. P. Hills, W. Duquet, and J. Bouckaert. 2003. Physical fitness and physical activity in obese and nonobese Flemish youth. *Obesity Research* 11(3):434-441.

Despres, J. P., A. Nadeau, A. Tremblay, M. Ferland, S. Moorjani, P. J. Lupien, G. Theriault, S. Pinault, and C. Bouchard. 1989. Role of deep abdominal fat in the association between regional adipose tissue distribution and glucose tolerance in obese women. *Diabetes* 38(3):304-309.

Dumith, S. C., D. Van Dusen, and H. W. Kohl. 2012. Physical fitness measures among children and adolescents: Are they all necessary? *Journal of Sports Medicine and Physical Fitness* 52(2):181-189.

Freedman, D. S., W. H. Dietz, S. R. Srinivasan, and G. S. Berenson. 1999a. The relation of overweight to cardiovascular risk factors among children and adolescents: The Bogalusa Heart Study. *Pediatrics* 103(6):1175-1182.

Freedman, D. S., M. K. Serdula, S. R. Srinivasan, and G. S. Berenson. 1999b. Relation of circumferences and skinfold thicknesses to lipid and insulin concentrations in children and adolescents: The Bogalusa Heart Study. *American Journal of Clinical Nutrition* 69(2):308-317.

Freedman, D. S., L. K. Khan, W. H. Dietz, S. R. Srinivasan, and G. S. Berenson. 2001. Relationship of childhood obesity to coronary heart disease risk factors in adulthood: The Bogalusa Heart Study. *Pediatrics* 108(3):712-718.

Freedman, D. S., W. H. Dietz, S. R. Srinivasan, and G. S. Berenson. 2009. Risk factors and adult body mass index among overweight children: The Bogalusa Heart Study. *Pediatrics* 123(3):750-757.

Freedman, D. S., A. Goodman, O. A. Contreras, P. DasMahapatra, S. R. Srinivasan, and G. S. Berenson. 2012. Secular trends in BMI and blood pressure among children and adolescents: The Bogalusa Heart Study. *Pediatrics* 130(1):e159-e166.

Going, S. B., T. G. Lohman, E. C. Cussler, D. P. Williams, J. A. Morrison, and P. S. Horn. 2011. Percent body fat and chronic disease risk factors in U.S. children and youth. *American Journal of Preventive Medicine* 41(4, Suppl. 2):S77-S86.

Goran, M. I., and B. A. Gower. 1998. Abdominal obesity and cardiovascular risk in children. *Coronary Artery Disease* 9(8):483-487.

Goran, M. I., D. B. Allison, and E. T. Poehlman. 1995. Issues relating to normalization of body-fat content in men and women. *International Journal of Obesity* 19(9):638-643.

Guo, S. S., W. Wu, W. C. Chumlea, and A. F. Roche. 2002. Predicting overweight and obesity in adulthood from body mass index values in childhood and adolescence. *American Journal of Clinical Nutrition* 76(3):653-658.

Gutin, B., M. Litaker, S. Islam, T. Manos, C. Smith, and F. Treiber. 1996. Body-composition measurement in 9-11-y-old children by dual-energy x-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. *American Journal of Clinical Nutrition* 63(3):287-292.

Gutin, B., Z. O. Yin, M. C. Humphries, and P. Barbeau. 2005. Relations of moderate and vigorous physical activity to fitness and fatness in adolescents. *American Journal of Clinical Nutrition* 81(4):746-750.

Haroun, D., S. J. C. Taylor, R. M. Viner, R. S. Hayward, T. S. Darch, S. Eaton, T. J. Cole, and J. C. K. Wells. 2010. Validation of bioelectrical impedance analysis in adolescents across different ethnic groups. *Obesity* 18(6):1252-1259.

Harrison, G. G., E. R. Buskirk, J. E. Lindsay Carter, F. E. Johnston, T. G. Lohman, M. L. Pollock, A. F. Roche, and J. Wilmore. 1988. Skinfold thicknesses and measurement technique. In *Anthropometric standardization reference manual*. Vol. 55-70, edited by T. G. Lohman, A. F. Roche, and R. Martorell. Champaign, IL: Human Kinetics Publishers.

He, Q., M. Horlick, J. Thornton, J. Wang, R. N. Pierson, S. Y. Heshka, and D. Gallagher. 2004. Sex-specific fat distribution is not linear across pubertal groups in a multiethnic study. *Obesity Research* 12(4):725-733.

Herman, K. M., C. L. Craig, L. Gauvin, and P. T. Katzmarzyk. 2009. Tracking of obesity and physical activity from childhood to adulthood: The Physical Activity Longitudinal Study. *International Journal of Pediatric Obesity* 4(4):281-288.

Heymsfield, S. B., T. G. Lohman, Z. M. Wang, and S. B. Going, eds. 2005. *Human body composition*. 2nd ed. Champaign, IL: Human Kinetics Publishers.

Himes, J. H., and W. H. Dietz. 1994. Guidelines for overweight in adolescent preventive services: Recommendations from an expert committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. *American Journal of Clinical Nutrition* 59(2):307-316.

Houtkooper, L. B., and S. B. Going. 1994. Body composition: How should it be measured? Does it affect sports performance? *Sports Science Exchange #52, Gatorade Sports Science Institute* 7(5):9.

Huang, Y. C., and R. M. Malina. 2010. Body mass index and individual physical fitness tests in Taiwanese youth aged 9-18 years. *International Journal of Pediatric Obesity* 5(5):404-411.

Huxley, R., S. Mendis, E. Zheleznyakov, S. Reddy, and J. Chan. 2010. Body mass index, waist circumference and waist: Hip ratio as predictors of cardiovascular risk: A review of the literature. *European Journal of Clinical Nutrition* 64(1):16-22.

Jekal, Y., E. S. Kim, J. A. Im, J. H. Park, M. K. Lee, S. H. Lee, S. H. Suh, S. H. Chu, E. S. Kang, H. C. Lee, and J. Y. Jeon. 2009. Interaction between fatness and fitness on CVD risk factors in Asian youth. *International Journal of Sports Medicine* 30(10):733-740.

Katzmarzyk, P. T., S. R. Srinivasan, W. Chen, R. M. Malina, C. Bouchard, and G. S. Berenson. 2004. Body mass index, waist circumference, and clustering of cardiovascular disease risk factors in a biracial sample of children and adolescents. *Pediatrics* 114(2):198-205.

Katzmarzyk, P. T., W. Shen, A. Baxter-Jones, J. D. Bell, N. F. Butte, E. W. Demerath, V. Gilsanz, M. I. Goran, V. Hirschler, H. H. Hu, C. Maffeis, R. M. Malina, M. J. Müller, A. Pietrobelli, and J. C. K. Wells. 2012. Adiposity in children and adolescents: Correlates and clinical consequences of fat stored in specific body depots. *Pediatric Obesity* [Epub ahead of print].

Kim, J., A. Must, G. M. Fitzmaurice, M. W. Gillman, V. Chomitz, E. Kramer, R. McGowan, and K. E. Peterson. 2005. Relationship of physical fitness to prevalence and incidence of overweight among schoolchildren. *Obesity Research* 13(7):1246-1254.

Kindblom, J. M., M. Lorentzon, E. Norjavaara, L. Lonn, J. Brandberg, J. E. Angelhed, A. Hellqvist, S. Nilsson, and C. Ohlsson. 2006. Pubertal timing is an independent predictor of central adiposity in young adult males: The Gothenburg Osteoporosis and Obesity Determinants Study. *Diabetes* 55(11):3047-3052.

Kuczmarski, R. J., C. L. Ogden, L. M. Grummer-Strawn, K. M. Flegal, S. S. Guo, R. Wei, Z. Mei, L. R. Curtin, A. F. Roche, and C. L. Johnson. 2000. CDC growth charts: United States. *Advance Data* (314):1-27.

Kuczmarski, R. J., C. L. Ogden, S. S. Guo, L. Grummer-Strawn, K. M. Flegal, Z. Mei, R. Wei, L. R. Curtin, A. F. Roche, and C. L. Johnson. 2002. 2000 CDC growth charts for the United States: Methods and development. *Vital and Health Statistics* 11(246): 1-190.

Labayen, I., F. B. Ortega, L. A. Moreno, C. Redondo-Figuero, G. Bueno, S. Gomez-Martinez, and M. Bueno. 2009. The effect of early menarche on later body composition and fat distribution in female adolescents: Role of birth weight. *Annals of Nutrition and Metabolism* 54(4):313-320.

Larsson, B., K. Svardsudd, L. Welin, L. Wilhelmsen, P. Bjorntorp, and G. Tibblin. 1984. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. *British Medical Journal (Clinical Research Edition)* 288(6428):1401-1404.

Lauer, R. M., and W. R. Clarke. 1989. Childhood risk-factors for high adult-blood pressure: The Muscatine Study. *Pediatrics* 84(4):633-641.

Lee, S. J., and S. A. Arslanian. 2007. Cardiorespiratory fitness and abdominal adiposity in youth. *European Journal of Clinical Nutrition* 61(4):561-565.

Lee, S., J. L. Kuk, Y. Kim, and S. A. Arslanian. 2011. Measurement site of visceral adipose tissue and prediction of metabolic syndrome in youth. *Pediatric Diabetes* 12(3, Pt. 2): 250-257.

Li, C., E. S. Ford, G. Zhao, and A. H. Mokdad. 2009. Prevalence of pre-diabetes and its association with clustering of cardiometabolic risk factors and hyperinsulinemia among U.S. adolescents: National Health and Nutrition Examination Survey 2005-2006. *Diabetes Care* 32(2):342-347.

Liu, A., A. P. Hills, X. Hu, Y. Li, L. Du, Y. Xu, N. M. Byrne, and G. Ma. 2010. Waist circumference cut-off values for the prediction of cardiovascular risk factors clustering in Chinese school-aged children: A cross-sectional study. *BMC Public Health* 10:82.

Lloyd, L. K., P. A. Bishop, J. L. Walker, K. R. Sharp, and M. T. Richardson. 2003. The influence of body size and composition on Fitnessgram® test performance and the adjustment of Fitnessgram® test scores for skinfold thickness in youth. *Measurement in Physical Education and Exercise Science* 7(4):205-226.

Lloyd, L. J., S. C. Langley-Evans, and S. McMullen. 2012. Childhood obesity and risk of the adult metabolic syndrome: A systematic review. *International Journal of Obesity* 36(1): 1-11.

Lobelo, F., R. R. Pate, M. Dowda, A. D. Liese, and S. R. Daniels. 2010. Cardiorespiratory fitness and clustered cardiovascular disease risk in U.S. adolescents. *Journal of Adolescent Health* 47(4):352-359.

Lohman, T. G. 1981. Skinfolds and body density and their relation to body fatness: A review. *Human Biology* 53(2):181-225.

Lohman, T. G. 1986. Applicability of body composition techniques and constants for children and youths. *Exercise and Sport Sciences Reviews* 14:325-357.

Malina, R. M. 1975. Anthropometric correlates of strength end motor performance. *Exercise and Sport Sciences Reviews* 3(1):249-274.

Malina, R. M. 1992. Physique and body composition: Effects on performance and effects on training, semistarvation, and overtraining In *Eating, body weight, and performance in athletes*, edited by K. D. Brownell, J. Rodin, and J. H. Wilmore. Philadelphia, PA: Lea & Febiger. Pp. 94-114.

Malina, R. M. 1995. Anthropometry. In *Physiological assessment of human fitness*, edited by P. J. Maud and C. Foster. Champaign, IL: Human Kinetics Publishers. Pp. 205-219.

Malina, R. M. 1996. Regional body composition: Age, sex, and ethnic variation. In *Human body composition*, edited by A.F. Roche, S. B. Heymsfield, and T. G. Lohman. Champaign, IL: Human Kinetics Publishers. Pp. 217-256.

Malina, R. M. 2000. Growth, maturation, and performance. In *Exercise and sport science*, edited by W. E. Garrett and D. T. Kirkendall. Philadelphia, PA: Lippincott Williams & Wilkins. Pp. 425-445.

Malina, R. M. 2005. Variation in body composition associated with sex and ethnicity. In *Human body composition*. 2nd ed., edited by S. B. Heymsfield, T. G. Lohman, Z. M. Wang and S. B. Going. Champaign, IL: Human Kinetics Publishers. Pp. 271-298, 473-476.

Malina, R. M., and C. Bouchard. 1988. Subcutaneous fat distribution during growth. In *Fat distribution during growth and later health outcomes*, edited by C. Bouchard and F. E. Johnston. New York: Alan R. Liss, Inc. Pp. 63-84.

Malina, R. M., and P. T. Katzmarzyk. 1999. Validity of the body mass index as an indicator of the risk and presence of overweight in adolescents. *American Journal of Clinical Nutrition* 70(1):131S-136S.

Malina, R. M., P. V. V. Hamill, and S. Lemeshow. 1973. *Selected body measurements of children 6-11 years*. Vital and Health Statistics, Series 11, No. 123. Washington, DC: U.S. Government Printing Office.

Malina, R. M., G. P. Beunen, A. L. Claessens, J. Lefevre, B. Vandeneynde, R. Renson, B. Vanreusel, and J. Simons. 1995. Fatness and physical-fitness of girls 7 to 17 years. *Obesity Research* 3(3):221-231.

Malina, R. M., P. T. Katzmarzyk, and S. R. Siegel. 1998. Overnutrition, undernutrition and the body mass index: Implications for strength and motor fitness. *Medicine and Sport Science* 43:13-26.

Malina, R. M., S. Koziel, and T. Bielicki. 1999. Variation in subcutaneous adipose tissue distribution associated with age, sex, and maturation. *American Journal of Human Biology* 11(2):189-200.

Malina, R., C. Bouchard, and O. Bar-Or. 2004. *Growth, maturation, and physical activity*. 2nd ed. Champaign, IL: Human Kinetics Publishers.

May, A. L., E. V. Kuklina, and P. W. Yoon. 2012. Prevalence of cardiovascular disease risk factors among US adolescents, 1999-2008. *Pediatrics* 129(6):1035-1041.

Maynard, L. M., W. Wisemandle, A. F. Roche, W. C. Chumlea, S. S. Guo, and R. M. Siervogel. 2001. Childhood body composition in relation to body mass index. *Pediatrics* 107(2):344-350.

McCarthy, H. D., K. V. Jarrett, and H. F. Crawley. 2001. The development of waist circumference percentiles in British children aged 5.0-16.9 y. *European Journal of Clinical Nutrition* 55(10):902-907.

McDowell, M. A., C. D. Fryar, R. Hirsch, and C. L. Ogden. 2005. Anthropometric reference data for children and adults: U.S. population, 1999-2002. *Advance Data* (361):1-5.

McDowell, M. A., C. D. Fryar, C. L. Ogden, and K. M. Flegal. 2008. Anthropometric reference data for children and adults: United States, 2003-2006. *National Health Statistics Report* 10.

Mei, Z., L. M. Grummer-Strawn, A. Pietrobelli, A. Goulding, M. I. Goran, and W. H. Dietz. 2002. Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *American Journal of Clinical Nutrition* 75(6):978-985.

Moreno, L. A., M. G. Blay, G. Rodríguez, V. A. Blay, M. I. Mesana, J. L. Olivares, J. Fleta, A. Sarría, M. Bueno, and A.-Z. S. Group. 2006. Screening performances of the International Obesity Task Force body mass index cut-off values in adolescents. *Journal of the American College of Nutrition* 25(5):403-408.

Mota, J., L. Flores, L. Flores, J. C. Ribeiro, and M. P. Santos. 2006. Relationship of single measures of cardiorespiratory fitness and obesity in young schoolchildren. *American Journal of Human Biology* 18(3):335-341.

Moulton, C. R. 1923. Age and chemical development in mammals. *Journal of Biological Chemistry* 57:79-97.

Mueller, W. H., and R. M. Malina. 1987. Relative reliability of circumferences and skinfolds as measures of body fat distribution. *American Journal of Physical Anthropology* 72(4):437-439.

Mueller, W. H., and R. Martorell. 1988. Reliability and accuracy of measurement. In *Anthropometric standardization reference manual*, edited by T. G. Lohman, A. F. Roche, and R. Martorell. Champaign, IL: Human Kinetics Publishers. Pp. 83-86.

Nassis, G. P., G. Psarra, and L. S. Sidossis. 2005. Central and total adiposity are lower in overweight and obese children with high cardiorespiratory fitness. *European Journal of Clinical Nutrition* 59(1):137-141.

Nguyen, Q. M., S. R. Srinivasan, J. H. Xu, W. Chen, and G. S. Berenson. 2008. Changes in risk variables of metabolic syndrome since childhood in pre-diabetic and type 2 diabetic subjects: The Bogalusa Heart Study. *Diabetes Care* 31(10):2044-2049.

ODPHP (Office of Disease Prevention and Health Promotion). 1985. *Summary of findings from National Children and Youth Fitness Study*. Washington, DC: U.S. Department of Health and Human Services.

ODPHP. 1987. *Summary of findings from National Children and Youth Fitness Study II*. Washington, DC: U.S. Department of Health and Human Services.

Pietrobelli, A., M. S. Faith, D. B. Allison, D. Gallagher, G. Chiumello, and S. B. Heymsfield. 1998. Body mass index as a measure of adiposity among children and adolescents: A validation study. *Journal of Pediatrics* 132(2):204-210.

Pollock, N. K., P. J. Bernard, K. Wenger, S. Misra, B. A. Gower, J. D. Allison, H. Zhu, and C. L. Davis. 2010. Lower bone mass in prepubertal overweight children with prediabetes. *Journal of Bone and Mineral Research* 25(12):2760-2769.

Reed, K. E., D. E. R. Warburton, H. M. Macdonald, P. J. Naylor, and H. A. McKay. 2008. Action Schools! BC: A school-based physical activity intervention designed to decrease cardiovascular disease risk factors in children. *Preventive Medicine* 46(6):525-531.

Reilly, J. J., A. R. Dorosty, N. M. Ghomizadeh, A. Sherriff, J. C. Wells, and A. R. Ness. 2010. Comparison of waist circumference percentiles versus body mass index percentiles for diagnosis of obesity in a large cohort of children. *International Journal of Pediatric Obesity* 5(2):151-156.

Ribeiro, J. C., S. Guerra, J. Oliveira, L. B. Andersen, J. A. Duarte, and J. Mota. 2004. Body fatness and clustering of cardiovascular disease risk factors in Portuguese children and adolescents. *American Journal of Human Biology* 16(5):556-562.

Rizzo, N. S., J. R. Ruiz, A. Hurtig-Wennlof, F. B. Ortega, and M. Sjostrom. 2007. Relationship of physical activity, fitness, and fatness with clustered metabolic risk in children and adolescents: The European Youth Heart Study. *Journal of Pediatrics* 150(4):388-394.

Roche, A. F., S. B. Heymsfield, and T. G. Lohman. 1996. *Human body composition*. Champaign, IL: Human Kinetics Publishers.

Rolland-Cachera, M. F., M. Deheeger, F. Bellisle, M. Sempé, M. Guilloud-Bataille, and E. Patois. 1984. Adiposity rebound in children: A simple indicator for predicting obesity. *American Journal of Clinical Nutrition* 39(1):129-135.

Ross, R., T. Berentzen, A. J. Bradshaw, I. Janssen, H. S. Kahn, P. T. Katzmarzyk, J. L. Kuk, J. C. Seidell, M. B. Snijder, T. I. A. Sorensen, and J. P. Despres. 2008. Diagnostic in obesity comorbidities: Does the relationship between waist circumference, morbidity and mortality depend on measurement protocol for waist circumference? *Obesity Reviews* 9(4):312-325.

Ryan, A. S., A. F. Roche, and R. J. Kuczmarski. 1999. Weight, stature, and body mass index data for Mexican Americans from the third National Health and Nutrition Examination Survey (NHANES III, 1988-1994). *American Journal of Human Biology* 11(5):673-686.

Safrit, M. J. 1990. The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science* 2:9-28.

Sandhu, J., Y. Ben-Shlomo, T. J. Cole, J. Holly, and G. D. Smith. 2006. The impact of childhood body mass index on timing of puberty, adult stature and obesity: A follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936-1964). *International Journal of Obesity* 30(1):14-22.

Savva, S. C., M. Tornaritis, M. E. Savva, Y. Kourides, A. Panagi, N. Silikiotou, C. Georgiou, and A. Kafatos. 2000. Waist circumference and waist-to-height ratio are better predictors of cardiovascular disease risk factors in children than body mass index. *International Journal of Obesity* 24(11):1453-1458.

Sekulic, D., N. Zenic, and G. Markovic. 2005. Non linear relationships between anthropometric and motor-endurance variables. *Collegium Antropologicum* 29(2):723-730.

Stigman, S., P. Rintala, K. Kukkonen-Harjula, U. Kujala, M. Rinne, and M. Fogelholm. 2009. Eight-year-old children with high cardiorespiratory fitness have lower overall and abdominal fatness. *International Journal of Pediatric Obesity* 4(2):98-105.

Stratton, G., D. Canoy, L. M. Boddy, S. R. Taylor, A. F. Hackett, and I. E. Buchan. 2007. Cardiorespiratory fitness and body mass index of 9-11-year-old English children: A serial cross-sectional study from 1998 to 2004. *International Journal of Obesity* 31(7): 1172-1178.

Strong, W. B., R. M. Malina, C. J. R. Blimkie, S. R. Daniels, R. K. Dishman, B. Gutin, A. C. Hergenroeder, A. Must, P. A. Nixon, J. M. Pivarnik, T. Rowland, S. Trost, and F. Trudeau. 2005. Evidence based physical activity for school-age youth. *Journal of Pediatrics* 146(6):732-737.

Taylor, R. W., I. E. Jones, S. M. Williams, and A. Goulding. 2000. Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy x-ray absorptiometry, in children aged 3-19 y. *American Journal of Clinical Nutrition* 72(2):490-495.

Taylor, R. W., A. M. Grant, S. M. Williams, and A. Goulding. 2010. Sex differences in regional body fat distribution from pre- to postpuberty. *Obesity* 18(7):1410-1416.

Tjonna, A. E., T. O. Stolen, A. Bye, M. Volden, S. A. Slordahl, R. Odegard, E. Skogvoll, and U. Wisloff. 2009. Aerobic interval training reduces cardiovascular risk factors more than a multitreatment approach in overweight adolescents. *Clinical Science* 116(3-4): 317-326.

Toprak, D., A. Toprak, W. Chen, J. H. Xu, S. Srinivasan, and G. S. Berenson. 2011. Adiposity in childhood is related to c-reactive protein and adiponectin in young adulthood: From the Bogalusa Heart Study. *Obesity (Silver Spring)* 19(1):185-190.

Wang, J., J. C. Thornton, S. Bari, B. Williamson, D. Gallagher, S. B. Heymsfield, M. Horlick, D. Kotler, B. Laferrere, L. Mayer, F. X. Pi-Sunyer, and R. N. Pierson. 2003. Comparisons of waist circumferences measured at 4 sites. *American Journal of Clinical Nutrition* 77(2):379-384.

Wang, Z. M., R. N. Pierson, Jr., and S. B. Heymsfield. 1992. The five-level model: A new approach to organizing body-composition research. *American Journal of Clinical Nutrition* 56(1):19-28.

Wang, Z. M., S. Heshka, R. N. Pierson, Jr., and S. B. Heymsfield. 1995. Systematic organization of body-composition methodology: An overview with emphasis on component-based methods. *American Journal of Clinical Nutrition* 61(3):457-465.

Wang, Z. M., W. Shen, R. T. Withers, and S. B. Heymsfield. 2005. Multicomponent molecular-level models of body composition analysis. In *Human body composition*, edited by S. B. Heymsfield, T. G. Lohman, Z. M. Wang, and S. B. Going. Champaign, IL: Human Kinetics Publishers. Pp. 163-176.

Wells, J. C. K., J. E. Williams, S. Chomtho, T. Darch, C. Grijalva-Eternod, K. Kennedy, D. Haroun, C. Wilson, T. J. Cole, and M. S. Fewtrell. 2010. Pediatric reference data for lean tissue properties: Density and hydration from age 5 to 20 y. *American Journal of Clinical Nutrition* 91(3):610-618.

Welon, Z., R. Jurynec, and W. Śliwak. 1988. Optimal body weight in men (ciężar należny mężczyzn) in polish. *Materiały i Prace Antropologiczny* 109:53-71.

WHO (World Health Organization). 1995. *Physical status: The use and interpretation of anthropometry*. Report of a WHO expert committee. Technical Report Series, No. 854. Geneva, Switzerland: WHO.

WHO. 2008. *WHO stepwise approach to surveillance (steps)*. Geneva, Switzerland: WHO.

WHO. 2011. *Waist circumference and waist-hip ratio: Report of a WHO expert consultation, Geneva, 8-11 December 2008*. Geneva, Switzerland: WHO.

Williams, D. P., S. B. Going, T. G. Lohman, D. W. Harsha, S. R. Srinivasan, L. S. Webber, and G. S. Berenson. 1992. Body fatness and risk for elevated blood-pressure, total cholesterol, and serum-lipoprotein ratios in children and adolescents. *American Journal of Public Health* 82(3):358-363.

5

Health-Related Fitness Measures for Youth: Cardiorespiratory Endurance

KEY MESSAGES

Although there is a well-known association between cardiorespiratory endurance and health outcomes in adults, the measurement of cardiorespiratory endurance in youth and of its relationship to health outcomes is relatively new to the literature. The committee's review revealed clear relationships between cardiorespiratory endurance and several health risk factors, including adiposity and cardiometabolic risk factors. Other studies point to a potential relationship between cardiorespiratory endurance and other, less studied risk factors, such as those related to pulmonary function, depression and positive self-concept, and bone health.

Limitations of the studies reviewed by the committee relate mainly to the design of the studies, specifically the lack of analysis of the independent effect of cardiorespiratory endurance on health. A paucity of studies explore the effects of several potential modifiers, such as age, gender, body composition, maturation status, and ethnicity, on performance on the various tests of cardiorespiratory endurance. While such effects have been suggested in the past, the committee could draw no conclusions based on the evidence reviewed.

The cardiorespiratory endurance tests most commonly associated with a positive change in a health marker are the shuttle run and tests conducted with the treadmill and cycle ergometer. Available evidence indicates that these three types of tests demonstrate acceptable validity and reliability. The health markers most frequently assessed are related to body weight or adiposity and cardiometabolic risk factors. Based on its relationship to health, as well as its reliability, validity, and feasibility, a timed or progressive shuttle run, such as the 20-meter shuttle run, is appropriate for measuring cardiorespiratory endurance in youth. If the test is to be administered in a setting with space limitations (e.g., a mobile test center for a national survey), a submaximal treadmill or cycle ergometer test should be used. The shuttle run is advantageous when there are time constraints and when cost may be a problem, such as in schools and other educational settings. Although the evidence for a relationship between distance/timed runs and health is insufficient at this time, this type of test is valid and reliable and could be an alternative in schools and other educational settings.

Until more data are collected with which to establish criterion-referenced cut-points (cutoff scores), interim cut-points corresponding to the lowest 20th percentile of the distribution of cardiorespiratory endurance should be used to interpret results of all cardiorespiratory endurance tests and to determine whether individuals are at risk of negative health outcomes.

Cardiorespiratory endurance has been recognized as a key component of physical fitness throughout the history of the field. This chapter presents the committee's review of the scientific literature that explores the relationship between specific field tests of cardiorespiratory endurance and health outcomes in youth. The committee's recommendations for the selection of fitness tests are based primarily on an extensive review of the literature provided by the Centers for Disease Control and Prevention (CDC) described in Chapter 3. In making its recommendations, the committee considered not only the evidence for a relationship to health, but also the scientific integrity (reliability and validity) of putative health-related tests, as well as the administrative feasibility of implementing these tests. After presenting these results, the chapter offers guidance for setting interim cut-points (cutoff scores) for the selected tests. The final section presents conclusions. Recommendations regarding specific tests for measuring cardiorespiratory endurance for national surveys and in schools and other educational settings are found in Chapters 8 and 9, respectively. Future research needs are addressed in Chapter 10.

DEFINITIONS

Cardiorespiratory endurance is the ability to perform large-muscle, whole-body exercise at moderate to high intensities for extended periods of time (Saltin, 1973). Numerous terms have been used to denote this component of physical fitness, including *aerobic fitness* and *aerobic capacity*. These terms are essentially synonymous with cardiorespiratory endurance, which is the term used in this report. Forms of exercise that depend on cardiorespiratory endurance include vigorous distance running, swimming, and cycling. This fitness component also affects a person's ability to perform, without undue fatigue, less intense, sustained whole-body activities, such as brisk walking, stair climbing, and home chores. People with good levels of cardiorespiratory endurance can perform large-muscle, whole-body exercise at high intensity for at least moderate durations before experiencing fatigue, and they can comfortably perform light- to moderate-intensity exercise for extended periods.

Cardiorespiratory endurance depends on the body's ability to support skeletal muscle activity through high rates of aerobic metabolism. The ability to produce energy at high rates through aerobic metabolism during exercise depends on three physiologic functions: (1) transport of oxygen from the atmosphere to the active muscles through the actions of the cardiorespiratory system, (2) consumption of oxygen in the aerobic metabolic process in the cells of the active muscles, and (3) removal of waste products. People with high levels of cardiorespiratory endurance typically have highly functional cardiorespiratory systems (i.e., heart, lungs, blood, blood vessels), and their skeletal muscles are well adapted to the use of oxygen in aerobic metabolism.

Higher levels of cardiorespiratory endurance have been associated with a wide range of health benefits in adults, including a lower risk of cardiovascular disease (Arraiz et al., 1992; Blair et al., 1989; Sandvik et al., 1993), type 2 diabetes (Colberg et al., 2010), hypertension (Blair et al., 1984), certain cancers (Oliveria et al., 1996), and premature mortality from all causes (Blair et al., 1989, 1993, 1995). The linkage between cardiorespiratory endurance and health in youth is discussed later in the chapter.

CARDIORESPIRATORY ENDURANCE TESTS

The gold standard measure of cardiorespiratory endurance is *maximal aerobic power* (VO_2max)—the greatest rate at which a person is able to consume oxygen during sustained, exhaustive exercise. In the laboratory, VO_2max is typically measured while a person performs maximal, graded exercise on a treadmill or cycle ergometer. VO_2max can be expressed in terms of liters of oxygen consumed per minute (l/min), or the values can be normalized for differences in body size and expressed as milliliters of oxygen consumed per kilogram of body weight per minute (ml/kg/min). VO_2max is known to be a key physiologic determinant of cardiorespiratory endurance and has typically been used as the criterion measure in the validation of field measures of cardiorespiratory endurance. Many field measures of this fitness component have been studied and used in various fitness test batteries around the world (see Table 2-6 in Chapter 2).

The most commonly used field tests involve *distance/timed runs* of varying length and graded-pace *shuttle runs*. Various types of distance/timed runs have been used to measure cardiorespiratory endurance in fitness test batteries since the advent of large-scale fitness testing in the post–World War II era. The tests vary in structure, some being based on a distance limitation in which performance is measured as time required to complete the specified distance (often 1 or 1.5 miles), and others on a time limitation in which performance is measured as the distance covered in the specified time (often 9 or 12 minutes). While runs as short as 600 yards were used in early versions of fitness test batteries, distance runs using the 1 mile or 9-minute format have been most common since the 1970s.

Shuttle runs measure cardiorespiratory endurance when an individual runs to and from two different points, usually around 20 meters apart, at a set pace. The progressive aerobic cardiovascular endurance run (PACER), a variation on the shuttle run, is a maximal cardiorespiratory endurance test in which lines are placed 15 or 20 meters apart, and the participant runs repeatedly between the two lines within prescribed times. The time decreases periodically while the distance remains the same until the participant cannot run fast enough to reach the finish line in the prescribed time.

Alternatively, some fitness surveys use quasi-laboratory tests (i.e., those that measure VO_2max but can be conducted in the field). These tests involve the performance of graded, submaximal exercise on a *treadmill* or *cycle ergometer*.

CARDIORESPIRATORY ENDURANCE AND HEALTH IN YOUTH

Literature Review Process

As noted, the evidence for the committee's recommendations for fitness tests for cardiorespiratory endurance was derived mainly from an extensive review of the literature provided by the CDC, which selected studies measuring the associations between various components of fitness and health. The CDC search strategy and data extraction procedures are described in detail in Chapter 3. For cardiorespiratory endurance, the CDC screened 4,795 studies; of these, only 260 longitudinal, experimental, and quasi-experimental studies satisfied the CDC's search criteria for further consideration. Of this subset, the committee reviewed 47 experimental studies, 24 longitudinal prospective studies, and 29 quasi-experimental studies. In addition to this review, the committee considered the integrity and the feasibility of the tests in a stepwise process, also described in Chapter 3. This section describes the committee's evaluation of the relationship between specific tests of cardiorespiratory endurance and health; the subsequent sections address the integrity and feasibility of the tests.

The committee selected only studies of high quality for review (see Chapter 3 for a list of general selection criteria). Studies for in-depth review were limited to those with designs appropriate to the committee's purpose, that is, only experimental, longitudinal, and quasi-experimental studies. (Cross-sectional studies or experimental studies with no control were excluded.) An additional literature search (utilizing search terms similar to those of the CDC review) was undertaken to cover studies published in 2011. The set of studies was further narrowed on the basis of the following criteria. First, the study provided important evidence linking a particular candidate measure of cardiorespiratory endurance—distance/timed run, shuttle run, treadmill, cycle ergometer—to a positive health outcome, marker, or risk factor in four categories (adiposity, metabolic risk, cognitive, and other). Studies also were categorized as presenting direct or associational evidence. A study was defined as presenting direct evidence when a change in a fitness measure resulted in a positive change in a health risk factor or outcome, and when the study used appropriate controls and statistical methods to analyze the independent effect of the intervention and potential confounders. When making its recommendations, the committee also considered associational evidence (i.e., from studies that did not consider all possible confounders) as it may constitute supporting evidence. In general, studies were excluded based on the following criteria: poor study design (e.g., no control population), inappropriate population (e.g., obese children with complex health issues), lack of power to detect changes (e.g., small sample size), inability to assess the independent effect of a dietary intervention or other important known confounder, or insufficient change in the fitness measure of interest.

The following sections review the strength of the evidence for a relationship between health outcomes and the four categories of fitness tests for cardiorespiratory endurance (distance/timed run, shuttle run, treadmill, and cycle ergometer). The discussion is organized by test because, in contrast with measures for other fitness components (i.e., musculoskeletal fitness and flexibility), the committee found sufficient evidence linking specific cardiorespiratory endurance tests to health markers, particularly cardiometabolic risk factors and body composition. The strength of the evidence is categorized as sufficient or insufficient based on the number of studies linking a measure to a particular category of health markers, the study designs (evidence from experimental and longitudinal studies having more weight than that from quasi-experimental studies), and the statistical significance of the association. The selected longitudinal, experimental, and quasi-experimental studies are summarized in Tables 5-1, 5-2, and 5-3, respectively. For each study, the tables include (1) the fitness test(s) used, (2) the health outcomes/markers examined, (3) the size and characteristics of the sample, and (4) a summary of the results and the quality and level of the evidence.

TABLE 5-1 Longitudinal Studies

		Health Outcome(s)/Marker(s)					
Reference	**Fitness Test(s)**	**Body Composition**	**Metabolic Health**	**Mental and Cognitive Function**	**Other**	**Sample Size and Characteristics**	**Study Summary, Quality and Level of Evidence**
Eisenmann et al., 2005	Treadmill (Balke, maximal)—maximal oxygen consumption (VO_2max)	Body mass index (BMI)/% fat/waist circumference (WC)	No relationship with blood pressure (BP), total cholesterol (TC), high-density lipoprotein (HDL), tryglycerides (TG), glucose			N = 48 Ages 15 and 26, male (M) and female (F)	The Aerobics Center Longitudinal Study (ACLS) tested subjects during adolescence and young adulthood. Statistics equal only partial correl ations; did not test for interactions. Adolescent treadmill time (TT) associated with lower adiposity as adults. No association with risk for cardiovascular disease (CVD). Level of evidence (LE): Associational
Johnson et al., 2000	Treadmill (walking)—VO_2max	Fat mass (FM) by dual-energy X-ray absorptiometry (DXA)				N = 115 Ages 4.5-11, M and F, white and African American	Progressive walking treadmill protocol. Assessed annually for 3-5 years post. Did not examine change in VO_2max relative to change in adiposity. Adjustments for Tanner, ethnicity, and gender. Relationship between baseline VO_2max and rate of increasing adiposity. Rate of increase of adiposity was lower for those who were fit at baseline. No relationship between ethnicity and VO_2max. LE: Associational

Byrd-Williams et al., 2008	Treadmill (maximal)—VO_2max	DXA			N = 160 Ages 8-13, M and F, Hispanic, overweight/obese	VO_2max at baseline and DXA and Tanner examined annually for up to 4 years. Linear mixed models measured gender-specific relationships between VO_2max and increases in adiposity. Adjustment for Tanner. Higher baseline fitness in overweight Hispanic boys was protective against increased adiposity. LE: Associational
Twisk et al., 2000	Treadmill (maximal)—VO_2 max	Skinfolds	TC/HDL, TC; no relationship with BP		N = 181 Ages 13-27, M and F, Amsterdam	Six repeated measurements from ages 13 to 27. Longitudinal linear regression for relationship between fitness and CVD risk factors, adjusting for time, gender, age, diet, and other lifestyle factors. Fitness inversely associated with TC/HDL, skinfolds, TC. Skinfolds mediated the relationship between fitness and CVD risk; no effect on BP. LE: Direct
Janz et al., 2002	Cycle ergometer (maximal)—VO_2max	Skinfolds/WC	TC/HDL, low-density lipoprotein (LDL)		N = 125 Ages 6-15, M and F	Examined whether fitness during first 4 years predicts CVD risk at 5 years, adjusted for age, gender, and maturation. Multiple linear

continued

TABLE 5-1 Continued

Reference	Fitness Test(s)	Health Outcome(s)/Marker(s): Body Composition	Metabolic Health	Mental and Cognitive Function	Other	Sample Size and Characteristics	Study Summary, Quality and Level of Evidence
							regression to examine change in fitness and CVD risk, adjusted for changes in fat-free mass (FFM) and maturation. Change in fitness associated with TC/HDL, LDL, central and total adiposity 4 years later. LE: Direct
McMurray et al., 2008	Cycle ergometer (submaximal)	Skinfolds/body mass index (BMI)	Metabolic syndrome			N = 389 Ages 7-10 to 14-17, M and F	Assessed baseline fitness and 7-year follow-up incidence of metabolic syndrome/logistic regression. Baseline fitness associated with incidence of metabolic syndrome 7 years later. LE: Direct
Ortega et al., 2011	Cycle ergometer (maximal)	BMI				N = 598 Ages 9.5 to 15, M and F, Estonian and Swedish	Normal-weight subjects at 9 years followed for 6 years. Examined change in fitness and incident overweight. Binary logistic regression, adjusted for sex, age, country, maturation.

				Improvements in fitness from childhood to adolescence associated with lower risk of overweight/obesity in adolescence. LE: Direct
Aires et al., 2010	Shuttle run (progressive aerobic cardiovascular endurance run [PACER])	BMI	N = 345 Ages 11-16 and 14-19, M and F	Fitnessgram® shuttle run grouped into healthy zone or under healthy zone. Linear mixed effects modeling to examine fitness and change in BMI. Unclear whether baseline fitness or change in fitness was used. No adjustment made for BMI, maturation. Under healthy zone predicted increase in BMI over 3 years. LE: Associational
Chen et al., 2007	Shuttle run (PACER)	BMI	N = 307 Ages 7-8, M and F, Chinese	Examined which factors were associated with changes in BMI over 1 year. Stepwise regression to examine predictors (fitness, mother's weight status, screen time, BMI at baseline) of BMI at 1-year follow-up. Cardiorespiratory endurance was independently associated with BMI after 1 year. LE: Direct

continued

TABLE 5-1 Continued

		Health Outcome(s)/Marker(s)					
Reference	Fitness Test(s)	Body Composition	Metabolic Health	Mental and Cognitive Function	Other	Sample Size and Characteristics	Study Summary, Quality and Level of Evidence
Kim et al., 2005	Shuttle run (20 m/6 min)	BMI				N = 2,927 Ages 5-14, M and F	Examined relationship between cardiorespiratory endurance and incident overweight over 1 year in schoolchildren. Multivariate logistic regression models. Adjusted for sociodemographics. Not passing cardiorespiratory endurance test associated with incident overweight. LE: Direct
Martins et al., 2009	Shuttle run (PACER)	BMI	BP, TC			N = 153 Ages 8-9, M and F, Portugal	Study examined the association between VO_2 max at baseline and changes in CVD risk over 5 years. Multilevel modeling to examine effect of fitness over time. Maturation measured; only age, gender adjusted for. No association between cardiorespiratory endurance and BP or TC. Lower level of cardiorespiratory endurance associated with higher BMI over 5 years. LE: Associational

Puder et al., 2011	Shuttle run (PACER)	BMI/WC, skinfolds	Homeostatic model assessment-insulin resistance (HOMA-IR), C-reactive protein (CRP)	N = 83 First and fifth graders, M and F	Association of adiposity and fitness with changes in inflammation and insulin resistance. Multiple linear regression, adjusted for age, gender, and puberty. No association with CRP. Low baseline fitness associated with increases in HOMA-IR. LE: Associational
Mota et al., 2009	9-min run	BMI		N = 135 Ages 6-12, M and F, Portugal	Baseline BMI and fitness with changes in BMI over 2 years; logistic regression; did not specify controlling for covariates. Baseline fitness associated with change in BMI over 2 years. Unfit children at baseline were 3.9 times more likely to be BMI gainers; however, a change in CRP was not associated with a change in BMI over time. LE: Associational

TABLE 5-2 Experimental Studies

		Health Outcome(s)/Marker(s)					
Reference	**Fitness Test(s)**	**Body Composition**	**Metabolic Health**	**Mental and Cognitive Function**	**Other**	**Sample Size and Characteristics**	**Study Summary, Quality and Level of Evidence**
Barbeau et al., 2002	Treadmill (submaximal test, but used a protocol with metabolic cart)—maximal oxygen consumption (VO_2max, heart rate 170)	Body fat (BF) (% BF, visceral adiposity)	Hemostatic factors			N = 74 Ages 12-16, male (M) and female (F), obese, African American and white	Lifestyle plus physical training intervention—students to study the effect on studied hemostatic markers and body composition. Conducted analysis for interactions between gender and ethnicity. Improvements in fitness and adiposity, but effect on hemostatic markers was null. Level of evidence (LE): Associational
Carrel et al., 2005	Treadmill (submaximal walk)	Body mass index (BMI)	Fasting insulin, glucose			N = 50 Mean age = 12.5, M and F, obese	A 9-month school-based fitness (gym) intervention to study the effect of a fitness program on body composition, fitness, and insulin sensitivity. No analysis was conducted on independent effects of adiposity and fitness on insulin sensitivity.

						Intervention resulted in improvements in BMI, fitness, and fasting insulin level. LE: Associational
Davis et al., 1985	Treadmill (exhaustion) (submaximal but used a protocol with metabolic cart)—modified cycle ergometer	BMI		Executive function	N = 94 Ages 7-11, M and F, overweight	A 15-week intervention to study the effect of aerobic exercise on executive function. Higher-dose training group improved over control in executive function and treadmill time, but no changes in BMI. LE: Associational
Farpour-Lambert et al., 2009	Treadmill—multistage (maximal)	BMI, abdominal fat	Blood pressure (BP), markers of atherosclerosis		N = 44 Ages 6-11, M and F, prepubertal, obese	Crossover design to study the effect of physical activity on health markers in obese youth. Improvements in fitness corresponded with decreased BMIs and abdominal fat and improvement in blood pressure and arterial function parameters. LE: Associational

continued

TABLE 5-2 Continued

Reference	Fitness Test(s)	Health Outcome(s)/Marker(s): Body Composition	Metabolic Health	Mental and Cognitive Function	Other	Sample Size and Characteristics	Study Summary, Quality and Level of Evidence
Walther et al., 2009	Treadmill—modified Bruce protocol	BMI, fat-free mass, total body water content	Number of endothelial progenitor cells, high-density lipoprotein (HDL)			N = 182 Mean age 11.1, M and F, normal distribution of weights	1–school year study to examine the effect of physical activity on health markers. No adjustment for puberty or other variables. Increase in endothelial progenitor cells with increase in fitness, but no change in BMI, fat-free mass, total body water content, or HDL. LE: Direct
Davis et al., 1985	Treadmill (maximal)—VO_2max	BF				N = 9 pairs of twins Ages 11-14, M, pubertal and prepubertal twins	Examined effect of 6-month training on body composition and fitness. Training resulted in an increase in VO_2 max in prepubertal but not in pubertal youth. Training resulted in a decrease in percent body fat, but VO_2max/kg body mass was higher than at baseline in both the control and the trained groups. LE: Direct

Murphy et al., 2006	Shuttle run—Loughborough Multistage Fitness Test 20-meter shuttles			Bone ultrasound	N = 90 Ages 16-17, F, obese	Examined effect of physical activity intervention on improving health risks. Randomized by schools. Self-taught and teacher-taught physical activity (PA) groups increased in fitness and bone health compared with controls. No increase in BMI was observed in the intervention group, while there was an increase in the control group. LE: Direct
Reed et al., 2008	Shuttle run—Legers 20-meter, incremental	BMI	BP Total cholesterol (TC), HDL, low-density lipoprotein (LDL), C-reactive protein (CRP), fibrinogen		N = 298 Ages 11-15, M and F, grades 4 and 5	Intervention randomized by schools. Maturity measured, but its effect was not analyzed. Intervention did not result in effects on cardiovascular disease (CVD) risk factors because of a smaller subsample size (N = 77) used for those risk factors. No change in BMI after intervention, but there was an improvement in fitness and BP. LE: Direct

continued

TABLE 5-2 Continued

Reference	Fitness Test(s)	Health Outcome(s)/Marker(s): Body Composition	Metabolic Health	Mental and Cognitive Function	Other	Sample Size and Characteristics	Study Summary, Quality and Level of Evidence
Ben Ounis et al., 2010	Cycle ergometer (submaximal but used breath by breath)—Brandon protocol	BMI, % BF	Glucose, tryglycerides (TG), HDL, BP, metabolic syndrome			N = 32 Ages 12-14, M and F, obese	An 8-week intervention to study the effect of exercise training on metabolic syndrome. No analysis of potential interactions or independent effects of fitness on health outcomes. Fitness and adiposity improved, as well as blood glucose, TG, HDL, and BP. LE: Associational
Kelly et al., 2004	Cycle ergometer—VO_2 peak, graded test	BMI	HDL, insulin sensitivity, endothelial function (vascular stiffness)			N = 20 Mean age 10.9, M and F, overweight	An 8-week intervention to study the effect of exercise on CVD risk factors. No adjustment for maturity. Fitness improved, along with HDL and endothelial function. Adiposity and fasting insulin did not change. LE: Direct

Lee et al., 2010	Cycle ergometer—VO_2max, Astrand protocol		Insulin resistance, visfatin			N = 38 Mean age 16.7, F, obese and normal	A 12-week to study the effect of a physical activity program on insulin and receptors. No independent analysis of the effects of exercise and diet on improvements in risk factors. Fitness improved, along with a decrease in plasma visfatin and insulin resistance. LE: Associational
Nourry et al., 2005	Cycle ergometer—incremental exercise/VO_2max	Body mass, fat mass (FM)			Pulmonary function	N = 24 Age 10, M and F, nonathletic	Examined effect of 8-week training program on resting and exercising lung function. Intervention led to enhanced resting pulmonary function and VO_2 max, but both control and intervention groups had an increased in body mass and no change in FM. LE: Direct

continued

TABLE 5-2 Continued

		Health Outcome(s)/Marker(s)					
Reference	**Fitness Test(s)**	**Body Composition**	**Metabolic Health**	**Mental and Cognitive Function**	**Other**	**Sample Size and Characteristics**	**Study Summary, Quality and Level of Evidence**
Chang et al., 2008	Distance run—50 meters and 800 meters (girls), 1,000 meters (boys)		TG, HDL, glucose, insulin			N = 49 Ages 12-14, M and F, obese	A 9-month exercise intervention in school to study the effect of exercise on health markers. No analysis of independent effects of fitness and BMI changes on CVD risk factors. No measure of puberty even though author recognized that testing is done during pubertal years. Intervention improved fitness, promoted weight loss and adiposity loss, decreased TG levels, maintained HDL, and decreased insulin levels. LE: Associational

Wong et al., 2008	Cycle ergometer—heart rate following submaximal test on cycle ergometer, Physical Working Capacity-170 (PWC-170) test	Lean body mass, BMI	Heart rate, systolic BP, TG, CRP		N = 24 Ages 13-14, M, obese	A 12-week physical education intervention to study the effects of the exercise on obesity and associated health risks. No analysis of independent effects of fitness. Compared with the control group, the intervention resulted in improvements in fitness; body composition; and CVD risk factors, such as heart rate, systolic BP, and TG levels. No effect on CRP, which was elevated at baseline. LE: Associational

TABLE 5-3 Quasi-experimental Studies

		Health Outcome(s)/Marker(s)					
Reference	Fitness Test(s)	Body Composition	Metabolic Health	Mental and Cognitive Function	Other	Sample Size and Characteristics	Study Summary, Quality and Level of Evidence
Gately et al., 2005	Treadmill walk (submaximal)	Body mass, % body fat (BF), body mass index (BMI)			Systolic and diastolic blood pressure (BP)	N = 185 Ages 9-14 and 15-18, male (M) and female (F), 14% overweight and 86% obese, English	A prospective cohort study to evaluate the effects of a physical activity and dietary intervention after an average of 29 days in the program. Improved BMI, BMI standard deviation (SD) score, % BF, fat mass (FM), waist circumference (WC) in overweight campers in intervention group vs. controls. Systolic and diastolic BP improved in overweight intervention groups vs. overweight and normal controls. Level of evidence (LE): Associational

Jekal et al., 2009	Cycle ergometer	Body mass, BMI, WC, body fatness, FM	High-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol, triglycerides, insulin resistance		Systolic and diastolic BP	N = 31 Ages 14-19, M, obese, fit (maximal oxygen consumption [VO_2max], 48 ml/min/kg), Korean	A 12-week exercise program with statistical analysis of cardiorespiratory endurance, fatness, and cardiovascular risk factors. Intervention decreased body weight, FM, % BF, and WC and increased fitness. There were also reductions in diastolic and systolic BP, total cholesterol, LDL, insulin resistance and cardiovascular disease (CVD) risk score and increases in HDL. There were interactions between fitness and fatness and insulin resistance and CVD risk factors. LE: Direct

continued

TABLE 5-3 Continued

		Health Outcome(s)/Marker(s)					
Reference	**Fitness Test(s)**	**Body Composition**	**Metabolic Health**	**Mental and Cognitive Function**	**Other**	**Sample Size and Characteristics**	**Study Summary, Quality and Level of Evidence**
Stella et al., 2005	Cycle ergometer	BMI, body mass, FM			Depression (Beck Depression Inventory) and anxiety (Trait State Anxiety Inventory)	N = 40 Ages 14-19, F, obese, Brazilian	Groups were aerobic (three ergonomic sessions/week, increasing from 40 to 60 min in third month), anaerobic training (same program but with load), leisure (60 min of physical activity/week), and control. Statistical analyses performed for intragroup variations and group comparison. Body mass, BMI, and FM were reduced in all groups; fat-free mass (FFM) increased in anaerobic training group. There were no differences in BMI, body mass, and FM among groups. Cardiorespiratory variables (VO_2peak, oxygen consumption at VO_2threshold, and anaerobic ventilator

					threshold load) increased in all groups except control. Depression scores were reduced only in aerobic group. No differences in anxiety scores except for control group. LE: Direct
Sidiropoulou et al., 2007	6-minute run		Pulmonary function	N = 29 Ages 11-15, M, Greek	Subjects had history of exercise-induced asthma. Intervention was soccer training for 8 weeks with a defined and controlled program based on maximal and resting heart rate. No statistical analysis conducted to evaluate association between fitness and pulmonary function. Both pulmonary function and distance covered in running test improved significantly in intervention group. LE: Associational

Review of the Scientific Literature

Distance/Timed Run Tests

The committee identified two experimental, one longitudinal, and one quasi-experimental quality studies that utilized various distance or timed runs to measure cardiorespiratory endurance and its association with health risk factors. There were fewer studies in this category of fitness tests than in the other categories. The distances and times varied, and the tests were primarily school based (Chang et al., 2008; Mota et al., 2009; Sidiropoulou et al., 2007).

The studies reviewed also varied in their purposes, and none were designed specifically to answer the questions the committee was tasked to address. One high-quality experimental study examined the associations between an intervention and a change in fitness (as measured by the specified distance run), adiposity, or cardiometabolic risk factors (Chang et al., 2008). The authors implemented school-based interventions and found increases in physical fitness, along with decreases in adiposity and improvements in cardiometabolic risk factors. This study did not examine in more depth whether the improvements in cardiometabolic risk factors were independently associated with improvements in fitness and/or adiposity.

One longitudinal study found associations between baseline fitness and gain in body mass index (BMI) over time, but did not find an association between changes in fitness and changes in BMI over time (Mota et al., 2009). Other studies included only a special population (Sidiropopulou et al., 2007).

Distance run protocols were used in children aged 7 and older, although the majority of studies were of youth aged 10 and older. The influence of gender on the relationship between fitness and health risk factors was not determined. Only one study specifically examined obese children (Chang et al., 2008). Although running is a weight-bearing activity, these studies did not adjust for BMI, which often was the major health outcome of interest (Chang et al., 2008; Mota et al., 2009).

As with the other categories of tests discussed below, these studies overall found that results of distance/timed run tests used to measure fitness in youth corresponded to health risk factors, especially body fatness and cardiometabolic risk, in youth.

Shuttle Run Tests

The committee identified three experimental, six longitudinal, and one quasi-experimental studies that used various shuttle run testing protocols to measure cardiorespiratory endurance and health outcomes. The protocols included the 6-minute, 20-meter shuttle run and various 20-meter multistage incremental tests. The majority of associations were found with adiposity (as measured by BMI), perhaps because many of the studies took place in a school setting, where it may be difficult to take more invasive measures of health risk factors, such as blood lipids (Aires et al., 2010; Chen et al., 2007; Kim et al., 2005; Martins et al., 2009).

One high-quality experimental study showed improvements in cardiorespiratory endurance in schools randomized to a comprehensive physical activity intervention (Reed et al., 2008). These students also showed improvements in blood pressure; however, there were no detectable changes in BMI. Blood cardiometabolic risk factors also were measured in a smaller subset of children, but changes in these risk factors were not significant. Similarly, Murphy and colleagues (2006) found that improvements in cardiorespiratory endurance corresponded to improvements in bone ultrasound results independent of any changes in BMI after a 6-month physical activity intervention.

Longitudinal studies by Martins and colleagues (2009) and Kim and colleagues (2005) demonstrated an inverse relationship between baseline cardiorespiratory endurance and increasing BMI and incidence of overweight over 1 year and 5 years, respectively. A few studies also found that improved performance on the shuttle run resulted in improvements in cardiometabolic risk factors (total cholesterol, low-density lipoprotein [LDL], high-density lipoprotein [HDL], and triglycerides) and blood pressure (Puder et al., 2011; Reed et al., 2008). While Puder and colleagues (2011) found that low baseline fitness was associated with an increase in homeostasic model assessment-insulin resistance (HOMA-IR) level over 1 year, changes in fitness over this time and the corresponding relationship to HOMA-IR were not examined and likely would have been difficult to detect with the study's small sample size (N = 83). In general, there were relatively few studies measuring cardiometabolic risk factors, and the relationships to health markers were not as strongly supported by the study design (e.g., the nonoptimal statistical approach of Martins et al. [2009]).

The shuttle run often is used as a measure of cardiorespiratory endurance in the school setting because it requires no special equipment, and training the administrators of the test is relatively easy. As the shuttle run is typically school based, studies of this type of test examined children aged 4-17. Many studies examined results by gender, but only Martins and colleagues (2009) specifically state that gender did not impact the relationship between fitness and BMI. Puberty was self-assessed in two studies (Martins et al., 2009; Puder et al., 2011), only one of which controlled for this factor (Puder et al., 2011) (an ommision that is understandable given that collecting data on pubertal status may be difficult in a school-based environment, where the majority of shuttle run testing takes place). Most of these studies also were inclusive of all weight categories and races/ethnicities and did not specifically examine the potential impact of these factors on the relationship between fitness and health risk.

Two quality experimental trials demonstrated improvements in fitness and corresponding improvements in bone health (Murphy et al., 2006) and blood pressure (Reed et al., 2008). Several quality studies also examined the longitudinal relationship between fitness as measured by the shuttle run and changes in adiposity in schoolchildren over time. Overall, the strongest established relationships between cardiorespiratory endurance as assessed by the shuttle run and health markers were with adiposity as measured by BMI.

Treadmill Tests

The committee identified six experimental, four longitudinal, and one quasi-experimental quality studies that used various treadmill protocols (both maximal and submaximal) to measure VO_2max and associated health risk factors. Longitudinal studies demonstrated a strong link between changes in cardiorespiratory endurance as measured by treadmill testing and changes in adiposity measures such as BMI, waist circumference, and adiposity (as measured by skinfold and dual-energy X-ray absorptiometry [DXA]) (Byrd-Williams et al., 2008; Eisenmann et al., 2005; Johnson et al., 2000; Twisk et al., 2000).

A strength of the longitudinal studies examined is that fitness levels in adolescence are related to body composition and total cholesterol levels in young adulthood (Eisenmann et al., 2005; Twisk et al., 2000). Twisk and colleagues (2000) showed that an increase in VO_2max over this period was positively associated with a healthy cardiovascular risk profile in adulthood. Controlling for body fatness, the authors demonstrated that fitness had an independent effect on the ratio of total cholesterol to HDL. Another longitudinal study spanning adolescence to adulthood (mean age 26) found that time to reach exhaustion on the treadmill in adolescents and the change in this time experienced from adolescence into adulthood were correlated with both adult body fatness and changes in body fatness from adolescence to adulthood, respectively (Eisenmann et al., 2005). However, this study did not demonstrate this positive relationship with adult risk factors for cardiovascular disease and may have been limited by its relatively small sample size (N = 48) (Eisenmann et al., 2005). Other longitudinal studies used simply baseline VO_2max in relationship to changes in adiposity (Byrd-Williams et al., 2008; Johnson et al., 2000) without examining the change in VO_2max over this period. Still, these studies highlight the important relationship between cardiorespiratory endurance and body fatness at a later time.

Experimental studies conducted in schools and laboratories also demonstrated positive relationships between changes in cardiorespiratory endurance and adiposity measures, as well as cardiometabolic risk, blood pressure, and executive functioning (Barbeau et al., 2002; Carrel et al., 2005; Davis et al., 1985; Farpour-Lambert et al., 2009; Gately et al., 2005). Several studies used well-controlled maximal treadmill test protocols (Farpour-Lambert et al., 2009; Walther et al., 2009), while others utilized submaximal protocols that involved use of a metabolic cart for measurement of gas exchange (Barbeau et al., 2002; Carrel et al., 2005; Davis et al., 1985).

Cardiorespiratory endurance and body fatness are highly interrelated, and both are factors in the risk for cardiovascular disease. Therefore, when examining associations of changes in cardiorespiratory endurance with body fatness and cardiometabolic risk factors, it is difficult to determine the independent causes and effects. Few of the well-controlled studies using the treadmill examined these effects independently (Twisk et al., 2000). After a 1-year school-based randomized physical activity intervention, Walther and colleagues (2009) demonstrated an increase in VO_2max and endothelial progenitor cells in youth without a corresponding decrease in adiposity. Even though fat mass has an inverse relationship to weight-bearing cardiorespiratory endurance measures, several studies did specifically examine overweight and obese youth using a treadmill protocol and observed measurable changes in cardiorespiratory endurance (Barbeau et al., 2002; Byrd-Williams et al., 2008; Carrel et al., 2005; Davis et al., 1985; Farpour-Lambert et al., 2009).

Some studies also adjusted for maturation stage (Byrd-Williams et al., 2008; Johnson et al., 2000), and some examined the interaction between specific ethnicities and the relationship of fitness to health markers (Byrd-Williams et al., 2008; Johnson et al., 2000). Although Johnson and colleagues did not find an effect of ethnicity, an interaction between gender and the relationship of baseline fitness to changes in adiposity was observed in Hispanic boys, but not Hispanic girls (Byrd-Williams et al., 2008).

Overall, the studies reviewed indicate that, among the health markers measured, adiposity and cardiometabolic risk factors show the strongest evidence for an association with treadmill-measured cardiorespiratory endurance.

Cycle Ergometry Tests

The committee identified four experimental, three longitudinal, and three quasi-experimental quality studies that utilized various cycle ergometry tests to measure cardiorespiratory endurance and health outcomes. The protocols in these studies varied widely, with the majority utilizing maximal exercise tests. Among those studies, several, both longitudinal and experimental, were of high quality (Ben Ounis et al., 2010; Janz et al., 2002; Kelly et al., 2004; Ortega et al., 2011). The weight of the evidence for an association between cardiorespiratory endurance as measured by cycle ergometry and health risk factors was particularly strong for measures of adiposity (BMI, waist circumference, percent body fat, fat mass), cardiometabolic risk factors (including total cholesterol, HDL, LDL, triglycerides, insulin resistance, glucose, and vascular stiffness), and blood pressure (Ben Ounis et al., 2010; Dunton et al., 2007; Janz et al., 2002; Kelly et al., 2004; Lee et al., 2010; McMurray et al., 2008; Ortega, et al., 2011; Stella et al., 2005).

The strongest evidence for a relationship between cycle ergometry test results and health risk factors is found with measures of adiposity and cardiometabolic risk. The strength of the evidence in cycle ergometry studies appears to be similar to that in the treadmill studies discussed above. Cycle ergometry studies include several well-executed experimental studies (Ben Ounis et al., 2010; Kelly et al., 2004; Wong et al., 2008) and two longitudinal studies that occurred over a period of 4-6 years (Janz et al., 2002; Ortega et al., 2011). Kelly and colleagues (2004) found improvements in cardiorespiratory endurance, HDL cholesterol, and endothelial function following exercise training in overweight youth without changes in adiposity, demonstrating the independent effects of changes in fitness on these cardiovascular disease markers (Kelly et al., 2004). Others found positive effects of exercise and improvements in cardiorespiratory endurance on cardiovascular risk markers and adiposity, but were unable to determine whether these were independent effects (Ben Ounis et al., 2010; Wong et al., 2008).

In longitudinal studies, cardiorespiratory endurance improvements over 6 years from childhood to adolescence were associated with a lower risk of becoming overweight/obese during adolescence (Ortega et al., 2011). Adjustment was made for confounding by baseline BMI, and no interactions by gender were identified. Similarly, Janz and colleagues (2002) demonstrated the relationship between changes in cardiorespiratory endurance over 5 years and health markers at the 5-year point. The authors demonstrated a relationship between changes in cardiorespiratory endurance and ratio of total cholesterol to HDL, LDL, and adiposity measures.

Jekal and colleagues (2009) designed their quasi-experimental study to evaluate the effect of a 12-week exercise program (Jekal et al., 2009). Although the study did not include a control group, measurements before and after the intervention in this small study demonstrated a significant association of cardiorespiratory endurance with fatness and risk factors for cardiovascular disease. Another study examined the effect of 12 weeks of aerobic training on plasma visfatin and insulin resistance in normal-weight and obese female adolescents; unfortunately, analyses were not conducted to elucidate whether the improvement in these risk factors was due to weight loss or improvements in fitness or both (Lee et al., 2010). The results suggest that the interaction between weight and cardiorespiratory endurance is important, even though the authors did not analyze the independent contributions of each of these variables to insulin resistance.

Two studies identified other health risk factors—depression and positive self-concept—that were mitigated by increased fitness (Dunton et al., 2007; Stella et al., 2005). Of interest is the fact that the population of Dunton and colleagues (2007) included various ethnic groups, even though there was no analysis of ethnic origin as a potential modifier of the relationship between performance on the test and self-concept.

As with the treadmill tests, few cycle ergometry studies evaluated interactions with modifiers such as age or gender. Studies utilizing cycle ergometry have focused mainly on children aged 10 and older, with one longitudinal study examining those aged 7-10 (McMurray et al., 2008). Given the non-weight-bearing nature of cycle ergometry tests, body weight is not a modifying factor for these tests; a number of cycle ergometry tests were conducted with overweight/obese children (Ben Ounis et al., 2010; Kelly et al., 2004; Lee et al., 2010; Stella et al., 2005).

In sum, the strength of the evidence from the use of cycle ergometry to measure cardiorespiratory endurance and associated health risk factors lies with adiposity, cardiometabolic risk factors, and blood pressure.

Limitations of the Scientific Literature

Among the four types of cardiorespiratory endurance tests evaluated in these studies, all but the distance/timed run tests showed significant relationships to health risk factors, specifically adiposity measures and cardiometabolic risk factors. The studies considered of highest quality for each of the tests were as follows:

- cycle ergometer (Janz et al., 2002; Kelly et al., 2004; Ortega et al., 2011);
- treadmill (Byrd-Williams et al., 2008; Davis et al., 1985; Farpour-Lambert et al., 2009; Johnson et al., 2000; Walther et al., 2009); and
- shuttle run (Kim et al., 2005; Puder et al., 2011; Reed et al., 2008).

The most important limitation of other studies reviewed was the lack of in-depth examination of confounders, specifically whether improvements in cardiometabolic risk were independently associated with improvements in cardiorespiratory endurance or were also due to a decrease in adiposity that is often experienced when individuals participate in physical activity. In several of the experimental studies, concurrent changes were seen in adiposity measures, cardiometabolic risk factors, and fitness (Barbeau et al., 2002; Ben Ounis et al., 2010; Carrel et al., 2005; Farpour-Lambert et al., 2009; Wong et al., 2009). In such cases, it is impossible to determine the independent effects of improvements in cardiorespiratory endurance on cardiometabolic risk factors.

The committee also considered whether studies examined the effects of various modifiers (e.g., age, gender) on the relationship between cardiorespiratory endurance and health. Many of the studies reviewed were of good quality but were not designed with these questions in mind. Only a subset of the studies specifically examined potential modifiers of the relationship between cardiorespiratory endurance and health risk factors. For example, studies included potential differences by gender (Barbeau et al., 2002; Johnson et al., 2000; Martins et al., 2009; Puder et al., 2011; Twisk et al., 2000), race/ethnicity (Barbeau et al., 2002; Byrd-Williams et al., 2008; Johnson et al., 2000; Tremblay and Lloyd, 2010), age/maturation stage (Byrd-Williams et al., 2008; Johnson et al., 2000; Puder et al., 2011), weight status (Twisk et al., 2000), and training status (Sidiropoulou et al., 2007). In most cases, however, studies included no analysis of these factors as modifiers of performance or as modifiers of the effect of performance on health.

In general, the committee considered experimental studies to be of higher quality because, by their design, such studies can demonstrate causes and effects in a more direct manner than is possible with other designs. Yet, it should be noted that many of the experimental studies measured health risk factors using methods requiring invasive (i.e., blood draws) and/or precise (e.g., body composition by DXA) techniques (Barbeau et al., 2002; Ben Ounis et al., 2010; Farpour-Lambert et al., 2009; Slaughter et al., 1988; Walther et al., 2009). Likewise, many studies measured fitness with laboratory protocols using either a treadmill or cycle ergometer and with small sample sizes (Ben Ounis et al., 2010; Farpour-Lambert et al., 2009; Kelly et al., 2004; Nourry et al., 2005). An experimental study using precision measurements would be more likely to detect relationships between changes in fitness and changes in adiposity and health risk factors, even with small sample sizes.

VALIDITY AND RELIABILITY OF SELECTED TEST ITEMS

As discussed above, a number of tests have been used to measure cardiorespiratory endurance in studies linking this component of physical fitness to indicators of health in youth. The most common tests used in large-scale surveys and youth fitness test batteries are distance/timed runs, shuttle runs to volitional fatigue, and graded-exercise heart rate extrapolation tests (treadmill or cycle ergometer). The validity and reliability of these tests have been studied extensively, and several authors have summarized the literature on their psychometric properties.

In reviewing this literature, the committee relied primarily on review articles identified through searches using PubMed and Web of Science with the following keywords: *fitness assessment, fitness testing, validity, reliability, children,* and *adolescents*. As necessary, the committee reviewed original research articles identified as above or from citations listed in review articles for the three categories of tests for which the committee found the strongest evidence for a relationship to health—the shuttle run as well as the treadmill and cycle ergometer (i.e., heart rate extrapolation) tests. The distance run also is reviewed here since it could be used as an alternative in schools and other educational settings, even though the literature on this type of test is sparser.

Shuttle Run Tests

Several permutations of graded-intensity shuttle runs to volitional fatigue have been used in youth fitness test batteries. The most common is the 20-meter shuttle run as developed by Léger, and it is this version of the shuttle run that has been examined most frequently in validity/reliability studies (Léger et al., 1988). Performance on the test is scored as 20-meter laps completed before the participant falls behind the pace set by an auditory timer, and validity has been examined as the correlation between laps completed and measured VO_2max. In Léger's developmental study, the correlation between laps completed and VO_2max was r = 0.71 in a group of 8- to 19-year-olds (Léger et al., 1988). Boreham and colleagues (1990) and Liu and colleagues (1992) completed similar validation studies and reported validity coefficients of r = 0.87 and r = 0.72, respectively. Clearly there is strong and consistent evidence that performance on the shuttle run in young people correlates highly with weight-relative VO_2max. This test also has been shown to be a highly reliable measure. In a recent review article, Artero and colleagues (2011) report that test-retest reliability coefficients for this test have ranged from r = 0.78 to r = 0.93. Overall, the available evidence suggests that the 20-meter shuttle run has excellent validity and reliability as a measure of cardiorespiratory endurance.

Treadmill and Cycle Ergometer (Heart Rate Extrapolation) Tests

It is well known that heart rate increases linearly with increasing intensity of endurance exercise; maximal heart rate and VO_2max tend to occur at the same exercise intensity, and therefore power output (e.g., exercise intensity) at a standard heart rate correlates highly with power output and VO_2 at maximal exercise. These relationships are the basis for tests of cardiorespiratory endurance that involve the performance of graded, submaximal exercise with heart rate monitoring. Perhaps the best known and most widely used of such tests is the Physical Working Capacity-170 (PWC-170) test (Wahlund, 1948). This test is performed on a cycle ergometer at three progressively increasing intensities. Performance on the test is quantified as power output at a heart rate of 170 beats per minute as estimated from the linear plot of heart rate versus power output. Similar treadmill tests based on the same principles have been developed (Gutin et al., 1990; National Center for Health Statistics, 2004). Performance on the PWC-170 test has been validated against VO_2max as a criterion measure. Rowland and colleagues (1993) found moderate correlations between absolute VO_2max and performance on the PWC-170 in boys and girls (r = 0.70 and 0.71, respectively), but relationships were weaker when VO_2 was expressed relative to body weight (r = 0.65 and 0.48, respectively). Boreham and colleagues (1990) reported a high correlation (r = 0.84) between performance on the PWC-170 and VO_2max in 48 adolescent boys and girls. Of interest, in the same study, Boreham and colleagues (1990) found that performance on the PWC-170 and 20-meter shuttle run was highly correlated (r = 0.89). The PWC-170 is highly reliable, with test-retest correlation coefficients ranging from 0.89 to 0.98 (Watkins and Ewing, 1983; Watson and Odonovan, 1976).

Distance/Timed Run Tests

The validity of distance/timed runs typically has been established by examining the correlation between a criterion measure—directly measured VO_2max (ml/kg/min) as determined during exhaustive treadmill running—and test performance (distance or time). The reviewers of this literature have consistently concluded that distance runs of 1 mile or greater demonstrate acceptable validity versus VO_2max. As noted by Safrit (1990) and Freedson and colleagues (2000), correlations between VO_2max and performance on distance/timed runs typically have been observed in the good to high range (r = –0.63 to –0.90; a negative correlation has been seen between time to complete and VO_2). Also, distance/timed runs have been found to be reliable based on test-retest correlations. In summarizing studies examining the reliability of distance runs, Freedson and colleagues (2000, pp. S80-S81) conclude that the "reliability of distance run tests has been generally high with correlation coefficients ranging from r = 0.61 to 0.92." A more recent review of studies examining the 1-mile run/walk test found intraclass correlation coefficients ranging from 0.39 to 0.90 in samples of children and adolescents (Artero et al., 2011).

ADMINISTRATIVE FEASIBILITY

Several factors should be considered with respect to administrative feasibility for tests that are to be used as part of a national survey or in schools and other educational settings. Although many of these factors apply to all settings (e.g., cost of the equipment), others relate more closely to schools specifically (e.g., whether the test is appropriate as part of the school curriculum). The latter considerations are discussed in more depth in Chapter 9.

The factors to be considered regarding administrative feasibility are summarized in the checklist in Box 3-2 in Chapter 3. In general, these factors are related to the test subject, the facility and equipment, the administrator of the test, and the parents of the test subject. The reader is referred to other publications that expand on these general factors (Mahar and Rowe, 2008). This section focuses on factors that are particularly relevant to conducting cardiorespiratory endurance tests and that apply to all settings.

Of interest is that 7 of 11 and 8 of 11 studies reviewed by the committee that used the treadmill and cycle ergometer tests, respectively, utilized maximal protocols. Maximal tests on either the treadmill or cycle ergometer are likely not to be administratively feasible in larger studies, especially if they are school based. Nonetheless, all three types of tests for which the committee found the strongest evidence for a relationship to health—the shuttle run, the treadmill, and the cycle ergometer—are generally feasible, and the setting will dictate the choice among these types. For example, if space is the major issue in test administration, such as in the case of a national survey, the treadmill and cycle ergometer tests will be preferred.

Facility factors are of particular importance as the different tests have different space and equipment requirements. For example, the shuttle run requires the most space—at least 20 meters for the test course; the treadmill and cycle ergometer tests require substantially less space. On the other hand, the treadmill and cycle ergometer tests require complex and expensive equipment. The different space requirements may have an impact on privacy for test subjects, the time required for testing, and the number of subjects who can be tested. Training of the test administrator in test protocols, test administration, and factors to consider is key to successful administration of a test and is another important consideration. For example, training for administration of the shuttle run is likely to be somewhat less complex than that required for the treadmill or cycle ergometer test. The cost of the equipment often is a major consideration in deciding which test should be used. The monetary cost of the equipment and of training the test administrators is relatively easy to assess. However, fitness testing may involve a wide range of additional direct and indirect costs. Ultimately, it is important to know the relative costs versus the relative benefits of using particular tests. No formal cost/benefit analyses have been performed for any of the available tests for cardiorespiratory endurance.

Parental factors include concerns about the impact of the test on the child. This may include fears regarding adverse events that could occur during testing, as well as concern about how the results and their interpretation will impact the child. Parents may be especially interested in the health implications of the results. These issues are probably equally important for all recommended cardiorespiratory endurance tests.

Adverse events, including injury during testing and the potential psychological effects of testing, should be considered. Adverse events of the various tests for assessment of cardiorespiratory endurance have not been systematically evaluated in the literature. The articles selected for this review do not report any injuries during testing. One recent manuscript (Ruiz et al., 2011) does address the safety of the 20-meter shuttle run, finding that no complications occurred during the testing, with only one report of a lower-body muscle cramp. The authors note that they have experienced no safety issues in more than 10,000 children they have tested.

GUIDANCE FOR INTERPRETATION OF TEST RESULTS

Chapter 3 presents a detailed discussion of the interpretation of fitness tests. Discussion of mathematical models for estimating cut-points, percentiles, or distribution curves is beyond the scope of this report.

Low cardiorespiratory endurance clearly is related to a variety of negative health outcomes, including obesity, elevated blood pressure, dyslipidemia, and cardiometabolic risk. There is also some evidence that cardiorespiratory endurance is associated with neurocognitive function. Some studies have suggested that the lowest third of the distribution of cardiorespiratory endurance is the group at highest risk for cardiometabolic risk factors/metabolic syndrome, but the relationship may be more of a continuous one, making specific cut-points more difficult to determine.

The committee recommends the use of interim cut-points based on data from both youth and adult populations on the relationship between treadmill performance and health outcomes until population-based evidence in youth is available for cardiorespiratory endurance tests. The bottom quintile of the distribution for cardiorespiratory endurance on a maximal treadmill test is associated with elevated morbidity and mortality (Blair et al., 1989) in adults. When interpreting test results, therefore, interim cut-points could be derived from low performers (e.g., 20th percentile) in the cardiorespiratory endurance distribution curve to identify youth at the highest risk of poor health outcomes and increase the likelihood that an individual identified as low fit is actually low fit. This is a more conservative approach than that taken by Lobelo and colleagues (2009) and Welk and colleagues (2011), who estimate approximately the 30th percentile to derive cut-points for cardiorespiratory endurance tests for youth. The committee's approach is based on its view that identifying a fit individual as low fit (potentially recommending an exercise intervention to a test taker who does not need it) is a more serious error than identifying an individual who is low fit as fit. It should be noted that this approach must take into account covariates such as age and sex, which allow standardization of the interpretation of test results across individuals and, more important, for an individual longitudinally across different ages. To derive the appropriate cut-points from percentiles, fitness data based on large populations for the test of interest are needed. If such data are not available, developers of cut-points should consult with statisticians to design a small study with a representative sample of U.S. youth to collect the necessary data.

Accurate interpretation and effective communication of test results are important when improved fitness is a goal of the test. As mentioned in Chapter 3, an individual's results can be presented against the background of a continuous distribution. The continuous background reflects the concept that improved fitness in general, even within a broader range, is associated with a lower risk of negative health outcomes. Ultimately, research should be conducted to evaluate the impact of this approach to classification and interpretation on test subjects, parents, test administrators, teachers, physicians, and others and on future health behaviors.

CONCLUSIONS

There is a well-known association between the fitness component cardiorespiratory endurance and health outcomes in adults. The measurement of cardiorespiratory endurance and its relationship to health outcomes in youth is relatively new to the literature. The committee's review revealed that sufficient relationships have been established between cardiorespiratory endurance and several health risk factors in youth, including adiposity and cardiometabolic risk factors (blood pressure, blood lipids and glucose, and insulin sensitivity). A few studies have established a relationship with other, less-studied pediatric health risk factors, such as pulmonary function, depression and positive self-concept, and bone health.

The literature review provided to the committee included 34 articles indicating a positive relationship between results of cardiorespiratory endurance tests in youth and health risk factors, independent of other interventions. The review included longitudinal, experimental, and quasi-experimental studies. There was substantial variability in the tests used, especially with the protocols for distance/timed runs and cycle ergometry. The characteristics of the subjects (e.g., age, gender, weight) varied as well.

The cardiorespiratory endurance tests most often associated with a positive change in a health risk factor were the shuttle run, treadmill, and cycle ergometer tests. The health markers most frequently assessed were related to body weight or adiposity and cardiometabolic risk factors. The shuttle run, treadmill, and cycle ergometer tests all showed strong relationships to health markers. Because of the paucity of studies addressing the influence of several potential modifiers of performance—age, gender, race/ethnicity, body composition, maturation status—on the various cardiorespiratory endurance tests, the committee was unable to examine this issue. Such influences have, however, been suggested in the past (Beets and Pitetti, 2004; Bovet et al., 2007; Chomitz et al., 2010; Cureton et al., 1997; Huang and Malina, 2007, 2010; Mahon and Vaccaro, 1989; Pate et al., 2006; Trowbridge et al., 1997).

The treadmill and cycle ergometer tests are quasi-laboratory tests that may be best suited to situations where space is a limitation. Field-based cardiorespiratory endurance tests include both distance/timed runs and the shuttle run. The shuttle run is advantageous when there are time constraints and the purchase of sophisticated equipment and use of expert testers may not be feasible.

The available evidence indicates that all of the approaches to measuring cardiorespiratory endurance examined in this chapter demonstrate acceptable validity and reliability. The validity and reliability coefficients for runs of varying distances and time limits are more variable and less consistently high than those reported for the shuttle run and heart rate extrapolation tests (treadmill and cycle ergometer).

Based on its relationship to health, as well as its reliability, validity, and feasibility, a timed or progressive shuttle run, such as the 20-meter shuttle run, is appropriate for measuring cardiorespiratory endurance in youth. If the test is to be administered in a setting where there are space limitations, a submaximal treadmill or cycle ergometer test should be used, even though several studies reviewed here were conducted with maximal tests. Submaximal protocols are recommended for feasibility reasons: maximal tests are not suitable for large samples or school settings because they require that participants meet certain criteria, such as reaching a certain number of beats/minute, respiratory quotient, and oxygen consumption. Moreover, there is a proven relationship between performance on a submaximal test and performance on a maximal test. Although the evidence for a relationship to health is not sufficient at this time for distance/timed runs, this test is valid and reliable and could be an alternative in schools and other educational settings.

Until population-based evidence in youth is available, the lowest 20th percentile of the distribution of cardiorespiratory endurance should be used to derive interim cut-points for determining whether individuals are at risk of cardiovascular-associated negative health outcomes. The committee's full recommendations on cardiovascular endurance tests for use in national youth fitness surveys and in schools and other educational settings are presented in Chapters 8 and 9, respectively.

REFERENCES

Aires, L., D. Mendonça, C. Silva, A. R. Gaya, M. P. Santos, J. C. Ribeiro, and J. Mota. 2010. A 3-year longitudinal analysis of changes in body mass index. *International Journal of Sports Medicine* 31(2):133-137.

Arraiz, G. A., D. T. Wigle, and Y. Mao. 1992. Risk assessment of physical-activity and physical-fitness in the Canada Health Survey mortality follow-up study. *Journal of Clinical Epidemiology* 45(4):419-428.

Artero, E. G., V. España-Romero, J. Castro-Piñero, F. B. Ortega, J. Suni, M. J. Castillo-Garzon, and J. R. Ruiz. 2011. Reliability of field-based fitness tests in youth. *International Journal of Sports Medicine* 32(3):159-169.

Barbeau, P., M. S. Litaker, K. F. Woods, C. R. Lemmon, M. C. Humphries, S. Owens, and B. Gutin. 2002. Hemostatic and inflammatory markers in obese youths: Effects of exercise and adiposity. *Journal of Pediatrics* 141(3):415-420.

Beets, M. W., and K. H. Pitetti. 2004. One-mile run/walk and body mass index of an ethnically diverse sample of youth. *Medicine and Science in Sports and Exercise* 36(10):1796-1803.

Ben Ounis, O., M. Elloumi, E. Makni, H. Zouhal, M. Amri, Z. Tabka, and G. Lac. 2010. Exercise improves the ApoB/ApoA-I ratio, a marker of the metabolic syndrome in obese children. *Acta Pædiatrica* 99(11):1679-1685.

Blair, S. N., N. N. Goodyear, L. W. Gibbons, and K. H. Cooper. 1984. Physical fitness and incidence of hypertension in healthy normotensive men and women. *Journal of the American Medical Association* 252(4):487-490.

Blair, S. N., H. W. Kohl, R. S. Paffenbarger, D. G. Clark, K. H. Cooper, and L. W. Gibbons. 1989. Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Journal of the American Medical Association* 262(17):2395-2401.

Blair, S. N., H. W. Kohl, and C. E. Barlow. 1993. Physical-activity, physical-fitness, and all-cause mortality in women: Do women need to be active? *Journal of the American College of Nutrition* 12(4):368-371.

Blair, S. N., H. W. Kohl, C. E. Barlow, R. S. Paffenbarger, L. W. Gibbons, and C. A. Macera. 1995. Changes in physical-fitness and all-cause mortality. A prospective-study of healthy and unhealthy men. *Journal of the American Medical Association* 273(14):1093-1098.

Boreham, C. A., V. J. Paliczka, and A. K. Nichols. 1990. A comparison of the PWC170 and 20-MST tests of aerobic fitness in adolescent schoolchildren. *Journal of Sports Medicine and Physical Fitness* 30(1):19-23.

Bovet, P., R. Auguste, and H. Burdette. 2007. Strong inverse association between physical fitness and overweight in adolescents: A large school-based survey. *International Journal of Behavioral Nutrition and Physical Activity* 4(24).

Byrd-Williams, C. E., G. Q. Shaibi, P. Sun, C. J. Lane, E. E. Ventura, J. N. Davis, L. A. Kelly, and M. I. Goran. 2008. Cardiorespiratory fitness predicts changes in adiposity in overweight Hispanic boys. *Obesity* 16(5):1072-1077.

Carrel, A. L., R. R. Clark, S. E. Peterson, B. A. Nemeth, J. Sullivan, and D. B. Allen. 2005. Improvement of fitness, body composition, and insulin sensitivity in overweight children in a school-based exercise program: A randomized, controlled study. *Archives of Pediatrics & Adolescent Medicine* 159(10):963-968.

Chang, C., W. Liu, X. Zhao, S. Li, and C. Yu. 2008. Effect of supervised exercise intervention on metabolic risk factors and physical fitness in Chinese obese children in early puberty. *Obesity Reviews* 9(Suppl. 1):135-141.

Chen, J. L., D. Wall, C. Kennedy, V. Unnithan, and C. H. Yeh. 2007. Predictors of increased body mass index in Chinese children. *Progress in Cardiovascular Nursing* 22(3): 138-144.

Chomitz, V. R., R. J. McGowan, J. M. Wendel, S. A. Williams, H. J. Cabral, S. E. King, D. B. Olcott, M. Cappello, S. Breen, and K. A. Hacker. 2010. Healthy Living Cambridge Kids: A community-based participatory effort to promote healthy weight and fitness. *Obesity* 18:S45-S53.

Colberg, S. R., R. J. Sigal, B. Fernhall, J. G. Regensteiner, B. J. Blissmer, R. R. Rubin, L. Chasan-Taber, A. L. Albright, and B. Braun. 2010. Exercise and type 2 diabetes. *Diabetes Care* 33(12):e147-e167.

Cureton, K. J., M. A. Sloniger, D. M. Black, W. P. McCormack, and D. A. Rowe. 1997. Metabolic determinants of the age-related improvement in one-mile run/walk performance in youth. *Medicine and Science in Sports and Exercise* 29(2):259-267.

Davis, P. O., C. O. Dotson, and A. V. Curtis. 1985. A simplified technique for the determination of percent body fat in adult males. *Journal of Sports Medicine and Physical Fitness* 25(4):255-261.

Dunton, G. F., M. Schneider, and D. M. Cooper. 2007. An investigation of psychosocial factors related to changes in physical activity and fitness among female adolescents. *Psychology & Health* 22(8):929-944.

Eisenmann, J. C., E. E. Wickel, G. J. Welk, and S. N. Blair. 2005. Relationship between adolescent fitness and fatness and cardiovascular disease risk factors in adulthood: The Aerobics Center Longitudinal Study (ACLS). *American Heart Journal* 149(1):46-53.

Farpour-Lambert, N. J., Y. Aggoun, L. M. Marchand, X. E. Martin, F. Herrmann, and M. Beghetti. 2009. Physical activity reduces systemic blood pressure and improves early markers of atherosclerosis in pre-pubertal obese children. *Journal of the American College of Cardiology* 54(25):2396-2406.

Freedson, P. S., K. J. Cureton, and G. W. Heath. 2000. Status of field-based fitness testing in children and youth. *Preventive Medicine* 31(2):S77-S85.

Gately, P. J., C. B. Cooke, J. H. Barth, B. M. Bewick, D. Radley, and A. J. Hill. 2005. Children's residential weight-loss programs can work: A prospective cohort study of short-term outcomes for overweight and obese children. *Pediatrics* 116(1):73-77.

Gutin, B., C. Basch, S. Shea, I. Contento, M. Delozier, J. Rips, M. Irigoyen, and P. Zybert. 1990. Blood-pressure, fitness, and fatness in 5-year-old and 6-year-old children. *Journal of the American Medical Association* 264(9):1123-1127.

Huang, Y. C., and R. M. Malina. 2007. BMI and health-related physical fitness in Taiwanese youth 9-18 years. *Medicine and Science in Sports and Exercise* 39(4):701-708.

Huang, Y. C., and R. M. Malina. 2010. Body mass index and individual physical fitness tests in Taiwanese youth aged 9-18 years. *International Journal of Pediatric Obesity* 5(5):404-411.

Janz, K. F., J. D. Dawson, and L. T. Mahoney. 2002. Increases in physical fitness during childhood improve cardiovascular health during adolescence: The Muscatine Study. *International Journal of Sports Medicine* 23(Suppl. 1):S15-S21.

Jekal, Y., E. S. Kim, J. A. Im, J. H. Park, M. K. Lee, S. H. Lee, S. H. Suh, S. H. Chu, E. S. Kang, H. C. Lee, and J. V. Jeon. 2009. Interaction between fatness and fitness on CVD risk factors in Asian youth. *International Journal of Sports Medicine* 30(10):733-740.

Johnson, M. S., R. Figueroa-Colon, S. L. Herd, D. A. Fields, M. Sun, G. R. Hunter, and M. I. Goran. 2000. Aerobic fitness, not energy expenditure, influences subsequent increase in adiposity in black and white children. *Pediatrics* 106(4):e50.

Kelly, A. S., R. J. Wetzsteon, D. R. Kaiser, J. Steinberger, A. J. Bank, and D. R. Dengel. 2004. Inflammation, insulin, and endothelial function in overweight children and adolescents: The role of exercise. *Journal of Pediatrics* 145(6):731-736.

Kim, J., A. Must, G. M. Fitzmaurice, M. W. Gillman, V. Chomitz, E. Kramer, R. McGowan, and K. E. Peterson. 2005. Relationship of physical fitness to prevalence and incidence of overweight among schoolchildren. *Obesity Research* 13(7):1246-1254.

Lee, K., Y. Shin, T. Jun, and W. Song. 2010. Aerobic exercise training-induced decrease in plasma visfatin and insulin resistance in obese female adolescents. *International Journal of Sport Nutrition & Exercise Metabolism* 20(4):275-281.

Léger, L. A., D. Mercier, C. Gadoury, and J. Lambert. 1988. The multistage 20 metre shuttle run test for aerobic fitness. *Journal of Sports Sciences* 6(2):93-101

Liu, N. Y. S., S. A. Plowman, and M. A. Looney. 1992. The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research Quarterly for Exercise and Sport* 63(4):360-365.

Lobelo, F., R. R. Pate, M. Dowda, A. D. Liese, and J. R. Ruiz. 2009. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Medicine and Science in Sports and Exercise* 41(6):1222-1229.

Mahar, M. T., and D. A. Rowe. 2008. Practical guidelines for valid and reliable youth fitness testing. *Measurement in Physical Education and Exercise Science* 12:126-145.

Mahon, A. D., and P. Vaccaro. 1989. Ventilatory threshold and V02max changes in children following endurance training. *Medicine and Science in Sports and Exercise* 21(4):425-431.

Martins, C., R. Santos, A. Gaya, J. Twisk, J. Ribeiro, and J. Mota. 2009. Cardiorespiratory fitness predicts later body mass index, but not other cardiovascular risk factors from childhood to adolescence. *American Journal of Human Biology* 21(1):121-123.

McMurray, R. G., S. I. Bangdiwala, J. S. Harrell, and L. D. Amorim. 2008. Adolescents with metabolic syndrome have a history of low aerobic fitness and physical activity levels. *Dynamic Medicine* 7:5.

Mota, J., J. Ribeiro, Carlos, J. Carvalho, M. P. Santos, and J. Martins. 2009. Cardiorespiratory fitness status and body mass index change over time: A 2-year longitudinal study in elementary school children. *International Journal of Pediatric Obesity* 4(4):338-342.

Murphy, N. M., M. Ni Dhuinn, P. A. Browne, and M. M. Orathaille. 2006. Physical activity for bone health in inactive teenage girls: Is a supervised, teacher-led program or self-led program best? *Journal of Adolescent Health* 39(4):508-514.

National Center for Health Statistics. 2004. *National Health and Nutrition Examination Survey (NHANES) cardiovascular fitness procedures manual.* Washington, DC: Centers for Disease Control and Prevention.

Nourry, C., F. Deruelle, C. Guinhouya, G. Baquet, C. Fabre, F. Bart, S. Berthoin, and P. Mucci. 2005. High-intensity intermittent running training improves pulmonary function and alters exercise breathing pattern in children. *European Journal of Applied Physiology* 94(4):415-423.

Oliveria, S. A., H. W. Kohl, D. Trichopoulos, and S. N. Blair. 1996. The association between cardiorespiratory fitness and prostate cancer. *Medicine and Science in Sports and Exercise* 28(1):97-104.

Ortega, F. B., I. Labayen, J. R. Ruiz, E. Kurvinen, H. M. Loit, J. Harro, T. Veidebaum, and M. Sjostrom. 2011. Improvements in fitness reduce the risk of becoming overweight across puberty. *Medicine and Science in Sports and Exercise* 43(10):1891-1897.

Pate, R. R., M. G. Davis, T. N. Robinson, E. J. Stone, T. L. McKenzie, and J. C. Young. 2006. Promoting physical activity in children and youth. *Circulation* 114(11):1214-1224.

Puder, J. J., P. Marques-Vidal, C. Schindler, L. Zahner, I. Niederer, F. Bürgi, V. Ebenegger, A. Nydegger, and S. Kriemler. 2011. Effect of multidimensional lifestyle intervention on fitness and adiposity in predominantly migrant preschool children (Ballabeina): Cluster randomised controlled trial. *British Medical Journal* 343(d6195).

Reed, K. E., D. E. R. Warburton, H. M. Macdonald, P. J. Naylor, and H. A. McKay. 2008. Action Schools! BC: A school-based physical activity intervention designed to decrease cardiovascular disease risk factors in children. *Preventive Medicine* 46(6):525-531.

Rowland, T. W., J. M. Rambusch, J. S. Staab, V. B. Unnithan, and S. F. Siconolfi. 1993. Accuracy of physical working capacity (PWC170) in estimating aerobic fitness in children. *Journal of Sports Medicine and Physical Fitness* 33(2):184-188.

Ruiz, J. R., J. Castro-Piñero, V. España-Romero, E. G. Artero, F. B. Ortega, M. M. Cuenca, D. Jimenez-Pavon, P. Chillon, M. J. Girela-Rejon, J. Mora, A. Gutierrez, J. Suni, M. Sjostrom, and M. J. Castillo. 2011. Field-based fitness assessment in young people: The alpha health-related fitness test battery for children and adolescents. *British Journal of Sports Medicine* 45(6):518-524.

Safrit, M. J. 1990. The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science* 2:9-28.

Saltin, B. 1973. Oxygen transport by the circulatory system during exercise in man. In *Limiting factors of physical performance,* edited by J. Keul. Stuttgart, Germany: Thieme Publishers. Pp. 235-252.

Sandvik, L., J. Erikssen, E. Thaulow, G. Erikssen, R. Mundal, and K. Rodahl. 1993. Physical-fitness as a predictor of mortality among healthy, middle-aged Norwegian men. *New England Journal of Medicine* 328(8):533-537.

Sidiropoulou, M. P., E. G. Fotiadou, V. K. Tsimaras, A. P. Zakas, and N. A. Angelopoulou. 2007. The effect of interval training in children with exercise-induced asthma competing in soccer. *The Journal of Strength & Conditioning Research* 21(2):446-450.

Slaughter, M. H., T. G. Lohman, R. A. Boileau, C. A. Horswill, R. J. Stillman, M. D. Vanloan, and D. A. Bemben. 1988. Skinfold equations for estimation of body fatness in children and youth. *Human Biology* 60(5):709-723.

Stella, S. G., A. P. Vilar, C. Lacroix, M. Fisberg, R. F. Santos, M. T. Mello, and S. Tufik. 2005. Effects of type of physical exercise and leisure activities on the depression scores of obese Brazilian adolescent girls. *Brazilian Journal of Medical and Biological Research* 38(11):1683-1689.

Tremblay, M., and M. Lloyd. 2010. Physical literacy measurement: The missing piece. *Physical and Health Education Journal* 76(1):26-30.

Trowbridge, C. A., B. A. Gower, T. R. Nagy, G. R. Hunter, M. S. Treuth, and M. I. Goran. 1997. Maximal aerobic capacity in African-American and Caucasian prepubertal children. *American Journal of Physiology-Endocrinology and Metabolism* 273(4):E809-E814.

Twisk, J. W. R., H. C. G. Kemper, and W. van Mechelen. 2000. Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Medicine and Science in Sports and Exercise* 32(8):1455-1461.

Wahlund, H. 1948. Determination of physical working capacity: Physiological and clinical-study with special reference to standardization of cardiopulmonary, functional tests. *Acta Medica Scandinavica* 132(Suppl. 215).

Walther, C., L. Gaede, V. Adams, G. Gelbrich, A. Leichtle, S. Erbs, M. Sonnabend, K. Fikenzer, A. Körner, W. Kiess, M. Bruegel, J. Thiery, and G. Schuler. 2009. Effect of increased exercise in school children on physical fitness and endothelial progenitor cells: A prospective randomized trial. *Circulation* 120(22):2251-2259.

Watkins, J., and B. G. Ewing. 1983. Physical working capacity and mile run performance in adolescent boys. *British Journal of Sports Medicine* 17(3):188-192.

Watson, A. W. S., and D. J. Odonovan. 1976. Physical working capacity of male adolescents in Ireland. *Irish Journal of Medical Science* 145(12):383-391.

Welk, G. J., K. R. Laurson, J. C. Eisenmann, and K. J. Cureton. 2011. Development of youth aerobic-capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine* 41(4, Suppl. 2):S111-S116.

Wong, P. C., M. Y. Chia, I. Y. Tsou, G. K. Wansaicheong, B. Tan, J. C. Wang, J. Tan, C. G. Kim, G. Boh, and D. Lim. 2008. Effects of a 12-week exercise training programme on aerobic fitness, body composition, blood lipids and C-reactive protein in adolescents with obesity. *Annals of the Academy of Medicine, Singapore* 37(4):286-293.

Wong, P., K. Chamari, A. Dellal, and U. Wisløff. 2009. Relationship between anthropometric and physiological characteristics in youth soccer players. *Journal of Strength & Conditioning Research (Lippincott Williams & Wilkins)* 23(4):1204-1210.

6

Health-Related Fitness Measures for Youth: Musculoskeletal Fitness

KEY MESSAGES

Musculoskeletal fitness is a multidimensional construct comprising the integrated function of muscle strength, muscle endurance, and muscle power. The link between musculoskeletal fitness and health in adults has extended beyond low-back health to other outcomes, such as personal independence and quality of life, cardiovascular disease, risk of fracture, and cognitive and functional ability. Although the relationship between musculoskeletal fitness and these health outcomes in youth is not as extensively or specifically studied as that in adults, this chapter summarizes what is currently known about this relationship in youth.

A thorough review of the literature revealed a lack of high-quality studies supporting a strong link between any specific musculoskeletal fitness test item and health outcomes in youth. This lack of evidence precluded the identification of any specific musculoskeletal fitness test items for inclusion in a national fitness survey for the general population of youth. Nonetheless, based predominantly on evidence indicating a relationship between musculoskeletal fitness and health outcomes in adults, the committee concluded that musculoskeletal fitness should be assessed in a national youth fitness survey. The handgrip strength and standing long jump tests (to measure upper- and lower-body musculoskeletal strength, respectively) should be included in such a survey based on their limited link to health and their acceptable validity, reliability, and feasibility. These tests should not, however, be interpreted in a health context until their relationships with health outcomes have been established more firmly in youth.

Limitations of the studies reviewed include that studies were not designed to answer questions about the relationship between the fitness tests studied and health, that interventions were inadequate, or that confounders were not considered. Although effects of age, gender, body composition, maturation status, and ethnicity on performance on the various tests have been suggested in the past, this review provided insufficient data for assessing the influence of such modifiers.

For school and other educational settings, administrators should consider the handgrip strength and standing long jump tests, as well as alternative tests that have not yet been shown to be related to health but are valid, reliable, and feasible. The modified pull-up and the push-up are possible alternatives for measuring upper-body musculoskeletal strength and power. The curl-up could also be considered for measuring an additional construct, core strength.

In the absence of criterion-referenced cut-points (cutoff scores) for youth or adults, interim cut-points corresponding to the lower percentile limit (20th percentile) should be used for tests of musculoskeletal fitness, analogous to the cut-points for cardiorespiratory endurance, until better evidence for criterion-referenced health-related cut-points is established by further research.

The functions and capacities of the neuromuscular and musculoskeletal systems play important roles in defining the physical fitness status of individuals and populations. Assessment of musculoskeletal fitness has traditionally included assessment of muscle strength, muscle endurance, flexibility, and bone health (Bouchard et al., 2007). With increasing interest in and study of the role of muscle power in the elderly, it is likely that muscle power will emerge as another important characteristic of musculoskeletal fitness worthy of inclusion in future youth fitness assessments (Ashe et al., 2008; Bonnefoy et al., 2007; Reid and Fielding, 2012).

This chapter addresses musculoskeletal fitness (muscle strength, endurance, and power) as it relates to health markers in youth; the flexibility component of musculoskeletal fitness is considered in Chapter 7. The committee's recommendations for selection of musculoskeletal fitness tests are based primarily on an extensive review of the literature provided by the Centers for Disease Control and Prevention (CDC). The CDC search strategy and data extraction procedures are described in Chapter 3. To make its recommendations on this fitness component, in addition to providing evidence for a relationship to health, the committee considered the scientific integrity (reliability and validity) of putative health-related test items, as well as the administrative feasibility of implementing these items. The committee also offers recommendations for setting cut-points (cutoff scores) for interpretation of performance on musculoskeletal fitness tests. Recommendations regarding specific tests for measuring musculoskeletal fitness for national surveys and in schools and other educational settings are presented in Chapter 8 and 9, respectively. Future research needs are addressed in Chapter 10.

DEFINITIONS

Musculoskeletal fitness is a multidimensional construct comprising the integrated function of muscle strength, muscle endurance, and muscle power to enable the performance of work against one's own body weight or an external resistance. No single measure of any of these dimensions adequately describes an individual's overall level of musculoskeletal fitness; rather, each of these dimensions must be assessed individually, compared with appropriate performance or health standards, and then interpreted in an integrated and unified assessment of overall musculoskeletal fitness.

Muscle strength is the ability of skeletal muscle (single or group) to produce measurable force, torque, or moment about a single or multiple joints, typically during a single maximal voluntary contraction and under a defined set of controlled conditions, which include specificity of movement pattern, muscle contraction type (concentric, isometric, or eccentric), and contraction velocity (Farpour-Lambert and Blimkie, 2008; Kell et al., 2001; Sale and Norman, 1982). In youth fitness assessments, this definition usually applies to the production of maximal muscle force during a single maximal voluntary effort on a particular strength test. Some musculoskeletal fitness tests, however (e.g., the pull-up test), allow the completion of more than one near-maximal effort (e.g., two to three repetitions) and have traditionally also been considered tests of muscle strength. Strength is typically measured as force registered by a dynamometer (e.g., handgrip dynamometer) or a measurable external load resisted or moved against (e.g., weight machine or external weight).

Muscle endurance is the ability of a muscle or group of muscles to perform repeated contractions against a constant external load for an extended period of time (Kell et al., 2001). The constant load can be either an absolute external resistance, which provides a measure of absolute endurance, or a relative load based on an individual's maximal strength, which provides a measure of relative endurance. In youth fitness assessments, this definition applies to voluntary submaximal efforts of variable force production and speed by a muscle or group of muscles during performance on a wide variety of tests. Muscle endurance is typically measured as elapsed time or number of paced or nonpaced repetitions of the muscle action within either a specified or unrestricted time period.

Muscle power is a physiological construct reflecting the rate at which work is performed (Knuttgen and Kraemer, 1987). It is derived from the product of the force production of a muscle or group of muscles and the velocity of the muscle contraction during a single- or multijoint action (Sale and Norman, 1982). Muscle power is a complex construct consisting of several subdomains, including average, peak, instantaneous, and contractile power (Moffroid and Kusiak, 1975). In youth fitness testing, different field tests probably assess different subdomains of muscle power, although the specific associations between individual fitness tests and the power subdomains are poorly defined. Peak muscle power is dependent on the velocity of the action and is inversely related to the external resistance against the action. Peak power is typically generated within the range of 40-90 percent of peak external resistance, or approximately 70 percent of an individual's one repetition maximum (1RM) (Reid and Fielding, 2012), and at submaximal velocity. Muscle power then can be defined as the product of force and velocity during execution of a maximal voluntary effort against a submaximal external resistance, and it can be measured directly in two ways: by setting a series of constant-velocity efforts and measuring muscle force at each velocity, or by setting a series of constant loads and measuring the velocity at each load, with power expressed in watts (W) being the product of force and velocity for each series effort. In practice, in youth fitness testing the velocity is either controlled or uncontrolled, and the external resistance is either the body weight or a resistance that is set below the peak force-producing capacity of the muscles involved in the action.

Field tests of muscle power typically involve assessment of upper-body (throwing distance) or lower-body (vertical squat jumps, vertical countermovement jumps, or long jump) muscle function, and usually measure height or distance covered. Performance on these tests is directly related to the attained velocity, which is proportional to the force generated during the action and provides an indirect measure of muscle power. Field tests of muscle power have been included as a surrogate measure of muscle strength even though physiologically, this extrapolation is valid only if the action is performed at a constant velocity, which is rarely the case in the field. For this review, the committee considered only power tests that incorporate a single maximal effort at a submaximal velocity and load (e.g., vertical or horizontal jumping tests). These tests require a high degree of neuromechanical coordination and are less dependent on the biochemical endurance capacities of the muscles compared with one of the most common measures of anaerobic power, the Wingate Anaerobic Test. Because of its unique physiological and neuromechanical characteristics, muscle power is considered one of three dimensions of musculoskeletal fitness in youth fitness assessments.

MUSCULOSKELETAL FITNESS TESTS

A plethora of fitness test batteries and items have been used over the past 55 years to assess musculoskeletal fitness in youth (see Table 2-6 in Chapter 2) (Castro-Piñero et al., 2010). The tests vary in their specific protocols, some purportedly assessing the muscle fitness of specific body regions (upper and lower body, trunk, abdomen, lower back) and some measuring isolated muscular function (e.g., muscle strength, endurance, or power) or combined strength and endurance function.

Since the mid-1970s, there has been growing interest in and development of health-related musculoskeletal fitness test batteries that have been based largely on theoretical construct validity and on health data from the adult population (AAHPERD, 1984; Jackson, 2006; Morrow et al., 2009; Plowman, 2008). The American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) Health Related Physical Fitness Test, the first of many subsequent international fitness test batteries to claim assessment of health-related fitness in youth, included the modified, timed (1-minute) sit-up as the sole measure of musculoskeletal fitness.

More than 11 different classes of fitness test items have since been used to assess the muscle strength, endurance, or power dimensions of musculoskeletal fitness, many of them evaluating similar dimensions (Table 6-1). For example, there are several variations on the pull-up test of differing durations (no time limit, 30- or 60-second limit), with different anatomical alignment of the body (full arm extension or right-angled pull-up), and with varying interpretations of what the test items actually measure (upper-body strength, upper-body endurance, combined upper-body strength and endurance, athletic ability, relative strength).

It is apparent that many of these test items do not satisfy the physiological definitions of the three dimensions of musculoskeletal fitness. Muscle endurance fitness test items arguably may be considered the most physiologically valid field tests in youth as opposed to those measuring muscle strength and power, which are more subject to velocity control, loads, and number of repetitions. Additionally, several of the currently used field-based fitness tests (e.g., curl-up and pull-up) purport to measure more than one musculoskeletal dimension concurrently. Because of their lower construct validity, results of muscle strength and power tests must be interpreted cautiously in youth.

TABLE 6-1 Summary of Muscle Strength, Endurance, and Power Fitness Test Items Used Historically to Assess Musculoskeletal Fitness in International Youth Fitness Test Batteries

Fitness Test Item	Fitness Component Evaluated	Variant Approaches
Flexed/Bent Arm Hang	• Arm and shoulder endurance • Functional strength	
Pull-up	• Upper-body strength and endurance • Upper-body endurance • Athletic ability • Relative strength	• Regular with full arm extension • Modified with right-angle limited extension • Untimed performance • Timed performance (30 and 60 seconds)
Push-up	• Upper-body strength and endurance	• Regular with full arm flexion/extension • Modified with limited arm flexion/extension to 90 degrees • Isometric hold position
Dip	• Upper-body endurance	
Sit-up	• Abdominal strength and endurance • Abdominal endurance	• Straight-legged • Bent-legged • Timed performance (30 and 60 seconds)
Curl-up	• Abdominal strength and endurance • Trunk strength	• Full flexion/extension • Partial flexion/extension • Untimed • Timed (30 seconds) • Cadence based
Trunk Lift	• Back extensor strength • Back flexibility	
Handgrip Strength Test	• Static/isometric strength	
Standing Broad/Long Jump	• Explosive power • Lower-body strength • Athletic ability	• Levels
Vertical Jump	• Explosive power • Lower-body strength • Athletic ability	• Jump with no countermovement • Countermovement jump
Throwing	• Upper-body explosive strength • Strength and endurance • Athletic ability	• Softball throw • Handball throw • Basketball throw • Medicine ball throw • Shotput (variable weights)

MUSCULOSKELETAL FITNESS AND HEALTH IN ADULTS

There is increasing evidence of the importance of musculoskeletal fitness as a determinant of health outcomes both in healthy young, middle-aged, and elderly adults and in adults with disability or chronic disease. A review of the relationship of early test batteries to health outcomes revealed that the evidence was limited, even though there was sound anatomical logical validity for a link between abdominal and back health and musculoskeletal fitness (Plowman, 1992). More recent evidence lends additional support to the idea that tests of abdominal and back extensor muscle endurance relate to back health status, as assessed by history of low-back pain, in adults (Payne et al., 2000).

In recent years, the link between musculoskeletal fitness and health in adults has extended beyond the initial focus on low-back health. Recent reviews have established positive associations between muscle strength and personal independence and quality of life, and inverse associations with cardiometabolic risk factors, frequency of cardiovascular disease events, risk of general morbidity for nonfatal diseases (e.g., fracture risk and cognitive decline), and all-cause mortality in middle-aged and elderly adults (Bohannon, 2008; Cooper et al., 2011; Garber et al., 2011; Warburton et al., 2001; Williams et al., 2007). Likewise, muscle endurance has been positively associated with overall quality of life and negatively associated with likelihood of falling and associated skeletal and soft tissue injuries (Warburton et al., 2001). Muscle power appears to decline more rapidly than muscle strength with aging, and loss of muscle power is strongly associated with decreases in functional ability (e.g., reduced ability to stand from sitting in a chair), and it may be predictive of decreased mobility and premature mortality in adults (Reid and Fielding, 2012; Warburton et al., 2001).

Skeletal muscle and its functional capacities may also be related to more health-related outcomes than has previously been appreciated. Reductions in skeletal muscle mass associated with acute or chronic illness may negatively impact musculoskeletal fitness as assessed by muscle strength, endurance, and power tests. Reduced muscle strength and function with accompanying loss of muscle mass in acute or chronic illness are related to increased recovery times, impaired patient quality of life, and likelihood of institutionalization (Wolfe, 2006). Further, skeletal muscle is a major regulator of glucose and fat metabolism and may play an important role in the development of the metabolic syndrome and perhaps even obesity (Jurca et al., 2005). The degree to which musculoskeletal fitness tests are predictive of the development of these conditions and their responsiveness to clinical management in adults remains an interesting yet untested question. Lastly, skeletal muscle may be an important determinant of bone and joint health in middle-aged and older adults as a result of direct muscle forces imparted to the skeleton during movement, as well as the effect of increased muscle mass on skeletal loading. While it is difficult to separate those two effects (Beck, 2009), suggestive evidence points to a positive association between measures of musculoskeletal fitness (especially muscle strength and power) and bone health in adults (Ashe et al., 2008; Cooper et al., 2011; von Stengel et al., 2005, 2007). Positive associations also have been reported between muscle strength and power and better quality of life, lower risk of falls and fractures, and reduced morbidity and mortality (Cooper et al., 2011; von Stengel et al., 2005, 2007). Likewise, muscle weakness has been identified as a risk factor for osteoarthritis in this population (Garber et al., 2011).

The validity of the relationships described above is further corroborated by evidence for the effect of resistance training programs on muscle strength, endurance, and power, along with changes in various health outcomes. Resistance training programs now are generally accepted as being effective at improving muscle strength, endurance, and power in both sexes, across all ages during adulthood, and for both healthy adults and those with chronic disease or disability (McCartney and Phillips, 2007; Reid and Fielding, 2012; Williams et al., 2007). These programs also have resulted in a multitude of adaptations that foster better health among adults, such as improved body composition, blood glucose and insulin regulation, systemic arterial blood pressure in prehypertensives, blood lipid and lipoprotein profiles, bone health and management of arthritic pain and disability, and prevention or improved management of the metabolic syndrome (Garber et al., 2011; McCartney and Phillips, 2007; Williams et al., 2007). Similarly, resistance training has resulted in enhanced exercise and functional capacity, improved balance, and decreased falls (Garber et al., 2011; McCartney and Phillips, 2007). Resistance training may also improve quality of life and self-efficacy and moderate levels of depression and anxiety among adults (Garber et al., 2011; McCartney and Phillips, 2007; Williams et al., 2007).

MUSCULOSKELETAL FITNESS AND HEALTH IN YOUTH

Literature Review Process

The CDC's systematic review of the literature included muscle strength and muscle endurance, but not muscle power, as components of fitness because they are the dimensions of musculoskeletal fitness that have been used most frequently in fitness test batteries. The muscle strength search screened 2,642 reports, only 63 of which satisfied the CDC search criteria for further consideration and were abstracted. Of this subset of 63 studies, 23 were classified as experimental, 22 as experimental with no control, 12 as quasi-experimental, and 6 as longitudinal.

The muscle endurance search screened 6,563 reports, 38 of which were retained for further consideration and were abstracted. Of this subset, 12 studies were experimental, 15 experimental with no control, 6 quasi-experimental, and 5 longitudinal. The committee chose to review only the experimental (including those with no control and quasi-experimental) and longitudinal prospective studies in making its recommendations. In addition to the CDC search strategy, the committee reviewed the reference lists in the selected articles and relevant studies published before 2000 or after 2010.

The committee developed a set of criteria with which to assess the scientific quality of the studies (see Chapter 3). Each study was evaluated against those criteria and categorized as of low, moderate, or high quality. Only those of high quality were reviewed further; they are described in Table 6-2. The evidence for a link between a test item and a health marker in the top high-quality studies was categorized as direct or associational based on the strength of the study design and the rigor of the statistical analysis. The strength of the evidence was categorized as sufficient or insufficient based on the number of studies with direct or indirect evidence, the study designs, and the statistical significance of the association.

Review of the Scientific Literature

Chronic, hypokinetic-related diseases are manifestations of latent progressive poor health over a protracted period of time. Because these diseases are relatively less prevalent in youth, there is substantially less scientific evidence supporting the association of musculoskeletal fitness with health outcomes in youth than in adults.

The relationship between health and musculoskeletal fitness in youth has been reviewed recently in relation to the development of the Fitnessgram®/Activitygram® (Welk and Blair, 2008) and a new health-related physical fitness test battery for European youth—the Assessing Levels of Physical Activity (ALPHA) study (Castro-Piñero et al., 2010; Ortega et al., 2008b; Ruiz et al., 2009). In a recent review, Ortega and colleagues (2008b) report significant inverse associations of lower-limb explosive strength (i.e., power) and abdominal endurance with lower-abdominal obesity in youth (e.g., $p < 0.001$ between performance on the standing long jump and waist circumference in 8-year-old males) (Brunet et al., 2007). The same review also highlights inverse associations ($p = 0.048$) between a composite muscle fitness index score and a standardized composite measure of cardiovascular risk among adolescent girls (Garcia-Artero et al., 2007), and between putative cardiovascular inflammatory markers and muscle strength in normal-weight and overweight adolescents (for C-reactive protein, $p = 0.02$ and $p = 0.09$, respectively) (Ruiz et al., 2008). Additionally, positive associations were found between muscle

TABLE 6-2 Summary of Top-Quality Studies Providing Best Evidence for Muscular Strength/Power

Reference and Study Type	Fitness Test(s)	Body Composition	Metabolic Health	Cardio-respiratory Health	Musculo-skeletal Health	Mental and Cognitive Health	Age, Gender, Maturity, Weight Status, Population	Study Summary, Quality, and Level of Evidence
Benson et al., 2008 Experimental	Bench press (BP), leg press (LP)	Waist circumference (WC), fat mass (FM), lean mass (LM), % body fat (BF), body mass index (BMI)	Homeostatic model assessment—insulin resistance (HOMA-IR), glucose,[b] insulin,[b] triglycerides (TG),[b] cholesterol and subfractions[b]				Ages 11-19, male and female (M and F), overweight, obese, New Zealand	An 8-week strength-training intervention program resulted in significant gains in both bench press and leg press strength and differential positive training effects on WC, % BF, FM, and BMI but not on any of the metabolic health markers. Level of evidence (LE): Direct
Janz et al., 2002 Longitudinal	Handgrip strength test	% BF, abdominal fat (AF)[a]	TG,[b] lipids,[b] cholesterol[b]	Systolic blood pressure (SBP), diastolic blood pressure (DBP)			Ages 10.5-15.5, M and F, normal weight	A 5-year prospective nonintervention study found significant negative correlations between changes in handgrip strength and changes in SBP, BF, and AF between 10 and 15 years of age, with those with high handgrip strength scores at the outset having the best health marker profiles 4-5 years later. LE: Associational

Ingle et al., 2006 Experimental	BP/LP, squat, standing long jump (SLJ), vertical jump (VJ)	% BF,[a] LM[a]		Ages 11-12, M, normal weight	A 12-week intervention program consisting of combined strength and plyometric training resulted in significant differential increases in BP, squat, and VJ, with a significant reduction in % BF and increase in LM, and then slight regression of these changes during detraining, while control values remained stable. LE: Direct
Heinonen et al., 2000 Experimental	SLJ/VJ/LJ, isometric leg extensor—90 degrees		Tibial bone mineral content (BMC),[b] femoral neck (FN) BMC,[a] lumbar spine (LS) BMC,[a] trochanter BMC,[a] tibia bone mineral density	F, pre- and postmenarcheal, normal weight	A 9-month intervention study of high-impact exercises found significant increases in LJ in both pre-and postmenarcheal girls, with no changes in leg extension strength in either age group; concurrent significant increases in FN and LS BMC in premenarcheal girls; and a differential positive effect in premenarcheal versus postmenarcheal girls for

continued

TABLE 6-2 Continued

Reference and Study Type	Fitness Test(s)	Body Composition	Metabolic Health	Cardio-respiratory Health	Musculo-skeletal Health	Mental and Cognitive Health	Age, Gender, Maturity, Weight Status, Population	Study Summary, Quality, and Level of Evidence
					(BMD),[b] tibia bone area[b] and strength[b]			FN and trochanter BMC but not in tibia peripheral qualitative computed tomography (PQCT) measures of BMD, bone area, or bone strength. LE: Associational
Kontulainen et al., 2002 Longitudinal	SLJ/VJ, isometric leg extension—90 degrees				LS BMC,[a] FN BMC,[b] trochanter BMC[b]		F, peri- and pubertal, normal weight	A prospective follow-up study 9 months after a jump training intervention program found significantly higher standing LJ scores but not leg extension scores and significantly higher LS BMC (but not BMC at other sites) in the trained group versus controls. LE: Associational

McGuigan et al., 2009 Quasi-experimental (no control)	One repetition maximum (1RM) squat, countermovement jump (CMJ), static VJ	% BF,[a] BMI,[b] LM,[a] FM[b]		Tibial BMC [b]	Ages 7-12, M and F, overweight, obese	An 8-week strength-training intervention study resulted in significant increases in countermovement jump height and relative counter-movement peak power and significant favorable changes in absolute % BF and lean tissue mass, but not BMI, FM, or whole-body BMC. Associational
Minck et al., 2000 Longitudinal	SLJ/VJ, maximal arm pull	% BF[a]			Ages 13-27, M and F, normal weight	A 14-year prospective nonintervention study found significant negative univariate correlations between changes in arm pull and VJ and changes in absolute BF and % BF corrected for height and weight in both sexes between 13 and 27 years of age. The strongest association was between changes in BF after adjustment for confounders and standing high jump (VJ). LE: Direct

continued

TABLE 6-2 Continued

Reference and Study Type	Fitness Test(s)	Body Composition	Metabolic Health	Cardio-respiratory Health	Musculo-skeletal Health	Mental and Cognitive Health	Age, Gender, Maturity, Weight Status, Population	Study Summary, Quality, and Level of Evidence
Morris et al., 1997 Experimental	Isokinetic shoulder and knee flexor/extensor and handgrip	LM,[a] FM[a]			Total body (TB), LS, FN, and proximal femur BMD,[a] multiple bone areas[a]		Age 9, F, pre-menarch, normal weight, Austrian	A 10-month intervention (schools randomized) that included 10 weeks of strength training (plus mixed aerobic activities) resulted in significant differential effects on shoulder flexor/extensor strength, knee extensor strength, and nondominant handgrip strength, with positive differential increases in LM and TB, LS, FN, and proximal femur BMD, as well as multiple bone areas, and a significant reduction in FM in the trained group. LE: Associational

Lubans et al., 2010 Experimental	BP/LP	BMI (girls),[a] % BF (girls),[a] LM (boys),[a] FM (boys)[a]		Physical self-worth	Ages 14-15, M and F, normal weight	An 8-week intervention program of traditional strength-training exercises and elastic tube training resulted in significant differential effects on BP and LP strength, with a significant decrease in FM and an increase in LM and fat-free mass (FFM) in boys and improvements in BMI and % BF in girls. There were significant negative correlations between changes in strength and fat loss in sex-pooled data and positive correlations between strength and physical self-worth changes. LE: Direct

continued

TABLE 6-2 Continued

Reference and Study Type	Fitness Test(s)	Body Composition	Metabolic Health	Cardio-respiratory Health	Musculo-skeletal Health	Mental and Cognitive Health	Age, Gender, Maturity, Weight Status, Population	Study Summary, Quality, and Level of Evidence
Naylor et al., 2008 Experimental	Combined BP and LP	LM,[a] % BF[a]		Myocardvelocity,[a] SBP,[a] left arterial pressure[a]			Ages 12-13, M and F, obese	An 8-week strength-training intervention program resulted in significant increases in "combined" 1RM dual leg press and bench press (largest gains in leg press strength) and significant decreases in SBP and increases in LM and % BF. Additional positive effects were found on selected cardiac function health markers reflecting transmitral peak early velocity (E), early diastolic peak myocardial velocity (E1), and the E/E1 ratio (a measure of left atrial pressure). LE: Associational
Nichols et al., 2001 Experimental	BP/LP	% BF,[b] FM,[b] FFM			Femoral neck BMC[a]/BMD,[a] whole-body BMC/BMD,[b] trochanter BMC/BMD,[b] LS BMC/		Ages 14-19, F, normal weight	A 15-month strength training intervention program resulted in significant differential training effects on LP strength only and increases in FN BMC/BMD only in normal

					BMD,[b] Ward's area BMC/BMD [b]			adolescent girls; however, only 5 subjects remained after 15 months in the trained group. LE: Direct
Shaibi et al., 2006 Experimental	BP/LP	BMI,[b] FFM,[a] % BF,[a] total fat mass (TFM[b])	Insulin and glucose response/ disposition index,[b] insulin sensitivity[a]				Age 15, M, overweight, Latino	A 16-week strength-training intervention resulted in a significant differential increase in BP and leg strength, with differential favorable effects on insulin sensitivity, total LM, and % BF (although the change in % BF was not significantly different from that in controls), but not on BMI or TFM. LE: Direct
Velez et al., 2010 Experimental	BP, seated row, shoulder press, squat	% BF,[a] FM,[a] LM[a]				Self-esteem[a]	Ages 14-18, M and F, normal weight, overweight, obese, Hispanic	A 12-week strength-training intervention program resulted in significant increases in BP, seated row, shoulder press, and squat strength, with significant reductions in % BF and FM and improved LM and self-esteem. LE: Associational

continued

TABLE 6-2 Continued

Reference and Study Type	Fitness Test(s)	Body Composition	Metabolic Health	Cardio-respiratory Health	Musculo-skeletal Health	Mental and Cognitive Health	Age, Gender, Maturity, Weight Status, Population	Study Summary, Quality, and Level of Evidence
van der Heijden et al., 2010 Quasi-experimental (no control)	Tension biceps curls, chest press, and triceps extensor, and flys with hand-held weights; hamstring curls, quadricep extensor, biodex squats, push-ups	FM,[b] LM,[a] % BF,[b] visceral fat,[b] BMI [b]	Hepatic insulin sensitivity,[a] peripheral insulin sensitivity,[b] blood lipids,[b] inflammatory markers,[b] TG,[b] hepatic glucose production[a]		Tibial BMD[a]		Ages 13+, M and F, obese, Hispanic	A 12-week strength-training intervention program resulted in increased strength for all strength measures in both sexes, with no changes in total or visceral fat mass, subcutaneous fat, or BMI but significant gains in FFM (LM) and tibial BMD and improvements in hepatic insulin sensitivity and glucose production rate and no changes in blood lipids, TG, inflammatory markers, or peripheral insulin sensitivity. LE: Associational
Witzke and Snow, 2000 Experimental	Isokinetic knee extensor—85-150 degrees	LM,[b] FM[b]			Tibial BMC,[b] greater trochanter BMC [a]		Ages 14-19, F, menarcheal, normal weight	A 9-month intervention study of plyometric jump training resulted in significant increases in knee extensor strength with a significant differential increase (with post hoc subgroup analysis but not with initial analysis of variance

							[ANOVA]) in trochanter BMC but not TB BMC and no changes in LM or FM. LE: Direct
Blimkie et al., 1996 Experimental	Biceps curl, triceps press, knee extensor/flex-or, squat, press	% BF[b]			TB and LS BMD/BMC[b]	Ages 14-18, F, menarcheal, normal weight	A 26-week strength-training intervention resulted in significant increases in all strength outcome measures, with a trend toward improvements but no statistically significant effect on TB and LS BMC/BMD or % BF (from skinfolds). There were weak to moderate significant positive correlations between most criterion strength measures and measures of BMC and BMD at baseline for the trained and control groups combined, but no significant correlations between changes in strength outcomes and changes in bone mineral measures in the resistance training group over the course of the study. LE: Direct

[a]Positive effect.
[b]Null effect.

strength and total and site-specific (e.g., lumbar spine) bone mineral status ($p < 0.001$ and $p < 0.001$, respectively) among prepubertal and adolescent youth (Ginty et al., 2005). Less convincing evidence hints at a positive association between improved muscular fitness and functional mobility and quality of life in youth with cancer (Ortega et al., 2008b). Other sources have reported significant positive univariate associations between various measures of muscle strength and bone health in normal nonathletic and athletic adolescents, but the strength of these relationships varies by skeletal site, muscle strength, and bone health measure and usually weakens when adjusted for body size or other known confounders (e.g., Blimkie et al., 1996; Duncan et al., 2002; Rice et al., 1993). A recent systematic review (Ruiz et al., 2009) for the period January 1990 to July 2008 concludes that there is strong evidence of a link between changes in muscle strength and changes in overall adiposity, but less strong (often inconsistent or nonexisting) evidence of a link with changes in central obesity, systolic blood pressure, blood lipid profiles, low-back pain, or quality of life and well-being in youth.

Body Composition

Among the high-quality studies included in the committee's review (Table 6-2), the relationships between measures of body composition and muscle strength/power were investigated most frequently. Six studies provide associational evidence supporting a link with an array of muscle strength/power fitness measures of the upper extremity (i.e., the handgrip, biceps curl, triceps press, shoulder press, shoulder flexion and extension), trunk (i.e., bench/chest press, seated row), and lower extremity (i.e., squat, leg press, knee/quadriceps extension, countermovement jump) (Janz et al., 2002; McGuigan et al., 2009; Morris et al., 1997; Naylor et al., 2008; van der Heijden et al., 2010; Velez et al., 2010). Body composition variables include body fat mass, percent body fat, abdominal fat, and lean or fat-free tissue mass, with most relationships observed in overweight/obese boys and girls within a rather narrow age range. Six high-quality studies provide direct evidence of a link between changes in muscle strength and power and favorable changes in health markers, including percent body fat, lean or fat-free mass, waist circumference, and body mass index (BMI) (Benson et al., 2008; Ingle et al., 2006; Lubans et al., 2010; Minck et al., 2000; Shaibi et al., 2006). Trunk (i.e., bench press) and lower-body (i.e., leg press, squat, and vertical jump) musculoskeletal measures are the most consistently related to these body composition outcomes, spanning the period from late childhood to adulthood in both normal-weight and overweight/obese youth of both sexes.

Bone Health

The second most frequently assessed relationship in the committee's review was between muscle strength and power measures and bone health outcomes. Four studies provide indirect evidence of positive associations between measures of upper-body (i.e., the handgrip, shoulder flexion and extension, biceps curl, chest fly), trunk (i.e., chest press), and lower-body (i.e., knee-quadriceps extension, hamstring curl, squat, long jump) strength and power using a variety of skeletal measures (i.e., total-body, lumbar spine, femoral neck, and proximal femur bone mineral content [BMC] or bone mineral density [BMD]), mainly in normal pre- and pubertal/menarcheal girls of normal weight (Heinonen et al., 2000; Kontulainen et al., 2002; Morris et al., 1997). A single study by van der Heijden and colleagues (2010) was the only one to investigate this relationship in both sexes in obese peripubertal youth. Only two high-quality studies provide direct evidence of a relationship between musculoskeletal strength and bone health. Witzke and Snow (2000) demonstrated a positive link between changes in knee extensor isokinetic strength and changes in trochanter BMC in menarcheal adolescent girls. Nichols and colleagues (2001) report a positive link between measures of bench press and leg press strength and femoral neck BMC/BMD in a small sample of five menarcheal girls. Few studies have investigated this relationship in boys and youth in early to middle childhood.

Metabolic Health

Only four high-quality studies investigated the relationship between musculoskeletal fitness and markers of metabolic health in youth (Benson et al., 2008; Janz et al., 2002; Shaibi et al., 2006; van der Heijden et al., 2010). These studies involved exclusively overweigh/obese youth of both sexes ranging in age from late childhood to late adolescence. One of these studies provides associational evidence of a link between multiple measures of upper-body (i.e., biceps curl, fly), trunk (i.e., chest press), and lower-body (i.e., hamstring curl, quadriceps extension, squat) muscle strength and hepatic insulin sensitivity and glucose production (van der Heijden et al., 2010). Another provides more direct evidence of a link between bench press (i.e., trunk) and leg press (i.e., lower-body) strength and improved insulin sensitivity in overweight adolescent boys (Shaibi et al., 2006).

Cardiorespiratory Health

Two high-quality studies provide indirect evidence for a link between musculoskeletal fitness and cardiorespiratory health markers in youth (Janz et al., 2002; Naylor et al., 2008). Increases in handgrip strength were associated with reductions in systolic blood pressure in normal-weight boys and girls aged 10-15 (Janz et al., 2002), and gains in a combined measure of bench press and leg press strength were associated with improvements in systolic blood pressure and markers of heart function (i.e., peak transmitral velocity of flow, diastolic myocardial velocity, and left atrial pressure) in obese boys and girls aged 12-13 (Naylor et al., 2008). Likewise, only two studies examined the relationship between muscle strength and power measures and measures of mental/cognitive health. Velez and colleagues (2010) report a positive association between measures of upper-body (i.e., shoulder press), trunk (i.e., bench press, seated row), and lower-body (i.e., squat) strength and power and self-esteem in normal-weight and overweight/obese adolescent boys and girls aged 14-18, whereas Lubans and colleagues (2010) provide direct evidence of a link between bench press and leg press strength and physical self-worth in normal-weight boys and girls aged 14-15.

Resistance Training Programs and Health Outcomes

Paralleling the adult literature, there is growing acceptance that appropriately prescribed and administered resistance training programs can improve muscle strength, endurance, and power in youth (Blimkie and Bar-Or, 2008; Faigenbaum et al., 2009; Malina, 2006). However, the health-related risks and benefits of this type of training and the relationship between improvements in musculoskeletal fitness and changes in health outcomes have not been as systematically investigated for youth. Effective resistance training programs may (1) reduce the risk of joint injury in adolescent athletes, (2) improve body composition specifically among children and adolescents who are obese or at risk of obesity, (3) improve insulin sensitivity in both normal-weight peripubertal children and obese adolescents, (4) reduce blood pressure in hypertensive adolescents, and (5) improve blood lipid profiles in both children and adolescents (Blimkie, 1993; Faigenbaum et al., 2009). The relationship between resistance training and improvements in musculoskeletal fitness and bone health in youth are controversial, mainly because of the complex manner and time frame in which bone responds to physical activities. In addition to very high forces on bone, bone adaptation may be regulated by other parameters of the activity (e.g., the loading rate) and muscle mass (Beck, 2009). Based on the limited number of good prospective controlled experimental studies, the link between improved musculoskeletal fitness and bone health remains tenuous in youth. Also paralleling the adult literature, the relationship between improved musculoskeletal fitness following resistance training and psychological health outcomes in youth is relatively weak (Faigenbaum et al., 2009).

Moderate- and high-intensity resistance training programs have been employed effectively, efficaciously, and safely with children as young as 8-10 years of age (Blimkie, 1993; Faigenbaum et al., 2009; Farpour-Lambert and Blimkie, 2008). Likewise, 1RM or relative repetition maximum (e.g., 10RM) strength testing has been employed safely with youth of this age. For younger youth, however, these forms of specialized training and testing have been used mainly in the research setting under the close supervision of experienced trainers and under closely controlled conditions. These activities are not risk-free, and age/developmental status should be considered carefully when they are being incorporated into youth fitness improvement/testing programs, especially those for preteen youth. Recommendations and guidelines for youth strength training and testing to mitigate risk were recently published by the National Strength and Conditioning Association (Faigenbaum et al., 2009) and the American Academy of Pediatrics and Council on Sports Medicine and Fitness (2008).

Limitations of the Scientific Literature

Most of the studies reviewed by the committee had limitations that precluded strong conclusions about the relationship between performance on musculoskeletal fitness tests and health outcomes or markers in youth. Many of the studies were not designed to answer questions about the relationship between the fitness tests employed and health. For example, primary study outcomes often were changes in diet, weight loss, or generalized physical activity rather than changes in musculoskeletal fitness characteristics. In many of the studies reviewed, either the nature of the intervention was not specific enough (e.g., a combination of endurance, strength, and power exercises without a focus on a particular dimension), or the dosage and duration of the exercise intervention were inadequate to elicit changes in musculoskeletal fitness, a requisite for establishing any relationships between a change in fitness and health.

Many of the reviewed studies were statistically underpowered to detect significant relationships, considered only very narrow gender-specific age ranges or discrete developmental groups, and often included unique subpopulations of overweight and obese youth. In addition, in many of the studies the analysis failed to consider the effects of potential confounders, and only indirect inferences could be drawn regarding the relationships between musculoskeletal fitness and health outcomes or markers. Quantifiable multivariate analyses, which were rarely conducted, would have permitted a more direct assessment of these relationships. Further, many studies related health outcome measures to the musculoskeletal fitness of isolated body regions, precluding generalization to whole-body musculoskeletal fitness status.

In summary, there is an insufficient body of high-quality literature to support a strong link between performance on any specific musculoskeletal fitness test by youth of either gender and across all ages and stages of development and any health outcomes or markers. The current literature in this area is too fragmented to permit identification of any specific musculoskeletal fitness test item that is unequivocally linked to health in the general population of healthy youth.

VALIDITY AND RELIABILITY OF SELECTED TEST ITEMS

Despite the limitations of the literature discussed above, the growing evidence in youth and stronger evidence in adults is suggestive of a fundamental relationship between musculoskeletal fitness and health outcomes across the life span. The committee finds that handgrip strength test and the standing long jump are two tests that globally represent musculoskeletal strength and power in youth and demonstrate adequate validity, reliability, and feasibility of administration for inclusion in fitness test batteries for all youth. This section reviews the validity and reliability of these two tests, for which there is some, albeit limited, evidence for a relationship to health in the literature reviewed. It also looks at the integrity of the modified pull-up and isometric leg extension tests, which also may be useful for assessing musculoskeletal fitness; however, the literature review provided very limited high-quality evidence for a link to health outcomes in youth for these two tests.

While numerous fitness tests purportedly measure muscle strength, endurance, and power in youth, information about their validity and reliability is limited. Nevertheless, an increasing body of literature pertaining to the validity and reliability of a few musculoskeletal fitness tests provides reasonable justification for including these tests in a test battery for assessment of musculoskeletal fitness in youth. As mentioned above, the committee's systematic literature review included muscle strength and endurance, but not muscle power, as components of fitness. Some of the tests reported, however, such as throwing and jumping tests, purportedly assess some aspects of muscle power (e.g., average, peak, instantaneous, and contractile power). Although the specific associations between individual fitness tests and aspects of muscle power are poorly defined, the committee's discussion of the validity of the tests takes account of the fact that the selected tests of musculoskeletal fitness could measure either muscle strength, endurance, or power.

The handgrip strength test is used extensively in European youth fitness testing. Based on the available literature, the handgrip strength test has moderate to strong construct validity ($r = 0.52$-0.84) with established upper-body (i.e., 1RM bench press) and lower-body (i.e., leg press and isokinetic knee extensor torque) strength tests (Holm et al., 2008; Milliken et al., 2008) and strong reliability (r = 0.71-0.90) in children and adolescents (Benefice et al., 1999; Brunet et al., 2007; Ruiz et al., 2006). It also has minimal test-retest learning and fatigue effects (Ortega et al., 2008a). Given differences in hand sizes among youth, optimal grip span adjustment, elbow angle, and device calibration are important for valid testing.

The standing long jump has been used extensively as a test of lower-body muscular strength, power, and explosive strength (see Table 2-6 in Chapter 2). Although not strictly a measure of power as that subdomain is defined, the standing long jump is the most widely used field-based test of muscle power/explosive strength. It demonstrates moderate to strong correlations with 1RM leg press/body weight (r = 0.39) (Milliken et al., 2008), isokinetic quadriceps torque (r = 0.50) (Holm et al., 2008), and total-body isometric strength (r = 0.77) (Castro-Piñero et al., 2010) in youth. In addition, the standing long jump correlates strongly (r = 0.70-0.91) with other lower- and upper-extremity field-based power tests (i.e., vertical jump, countermovement vertical jump, upper-body explosive throw) in youth (ages 6-17), controlling for age, gender, and BMI and/or weight (Castro-Piñero et al., 2010; Milliken et al., 2008). The standing long jump also has been found to have acceptable reliability in youth (r = 0.52-0.99) (Benefice et al., 1999; España-Romero et al., 2010; Malina et al., 2004; Pena Reyes et al., 2003; Safrit, 1995; Simons et al., 1990). In addition, the reliability of this test appears not to be affected by either systematic bias or sex differences among adolescents (Ortega et al., 2008a), although reliability estimates generally increase with age. Differences in gross motor coordination and experience with jumping across developmental time may influence the degree of test-retest reliability for the standing long jump. Controlling individually for anthropometric variables (i.e., height and body mass) provides a more valid assessment of lower-body strength and power for this test across ages (Castro-Piñero et al., 2010; Milliken et al., 2008).

The modified pull-up and isometric knee extension tests also are valid and reliable tests of upper- and lower-body musculoskeletal fitness, respectively; however, insufficient scientific evidence supports the link between these two tests and health outcomes in youth. The modified pull-up has demonstrated moderate to strong construct validity (r = 0.60-0.79) with other upper-body criterion strength measures (i.e., 1RM bench press, pull-down, arm curl) in boys and girls when measured per unit body weight (Pate et al., 1993). The highest correlation demonstrated (r = 0.75) comprises the sum of the multiple upper-extremity strength tests, which demonstrates strong construct validity for a composite measure of upper-body strength. The modified pull-up also is moderately to strongly correlated (r = 0.64-0.79) with push-ups, thus demonstrating a crossover effect with muscle strength and endurance. Moderate to high test-retest reliability (r = 0.52-0.99) (Engelman and Morrow, 1991; Erbaugh, 1990; Kollath et al., 1991; Saint Romain and Mahar, 2001) and modified Kappa coefficients (0.87-0.94) (Saint Romain and Mahar, 2001) have been demonstrated for the modified pull-up.

The isometric knee extension test is a criterion measure of lower-extremity quadriceps strength (and a general measure of lower-extremity leg extension strength) used primarily in clinical or laboratory settings. Many different methods (e.g., supine and upright sitting) and instruments (e.g., hand-held dynomometer, Cybex isokinetic equipment) have been used to test knee extension strength, and an optimal knee angle is an important consideration for adequately measuring maximum torque at the knee. Optimal knee angle remains relatively untested in children; however, knee angles of 80-90 degrees may produce the highest torque levels in this population (Marginson and Eston, 2001). Validation of isometric knee extension tests with other criterion lower-body strength measures (e.g., leg press and bilateral squat exercises) in youth is limited. The strength of the relationship between isometric and isokinetic knee extension has been shown to decrease with increasing isokinetic angular speeds (Hill et al., 1996). The isometric knee extension test demonstrates strong reliability (r = 0.76-0.97) for single- and double-leg tests (Escolar et al., 2001; Hill et al., 1996; Mercer and Lewis, 2001; Teeple et al., 1975) in both normal and disabled children. As with other fitness tests, familiarization with the testing procedures is advisable to optimize the validity of the test results (Farpour-Lambert and Blimkie, 2008).

ADMINISTRATIVE FEASIBILITY

In addition to validity and reliability, the selection of musculoskeletal fitness test items for inclusion in a youth fitness survey will depend on their administrative feasibility and practicality in the field. Principles relating to administrative feasibility for fitness testing for all fitness components are discussed in general in Chapter 3 and more specifically for application in school settings in Chapter 9. Developers and administrators of fitness surveys should carefully consider the issues outlined in Box 3-2 in Chapter 3 and in Chapter 9 when selecting specific musculoskeletal fitness test items for inclusion in a youth fitness test battery. The two specific musculoskeletal fitness tests discussed in the previous section and highlighted for their potential relationship to health in youth (i.e., handgrip strength and standing long jump tests) are among the most practical and feasible of a plethora of muscle strength, endurance, and power tests for field-based physical fitness assessment in this population. These tests can be taught effectively and administered safely to most school-aged youth, with consistency and reliability likely improving with increasing age and maturity from age 5 until the onset of puberty.

Administrative considerations for musculoskeletal fitness tests applicable to schools and other educational settings are described in Chapter 9. These tests (modified pull-up, push-up, and curl-up) generally require more skill and coordination than the handgrip and standing long jump tests and are perhaps more susceptible to learning and other effects. Thus these tests may be taught to all school-aged youth; however, performance may be less reliable than is the case for the handgrip strength and standing long jump tests for younger ages, but as with those tests, may improve with advancing age and maturity.

GUIDANCE FOR INTERPRETATION OF TEST RESULTS

Chapter 3 recommends several strategies that developers of fitness test batteries can employ to ensure accurate interpretation of health-fitness relationships in youth. The most robust approach requires establishing a strong link between some fitness parameter and one or several putative health markers in a broad population of youth and identifying health-related cut-points. However, the literature contains no recent (within the past 10 years) national normative data for the muscle strength, endurance, and power tests discussed in this chapter for U.S. youth, and there is scant evidence of any link between these tests and possible health markers for this population. At present, therefore, empirically determined health-related cut-points cannot be established for these tests. In the absence of criterion-referenced cut-points in youth or adults, interim cut-points corresponding to the 20th percentile should be used for tests of musculoskeletal fitness, analogous to the cut-points for cardiorespiratory endurance, until better evidence for criterion-referenced health-related cut-points is established by further research. Experts who will establish cut-points for musculoskeletal fitness tests in youth should follow the guidance in this report (Chapter 3) and base the cut-points on the unique purposes of the testing (e.g., cut-points for special populations such as athletes or people with disabilities).

CONCLUSIONS

Based on the CDC literature and the supplemental literature reviewed for this report, the committee concludes that there currently is sufficient evidence affirming that musculoskeletal fitness is related to health in humans. This conclusion is based mainly on increasing evidence for the importance of musculoskeletal fitness, especially muscle strength and power, to health outcomes in adults. There is some, albeit much more limited, support for this link among youth. At this time, however, there is insufficient high-quality evidence supporting an association between any single musculoskeletal fitness test item and health markers in youth. Studies reviewed also provide insufficient data for assessing the influence of several potential modifiers—age, gender, race/ethnicity, body composition, maturation status—on performance on musculoskeletal fitness tests.

The committee found growing evidence supporting the handgrip and standing long jump tests as putative health-related (i.e., bone health and body composition) musculoskeletal fitness test items in youth. The handgrip strength test demonstrates moderate to strong validity with both upper- and lower-body criterion strength measures. The standing long jump, although not strictly a measure of pure muscle strength, demonstrates acceptable concurrent validity with lower- and upper-body criterion strength measures and lower-body power measures in youth. The handgrip strength and standing long jump tests demonstrate strong and moderate reliability, respectively. Both are applicable across a broad age range, in both sexes, and in both normal and special pediatric subpopulations. These two tests also are currently included in the ALPHA test battery for musculoskeletal fitness assessment in European youth. Test administrators may wish to include these tests in a national youth fitness survey based on their integrity and feasibility; however, the results of these tests should not be interpreted in a health context until such relationships are more firmly established. The committee found no evidence of adverse events associated with the administration of these tests in the studies reviewed.

Other tests, such as the modified pull-up and isometric knee extension, also are being used as measures of muscular strength in current fitness test batteries in the United States but are linked only weakly with health markers in youth at this time. Therefore, despite their acceptable validity, reliability, and feasibility, the committee does not recommend these tests for a national youth fitness survey until such health links are more firmly established. In addition, although the bench press and leg press tests are viewed as standard criterion measures of strength or endurance (based on the number of repetitions demanded) in adults, they cannot be recommended for inclusion in a national youth fitness survey at this time because of the limited quality and level of the scientific evidence for the relationship of these tests to health outcomes in youth; the paucity of information on their reliability across childhood; and concerns regarding their administrative feasibility, practicality, and safety.

For schools and other educational settings, administrators should consider the hand grip strength and standing long jump tests as well as alternative tests that have not yet been shown to be related to health, but are valid, reliable, and feasible. The modified pull-up and push-up tests are possible alternatives for measuring upper-body musculoskeletal strength. The curl-up could also be considered for measuring an additional construct, core strength. The committee found no evidence of adverse events associated with the administration of these tests in the studies reviewed. The committee's full recommendations on musculoskeletal fitness tests for use in national youth fitness surveys and in schools and other educational settings are presented in Chapters 8 and 9, respectively.

Moderate to strong tracking of selected measures of muscle strength and power both during adolescence (Maia et al., 2001; Malina, 1996; Pate et al., 1999) and from adolescence into adulthood (Beunen et al., 1992; Malina, 1996; Mikkelsson et al., 2006; Twisk et al., 2000) suggests that measures of musculoskeletal fitness in youth may prove to be useful predictors of future adult health. Tracking relationships appear to be weaker during the preadolescent years and more stable for lower- versus upper-body strength/power measures (Malina et al., 2004). Tracking variability in youth may be explained by age-related differences in the development of inter- and intramuscular coordination and differing levels of experience with specific fitness tests. Further, there is increasing evidence of moderate tracking of biologic health markers, especially for coronary heart disease, from childhood/adolescence into adulthood that in the future may be shown to be related to musculoskeletal fitness in youth (Bao et al., 1995; Froberg and Andersen, 2005; Malina et al., 2004; Twisk et al., 1995, 1997). Whether changes in muscle strength, endurance, and power during youth are predictive of adult health outcomes in later life, however, remains to be determined.

REFERENCES

AAHPERD (American Alliance for Health, Physical Education, Recreation and Dance). 1984. *Health related physical fitness test: Technical manual.* Reston, VA: AAHPERD.

American Academy of Pediatrics, and Council on Sports Medicine and Fitness. 2008. Strength training by children and adolescents. *Pediatrics* 121(4):835-840.

Ashe, M. C., T. Y. L. Liu-Ambrose, D. M. L. Cooper, K. M. Khan, and H. A. McKay. 2008. Muscle power is related to tibial bone strength in older women. *Osteoporosis International* 19(12):1725-1732.

Bao, W., S. A. Threefoot, S. R. Srinivasan, and G. S. Berenson. 1995. Essential hypertension predicted by tracking of elevated blood pressure from childhood to adulthood: The Bogalusa Heart Study. *American Journal of Hypertension* 8(7):657-665.

Beck, B. R. 2009. Muscle forces or gravity—what predominates mechanical loading on bone? Introduction. *Medicine and Science in Sports and Exercise* 41(11):2033-2036.

Benefice, E., T. Fouere, and R. M. Malina. 1999. Early nutritional history and motor performance of Senegalese children, 4-6 years of age. *Annals of Human Biology* 26(5):443-455.

Benson, A. C., M. E. Torode, and M. A. Fiatarone Singh. 2008. The effect of high-intensity progressive resistance training on adiposity in children: A randomized controlled trial. *International Journal of Obesity* 32(6):1016-1027.

Beunen, G., J. Lefevre, A. L. Claessens, R. Lysens, H. Maes, R. Renson, J. Simons, B. Vanden Eynde, B. Vanreusel, and C. Van den Bossche. 1992. Age-specific correlation analysis of longitudinal physical fitness levels in men. *European Journal of Applied Physiology and Occupational Physiology* 64(6):538-545.

Blimkie, C. J. R. 1993. Benefits and risks of resistance training in children. In *Intensive participation in children's sports*, edited by B. R. Cahill and A. J. Pearl. Champaign, IL: Human Kinetics Publishers. Pp. 133-165.

Blimkie, C. J. R., and O. Bar-Or. 2008. Muscle strength, endurance, and power: Trainability during childhood. In *The young athlete*, edited by H. Heibestret and O. Bar-Or. Oxford, UK: Blackwell Publishing, Ltd.

Blimkie, C. J., S. Rice, C. E. Webber, J. Martin, D. Levy, and C. L. Gordon. 1996. Effects of resistance training on bone mineral content and density in adolescent females. *Canadian Journal of Physiology and Pharmacology* 74(9):1025-1033.

Bohannon, R. W. 2008. Hand-grip dynamometry predicts future outcomes in aging adults. *Journal of Geriatric Physical Therapy* 31(1):3-10.

Bonnefoy, M., M. Jauffret, and J. F. Jusot. 2007. Muscle power of lower extremities in relation to functional ability and nutritional status in very elderly people. *Journal of Nutrition Health & Aging* 11(3):223-228.

Bouchard, C., S. N. Blair, and W. L. Haskell. 2007. Why study physical activity and health? In *Physical activity and health*, edited by C. Bouchard, S. N. Blair, and W. L. Haskell. Champaign, IL: Human Kinetics Publishers. Pp. 3-19.

Brunet, M., J. P. Chaput, and A. Tremblay. 2007. The association between low physical fitness and high body mass index or waist circumference is increasing with age in children: The 'Quebec en Forme' Project. *International Journal of Obesity* 31(4):637-643.

Castro-Piñero, J., E. G. Artero, V. España-Romero, F. B. Ortega, M. Sjostrom, J. Suni, and J. R. Ruiz. 2010. Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* 44(13):934-943.

Cooper, R., D. Kuh, C. Cooper, C. R. Gale, D. A. Lawlor, F. Matthews, R. Hardy, and the FALCon and HALCyon Study Teams. 2011. Objective measures of physical capability and subsequent health: A systematic review. *Age Ageing* 40(1):14-23.

Duncan, C. S., C. J. Blimkie, C. T. Cowell, S. T. Burke, J. N. Briody, and R. Howman-Giles. 2002. Bone mineral density in adolescent female athletes: Relationship to exercise type and muscle strength. *Medicine and Science in Sports and Exercise* 34(2):286-294.

Engelman, M. E., and J. R. Morrow, Jr. 1991. Reliability and skinfold correlates for traditional and modified pull-ups in children grades 3-5. *Research Quarterly for Exercise and Sport* 62(1):88-91.

Erbaugh, S. J. 1990. Reliability of physical fitness tests administered to young children. *Perceptual and Motor Skills* 71(3, Pt. 2):1123-1128.

Escolar, D. M., E. K. Henricson, J. Mayhew, J. Florence, R. Leshner, K. M. Patel, and P. R. Clemens. 2001. Clinical evaluator reliability for quantitative and manual muscle testing measures of strength in children. *Muscle & Nerve* 24(6):787-793.

España-Romero, V., E. G. Artero, D. Jimenez-Pavon, M. Cuenca-Garcia, F. B. Ortega, J. Castro-Piñero, M. Sjostrom, M. J. Castillo-Garzon, and J. R. Ruiz. 2010. Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine* 31(7):490-497.

Faigenbaum, A. D., W. J. Kraemer, C. J. Blimkie, I. Jeffreys, L. J. Micheli, M. Nitka, and T. W. Rowland. 2009. Youth resistance training: Updated position statement paper from the national strength and conditioning association. *Journal of Strength and Conditioning Research* 23(5):S60-S79.

Farpour-Lambert, N. J., and C. J. R. Blimkie. 2008. Muscle strength. In *Paediatric exercise science and medicine*. 2nd ed., edited by N. Armstrong and W. van Mechelen. Oxford, UK: Oxford University Press. Pp. 37-53.

Froberg, K., and L. B. Andersen. 2005. Mini review: Physical activity and fitness and its relations to cardiovascular disease risk factors in children. *International Journal of Obesity* 29(Suppl. 2):S34-S39.

Garber, C. E., B. Blissmer, M. R. Deschenes, B. A. Franklin, M. J. Lamonte, I. M. Lee, D. C. Nieman, D. P. Swain, and American College of Sports Medicine. 2011. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise* 43(7):1334-1359.

Garcia-Artero, E., F. B. Ortega, J. R. Ruiz, J. L. Mesa, M. Delgado, M. Gonzalez-Gross, M. Garcia-Fuentes, G. Vicente-Rodriguez, A. Gutierrez, and M. J. Castillo. 2007. Lipid and metabolic profiles in adolescents are affected more by physical fitness than physical activity (AVENA Study). *Revista Espanola De Cardiologia* 60(6):581-588.

Ginty, F., K. L. Rennie, L. Mills, S. Stear, S. Jones, and A. Prentice. 2005. Positive, site-specific associations between bone mineral status, fitness, and time spent at high-impact activities in 16- to 18-year-old boys. *Bone* 36(1):101-110.

Heinonen, A., H. Sievanen, P. Kannus, P. Oja, M. Pasanen, and I. Vuori. 2000. High-impact exercise and bones of growing girls: A 9-month controlled trial. *Osteoporosis International* 11 (12):1010-1017.

Hill, C., R. Croce, J. Miller, and F. Cleland. 1996. Muscle torque relationships between hand-held dynamometry and isokinetic measurements in children ages 9 to 11. *Journal of Strength and Conditioning Research* 10(2):77-82.

Holm, I., P. Fredriksen, M. Fosdahl, and N. Vollestad. 2008. A normative sample of isotonic and isokinetic muscle strength measurements in children 7 to 12 years of age. *Acta Paediatrica* 97(5):602-607.

Ingle, L., M. Sleap, and K. Tolfrey. 2006. The effect of a complex training and detraining programme on selected strength and power variables in early pubertal boys. *Journal of Sports Sciences* 24(9):987-997.

Jackson, A. S. 2006. The evolution and validity of health-related fitness. *Quest* 58(1):160-175.

Janz, K. F., J. D. Dawson, and L. T. Mahoney. 2002. Increases in physical fitness during childhood improve cardiovascular health during adolescence: The Muscatine Study. *International Journal of Sports Medicine* 23(Suppl. 1):S15-S21.

Jurca, R., M. J. Lamonte, C. E. Barlow, J. B. Kampert, T. S. Church, and S. N. Blair. 2005. Association of muscular strength with incidence of metabolic syndrome in men. *Medicine and Science in Sports and Exercise* 37(11):1849-1855.

Kell, R. T., G. Bell, and A. Quinney. 2001. Musculoskeletal fitness, health outcomes and quality of life. *Sports Medicine* 31(12):863-873.

Knuttgen, H. G., and W. J. Kraemer. 1987. Terminology and measurement in exercise performance. *Journal of Strength and Conditioning Research* 1(1):1-10.

Kollath, J. A., M. J. Safrit, W. Zhu, and L. G. Gao. 1991. Measurement errors in modified pull-ups testing. *Research Quarterly for Exercise and Sport* 62(4):432-435.

Kontulainen, S. A., P. A. Kannus, M. E. Pasanen, H. T. Sievanen, A. O. Heinonen, P. Oja, and I. Vuori. 2002. Does previous participation in high-impact training result in residual bone gain in growing girls? One year follow-up of a 9-month jumping intervention. *International Journal of Sports Medicine* 23(8):575-581.

Lubans, D. R., E. J. Aguiar, and R. Callister. 2010. The effects of free weights and elastic tubing resistance training on physical self-perception in adolescents. *Psychology of Sport and Exercise* 11(6):497-504.

Maia, J. A., J. Lefevre, A. Claessens, R. Renson, B. Vanreusel, and G. Beunen. 2001. Tracking of physical fitness during adolescence: A panel study in boys. *Medicine and Science in Sports and Exercise* 33(5):765-771.

Malina, R. M. 1996. Tracking of physical activity and physical fitness across the lifespan. *Research Quarterly for Exercise and Sport* 67(Suppl. 3):48-57.

Malina, R. M. 2006. Weight training in youth-growth, maturation, and safety: An evidence-based review. *Clinical Journal of Sport Medicine* 16(6):478-487.

Malina, R., C. Bouchard, and O. Bar-Or. 2004. *Growth, maturation, and physical activity.* 2nd ed. Champaign, IL: Human Kinetics Publishers.

Marginson, V., and R. Eston. 2001. The relationship between torque and joint angle during knee extension in boys and men. *Journal of Sports Sciences* 19(11):875-880.

McCartney, N., and S. Phillips. 2007. Physical activity, muscular fitness, and health. In *Physical activity and health*, edited by C. Bouchard, S. N. Blair and W. L. Haskell. Champaign, IL: Human Kinetics Publishers. Pp. 231-246.

McGuigan, M. R., M. Tatasciore, R. U. Newton, and S. Pettigrew. 2009. Eight weeks of resistance training can significantly alter body composition in children who are overweight or obese. *Journal of Strength and Conditioning Research* 23(1):80-85.

Mercer, V. S., and C. L. Lewis. 2001. Hip abductor and knee extensor muscle strength of children with and without down syndrome. *Pediatric Physical Therapy* 13(1):18-26.

Mikkelsson, L., J. Kaprio, H. Kautlainen, U. Kujala, M. Mikkelsson, and H. Nupponen. 2006. School fitness tests as predictors of adult health-related fitness. *American Journal of Human Biology* 18(3):342-349.

Milliken, L. A., A. D. Faigenbaum, R. L. Loud, and W. L. Westcott. 2008. Correlates of upper and lower body muscular strength in children. *Journal of Strength and Conditioning Research* 22(4):1339-1346.

Minck, M. R., L. M. Ruiter, W. Van Mechelen, H. C. G. Kemper, and J. W. R. Twisk. 2000. Physical fitness, body fatness, and physical activity: The Amsterdam Growth and Health Study. *American Journal of Human Biology* 12(5):593-599.

Moffroid, M. T., and E. T. Kusiak. 1975. The power struggle: Definition and evaluation of power of muscular performance. *Journal of Physical Therapy* 55(10):1098-1104.

Morris, F. L., G. A. Naughton, J. L. Gibbs, J. S. Carlson, and J. D. Wark. 1997. Prospective ten-month exercise intervention in premenarcheal girls: Positive effects on bone and lean mass. *Journal of Bone and Mineral Research* 12(9):1453-1462.

Morrow, J. R., W. Zhu, B. D. Franks, M. D. Meredith, and C. Spain. 2009. 1958-2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise and Sport* 80(1):1-11.

Naylor, L. H., K. Watts, J. A. Sharpe, T. W. Jones, E. A. Davis, A. Thompson, K. George, J. M. Ramsay, G. O'Driscoll, and D. J. Green. 2008. Resistance training and diastolic myocardial tissue velocities in obese children. *Medicine and Science in Sports and Exercise* 40(12):2027-2032.

Nichols, D. L., C. F. Sanborn, and A. M. Love. 2001. Resistance training and bone mineral density in adolescent females. *Journal of Pediatrics* 139(4):494-500.

Ortega, F. B., E. G. Artero, J. R. Ruiz, G. Vicente-Rodriguez, P. Bergman, M. Hagstromer, C. Ottevaere, E. Nagy, O. Konsta, J. P. Rey-Lopez, A. Polito, S. Dietrich, M. Plada, L. Beghin, Y. Manios, M. Sjostrom, M. J. Castillo, and HELENA Study Group. 2008a. Reliability of health-related physical fitness tests in European adolescents. The HELANA Study. *International Journal of Obesity* 32(Suppl. 5):S49-S57.

Ortega, F. B., J. R. Ruiz, M. J. Castillo, and M. Sjostrom. 2008b. Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity* 32(1):1-11.

Pate, R. R., M. L. Burgess, J. A. Woods, J. G. Ross, and T. Baumgartner. 1993. Validity of field-tests of upper-body muscular strength. *Research Quarterly for Exercise and Sport* 64(1):17-24.

Pate, R. R., S. G. Trost, M. Dowda, A. E. Ott, D. S. Ward, R. Saunders, and G. Felton. 1999. Tracking of physical activity, physical inactivity, and health-related physical fitness in rural youth. *Pediatric Exercise Science* 11(4):364-376.

Payne, N., N. Gledhill, P. T. Katzmarzyk, and V. Jamnik. 2000. Health-related fitness, physical activity, and history of back pain. *Canadian Journal of Applied Physiology* 25(4):236-249.

Pena Reyes, M. E., S. K. Tan, and R. M. Malina. 2003. Urban-rural contrasts in the physical fitness of school children in Oaxaca, Mexico. *American Journal of Human Biology* 15(6):800-813.

Plowman, S. A. 1992. Physical activity, physical fitness, and low back pain. *Exercise and Sport Sciences Reviews* 20:221-242.

Plowman, S. A. 2008. Muscular strength, endurance, and flexibility assessments. In *Fitnessgram/Activitygram reference guide*, edited by G. J. Welk and M. D. Meredith. Dallas, TX: Cooper Institute. Pp. 11.11-11.40.

Reid, K. F., and R. A. Fielding. 2012. Skeletal muscle power: A critical determinant of physical functioning in older adults. *Exercise and Sport Sciences Reviews* 40(1):4-12.

Rice, S., C. J. Blimkie, C. E. Webber, D. Levy, J. Martin, D. Parker, and C. L. Gordon. 1993. Correlates and determinants of bone mineral content and density in healthy adolescent girls. *Canadian Journal of Physiology and Pharmacology* 71(12):923-930.

Ruiz, J. R., F. B. Ortega, A. M. Gutierrez, D., M. Sjöström, and M. J. Castillo. 2006. Health-related fitness assessment in childhood and adolescence: A European approach based on the AVENA, EYHS and HELENA studies. *Journal of Public Health* 14(5):269-277.

Ruiz, J. R., F. B. Ortega, J. Warnberg, L. A. Moreno, J. J. Carrero, M. Gonzalez-Gross, A. Marcos, A. Gutierrez, and M. Sjöström. 2008. Inflammatory proteins and muscle strength in adolescents. *Archives of Pediatrics & Adolescent Medicine* 162(5):462-468.

Ruiz, J. R., J. Castro-Piñero, E. G. Artero, F. B. Ortega, M. Sjöström, J. Suni, and M. J. Castillo. 2009. Predictive validity of health-related fitness in youth: A systematic review. *British Journal of Sports Medicine* 43(12):909-923.

Safrit, M. J. 1995. *Complete guide to youth fitness testing.* Champaign, IL: Human Kinetics Publishers.

Saint Romain, B., and M. T. Mahar. 2001. Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science* 5(2):67-80.

Sale, D. G., and R. W. Norman. 1982. Testing strength and power. In *Physiological testing of the elite athlete*, edited by J. D. Macdougall, H. A. Wenger and H. J. Green. Canada: Mutual Press. Pp. 7-37.

Shaibi, G. Q., M. L. Cruz, G. D. C. Ball, M. J. Weigensberg, G. J. Salem, N. C. Crespo, and M. I. Goran. 2006. Effects of resistance training on insulin sensitivity in overweight Latino adolescent males. *Medicine and Science in Sports and Exercise* 38(7):1208-1215.

Simons, J., G. P. Beunen, R. Renson, A. L. M. Claessens, B. Vanreusel, and J. Lefevre, eds. 1990. *Growth and fitness of Flemish girls: The Leuven Growth Study.* Champaign, IL: Human Kinetics Publishers.

Teeple, J. B., T. G. Lohman, J. E. Misner, R. A. Boileau, and B. H. Massey. 1975. Contribution of physical development and muscular strength to the motor performance capacity of 7 to 12 year old boys. *British Journal of Sports Medicine* 9(3):122-129.

Twisk, J. W., H. C. G. Kemper, and J. Snel. 1995. Tracking of cardiovascular risk factors in relation to lifestyle. In *The Amsterdam Growth Study: A longitudinal analysis of health, fitness, and lifestyle*, edited by H. C. G. Kemper. Champaign, IL: Human Kinetics Publishers. Pp. 203-224.

Twisk, J. W., H. C. Kemper, W. van Mechelen, and G. B. Post. 1997. Tracking of risk factors for coronary heart disease over a 14-year period: A comparison between lifestyle and biologic risk factors with data from the Amsterdam Growth and Health Study. *American Journal of Epidemiology* 145(10):888-898.

Twisk, J. W., H. C. Kemper, and W. van Mechelen. 2000. Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Medicine and Science in Sports and Exercise* 32(8):1455-1461.

van der Heijden, G. J., Z. Y. J. Wang, Z. L. Chu, G. Toffolo, E. Manesso, P. J. J. Sauer, and A. L. Sunehag. 2010. Strength exercise improves muscle mass and hepatic insulin sensitivity in obese youth. *Medicine and Science in Sports and Exercise* 42(11):1973-1980.

Velez, A., D. L. Golem, and S. M. Arent. 2010. The impact of a 12-week resistance training program on strength, body composition, and self-concept of Hispanic adolescents. *Journal of Strength and Conditioning Research* 24(4):1065-1073.

von Stengel, S., W. Kemmler, R. Pintag, C. Beeskow, J. Weineck, D. Lauber, W. A. Kalender, and K. Engelke. 2005. Power training is more effective than strength training for maintaining bone mineral density in postmenopausal women. *Journal of Applied Physiology* 99(1):181-188.

von Stengel, S., W. Kemmler, W. A. Kalender, K. Engelke, and D. Lauber. 2007. Differential effects of strength versus power training on bone mineral density in postmenopausal women: A 2-year longitudinal study. *British Journal of Sports Medicine* 41(10):649-655; discussion 655.

Warburton, D. E., N. Gledhill, and A. Quinney. 2001. Musculoskeletal fitness and health. *Canadian Journal of Applied Physiology* 26(2):217-237.

Welk, G. J., and S. N. Blair. 2008. Health benefits of physical activity and fitness in children. In *Fitnessgram/Activitygram reference guide*, edited by G. J. Welk and D. M. Meredith. Dallas, TX: Cooper Institute. http://www.cooperinstitute.org/vault/2440/web/files/662.pdf (accessed June 13, 2012).

Williams, M. A., W. L. Haskell, P. A. Ades, E. A. Amsterdam, V. Bittner, B. A. Franklin, M. Gulanick, S. T. Laing, K. J. Stewart, American Heart Association Council on Clinical Cardiology, American Heart Association Council on Nutrition, and Physical Activity, and Metabolism. 2007. Resistance exercise in individuals with and without cardiovascular disease: 2007 update: A scientific statement from the American Heart Association Council on Clinical Cardiology and Council on Nutrition, Physical Activity, and Metabolism. *Circulation* 116(5):572-584.

Witzke, K. A., and C. M. Snow. 2000. Effects of plyometric jump training on bone mass in adolescent girls. *Medicine and Science in Sports and Exercise* 32(6):1051-1057.

Wolfe, R. R. 2006. The underappreciated role of muscle in health and disease. *American Journal of Clinical Nutrition* 84(3):475-482.

7

Health-Related Fitness Measures for Youth: Flexibility

KEY MESSAGES

Flexibility has been defined as the range of motion of muscle and connective tissues at a joint or group of joints. In contrast to other, more general or systemic fitness components, flexibility is highly specific to each of the joints of the body. For this reason, although flexibility has been included in national fitness test batteries, linking it to one or more health outcomes is difficult, and few data support such an association. Future efforts to study the relationship of flexibility to health will require a multivariate approach.

The principal health outcomes hypothesized to be associated with flexibility are prevention of and relief from low-back pain, prevention of musculoskeletal injury, and improved posture. These associations have been studied most frequently in adults, and the strength of any associations for specific flexibility tests in youth is minimal. Various reasons may explain the difficulty of establishing a link between flexibility and health. First, in contrast with other fitness components, no large-scale studies have been specifically designed to assess the relationship between flexibility and health. Second, flexibility may be associated with health in combination with other musculoskeletal variables. Finally, studies addressing flexibility have varied substantially in the tests used, the study designs, and the characteristics of participants (e.g., age, gender, weight).

Although the evidence is not yet clear, flexibility in youth may in fact be linked to various health outcomes, such as back pain, injury prevention, and posture, and appropriate studies are needed to explore such associations. The limitations described above led the committee not to recommend a flexibility test for a national youth fitness survey. Instead, the committee recommends conducting further research on this fitness component, as well as considering the use of flexibility tests in schools and other educational settings for educational purposes.

Until the relationship to health is confirmed and national normative data and health data are collected for youth, the comparatively relative position method should be used for setting cut-points (cutoff scores) for performance on flexibility tests. With this method, percentiles established for other fitness measures are used to establish interim cut-points for the measure of interest. For example, interim cut-points corresponding to the 20th percentile should be used for flexibility tests, analogous to the cut-points for cardiorespiratory endurance tests.

Flexibility as a component of fitness first gained prominence in the early 1900s as the field of physical therapy emerged (Linker, 2011). Later in that century, circumstances (i.e., two world wars and a polio epidemic) provided further impetus for growth in the professions of occupational and physical therapy and a rise in schools for preparing therapists. In 1980 the first health-related physical fitness test was published (AAHPERD, 1980), and it included a test of flexibility (sit-and-reach). Subsequent U.S. and international health-related test batteries—including the President's Council on Fitness, Sports and Nutrition (PCFSN) and Fitnessgram® batteries—have included items to measure flexibility.

This chapter reviews existing data on the relationship between flexibility and health outcomes in youth. The focus is on the extent to which flexibility is associated with better health and function, excluding those outcomes related to athletic performance. The chapter begins by defining flexibility and describing the relevant physiology as a basis for explaining the challenges involved in identifying an association between a single flexibility test and a health outcome. The most frequently used flexibility tests are then described. Next, the chapter presents findings from the literature on what is known about the relationship between flexibility and health in adults and in youth, which serve as the basis for the committee's guidance for interpreting results of flexibility tests, as well as for its conclusions about the associations between flexibility tests and health outcomes in youth. The validity and reliability of these tests are also examined. The process for selecting the studies included here is described briefly in this chapter and in more detail in Chapter 3. Based on its conclusions about the relationship between flexibility tests and health, the committee makes no recommendation for including a flexibility test in national fitness surveys, but only recommendations regarding the use of specific flexibility tests with educational value in schools and other educational settings (see Chapter 9). These recommendations are based on the validity and reliability of the tests and on additional factors that should be considered when implementing fitness tests in schools (also described in Chapter 9). Future research needs related to this fitness component are addressed in Chapter 10.

DEFINITIONS

Flexibility has been defined in many different ways, although the focus has consistently been on the characteristics and functioning of muscle. Kraus and Hirschland (1954), whose research in the 1950s precipitated the U.S. national youth fitness testing movement, referred to flexibility as a muscle fitness component associated with "muscle stiffness" and "tension." Kraus and Raab (1961) referred to muscle "tension" and "tightness" when discussing flexibility in their classic book *Hypokinetic Disease*. Fleischman (1964) identified two flexibility components using factor analysis: extent flexibility and dynamic flexibility. Extent flexibility was defined as "the ability to flex or stretch the trunk and back muscles as far as possible" (p. 77) (e.g., twist and touch tests). Dynamic flexibility was defined as "the ability to make repeated, rapid, flexing movements" (p. 79) (e.g., rapid bending, twisting, and touching movements).

According to Cureton, an early fitness pioneer, "Flexibility indicates that joints are not muscle bound or stiff for some other reason" (Cureton, 1965, p. 42). It is important that his definition included reference to joints and not just muscles, consistent with clinical definitions that evolved from the development of the field of physical therapy and focused on "range of joint motion" as the key component of flexibility. Textbooks on physical fitness also focused on joint range of motion. For example, Johnson and colleagues (1966, p. 23) defined flexibility as "the functional capacity of the joints to move through a full range of motion." Corbin and colleagues (1968) defined flexibility as "the wide range of movement or the ability to bend in many directions" (p. 6). And an early high school fitness text (Corbin and Lindsey, 1979, p. 14) referred to flexibility as "the ability to use your joints fully."

The U.S. surgeon general's report on physical activity and health (HHS, 1996) defined flexibility as "a health-related component of physical fitness that relates to the range of motion available at a joint," a definition originally proposed by Wilmore and Costill (1994). The most recent national youth fitness test batteries have used similar definitions. Fitnessgram (Plowman, 2008, pp. 11-13) defines flexibility as the ability to "move freely through a full range of motion." A review of issues related to flexibility (Knudson et al., 2000) uses the definition of Holt and colleagues (1996, p. 172). Flexibility is defined as "the intrinsic property of body tissues which determines the range of motion achievable without injury at a joint or group of joints."

The definition of flexibility used in this report is an adaptation of that of Holt and colleagues (1996). In this report, flexibility is operationally defined as "the intrinsic property of body tissues, including muscle and connective tissues that determines the range of motion achievable without injury at a joint or group of joints." Flexibility is highly specific to each joint.

Fitness is considered to be a "state of being" (Corbin et al., 2000), which is different from the behavior that produces that state. In the case of flexibility, stretching is a physical activity behavior or exercise typically performed to increase muscle-tendon unit (MTU) length and to allow improved joint range of motion. Common forms of *stretching* include static stretch (passive and active), proprioceptive neuromuscular facilitation, ballistic stretch, and dynamic stretch (see ACSM, 2010; Garber et al., 2011). Other forms of physical activity that require stretching of the MTU (e.g., gymnastics, dance) can also result in improved flexibility.

FLEXIBILITY FITNESS TESTS

Flexibility tests measure joint range of motion and can in general be classified into two categories: *laboratory tests* and *field tests*. Laboratory tests are those often used in controlled settings and are administered to patients or study participants on a one-to-one basis with specifically designed devices. As a result, the administration of laboratory tests can be expensive and time-consuming. Field tests, in contrast, are used in schools, fitness clubs, or similar practical group settings and can be administered to more participants at a relatively lower cost and in a relatively shorter time. Characteristics of laboratory and field tests are briefly described below.

Laboratory Tests

Most clinical assessments of flexibility fall within the category of "goniometry," which is derived from the Greek words "gonia" (i.e., angle) and "netron" (measure) (Eston and Reilly, 1966; Norkin and White, 2003). Thus, measuring flexibility can simply be viewed as measuring the angle of joints or their range of motion (ROM). The devices used for the assessments are called goniometers. Although they vary in size, shape, and material used, goniometers usually consist of three parts—the body and two thin extensions called "arms." The body resembles a protractor that forms a half

(0 to 180 degrees) or full (0 to 360 degrees) circle. One arm is called the "stationary arm" and other the "moving arm." During the assessment, the examiner determines the range of motion by placing the goniometer along the bone immediately proximal and distal to the joint being measured.

Field Tests

Field tests for flexibility have been used in a number of fitness test batteries. In the United States, the shoulder stretch (sometimes called the zipper), trunk lift (assesses both flexibility and muscle fitness), and sit-and-reach (assesses low-back and hamstring flexibility) have been used, as have modifications of these tests. There are also several other tests not used in national batteries, such as the Schober test, the modified Schober test, and the straight leg raise (see also Table 2-6 in Chapter 2).

In the *shoulder stretch*, the person being tested reaches over the shoulder and down the back with one hand, and reaches behind the back and upward with the other hand, trying to touch the fingers of the hands together. The distance between the hands or distance of overlap is measured on both sides of the body (Meredith and Welk, 2010, pp. 59-60).

The *trunk lift* is presumed to measure both muscle strength and flexibility. In this test, the person being tested lies prone on the floor, lifts the upper body (trunk) off the floor, and holds the position while the height of the chin from the floor is measured (Meredith and Welk, 2010, pp. 49-50).

Sit-and-reach and other similar tests that require a person to flex the hip to touch the toes are the most common field tests of flexibility. Such tests are designed to assess low-back and upper hamstring (complex of three posterior thigh muscles) flexibility.

The first U.S. health-related fitness battery used a *bilateral sit-and-reach* test (AAHPERD, 1980). Sitting on the floor or a mat, legs straight and feet 8-12 inches apart, the person being tested reaches forward with the arms (hands overlapping). The distance of reach is measured in inches using a measuring line marked on the floor (PCPFS, 2012).

An alternative to the bilateral sit-and-reach test is the unilateral test called the *backsaver sit-and-reach* (Meredith and Welk, 2010, pp. 57-59). The Fitnessgram test manual (Meredith and Welk, 2010) outlines the reasons for including this test. A flexibility box with a ruler extension is used. The person being tested sits on the floor or a mat with one leg straight. The other leg is bent to the side, foot near the knee of the straight leg. The person being tested reaches forward with the arms (hands overlapping). The distance reached in centimeters or inches (on the flexibility box ruler) determines the person's score. The test is then repeated with the other leg extended.

FLEXIBILITY AND HEALTH

Flexibility is associated with length of muscle and connective tissue, joint structure, age, disease state, and gender. MTU length is typically the prime focus of flexibility testing in the field setting. Factors such as MTU stiffness/compliance, elasticity, and viscoelasticity relate to flexibility and MTU function (Alter, 2004; Knudson et al., 2000). The utility of flexibility as a component of physical fitness has its roots in sports performance, and considerable research has investigated the association between acute stretch and muscle cramps (DeVries, 1967), injury (McHugh and Cosgrave, 2010), performance (Kay and Blazevich, 2012; McHugh and Nesse, 2008), postural stability (Nelson et al., 2011), and delayed muscle soreness (Henschke and Lin, 2011; Herbert et al., 2011). This chapter, however, focuses on outcomes related to better general function and health, not athletic performance.

In contrast to other fitness components that are general or systemic in nature, flexibility is highly specific to each of the joints of the body. For example, a person can be very flexible with a good range of motion in and around the shoulder joint but tight and lacking range of motion in the hip. The specificity of flexibility to joints of the body makes it difficult to isolate a single flexibility-related factor contributing to a health outcome. The ability to touch the toes in a sit-and-reach test, for example, involves many joints and MTUs. MTU length in one area of the body may contribute to poor performance on the test but not account for a large amount of variance in total test performance. As a result, establishing a relationship between flexibility and health outcomes is likely to require a multisite, multivariate approach specific to each health outcome. Accordingly, establishing a link to one or more health outcomes for one specific flexibility test item is difficult.

An added complication is that field tests used to assess flexibility may not have the specificity to isolate particular joints of interest to health-related outcomes. For example, although low-back pain has been hypothesized to be associated with flexibility, the sit-and-reach test that is commonly used to assess low-back and hip flexibility has been shown to measure hip flexibility rather than low-back flexibility (Chillon et al., 2010). The extent to which range of motion around the hip joints is a better predictor of low-back pain than range of motion around the lumbar region is not known. Results of a study by Cornbleet and Woolsey (1996) indicate that the sit-and-reach test is correlated with hamstring length. However, attention must be paid to the final position of the hip joint rather than the final position of the fingertips and any mobility in the spine in assessment of hamstring length. Of interest, this study also suggests that hamstring length differs between boys and girls.

Although flexibility may be associated with health outcomes, strong evidence of a health link to an individual field test is not apparent. Flexibility is not necessarily linearly related to health outcomes. Excess range of motion around a joint or joints (e.g., joint hypermobility syndrome [JHS]) is characterized by excessive movement and wear of joints that can lead to injury and disability (Wolf et al., 2011). JHS affects children more than adults and females more than males (Remvig et al., 2007). Accordingly, care must be taken in interpreting an individual's range of flexibility in terms of health outcomes.

Flexibility and Health in Adults

The evidence relating flexibility to health outcomes among adults is equivocal. The American College of Sports Medicine's (ACSM's) position statement (Garber et al., 2011) indicates that flexibility exercises may enhance postural stability and balance (see also Bird et al., 2011). Plowman (2008) reports that some studies show an association between flexibility and low-back pain, while others do not. Recent studies using Functional Movement Screening (FMS), a multi-item musculoskeletal screening battery, have shown promise for predicting injuries among military personnel (O'Connor et al., 2011), firefighters (Peate et al., 2007), and professional athletes (Kiesel et al., 2007, 2011). These preliminary studies suggest that batteries of musculoskeletal test items may prove to be better predictors of injury than single musculoskeletal test items (including items designed to test flexibility), at least in people for whom high-intensity exercise and vigorous-intensity physical activity are important job features.

The association between flexibility and functional capacity among adults is unclear, although several recent studies have investigated exercise training and functional capacity. Studies in cancer survivors (Eyigor et al., 2010) and people with Parkinson's disease (Reuter et al., 2011), fibromyalgia (Carbonell-Baeza et al., 2012), and other conditions have sought to determine the effect of multimodal exercise on various aspects of functional capacity. These studies are rooted largely in the physical therapy literature, where a goal of patient care is increasing or returning musculoskeletal function. Reuter and colleagues (2011) compared a stretching and relaxation treatment (ostensibly a control condition) with a walking or gym-based exercise treatment in a randomized study of 90 Parkinson's patients. After 6 months, the control patients showed improvements in their reported pain, balance, and health-related quality-of-life measures equal to those of the exercise treatment groups. As with the bulk of the literature on flexibility and health outcomes, few studies have focused specifically on stretching (and changes in flexibility) as the key exposure as it may relate to functional capacity. Moreover, the heterogeneity of populations and conditions studied makes general conclusions tenuous.

Stretching as Part of a Regular Exercise Program

There is some evidence that stretching, if included as part of a regular program of exercise, results in improved flexibility. The ACSM (Garber et al., 2011) found there were limited randomized controlled trials showing the effect of frequency, type, volume, and pattern, and only observational or nonrandomized trials showing the effect of intensity and time (length of stretch). However, the ACSM notes that "no consistent link has been shown between regular flexibility exercise and reduction of musculotendinous injuries, prevention of low back pain, or DOMS [delayed onset muscle soreness]" (Garber et al., 2011, p. 1344). Yet, it is important to note that stretching has been used in physical therapy for injury rehabilitation, treatment of neuromuscular symptoms of disease, and restoration of functional capacity for daily living, although the need for solid scientific support continues (Reurink et al., 2012). Stretching also has been used for improving/correcting posture (Nelson et al., 2011) and for treating neck, back, and other types of pain (Franca et al., 2012). Stretching is useful in relieving muscle cramps (Schwellnus et al., 2008) associated with muscle pain.

Other activities that involve stretching (i.e., Tai Chi, Qigong, yoga) have been associated with health outcomes as well. But because they also rely on strength, muscular endurance, balance, and other neuromuscular factors, it is impossible to quantify the independent effect of stretching (and resultant flexibility). Three different literature reviews (Chang et al., 2010; Jahnke et al., 2010; Wang et al., 2004) indicate that Tai Chi and Qigong have a variety of associated health benefits (e.g., bone health, cardiopulmonary fitness, some aspects of physical function, quality of life, self-efficacy, and factors related to prevention of falls [Jahnke et al., 2010, p. 22]), especially among older adults. Yoga has been associated with benefits in treating low-back pain (Sherman et al., 2005, 2011; Tilbrook et al., 2011) and with psychological health benefits among cancer survivors (Lin et al., 2011).

Acute Stretch

The stretching warm-up (acute static stretch) has long been considered important in preparing for vigorous-intensity physical activity, including sports, dance, and various forms of fitness training. Recent research, however, has questioned some of the purported performance and health benefits, including prevention of soreness and injury. In a recent systematic review, Kay and Blazevich (2012) cite 18 studies and indicate that static stretching can reduce strength, power, and speed. However, they also note that strength, power, and speed are not compromised after short-duration stretches (45 seconds or less). Another recent meta-analysis (Simic et al., 2012) that includes 104 studies published from 1966 to 2010 suggests that static stretching should be avoided as the sole warm-up routine for strength, power, and explosive strength performance, but notes that negative effects are greatest for stretches lasting more than 45 seconds. After reviewing 12 relevant studies, Herbert and colleagues (2011, p. 2) found that "there was little or no effect on muscle soreness experienced in the week after physical activity." There is evidence, however, that acute static stretching decreases musculoskeletal stiffness (Kay and Blazevich, 2009).

Witvrouw and colleagues (2004) and Thacker and colleagues (2004) report no association between acute static stretching and injury reduction. A recent review (McHugh and Cosgrave, 2010) indicates that acute stretching can reduce the risk of acute muscle strain injuries, but also reports no reduction in overuse injuries after a static stretch warm-up. Pereles and colleagues (2010) note that there were no differences in injury risk between prerun stretching and nonstretching groups of teens and adults and suggest that an immediate shift in a regimen (i.e., from stretching to no stretching) may be more important than the regimen itself.

Flexibility and Health in Youth

Literature Review Process

The majority of the studies cited come from a literature review by the Centers for Disease Control and Prevention (CDC). This literature search screened a total of 6,016 studies addressing flexibility. As mentioned in Chapter 3, the CDC did not abstract these articles because of time and resource limitations. However, when flexibility was measured in studies that were identified from the aerobic, muscular endurance, and muscular strength search libraries, that information was coded and extracted into a central database. Of these studies, seven were classified as experimental, five as quasi-experimental, and four as longitudinal. In addition, the committee reviewed studies provided through a public information gathering session. Because there were so few relevant studies, the committee also examined cross-sectional studies to gain further insight; however, these studies yielded no findings relevant to the committee's task. The criteria used to select high-quality studies are discussed in Chapter 3. Given the paucity of studies and the lack of evidence, this section presents findings from all the studies reviewed regardless of the quality of the evidence in support of a relationship of flexibility to health, as a basis for the committee's conclusions on flexibility.

Review of the Scientific Literature

A variety of forms of stretching (e.g., static stretch, active stretch, passive stretch, proprioceptive neuromuscular facilitation [PNF]) produce increases in flexibility. Results of studies included in this report indicate that programs of physical activity for youth, even those not designed primarily to improve flexibility (Cheung and Ng, 2003; Dorgo et al., 2009; Faude et al., 2010; Katz et al., 2010; Serbescu et al., 2006), result in improved flexibility (Ahlqwist et al., 2008; Jones et al., 2007).

It should be noted, however, that there are differences in flexibility based on gender and ethnicity. Alter's (2004) text the *Science of Flexibility* indicates that in general, girls are more flexible than boys, younger youth are more flexible than older youth, and youth are more flexible than adults. More recently, Tremblay and colleagues (2010) found that girls were more flexible than boys across all age groups during the school years, but found no differences across age groups for either boys or girls. In a large cross-sectional study of youth fitness in Texas, Welk and colleagues (2010) found higher sit-and-reach scores for girls than boys at the high school level but not at lower school levels. The study also found that boys had better sit-and-reach scores in high school than in elementary or middle school, and that girls had lower sit-and-reach scores in high school than in elementary or middle school (Welk et al., 2010). Results of the most recent California physical fitness test indicate that the percentage of students meeting sit-and-reach standards is higher among girls than boys and that for both sexes, more youth meet the standards at upper than at lower grades.[1] Finally, results of a statewide fitness survey of students in fifth and seventh grades in Georgia suggest that 21 percent of students failed to meet flexibility standards (as measured by the sit-and-reach test) (The Philanthropic Collaborative for a Healthy Georgia, 2008). No gender differences were noted among the younger (fifth-grade) students, but the percentage of older girls meeting the standards was higher than that of older boys (25 percent versus 20 percent). Results of the Georgia survey also suggest differences by race/ethnicity, with Hispanic students being less likely to reach flexibility standards than their white or African American peers.

In terms of secular changes, a longitudinal study of the fitness of Canadian youth compared fitness scores (cardiorespiratory endurance, body composition, flexibility, muscle fitness) collected between 2007 and 2009 with scores from 1981. Sit-and-reach scores for boys and girls in all age groups were lower in 2007-2009 than in 1981 (Tremblay et al., 2010). In a study by McMillan and Erdmann (2010), girls improved in sit-and-reach performance over a 6-year period, but performance among boys decreased.

[1] Available at http://www.cde.ca.gov/ta/tg/pf/pftresults.asp (accessed June 18, 2012).

Pain and injury Of seven experimental studies in the CDC review dealing with flexibility, only one (Ahlqwist et al., 2008) looked directly at health outcomes commonly associated with flexibility (e.g., pain, injury). Its results suggest that back pain scores in teens improved as flexibility (as measured by the sit-and-reach test) improved. Improvements in flexibility were greater in the physical therapy group than in the home exercise and educational materials groups. In some of the studies, the intervention did not result in the desired change in flexibility. For example, Faude and colleagues (2010) compared children in a soccer intervention group with controls. Both groups improved in sit-and-reach performance, as well as in body mass.

Of the five studies in the CDC review classified as quasi-experimental, one focused specifically on a dependent variable associated with flexibility. Jones and colleagues (2007) studied a small group of teens with back pain who were exposed to 8 weeks of rehabilitation versus no-exercise controls. Side bending, hip flexion (sit-and-reach), and sit-up performance increased in the rehabilitation group but not the controls. Pain intensity decreased in the intervention group.

In an observational study reviewed, Feldman and colleagues (2001) tracked adolescents over 1 year and found tight quadriceps and tight hamstrings to be associated with back pain. An initial study by Kujala and colleagues (1992) found that flexibility measures were not associated with back pain. However, a 3-year follow-up found that poor lumbar flexion was part of a multivariate profile that predicted pain for boys, and that decreased range of motion in the lower lumbar segments, low maximal lumbar extension, and high body weight at baseline predicted low-back pain for the following 3 years (Kujala et al., 1997). A retrospective study of 1,025 men and women for whom sit-and-reach and sit-up performance was measured as teens found that good flexibility (sit-and-reach) in boys and good endurance strength (sit-up) in girls were associated with decreased risk of neck tension (Mikkelsson et al., 2006). Neither sit-and-reach nor sit-up performance was associated with back pain. A high body mass index (BMI) was associated with increased neck tension, and the authors speculate that it may be related to poor hamstring length and back stiffness. In a study with 402 subjects (6-18 years old), Miereau and colleagues (1989) found that adolescent males with a history of low-back pain also had decreased hamstring length; the same relationship was not found in girls. Lower straight leg raise scores were found among older teens. Salminen and colleagues (1992) studied 15-year-olds with and without back pain and found lumber extension and hamstring length to be associated with back pain, but no relationship was found between back pain and trunk flexion. A later study by Salminen and colleagues (1995) found no association between low-back pain and flexibility measures, but showed low activity levels to be a risk factor for low-back pain. Bloemers and colleagues (2012) also found an increased risk of injury among inactive youth, but no direct link to flexibility or other fitness parameters was established. Finally, Burton and colleagues (1996) tracked 11-year-olds over 4 years (to age 15) and found that flexibility measures were not predictive of back pain. Lower flexibility was reported between ages 11 and 15, and girls were more flexible than boys.

Body composition and cardiometabolic health Two experimental studies (Manios et al., 2002; Serbescu et al., 2006) found that after an exercise training intervention, improvements were seen in body composition or lipids and lipoproteins that were measured as health outcomes, which in theory are not physiologically linked to flexibility. It should be noted, however, that in one of the studies (Manios et al., 2002), the exercise intervention did not change the flexibility of the participants. Five prospective studies provide information relevant to flexibility and health outcomes. Inconsistent results were found with regard to the association between flexibility (as measured by sit-and-reach) and body composition. Two studies showed an association between decreases in flexibility and higher skinfold measurements (Matton et al., 2006) or BMI (Kim et al., 2005). Others (Aires et al., 2010; Chen et al., 2007), however, found no association between performance on the sit-and-reach test and BMI. These inconsistencies could be due to differences in study designs, such as the length of the studies, the ages of the children, or the appropriateness of the health outcome itself (body composition).

Limitations of the Scientific Literature

The committee notes that the quality of the research reviewed was less than optimal for several reasons. In some cases, there were problems with the design of the study (e.g., no controls). There have been no large trials with adequate statistical power to demonstrate a relationship between flexibility and any health outcome or marker. Moreover, studies typically were not designed to test hypotheses central to flexibility. For example, flexibility measures often were included as one of the fitness components measured, but the health outcomes assessed were chosen because of their hypothesized association with fitness variables other than flexibility, such as BMI. Early studies that influenced eventual large-scale fitness testing of youth focused on the importance of flexibility to back health. The six-item Kraus-Weber test, which was clinically derived, was thought to predict future back pain in adults and was subsequently used as a fitness test for youth (Kraus and Hirschland, 1954). Flexibility has not been theoretically linked to metabolic markers as have cardiorespiratory endurance and body composition, nor has it typically been linked to bone density, as has musculoskeletal fitness.

As noted earlier, the principal health outcomes thought to be associated with flexibility have been relief from back pain symptoms, as well as prevention of injury and posture problems.

VALIDITY, RELIABILITY, AND FEASIBILITY OF SELECTED FLEXIBILITY TEST ITEMS

Evidence on the validity and reliability of the commonly used field tests of flexibility discussed here has been reported (see, e.g., Castro-Piñero et al., 2010; España-Romero et al., 2010; Freedson et al., 2000; Plowman, 2008; Safrit, 1990). In general, the test-retest reliability of the tests is consistently high. Validity, on the other hand, ranges from low to moderate depending on the criterion used. Using the sit-and-reach test as an example, a reliability of 0.99 was reported for 13- to 15-year-old girls, of 0.94-0.97 for 11- to 14-year-old boys, and of 0.80-0.96 for 11- to 14-year-old girls. However, validity was moderate (0.60-0.73) when hamstring flexibility testing was used as the criterion, and was only 0.27-0.30 when goniometer-measured low-back flexibility was used (Plowman, 2008; Safrit, 1990). The finding of moderate validity with hamstring and lumber flexibility tests was recently affirmed in a systematic literature review (Castro-Piñero et al., 2009).

A list of questions to be addressed in assessing the feasibility of a test is presented in Box 3-2 in Chapter 3. While a compelling link between health and flexibility measures has not been established, the widely used sit-and-reach test has been the most frequently studied. The backsaver sit-and-reach is also widely used and has acceptable feasibility based on the criteria in Box 3-2. Additional factors to consider when implementing fitness tests in schools are described in Chapter 9.

GUIDANCE FOR INTERPRETATION OF TEST RESULTS

This report provides guidance to assist those interpreting health-fitness relationships in youth (Chapter 3). Ideally, once there is enough evidence of an association between a test and a health outcome or health marker in youth, cut-points (cutoff scores) for a specific test can be established by mining data on that association collected from a broad population of youth. However, national normative data from flexibility tests for U.S. youth and concurrent data on possible associated health outcomes or health markers are not available. Further, cut-points for adults have not been established for flexibility tests. Until the relationship to health is confirmed and population-based data are collected, the comparatively relative position method should be used in interpreting the results of flexibility tests. With this method, percentiles established for other fitness measures are used to establish interim cut-points for the measure of interest.

CONCLUSIONS

Exercises designed to produce changes in flexibility have been shown to be effective in increasing flexibility, and youth who participate in active sports generally have better flexibility than those who do not. There has been a decrease in flexibility among youth in the past 20 years, at a time when body weight has increased dramatically.

Flexibility is specific to joints, and relationships to general systemic health outcomes or health markers are therefore less likely to exist than is the case for other fitness components, such as cardiorespiratory endurance. Clinical theory suggests that complex interaction among multiple musculoskeletal factors (e.g., flexibility, strength, muscular endurance, and neuromuscular factors), rather than any individual variable, is most likely to show a relationship to health. Therefore, establishing an association with health outcomes (e.g., back pain, risk of injury, posture problems) and a single flexibility test item is challenging. Further, possible associations are complicated by the fact that the relationship between flexibility and health outcomes is not linear; that is, risk may be higher for both those with low flexibility and those with exceptionally high flexibility than for those in the middle ranges.

The strength of any association between specific flexibility tests and health outcomes in youth is minimal. There may be various reasons for this. First, in contrast with other fitness variables, there have been no large-scale studies of flexibility and health. Second, flexibility may be associated with health when other musculoskeletal variables are taken into account. Finally, the tests used to measure flexibility, the study designs, and the characteristics of the subjects (e.g., age, gender, weight) have varied substantially, making it difficult to establish any possible link between flexibility and various health outcomes. Data were insufficient to permit assessment of the influence of several potential modifiers, such as age, gender, race/ethnicity, body composition, and maturation status, on performance on the various flexibility tests.

The validity and reliability of some of the flexibility tests used in youth fitness test batteries in the United States and abroad have been confirmed. Among the tests reviewed, the various forms of the sit-and-reach have reasonable validity and reliability when used in both survey and school settings. The degree to which the sit-and-reach test is an indicator of overall systemic flexibility is unclear, however.

Based on the lack of evidence for an association between flexibility tests and health outcomes in youth, the committee does not recommend including such tests in a national survey at this time. At the same time, the committee recognizes that, although the evidence is not yet clear, flexibility in youth may in fact be linked to various health outcomes, such as back pain, injury risk, and posture problems. Further, the committee found no evidence of adverse events observed in studies of flexibility, nor is there reported evidence of adverse events in field testing of flexibility using popular test items. With this in mind, the committee suggests that in schools and other educational settings, flexibility test items may be included to educate youth and their parents about flexibility as a component of overall musculoskeletal fitness, function, and performance. The selection of such a test should be based on its validity, reliability, and feasibility. To establish interim cut-points for such tests, the guidance provided in Chapter 3 of this report should be followed. Full recommendations on the use of these tests in schools and other educational settings are presented in Chapter 9.

REFERENCES

AAHPERD (American Alliance for Health, Physical Education, Recreation and Dance). 1980. *Health related physical fitness test: Technical manual.* Reston, VA: AAHPERD.

ACSM (American College of Sports Medicine). 2010. *ACSM's guidelines for exercise testing and prescription.* 8th ed. Philadelphia, PA: Lippincott, Williams, and Wilkins.

Ahlqwist, A., M. Hagman, G. Kjellby-Wendt, and E. Beckung. 2008. Physical therapy treatment of back complaints on children and adolescents. *Spine* 33(20):E721-E727.

Aires, L., D. Mendonca, G. Silva, A. R. Gaya, M. P. Santos, J. C. Ribeiro, and J. Mota. 2010. A 3-year longitudinal analysis of changes in body mass index. *International Journal of Sports Medicine* 31(2):133-137.

Alter, M. J. 2004. *Science of flexibility.* 3rd ed. Champaign, IL: Human Kinetics Publishers.

Bird, M., K. D. Hill, M. Ball, S. Hetherington, and A. D. Williams. 2011. The long-term benefits of a multi-component exercise intervention to balance and mobility in healthy older adults. *Archives of Gerontology and Geriatrics* 52(2):211-216.

Bloemers, F., D. Collard, M. C. A. Paw, W. Van Mechelen, J. Twisk, and E. Verhagen. 2012. Physical inactivity is a risk factor for physical activity-related injuries in children. *British Journal of Sports Medicine* 46(9):669-674.

Burton, A. K., R. D. Clarke, T. D. McClune, and K. M. Tillotson. 1996. The natural history of low back in adolescents. *Spine* 21(20):2323-2328.

Carbonell-Baeza, A., J. R. Ruiz, V. A. Aparicio, F. B. Ortega, D. Munguia-Izquierdo, I. C. Alvarez-Gallardo, V. Segura-Jimenez, D. Camiletti-Moiron, A. Romero, F. Estevez-Lopez, B. Samos, A. J. Casimiro, A. Sierra, P. A. Latorre, M. Pulido-Martos, P. Femia, I. J. Perez-Lopez, P. Chillon, M. J. Girela-Rejon, P. Tercedor, A. Lucia, and M. Delgado-Fernandez. 2012. Land- and water-based exercise intervention in women with fibromyalgia: The al-Andalus physical activity randomised controlled trial. *BMC Musculoskeletal Disorders* 13:18.

Castro-Piñero, J., P. Chillon, F. B. Ortega, J. L. Montesinos, M. Sjostrom, and J. R. Ruiz. 2009. Criterion-related validity of sit-and-reach and modified sit-and-reach test for estimating hamstring flexibility in children and adolescents aged 6-17 years. *International Journal of Sports Medicine* 30(9):658-662.

Castro-Piñero, J., E. G. Artero, V. España-Romero, F. B. Ortega, M. Sjostrom, J. Suni, and J. R. Ruiz. 2010. Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* 44(13):934-943.

Chang, Y. K., Y. H. Nien, C. L. Tsai, and J. L. Etnier. 2010. Physical activity and cognition in older adults: The potential of Tai Chi Chuan. *Journal of Aging and Physical Activity* 18(4):451-472.

Chen, J. L., D. Wall, C. Kennedy, V. Unnithan, and C. H. Yeh. 2007. Predictors of increased body mass index in Chinese children. *Progress in Cardiovascular Nursing* 22(3):138-144.

Cheung, C. Y. W., and G. Y. F. Ng. 2003. An eight-week exercise programme improves physical fitness of sedentary female adolescents. *Physiotherapy* 89(4):249-255.

Chillon, P., J. Castro-Piñero, J. R. Ruiz, V. M. Soto, A. Carbonell-Baeza, J. Dafos, G. Vicente-Rodriguez, M. J. Castillo, and F. B. Ortega. 2010. Hip flexibility is the main determinant of the back-saver sit-and-reach test in adolescents. *Journal of Sports Sciences* 28(6):641-648.

Corbin, C. B., and R. Lindsey. 1979. *Fitness for life*. Glenview, IL: Scott, Foresman and Company.

Corbin, C. B., L. J. Dowell, H. Tolson, and C. Landis. 1968. *Concepts and experiments in physical education*. Dubuque, IA: William C. Brown Book Company.

Corbin, C. B., R. P. Pangrazi, and B. D. Franks. 2000. Definitions: Health, fitness and physical activity. *President's Council on Physical Fitness and Sports Research Digest* 3(9):1-8.

Cornbleet, S. L., and N. B. Woolsey. 1996. Assessment of hamstring muscle length in school-aged children using the sit-and-reach test and the inclinometer measure of hip joint angle. *Physical Therapy* 76(8):850-855.

Cureton, T. K. 1965. *Physical fitness and dynamic health*. New York: Dial Press.

DeVries, H. A. 1967. Quantitative electromyographic investigation of the spasm theory of muscle pain. *American Journal of Physical Medicine* 45(3):119-134.

Dorgo, S., G. A. King, N. G. Candelaria, J. O. Bader, G. D. Brickey, and C. E. Adams. 2009. Effects of manual resistance training on fitness in adolescents. *Journal of Strength and Conditioning Research* 23(8):2287-2294.

España-Romero, V., E. G. Artero, D. Jimenez-Pavon, M. Cuenca-Garcia, F. B. Ortega, J. Castro-Piñero, M. Sjostrom, M. J. Castillo-Garzon, and J. R. Ruiz. 2010. Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine* 31(7):490-497.

Eston, R. G., and T. Reilly. 1966. *Kinanthropometry and exercise physiology laboratory workbook, volume 1: Anthropometry.* Albingdon, England: E & FN Spon.

Eyigor, S., H. Karapolat, H. Yesil, R. Uslu, and B. Durmaz. 2010. Effects of pilates exercises on functional capacity, flexibility, fatigue, depression and quality of life in female breast cancer patients: A randomized controlled study. *European Journal of Physical and Rehabilitation Medicine* 46(4):481-487.

Faude, O., O. Kerper, M. Multhaupt, C. Winter, K. Beziel, A. Junge, and T. Meyer. 2010. Football to tackle overweight in children. *Scandinavian Journal of Medicine and Science in Sports* 20(Suppl. 1):103-110.

Feldman, D. E., I. Shrier, M. Rossignol, and L. Abenhaim. 2001. Risk factors for the development of low back pain in adolescence. *American Journal of Epidemiology* 154(1):30-36.

Fleishman, E. A. 1964. *The structure and measurement of physical fitness*. Englewood Cliffs, NJ: Prentice-Hall.

Franca, F. R., T. N. Burke, R. R. Caffaro, L. A. Ramos, and A. P. Marques. 2012. Effects of muscular stretching and segmental stabilization on functional disability and pain in patients with chronic low back pain: A randomized, controlled trial. *Journal of Manipulative and Physiological Therapeutics* 35(4):279-285.

Freedson, P. S., K. J. Cureton, and G. W. Heath. 2000. Status of field-based fitness testing in children and youth. *Preventive Medicine* 31(2):S77-S85.

Garber, C. E., B. Blissmer, M. R. Deschenes, B. A. Franklin, M. J. Lamonte, I. M. Lee, D. C. Nieman, D. P. Swain, and American College of Sports Medicine. 2011. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise. *Medicine and Science in Sports and Exercise* 43(7):1334-1359.

Henschke, N., and C. C. Lin. 2011. Stretching before or after exercise does not reduce delayed-onset muscle soreness. *British Journal of Sports Medicine* 45(15):1249-1250.

Herbert, R. D., M. de Noronha, and S. J. Kamper. 2011. Stretching to prevent or reduce muscle soreness after exercise. *Cochrane Database of Systematic Reviews* (7):CD004577.

HHS (U.S. Department of Health and Human Services). 1996. *Physical activity and health: A report by the Surgeon General.* Atlanta, GA: HHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

Holt, J., L. E. Holt, and T. W. Pelham. 1996. Flexibility redefined. In *Biomechanics in sports XIII*, edited by T. Bauer. Thunder Bay, Ontario: Lakehead University. Pp. 170-174.

Jahnke, R., L. Larkey, C. Rogers, J. Etnier, and F. Lin. 2010. A comprehensive review of health benefits of qigong and tai chi. *American Journal of Health Promotion* 24(6):e1-e25.

Johnson, P. B., W. F. Updyke, D. C. Stolberg, and M. Schaefer. 1966. *Physical education: A problem-solving approach to health and fitness*. New York: Holt, Rinehart and Winston.

Jones, M., G. Stratton, T. Reilly, and V. Unnithan. 2007. The efficacy of exercise as an intervention to treat recurrent nonspecific low back pain in adolescents. *Pediatric Exercise Science* 19(3):349-359.

Katz, D. L., D. Cushman, J. Reynolds, V. Njike, J. A. Treu, J. Walker, E. Smith, and C. Katz. 2010. Putting physical activity where it fits in the school day: Preliminary results of the ABC (Activity Bursts in the Classroom) for fitness program. *Preventing Chronic Disease* 7(4):A82.

Kay, A. D., and A. J. Blazevich. 2009. Moderate-duration static stretch reduces active and passive plantar flexor moment but not achilles tendon stiffness or active muscle length. *Journal of Applied Physiology: Respiratory, Environmental and Exercise Physiology* 106(4):1249-1256.

Kay, A. D., and A. J. Blazevich. 2012. Effect of acute static stretch on maximal muscle performance: A systematic review. *Medicine and Science in Sports and Exercise* 44(1):154-164.

Kiesel, K., P. J. Plisky, and M. L. Voight. 2007. Can serious injury in professional football be predicted by a preseason functional movement screen? *North American Journal of Sports Physical Therapy* 2(3):147-158.

Kiesel, K., P. Plisky, and R. Butler. 2011. Functional movement test scores improve following a standardized off-season intervention program in professional football players. *Scandinavian Journal of Medicine and Science in Sports* 21(2):287-292.

Kim, J., A. Must, G. M. Fitzmaurice, M. W. Gillman, V. Chomitz, E. Kramer, R. McGowan, and K. E. Peterson. 2005. Relationship of physical fitness to prevalence and incidence of overweight among schoolchildren. *Obesity Research* 13(7):1246-1254.

Knudson, D. V., P. Magnusson, and M. McHugh. 2000. Current issues in flexibility fitness. *President's Council on Physical Fitness and Sports Research Digest* 3(10):1-8.

Kraus, H., and R. P. Hirschland. 1954. Minimum muscular fitness tests in school children. *Research Quarterly* 25:178-188.

Kraus, H. and W. Raab. 1961. *Hypokinetic Disease.* Springfield, IL: Charles C. Thomas Publishers.

Kujala, U. M., J. J. Salminen, S. Taimela, A. Oksanen, and L. Jaakkola. 1992. Subject characteristics and low back pain in young athletes and nonathletes. *Medicine and Science in Sports and Exercise* 24(6):627-632.

Kujala, U. M., S. Taimela, A. Oksanen, and J. J. Salminen. 1997. Lumbar mobility and low back pain during adolescence. A longitudinal three-year follow-up study in athletes and controls. *American Journal of Sports Medicine* 25(3):363-368.

Lin, K. Y., Y. T. Hu, K. J. Chang, H. F. Lin, and J. Y. Tsauo. 2011. Effects of yoga on psychological health, quality of life, and physical health of patients with cancer: A meta-analysis. *Evidence-based Complementary and Alternative Medicine* 2011:659876.

Linker, B. 2011. *War's waste: Rehabilitation in World War I America*. Chicago, IL: University of Chicago Press.

Manios, Y., J. Moschandreas, C. Hatzis, and A. Kafatos. 2002. Health and nutrition education in primary schools of Crete: Changes in chronic disease risk factors following a 6-year intervention programme. *British Journal of Nutrition* 88(3):315-324.

Matton, L., M. Thomis, K. Wijndaele, N. Duvigneaud, G. Beunen, A. L. Claessens, B. Vanreusel, R. Philippaerts, and J. Lefevre. 2006. Tracking of physical fitness and physical activity from youth to adulthood in females. *Medicine and Science in Sports and Exercise* 38(6):1114-1120.

McHugh, M. P., and C. H. Cosgrave. 2010. To stretch or not to stretch: The role of stretching in injury prevention and performance. *Scandinavian Journal of Medicine and Science in Sports* 20(2):169-181.

McHugh, M. P., and M. Nesse. 2008. Effect of stretching on strength loss and pain after eccentric exercise. *Medicine and Science in Sports and Exercise* 40(3):566-573.

McMillan, C. S., and L. D. Erdmann. 2010. Tracking adiposity and health-related physical fitness test performances from early childhood through elementary school. *Pediatric Exercise Science* 22(2):231-244.

Meredith, M. D., and G. J. Welk, eds. 2010. *Fitnessgram/Activitygram test administration manual.* 4th ed. Champaign, IL: Human Kinetics Publishers.

Mierau, D., J. D. Cassidy, and K. Yonghing. 1989. Low-back-pain and straight leg raising in children and adolescents. *Spine* 14(5):526-528.

Mikkelsson, L. O., H. Nupponen, J. Kaprio, H. Kautiainen, M. Mikkelsson, and U. M. Kujala. 2006. Adolescent flexibility, endurance strength, and physical activity as predictors of adult tension neck, low back pain, and knee injury: A 25 year follow up study. *British Journal of Sports Medicine* 40(2):107-113.

Nelson, A. G., J. Kokkonen, D. A. Arnall, and L. Li. 2011. Acute stretching increases postural stability in non-balance trained individuals. *Journal of Strength & Conditioning Research* [Epub ahead of print].

Norkin, C. C., and D. J. White. 2003. *Measurement of joint motion: A guide to goniometry.* 3rd ed. Philadelphia, PA: F. A. Davis Company.

O'Connor, F. G., P. A. Deuster, J. Davis, C. G. Pappas, and J. J. Knapik. 2011. Functional movement screening: Predicting injuries in officer candidates. *Medicine and Science in Sports and Exercise* 43(12):2224-2230.

PCPFS (President's Council on Physical Fitness and Sports). 2012. *The president's challenge physical fitness test: V-sit reach*. https://www.presidentschallenge.org/challenge/physical/activities/v-sit-reach.shtml (accessed June 18, 2012).

Peate, W. F., G. Bates, K. Lunda, S. Francis, and K. Bellamy. 2007. Core strength: A new model for injury prediction and prevention. *Journal of Occupational Medicine and Toxicology* 2:3.

Pereles, D., A. Roth, and D. J. S. Thompson. 2010. A large, randomized, prospective study of the impact of a pre-run stretch on the risk of injury in teenage and older runners. *USA Track and Field*. http://www.usatf.org/stretchStudy/StretchStudyReport.pdf (accessed June 18, 2012).

The Philanthropic Collaborative for a Healthy Georgia. 2008. *Georgia youth fitness assessment 2006*. Atlanta, GA: The Philanthropic Collaborative for a Healthy Georgia.

Plowman, S. A. 2008. Muscular strength, endurance, and flexibility assessments. In *Fitnessgram/Activitygram Reference Guide*, edited by G. J. Welk and D. M. Meredith. Dallas, TX: Cooper Institute. http://www.cooperinstitute.org/vault/2440/web/files/662.pdf (accessed June 18, 2012).

Remvig, L., D. V. Jensen, and R. C. Ward. 2007. Epidemiology of general joint hypermobility and basis for the proposed criteria for benign joint hypermobility syndrome: Review of the literature. *Journal of Rheumatology* 34(4):804-809.

Reurink, G., G. J. Goudswaard, J. L. Tol, J. A. Verhaar, A. Weir, and M. H. Moen. 2012. Therapeutic interventions for acute hamstring injuries: A systematic review. *British Journal of Sports Medicine* 46(2):103-109.

Reuter, I., S. Mehnert, P. Leone, M. Kaps, M. Oechsner, and M. Engelhardt. 2011. Effects of a flexibility and relaxation programme, walking, and nordic walking on Parkinson's disease. *Journal of Aging Research* 2011:232473.

Safrit, M. J. 1990. The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science* 2:9-28.

Salminen, J. J., P. Maki, A. Oksanen, and J. Pentti. 1992. Spinal mobility and trunk muscle strength in 15-year-old schoolchildren with and without low-back-pain. *Spine* 17(4):405-411.

Salminen, J. J., M. Erkintalo, M. Laine, and J. Pentti. 1995. Low-back-pain in the young: A prospective 3-year follow-up-study of subjects with and without low-back-pain. *Spine* 20(19):2101-2107.

Schwellnus, M. P., N. Drew, and M. Collins. 2008. Muscle cramping in athletes—risk factors, clinical assessment, and management. *Clinics in Sports Medicine* 27(1):183-194, ix-x.

Serbescu, C., D. Flora, I. Hantiu, D. Greene, C. L. Benhamou, and D. Courteix. 2006. Effect of a six-month training programme on the physical capacities of Romanian schoolchildren. *Acta Paediatrica* 95(10):1258-1265.

Sherman, K. J., D. C. Cherkin, J. Erro, D. L. Miglioretti, and R. A. Deyo. 2005. Comparing yoga, exercise, and a self-care book for chronic low back pain: A randomized, controlled trial. *Annals of Internal Medicine* 143(12):849-856.

Sherman, K. J., D. C. Cherkin, R. D. Wellman, A. J. Cook, R. J. Hawkes, K. Delaney, and R. A. Deyo. 2011. A randomized trial comparing yoga, stretching, and a self-care book for chronic low back pain. *Archives of Internal Medicine* 171(22):2019-2026.

Simic, L., N. Sarabon, and G. Markovic. 2012. Does pre-exercise static stretching inhibit maximal muscular performance? A meta-analytical review. *Scandinavian Journal of Medicine and Science in Sports* [Epub ahead of print].

Thacker, S. B., J. Gilchrist, D. F. Stroup, and C. D. Kimsey, Jr. 2004. The impact of stretching on sports injury risk: A systematic review of the literature. *Medicine and Science in Sports and Exercise* 36(3):371-378.

Tilbrook, H. E., H. Cox, C. E. Hewitt, A. R. Kang'ombe, L. H. Chuang, S. Jayakody, J. D. Aplin, A. Semlyen, A. Trewhela, I. Watt, and D. J. Torgerson. 2011. Yoga for chronic low back pain: A randomized trial. *Annals of Internal Medicine* 155(9):569-578.

Tremblay, M. S., M. Shields, M. Laviolette, C. L. Craig, I. Janssen, and S. C. Gorber. 2010. Fitness of Canadian children and youth: Results from the 2007-2009 Canadian Health Measures Survey. *Health Reports* 21(1):7-20.

Wang, C., J. P. Collet, and J. Lau. 2004. The effect of tai chi on health outcomes in patients with chronic conditions: A systematic review. *Archives of Internal Medicine* 164(5):493-501.

Welk, G. J., M. D. Meredith, M. Ihmels, and C. Seeger. 2010. Distribution of health-related physical fitness in Texas youth: A demographic and geographic analysis. *Research Quarterly for Exercise and Sport* 81(Suppl. 3):S6-S15.

Wilmore, J. H., and D. L. Costill. 1994. *Physiology of sport and exercise*. Champaign, IL: Human Kinetics Publishers.

Witvrouw, E., N. Mahieu, L. Danneels, and P. McNair. 2004. Stretching and injury prevention: An obscure relationship. *Sports Medicine* 34(7):443-449.

Wolf, J. M., K. L. Cameron, and B. D. Owens. 2011. Impact of joint laxity and hypermobility on the musculoskeletal system. *Journal of the American Academy of Orthopaedic Surgeons* 19(8):463-471.

8

Fitness Measures for a National Youth Survey

KEY MESSAGES

A substantial body of evidence supports the idea that specific tests measuring body composition and cardiorespiratory endurance are related to health markers in youth; the evidence for musculoskeletal fitness is less extensive. A national survey of health-related physical fitness in youth should include the following fitness test items: (1) measures of body mass index, waist circumference, and skinfold thickness (triceps and subscapular sites) to assess body composition; (2) a progressive shuttle run, such as the 20-meter shuttle run, or a submaximal cycle ergometer or treadmill test if there are space limitations, to measure cardiorespiratory endurance; and (3) handgrip strength and standing long jump tests to measure musculoskeletal fitness.

Assuming that a national youth fitness survey would be implemented in school-based settings, survey administrators should distribute the equipment needed to conduct the recommended test items such that the survey participants have the opportunity to receive sufficient training in the measurement protocols and to practice the fitness tests. Likewise, survey administrators and those interpreting and communicating the results should receive the appropriate training in conducting and interpreting tests to minimize measurement and classification errors and prevent adverse events.

> Developers of fitness test batteries should calculate age- and gender-specific cut-points (cutoff scores) to determine whether individuals are at risk of fitness-related poor health outcomes by applying the general guidance in Chapter 3 to specific fitness components.

One of the uses of fitness measures is in the assessment of youth populations through statewide or national surveys. Chapter 2 provides a brief history of fitness testing and an overview of national fitness surveys (Tables 2-1 through 2-5) and a list of surveys currently in use in the United States and other countries (Table 2-6).

As its statement of task requested, the committee reviewed the evidence for tests for four components of fitness that historically have been recognized as health related using the methodology described in Chapter 3. The scientific rationale for the committee's selection of tests based on their relationship to health and their validity, reliability, and feasibility is given in Chapters 4 (body composition), 5 (cardiorespiratory endurance), 6 (musculoskeletal fitness), and 7 (flexibility). This chapter provides the committee's conclusions and recommendations regarding fitness test items to be included in a battery for use in a national survey. In presenting these conclusions and recommendations, the committee emphasizes that a national youth fitness survey should be implemented in schools by skilled national survey administrators (i.e., those familiar with the procedures for conducting large surveys and the protocols for fitness testing). Although this report does not include recommendations for specific fitness test protocols, the committee recognizes the need to develop standardized protocols for field-based tests for youth to enable meaningful comparisons of results from different studies and surveys.

CURRENT STATUS OF NATIONAL YOUTH FITNESS TESTING

As discussed in Chapter 2, no national fitness survey has been conducted since the 1980s. However, the currently active National Health and Nutrition Examination Survey (NHANES) includes components pertinent to physical fitness, such as body composition, cardiovascular fitness, and physical activity (Morrow et al., 2009). This set of fitness tests was recently extended for youth, and a 2012 NHANES Youth Fitness Survey is currently under way. The survey includes the following test items: body mass index (BMI), arm length and circumference, waist circumference, skinfolds, and whole-body dual-energy X-ray absorptiometry (DXA) scans (body composition); the handgrip, plank, and knee extension (muscle strength); the treadmill (cardiovascular endurance); and a test of gross motor development. Selection of these items was based on expert opinion and feasibility.[1]

CONCLUSIONS AND RECOMMENDATIONS

Based on its review of the scientific literature, as presented in Chapters 4 through 7, the committee concluded that there is enough scientific evidence to support recommending specific health-related items for youth fitness tests for the components of body composition, cardiorespiratory endurance, and musculoskeletal fitness. The committee concluded enough evidence has not been accumulated to support recommending a measure of flexibility for a national survey at this time. While the literature reviewed also did not provide enough information to support specific conclusions on the relationship between age and performance on fitness test items, the committee based its recommendations in part on feasibility of administration, selecting test items that are the most practical and relatively independent of age. The committee generally recognized that age can affect the validity, reliability, and safety of test items, so considering the age-appropriateness of test items is important when developing a national survey. The committee also concluded that there are many modifiers, such as demographic factors, that must be considered to avoid introducing errors in the interpretation of performance on fitness tests. In addition to age and gender, which are often recorded and used to interpret results, it is essential for survey developers to include a race/ethnicity questionnaire as part of a national survey. Although the evidence is not as substantial for maturation status and motor skill, measurements of these factors should also be considered in the survey design to enhance understanding of their mediating effects on performance on fitness tests. Variables that are more challenging to account for are those related to differences in the training of administrators and in the previous level of practice or physical activity of participants. To minimize those differences, care should be taken in training administrators and in providing opportunities for participants to become familiar with the tests when feasible. It is also important for the trained administrators to record adverse events associated with test taking so that safety issues can be better understood and addressed appropriately.

The committee's statement of task included determining how scores should be interpreted for test items selected for inclusion in a national

[1]Personal communication, L. G. Borrud and V. L. Burt, National Center for Health Statistics, January 5, 2012. Additional information about the NHANES National Youth Fitness Survey is available online at http://www.cdc.gov/nchs/nyfs.htm (accessed August 16, 2012).

youth fitness survey and in youth fitness tests conducted in other settings. Cut-points (cutoff scores) are critical to interpretation of the results of health-related fitness tests since they serve as a way to distinguish individuals and populations that may be at risk of poor health outcomes from those that are not. As explained in Chapter 3, the committee concluded that a criterion-referenced method using cut-points associated with health outcomes or health markers is the ideal approach. Except for BMI, however, insufficient evidence exists with which to develop age- and gender-specific criterion-referenced cut-points related to health outcomes for any of the recommended tests; instead, age- and gender-specific interim cut-points corresponding to percentiles for youth or for adults should be used until enough data have been collected to enable establishing age- and gender-specific criterion-referenced cut-points. The committee provides general guidance for developing cut-points and interpreting performance results in Chapter 3; specific recommendations for developing cut-points depend on data available for each of the components of fitness. The following is a summary of the basis for the committee's recommendations for deriving interim cut-points[2] (additional explanation can be found in the chapters on the fitness components [Chapters 4-7]):

- *Body composition*—BMI cut-points were set based on the already established 2000 Centers for Disease Control and Prevention (CDC) growth charts and percentiles. Until population-based evidence in youth is available for skinfolds and waist circumference, the 85th percentile (borrowed from the BMI percentiles) should be used to derive interim cut-points for these measures.
- *Cardiorespiratory endurance*—Until population-based evidence in youth is available, the recommended interim cut-points should be based on data from both youth and adult populations on the relationship between treadmill test performance and health outcomes. For adults, the lowest quintile has been determined as appropriate (Blair et al., 1989). For youth, the 30th percentile has been established as identifying those at risk of poor health outcomes (Lobelo et al., 2009; Welk et al., 2011). Based on those two determinations, the committee recommends that interim cut-points be derived from

[2]Depending on the nature of the test, the risk of a poor health outcome is defined by an individual's being either below or above a specific percentile of a population. For BMI, for example, individuals above the 85th percentile of the 2000 Centers for Disease Control and Prevention (CDC) growth charts are considered to be overweight or obese and therefore at risk of a poor health outcome. In contrast, for cardiorespiratory endurance tests, those below a certain percentile (e.g., 20th percentile) of the population are the ones who are less able to perform exercise without fatigue and therefore at risk of a poor health outcome.

the lowest performers (e.g., 20th percentile) to ensure a high probability that an individual identified as being low fit is really low fit and thus at risk for poor health outcomes.

- *Musculoskeletal fitness*—Until population-based evidence in youth is available, interim cut-points borrowed from the interim cardiorespiratory endurance percentile (e.g., the 20th percentile) are recommended. There are no data sets from youth or adult populations for any musculoskeletal fitness test. As stated in Chapter 3, in such cases, cut-points should be derived by borrowing the percentile from a test that is as comparable as possible in nature (e.g., requires movement of the body) and in the dimension it measures (e.g., upper-body strength) to the test of interest.

Further research is needed to better understand the relationship between fitness and health and to support selection and validation of the ideal criterion-referenced cut-points for fitness measures. This research should involve the implementation of national surveys that include measures of health outcomes. This and other research recommendations are presented in Chapter 10.

While the use of percentiles in establishing interim cut-points is appropriate until more evidence is collected, the committee recommends that once age- and gender-specific interim cut-points have been established, those cut-points rather than percentiles be used in communicating test results to those being tested, health and school officials, and parents. Doing so will minimize the confusion that might arise in communicating in terms of the percentiles used to derive the cut-points. For example, the CDC has used the 95th percentile from the 1960s to the 1990s as a cut-point for obesity in children, yet more than 15 percent of youth currently exceed the 95th percentile. In instances where percentiles may allow a clearer presentation of the results, as for BMI, the year of data collection should be reported with the percentile. The committee also notes that test administrators, those interpreting and communicating the results, and researchers should be fully familiar with the derivation of cut-points so they can interpret distribution changes in the population. To this end, survey administrators and those communicating the results should receive the appropriate training in interpreting tests to minimize classification errors. In addition, researchers developing percentile data for use in deriving interim cut-points should report the year of data collection. Fitness data from large populations are needed to derive the appropriate cut-points from percentiles. If such data are not available, developers of interim cut-points should consult with statisticians to design a small study with a representative sample of U.S. youth to provide such data.

RECOMMENDATION 8-1.[3] A national survey of health-related physical fitness in youth should include measures of cardiorespiratory endurance, body composition, and musculoskeletal fitness. The survey should include the following fitness test items: (1) measures of BMI, waist circumference, and skinfold thickness (triceps and subscapular sites) to assess body composition; (2) a progressive shuttle run, such as the 20-meter shuttle run (or a submaximal treadmill or cycle ergometer test if there are space limitations) to measure cardiorespiratory endurance; and (3) handgrip strength and standing long jump tests to measure musculoskeletal fitness.

RECOMMENDATION 8-2. Standard protocols for the administration of measures of youth fitness in national surveys should be developed and implemented. The focus should be on maximizing the measures' reliability, validity, and safety. Trained personnel should be used for test administration and data collection.

RECOMMENDATION 8-3. Developers of fitness test batteries should use age- and sex-specific cut-points to determine which individuals are at risk of poor fitness-related health outcomes. Optimum cut-points should be based on criterion values when population-based evidence is available on the relationship between the level of performance on a fitness test and a health outcome or marker. In the absence of criterion values, interim population-based percentile values should be applied. These values might be derived from adults on tests for the same component or from youth on tests for a different or the same component. Specifically, the guidance of the committee should be applied as follows:

- **Body composition: For BMI, the CDC-established cut-points for underweight, overweight, and obesity evaluations should be used. Interim cut-points for skinfold and waist circumference measures could be derived from the CDC-established percentiles for BMI.**
- **Cardiorespiratory endurance: For measures of cardiorespiratory endurance, interim cut-points could be derived from the lowest performers (e.g., 20th percentile) on the cardiorespiratory endurance distribution curve.**
- **Musculoskeletal fitness: For musculoskeletal fitness tests, interim cut-points could be derived by borrowing the percentile (e.g., 20th percentile) from the cardiorespiratory endurance tests.**

[3]The committee's recommendations are numbered according to the chapter of the main text in which they appear.

REFERENCES

Blair, S. N., H. W. Kohl, III, R. S. Paffenbarger, Jr., D. G. Clark, K. H. Cooper, and L. W. Gibbons. 1989. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *Journal of the American Medical Association* 262(17):2395-2401.

Lobelo, F., R. R. Pate, M. Dowda, A. D. Liese, and J. R. Ruiz. 2009. Validity of cardiorespiratory fitness criterion-referenced standards for adolescents. *Medicine and Science in Sports and Exercise* 41(6):1222-1229.

Morrow, J. R., Jr., W. Zhu, B. D. Franks, M. D. Meredith, and C. Spain. 2009. 1958-2008: 50 years of youth fitness tests in the United States. *Research Quarterly for Exercise and Sport* 80(1):1-11.

Welk, G. J., K. R. Laurson, J. C. Eisenmann, and K. J. Cureton. 2011. Development of youth aerobic-capacity standards using receiver operating characteristic curves. *American Journal of Preventive Medicine* 41(4, Suppl. 2):S111-S116.

9

Fitness Measures for Schools and Other Educational Settings

KEY MESSAGES

Conducting fitness tests in schools and other educational settings may result in benefits for both individuals and groups beyond improving fitness. Examples include tracking fitness and disease risk and using test results to set health goals, planning for enduring healthy behaviors, and driving physical education instruction.

To plan and conduct fitness testing in schools effectively and appropriately, test developers and administrators should consider the following four factors:

- Test items should be selected with consideration of contextual variables, such as access to high-quality equipment, space, cost, privacy, and availability of volunteers, as well as cultural and racial/ethnic factors.
- When administering tests, consideration should be given to the safety of participants, the presence of pre-existing conditions, the effects of body composition and other modifiers on test results, and the confidentiality of results.
- School-based professional development that is applicable to the daily routine of teachers and includes instruction in how to integrate fitness testing into the curriculum should be provided.
- Professional development should include training in the administration of protocols and interpretation and communication of test

results, with emphasis on educating participants about the importance of fitness, supporting the achievement of fitness goals, and developing healthy living habits. Those interpreting and communicating test results should ensure confidentiality, consider each individual's demographic characteristics, provide for the involvement of parents, and offer positive feedback and recommendations to students and parents.

Based on their relationship to health, their integrity, and their feasibility (e.g., ease of administration and interpretation, need for and cost of special equipment, privacy concerns), developers and administrators of fitness test batteries for schools and other educational settings should consider including a set of test items similar to those recommended for surveys:

- standing height (a measure of linear growth status) and weight (a measure of body mass) to calculate body mass index as an indicator of body composition;
- a progressive shuttle run, such as the 20-meter shuttle run, to measure cardiorespiratory endurance; and
- the handgrip strength and standing long jump tests to measure upper- and lower-body musculoskeletal strength and power, respectively.

Additional tests not yet shown to be related to health, such as distance and/or timed runs, the modified pull-up, the push-up, and the curl-up, may also be considered as supplemental educational tools. The guidance provided in Chapter 8 should be followed in establishing cutpoints (cutoff scores) for interpreting performance on the selected fitness test items.

The preceding chapters highlight the importance of measuring and monitoring the prevalence of physical fitness during childhood and address questions relating to the core purpose of this report—identifying fitness tests that are related to health and are valid and reliable. Tests for a national youth fitness survey are recommended based on those criteria and on practical considerations related to the feasibility of their administration in a national survey. In addition to national surveys, fitness tests often are conducted in schools (and other educational settings) for a variety of reasons. Examples include uses associated with state or local physical fitness testing mandates in schools and with physical education curricula and instruction.

Along with improving the fitness performance of individuals, fitness tests in educational settings can yield other benefits when appropriately conducted and interpreted. One benefit is that, when integrated into physical education programs in school settings, fitness testing can provide clear technical performance expectations and minimize the effect of practice on test performance in a national survey. Another benefit is that fitness testing in schools allows for group and individual tracking of physical fitness trends and disease risk. Fitness test results can also be used for assessing learning outcomes and physical education content standards. Given the connection between physical activity/fitness and cognitive performance (Castelli et al., 2007; Hillman et al., 2009; Kamijo et al., 2011; Welk et al., 2010), moreover, it becomes important for knowledge, attributes, and awareness of fitness to be promoted in educational settings as part of fostering healthy lifestyle choices across the life span. When the primary objectives of physical education or physical activity programming are achieved as intended, such programming can lead to the development of habitual healthy behaviors. The inclusion of fitness testing in physical education provides a forum for supporting and measuring the attainment of learning standards associated with physical fitness (Tremblay and Lloyd, 2010).

Accordingly, physical fitness is a focus of six national standards for physical education that reflect the skills, knowledge, and abilities resulting from participation in effective physical education and physical activity programming in schools (NASPE, 2004). As of June 2011, all 50 states had learning standards centered on health-related fitness (Centeio and Keating, 2011); 14 states mandated direct measurement of physical fitness (NASPE, 2010). Physical education and the implementation of models such as Coordinated School Health and Comprehensive School Physical Activity Programs have outcomes concentrated on both the achievement and maintenance of health-enhancing levels of fitness and regular engagement in physical activity, as these variables are independent risk factors associated with health (Plowman, 2005).

To administer fitness testing appropriately in schools, it is necessary to consider such factors as ensuring that the testing has clear ties to improved physical fitness and fostering increased engagement in physical activity among students (Keating, 2003). Although some have questioned how the inclusion of physical fitness testing may influence time for academic learning, there is evidence that fitness may have positive effects on both health and learning (CDC, 2010; Rasberry et al., 2011; Rosas et al., 2009), including evidence for a causal relationship between children's fitness and cognitive performance (Kamijo et al., 2011).

The committee's literature review included only studies that involved populations of healthy and obese youth and excluded studies of youth with congenital diseases or disabilities. The fitness testing recommendations in this report, therefore, are driven by the evidence for healthy study populations. Nonetheless, it is important for students with disabilities to be included in fitness testing whenever possible and for the interpretation of test results to be modified accordingly. Specifically, those students with personal fitness goals should be encouraged to participate in fitness testing as a means of tracking progress toward their goals. The Brockport Fitness Test is an example of how specific fitness tests can be modified for students with disabilities, and the *Brockport Physical Fitness Technical Manual* provides criterion-referenced cut-points (cutoff scores) for a variety of disabilities (Winnick and Short, 1999). While the relationship between health outcomes and physical activity in people with disabilities is not the focus of this report, other reviews, such as the *Physical Activity Guidelines Advisory Committee Report* (Physical Activity Guidelines Advisory Committee, 2008), specifically examine this issue.

Given the potential benefits of fitness testing, the committee recommends the use of some measures in schools (and other educational settings) even though the evidence for their relationship with health is only promising at this time. The committee recommends these additional measures with the expectation that future research will elucidate whether they are related to health in youth.

This chapter examines factors related to the following issues in school and other educational settings: the selection and implementation of test items, the administration of the test items, the interpretation of test results, and the incorporation of fitness testing into a curriculum or program. It then briefly reviews appropriate and inappropriate fitness testing practices in these settings. The final section presents the committee's conclusions and recommendations for school-based fitness testing.

FACTORS RELATED TO SELECTING AND IMPLEMENTING TEST ITEMS

Children enrolled in regularly scheduled physical education classes participate in significantly more physical activity than those who attend physical education infrequently (Cawley et al., 2007; Gordon-Larsen et al., 2000; Pate et al., 2007). On the other hand, participation in physical education alone cannot facilitate high levels of fitness in every child (Dale and Corbin, 2000), given heredity effects on fitness (Bouchard and Shephard, 1994), a lack of instructional time dedicated to physical activity (NASPE, 2010; Pate et al., 2011), and low to moderate relationships between physical activity and fitness (Morrow and Freedson, 1994; Pate et al., 1990;

Payne and Morrow, 1993). Furthermore, the lack of national fitness surveys since the 1980s makes it difficult to establish relationships between physical activity and fitness measures over time (Corbin and Pangrazi, 1992; Flegal et al., 1998).

Physical activity leaders and teachers selecting fitness test items for schools need to consider contextual variables such as access to high-quality equipment, space, cost, privacy, and the availability of volunteers, as schools differ greatly on these variables (Martin et al., 2010). Box 3-2 in Chapter 3 includes a general list of criteria for evaluating administrative feasibility. The use of only high-quality equipment is critical to avoid measurement and interpretation errors. To ensure that performance on a fitness test is actually a reflection of physical fitness, it is also necessary to consider the reliability, validity, and feasibility of test items; the standardization of test protocols; and the confidentiality of test results. It is vital as well for administrators to ensure the safety of fitness test participants by being sensitive to such variables as participants' pre-existing disease(s), body composition, and maturation stage. Age is a particularly critical consideration for ensuring the validity, reliability, and safety of selected test items, as performance on some items may improve with age and maturity. Cultural relevance and potential racial/ethnic bias also are related to test performance and therefore should be considered in the selection of test items (Miech et al., 2006).

The educational value of a test item and its corresponding health-related fitness component should carry weight in the selection process. Specifically, how does the identified test item align with the existing curriculum goals, and to what degree can fitness education be carried out as a valued part of instruction?[1] When such evaluations and corresponding instruction occur, there is a high likelihood that health-related fitness knowledge will increase (Kulinna, 2004; Stewart and Mitchell, 2003) and that youth's misconceptions about fitness will be addressed (Keating et al., 2009). Further, studies have shown that conceptual physical education may lead to less sedentary behavior after students complete their schooling (Dale and Corbin, 2000; Dale et al., 1998). If educators and physical activity leaders avoid teaching to the test and instead allow the results of fitness tests to drive instruction and create educational opportunities, the potential exists for youth to change their behaviors through self-management and goal setting. It is, however, important for students to know how to perform fitness tests and be given the opportunity to practice the tests prior to the testing session (see Chapter 8). Finally, despite evidence that augmented knowledge about health-related fitness may increase engagement in physical activity (Kulinna and Silverman, 2000), it remains unclear whether enhanced knowledge

[1]Available at http://www.aahperd.org/naspe/publications/upload/Instructional-Framework-for-Fitness-Education-in-PE-2012-2.pdf (accessed August 17, 2012).

(Ennis, 2007; Placek et al., 2001) and such learning experiences linked to fitness testing (Cale and Harris, 2009) will directly change behavior to a degree that will reduce health risk, as this area is understudied.

FACTORS RELATED TO TEST ADMINISTRATION

In a society where childhood obesity is a growing concern (Ogden et al., 2010), teachers are being asked to fulfill multiple responsibilities related to physical fitness and activity as part of their job description.[2] Since physical education teachers have limited opportunities for professional development (Doutis and Ward, 1999; O'Sullivan and Deglau, 2006; Stroot et al., 1994), learning experiences for these teachers in administering fitness tests are most valuable when they are school based and applicable to the teachers' daily routine (e.g., how to manage the class while administering the tests), collaborative in nature, and centered on how to achieve the desired student outcomes (i.e., enhancing student understanding and progress toward attaining physical fitness standards) (Armour and Yelling, 2007). Professional development aimed at preparing physical education teachers to administer a battery of fitness tests can include a combination of the following components:

- how to integrate fitness testing into the curriculum;
- protocols and use of proper equipment for fitness test items;
- how to familiarize participants with the test, together with specifications regarding the amount and type of practice;
- how to communicate consistently with the students in ways that create a positive and encouraging environment for learners of all ability levels;
- teacher burden;
- participant burden;
- the validity and reliability of test items;
- class management during test periods; and
- how to interpret and communicate test results.

An extensive body of literature expands on components of effective and sustainable professional development, a topic that is beyond the scope of this report. In general, however, professional development enables physical education teachers to administer physical fitness tests accurately and with minimal bias (Morrow et al., 2010) while providing physical activity oppor-

[2]For example, see the National Association for Sport and Physical Education (NASPE) Director of Physical Activity Certification Program, available at http://www.aahperd.org/naspe/professionaldevelopment/dpa/index.cfm (accessed June 26, 2012).

tunities that enhance fitness (Kibbe et al., 2011). A recent meta-analysis suggests that in general, students are motivated to participate and to learn in physical education (Chen et al., 2012). Yet student motivation is influenced by the school climate, specifically the task or ego orientation of the activities offered during physical education (Parish and Treasure, 2003; Standage et al., 2003). Teachers who develop a positive and mastery-oriented climate are more likely to have students who perform better on assessments such as fitness testing. When introducing students to fitness testing, for example, the use of instructions that provide personal relevance and meaning for a student can lead to enhanced performance (Simons et al., 2003). Accordingly, it is important for teachers to be consistent in the delivery of content related to fitness testing, as well as to be equally supportive to learners of all ability levels, or the test may be biased. Teachers need professional development to apply these principles of fitness testing in schools (Corbin, 2010; Keating, 2003).

FACTORS RELATED TO INTERPRETATION OF TEST RESULTS

Perhaps the most important element of fitness testing, the interpretation and dissemination of results must be planned for if the goals of the testing are to be achieved. Identifying the level of health risk associated with the established cut-points for a particular age is a way to involve and educate parents as well as children by providing personalized feedback, including comparison of current and previous test results. Note that in communicating test results, it is essential to ensure confidentiality to avoid the reduced self-esteem that can occur if the results (low performance on fitness tests or high body composition numbers) are shared with others (Fox, 1988). In addition to information about health risks, positive feedback and recommendations, including information and resources related to health care, are valuable characteristics of fitness test reports. Fitness education thereby has the potential to build fitness competence, create a sense of the importance of fitness, and provide motivation and opportunities to support the achievement of fitness goals (Fox, 1988). Administrators and those communicating results of fitness tests in schools should be trained in these areas.

Test scores are estimations of health-related fitness at a point in time. To maintain or improve scores, continued participation in physical activity is necessary. The Fitnessgram® program suggests that, when interpreting performance on fitness tests, the following characteristics are essential in educational settings:

- “The physical fitness experience should always be fun and enjoyable.
- Physical fitness testing should not become a competitive sport.

- The performance of one student should not be compared to that of another student.
- The primary reason for testing is to provide the [participant] with personal information that may be used in planning a personal fitness program.
- The performance level on fitness tests should not be used as a basis for grading." (Meredith and Welk, 2010, p. 58)

Test administrators and those interpreting and communicating results should be fully familiar with the meaning of cut-points and the effects of modifiers (e.g., maturation status, race/ethnicity) for each test. Other variables, such as biology, the emotional investment of the participant, tester error, equipment, the amount of practice, and testing conditions, also affect performance on a fitness test. As part of test interpretation, the test administrator and those interpreting and communicating results must decide whether the scores are valid or their deviation from expected results is beyond these sources of error. For example, having some participants engage repeatedly in a shuttle run in an effort to understand the technical protocol and necessary adherence not only allows for more valid and reliable test administration, but also may enable these participants to achieve a higher level of cardiorespiratory fitness than those who have not had the opportunity to practice the test. It is important, then, that an educational component be integrated into the physical education program to provide clear technical performance expectations for fitness testing and minimize practice effects. When physical education teachers and physical activity leaders adhere to these principles, fitness tests can help identify risk for developing chronic disease while helping participants better understand the concepts of fitness through comprehensive fitness education (Freedson et al., 2000).

FACTORS RELATED TO THE INCORPORATION OF FITNESS TESTING INTO A CURRICULUM OR PROGRAM

When fitness testing is integrated into educational programs or curricula, it provides a mechanism for longitudinally tracking and monitoring physical fitness trends and risk for disease among individuals and groups. In an educational setting, individual tracking is most relevant as school is one of the few places where feedback can be provided to both participants and their parents. However, group tracking over time also can be useful for physical education teachers, enabling them to utilize trends to inform instruction by identifying the needs of the current student body.

It has been suggested that, regardless of developmental stage, the benefits of being able to monitor progress, set goals, provide feedback, give incentives, and design a personalized physical activity plan outweigh the risks of participation in physical fitness testing (Safrit, 1995). Clearly communicating to participants the meaning of each test item and discussing the training principle of specificity (i.e., the activity's association with an identified joint or muscle group) is important. Participants then can set personalized goals and create an individualized plan for achieving those goals that purposefully links modes of physical activity to health-related fitness components. Learning experiences that apply knowledge to authentic situations increase the likelihood that conceptual learning will lead to enhanced participation in physical activity.

The use of fitness awards in schools has been the subject of ongoing controversy. Although fitness awards were created to motivate youth to be fit, questions have been raised about their motivational value. For example, reports have suggested that the Presidential Physical Fitness Awards may be awarded to youth who are already athletically successful (Corbin et al., 1990), that they are not motivating to youth with low fitness (Corbin et al., 1988), and that they may reduce rather than enhance intrinsic motivation (Whitehead and Corbin, 1991). To date, evidence has not been presented to support the use of fitness awards. While it is beyond the scope of this report to make suggestions about fitness awards, the committee believes a comprehensive study of such awards, similar to this study of fitness test items, should be conducted to determine whether there is sufficient scientific evidence to warrant their use.

APPROPRIATE AND INAPPROPRIATE PRACTICES

If physical fitness tests are to be used effectively in schools and other educational settings, appropriate practices must be employed in their administration. Appropriate practice varies by maturation stage; thus what may be suitable for elementary school students may be inappropriate for adolescents. Numerous authors have outlined appropriate practices (Corbin, 2009; Corbin and Pangrazi, 2008; Ernst et al., 2006), and regardless of stage of maturation, some basic tenets apply, as summarized and supported by national organizations (Table 9-1). These include the following:

- Health-related fitness activities are integrated into an existing curriculum.
- Fitness test results are used to set individual goals and develop fitness plans.
- Fitness assessments are part of the ongoing process of helping students understand, enjoy, improve, and maintain their physical fitness and well-being.
- Youth are physically prepared to participate in fitness testing.

TABLE 9-1 Appropriate and Inappropriate Practices Related to Fitness Testing in Schools and Other Educational Settings

Appropriate Practice	Inappropriate Practice
In elementary school, motor skills are the focus of instruction, with health-related fitness components being integrated into the curriculum and lessons focused on fitness education.	Health-related fitness is rarely integrated into instruction. Students fail to understand the benefits of health-related fitness and know little about how to develop a fitness plan.
Fitness testing is used to set individual goals as part of fitness education. At the secondary level, students use fitness test data to design and apply a personal fitness plan.	Fitness testing is conducted without meaningful understanding, interpretation, and application.
Physical educators use fitness assessment as part of the ongoing process of helping students understand, enjoy, improve, and maintain their physical fitness and well-being (e.g., students set fitness goals for improvement that are revisited during the school year).	Physical educators use fitness test results to assign a grade.
Children are physically prepared to participate in fitness testing.	Children are required to participate in fitness testing without proper preparation.

SOURCES: Adapted from NASPE, 2009a,b,c.

Two specific unacceptable practices warrant further discussion: the use of fitness test scores for academic grading and for high-stakes accountability. Both of these practices are considered improper applications of fitness test results (NASPE, 2009a,b,c). It is inappropriate to include fitness test scores in academic grades or grade point averages (NASPE, 2009a,b,c). Although physical fitness can be increased through engagement in specific types of physical activity, factors other than physical activity affect a student's fitness that are beyond the control of the student and physical education teacher. Examples include heredity, caloric consumption, access to opportunities to be physically active both within and beyond the school day, and possibly socioeconomic status. For similar reasons, physical fitness testing for the purpose of teacher and school accountability is also inappropriate. Even though physical fitness may be a primary goal of a given program, confounding, uncontrollable variables remain (e.g., heredity [Bouchard and Shephard, 1994], socioeconomic status, and other school contextual variables [Mitchell et al., 2003]); therefore, this practice is a misstep in the interpretation and utilization of fitness testing (Harris and Cale, 2007).

CONCLUSIONS AND RECOMMENDATIONS

In general, active children display healthier physical fitness profiles—including higher cardiorespiratory and musculoskeletal fitness and bone mass and lower body fat—than their inactive peers (Boreham and Riddoch, 2001). Because these trends often track to adulthood (Baranowski et al., 2000; Boreham et al., 2002; Hasselstrom et al., 2002; Janz et al., 2002; Lefevre et al., 2002; Twisk et al., 2002), the committee has highlighted the importance of measuring and monitoring the prevalence and specificity of physical fitness during youth in the preceding chapters. Using many of the factors outlined above, the committee considered the strengths and weaknesses of the test items recommended for a national survey (Chapter 8) with regard to their practicality in schools and other educational settings. The sections below detail the committee's adaptations for fitness tests when the testing is to be conducted in such settings. Note that, regardless of the setting, it is important to develop and use standardized test protocols so comparisons can be made among surveys and studies.

Body Composition

Since body composition is an important health outcome, the committee recommends that it be measured to track health risk and long-term health relationships in youth (see Chapter 8). As a health marker, moreover, body composition—specifically being overweight or obese—is negatively related to academic achievement (Averett and Stifel, 2007; Bagully, 2006; Castelli et al., 2007) and inversely related to reaction time and accuracy of memory tasks (Kamijo et al., 2011).

The committee recommends that in educational settings, standing height and body weight be measured and transformed into body mass index (BMI) as a mediator of physical fitness and a measure of health risk. These data should remain private and be shared only with the child and parent(s). The already established Centers for Disease Control and Prevention (CDC) cut-points based on the 2000 CDC growth charts for children and adolescents should be applied when interpreting BMI data.

Although the committee recommends skinfold and waist circumference measurements for a national survey, their implementation in a school setting raises concerns. First, measuring skinfolds and waist circumference requires specific and intense training to avoid the introduction of errors (e.g., intra- and interobserver errors). Second, these two measurements are not free of potential motivational or self-esteem influence; self-esteem may be impacted by the interpretation of results for estimated body composition. Third, it is difficult to ensure the privacy of the measurement process given that measurement of skinfolds and waist circumference is more invasive than measurements for BMI because it requires exposure of the trunk to allow the test administrator to access the subscapular and waist regions. As a result, conducting these tests likely requires the presence of two test administrators, thus increasing the administrative burden. By contrast, less effort is required for measurement of height and weight (see Chapter 4).

Cardiorespiratory Endurance

A large body of evidence suggests that cardiorespiratory endurance is related to health outcomes such as adiposity and cardiometabolic risk factors (e.g., blood pressure, blood lipids and glucose, insulin sensitivity) during childhood and adulthood (see Chapter 5). Cardiorespiratory endurance is believed to be lower in sedentary and overweight female youth. Approximately one-third of U.S. youth (males and females) aged 12 to 19 fail to meet the standards for cardiorespiratory endurance (Pate et al., 2006). As indicated earlier in this chapter, emerging evidence also identifies a link between cardiorespiratory endurance and academic achievement (Donnelly and Lambourne, 2011; Hillman et al., 2009), as well as working memory and attention—essential antecedents of learning (Kamijo et al., 2011).

Among the valid and reliable tests for which strong evidence indicates a relationship to health, the shuttle run has the feasibility required for implementation in a school setting, requiring no expensive equipment. An alternative to the shuttle run is any of the distance runs that, as noted in Chapter 5, have been used to measure cardiorespiratory endurance in fitness test batteries since the advent of large-scale fitness testing in the post-World War II era. Numerous studies have assessed the validity of distance run tests by examining the correlation between a criterion measure—directly measured maximal oxygen uptake (VO_2max) (ml/kg/min)—and time or distance on the run, and have concluded that distance runs of 1 mile or greater demonstrate acceptable validity (Freedson et al., 2000; Safrit, 1990). Also, distance runs have been found to be reliable based on test-retest correlations (Artero et al., 2011; Freedson et al., 2000).

Musculoskeletal Fitness

Musculoskeletal fitness, including muscle strength, muscle endurance, and muscle power, has been positively associated with quality of life and inversely linked to risk for cardiovascular disease in adults (see Chapter 6). In children, the link between musculoskeletal fitness and health is less clear given developmental and maturational influences and the paucity of recent normative data. However, the committee concluded that musculoskeletal fitness during childhood is likely linked to health during adulthood; thus, musculoskeletal fitness tests should be used in schools as a tool for educating about their potential health benefits.

In Chapter 8, the handgrip strength and standing long jump tests are recommended for a national youth fitness survey as measures of musculoskeletal fitness because of this component's suggestive relationship to health (particularly in adults), sufficient validity and reliability of both tests in youth, and feasibility (e.g., equipment cost, equipment calibration, administrator training). For schools, in addition to these two tests, the committee considered the value of other musculoskeletal fitness tests that are valid and that may have adequate reliability because of their wide use and familiarity to students and test administrators.

For example, the modified pull-up, which is currently used in school-based fitness test batteries in the United States, has moderate reliability and sound logical construct validity as a measure of upper-body strength (Engelman and Morrow, 1991; Erbaugh, 1990; Kollath et al., 1991; Pate et al., 1995; Saint Romain and Mahar, 2001). While there is scant evidence of this test's relationship to health in youth, it does provide a valid assessment of an individual's or group's musculoskeletal fitness status (see Chapter 6) and is feasible for use in schools and other educational settings. Also used frequently in schools, the curl-up and push-up may have value as fitness educational tools. Both have been shown to have reasonable reliability and validity when administered in a large school-based survey; however, these values are lower than for cardiorespiratory endurance and body composition tests (Morrow et al., 2010; Plowman, 2008). Because the curl-up test measures a different construct of musculoskeletal fitness from the handgrip strength and standing long jump tests (i.e., core strength and endurance), it should not be considered as an alternative to those tests. It is important to stress that none of the musculoskeletal fitness tests should be interpreted in a health context until such relationships are more firmly established in the future.

Flexibility

As described in Chapter 7, information is lacking about the association between flexibility and health outcomes in youth and is inconsistent in adults. For this reason, the committee does not recommend flexibility testing as a foundational item in school-based fitness testing for youth. Although the evidence is not yet clear, however, flexibility may be linked to various health outcomes in youth, such as prevention of back pain, injury, and posture-related problems. Schools may therefore wish to include flexibility testing to help educate youth and their parents about flexibility as a component of overall musculoskeletal health, function, and performance.

Administrators can select flexibility tests to be implemented in schools and physical education settings based on their validity, reliability, and feasibility, for which evidence has been reported (see, e.g., Castro-Piñero et al., 2010; España-Romero et al., 2010; Freedson et al., 2000; Plowman, 2008; Safrit, 1990). Although the degree to which the sit-and-reach test is an indicator of overall systemic flexibility is unclear, only that test, including its alternatives (e.g., backsaver sit-and-reach), among the measures commonly used to assess flexibility in youth has been used widely, and it also has been the most frequently studied. The sit-and-reach test has reasonable validity and reliability when used in school settings.

Recommended Fitness Tests for Schools

The committee found strong evidence linking cardiorespiratory endurance and body composition to health in youth and evidence in adults to support a link between musculoskeletal fitness and health. Given the connections to health and the benefits of promoting a physically active lifestyle through physical fitness education, the committee selected measures of body composition, cardiorespiratory endurance, and musculoskeletal fitness that should be included in a fitness test battery for use in schools and other educational settings. As with national surveys, test administrators should distribute the equipment needed to conduct the recommended test items such that the students have the opportunity to receive sufficient training in the measurement protocols and to practice the tests. Likewise, both those administering the tests and those interpreting and communicating the test results should receive the appropriate training to prevent adverse events, minimize measurement and classification errors, create an encouraging environment for students, and ensure the confidentiality of the results.

RECOMMENDATION 9-1. Developers and administrators of fitness test batteries in schools and other educational settings should consider including the following test items:

- **standing height (measure of linear growth status) and weight (measure of body mass) to calculate BMI as an indicator of body composition;**
- **a progressive shuttle run, such as the 20-meter shuttle run, to measure cardiorespiratory endurance; and**
- **handgrip strength and standing long jump tests to measure upper- and lower-body musculoskeletal strength and power, respectively.**

Additional tests that have not yet been shown to be related to health but that are valid, reliable, and feasible may also be considered as supplemental educational tools. For cardiorespiratory endurance, alternatives to the shuttle run include distance and/or timed runs, such as the 9-minute or 1-mile run, while the modified pull-up and push-up are possible alternatives for measuring upper-body musculoskeletal strength. The curl-up may be considered in addition to the suggested musculoskeletal fitness tests for measuring core strength and endurance. Although the committee does not recommend a flexibility measure as a core component of a fitness test battery, administrators in schools and other educational settings may wish to include the sit-and-reach test or its alternatives (e.g., backsaver sit-and-reach) to measure flexibility. Experts who establish cut-points for interpreting performance on these fitness test items should follow the guidance provided in Chapter 3.

REFERENCES

Armour, K. M., and M. Yelling. 2007. Effective professional development for physical education teachers: The role of informal, collaborative learning. *Journal of Teaching in Physical Education* 26(2):177-200.

Artero, E. G., V. España-Romero, J. Castro-Piñero, F. B. Ortega, J. Suni, M. J. Castillo-Garzon, and J. R. Ruiz. 2011. Reliability of field-based fitness tests in youth. *International Journal of Sports Medicine* 32(3):159-169.

Averett, S., and D. Stifel. 2007. *Food for thought: The cognitive effects of childhood malnutrition in the United States*. http://sites.lafayette.edu/stifeld/files/2010/05/averett_stifel_obesity_june_2007.pdf (accessed June 26, 2012).

Bagully, M. D. 2006. *The impact of childhood obesity on academic performance*. MPP thesis, Georgetown Public Policy Institute, Georgetown University, Washington, DC.

Baranowski, T., J. Mendlein, K. Resnicow, E. Frank, K. W. Cullen, and J. Baranowski. 2000. Physical activity and nutrition in children and youth: An overview of obesity prevention. *Preventive Medicine* 31(2):S1-S10.

Boreham, C., and C. Riddoch. 2001. The physical activity, fitness and health of children. *Journal of Sports Sciences* 19(12):915-929.

Boreham, C., J. Twisk, C. Neville, M. Savage, L. Murray, and A. Gallagher. 2002. Associations between physical fitness and activity patterns during adolescence and cardiovascular risk factors in young adulthood: The Northern Ireland Young Hearts Project. *International Journal of Sports Medicine* 23:S22-S26.

Bouchard, C., and R. J. Shephard. 1994. Physical activity, fitness and health: The model and key concepts: International proceedings and consensus statement. In *Physical activity, fitness and health*, edited by C. Bouchard, R. J. Shephard, and T. Stephens. Champaign, IL: Human Kinetics Publishers. Pp. 11-22.

Cale, L., and J. Harris. 2009. Fitness testing in physical education: A misdirected effort in promoting healthy lifestyles and physical activity? *Physical Education and Sport Pedagogy* 14(1):89-108.

Castelli, D. M., C. H. Hillman, S. M. Buck, and H. E. Erwin. 2007. Physical fitness and academic achievement in third- and fifth-grade students. *Journal of Sport and Exercise Psychology* 29(2):239-252.

Castro-Piñero, J., E. G. Artero, V. España-Romero, F. B. Ortega, M. Sjostrom, J. Suni, and J. R. Ruiz. 2010. Criterion-related validity of field-based fitness tests in youth: A systematic review. *British Journal of Sports Medicine* 44(13):934-943.

Cawley, J., C. Meyerhoefer, and D. Newhousec. 2007. The impact of state physical education requirements on youth physical activity and overweight. *Health Economics* 16(12):1287-1301.

CDC (U.S. Centers for Disease Control and Prevention). 2010. *The association between school-based physical activity, including physical education, and academic performance*. Altanta, GA: U.S. Department of Health and Human Services.

Centeio, E. F., and X. D. Keating. 2011. How phat is your fitness knowledge? *Research Quarterly for Exercise and Sport* 82(1):A3.

Chen, S., A. Chen, and X. Zhu. 2012. Are K12 learners motivated in physical education? A meta-analysis. *Research Quarterly for Exercise and Sport* 83(1):36-48.

Corbin, C. B. 2009. Appropriate use of physical fitness tests. *The CAHPERD Journal* 72(1): 20-25, 40.

Corbin, C. B. 2010. Texas Youth Fitness Study: A commentary. *Research Quarterly for Exercise and Sport* 81(Suppl. 3):S75-S78.

Corbin, C. B., and R. P. Pangrazi. 1992. Are American children and youth fit? *Research Quarterly for Exercise and Sport* 63(2):96-106.

Corbin, C., and R. P. Pangrazi. 2008. Appropriate uses of Fitnessgram. In *Fitnessgram/Activitygram reference guide*. 3rd ed., edited by G. J. Welk and M. D. Meredith. Dallas, TX: Cooper Institute.

Corbin, C. B., J. R. Whitehead, and P. Y. Lovejoy. 1988. Youth physical fitness awards. *Quest* 40(3):200-218.

Corbin, C. B., P. Y. Lovejoy, P. Steingard, and R. Emerson. 1990. Fitness awards: Do they accomplish their intended objectives? *American Journal of Health Promotion* 4:345-351.

Dale, D., and C. B. Corbin. 2000. Physical activity participation of high school graduates following exposure to conceptual or traditional physical education. *Research Quarterly for Exercise and Sport* 71(1):61-68.

Dale, D., C. B. Corbin, and T. F. Cuddihy. 1998. Can conceptual physical education promote physically active lifestyles? *Pediatric Exercise Science* 10(2):97-109.

Donnelly, J. E., and K. Lambourne. 2011. Classroom-based physical activity, cognition, and academic achievement. *Preventive Medicine* 52:S36-S42.

Doutis, P., and P. Ward. 1999. Teachers' and administrators' perceptions of the Saber-tooth Project reform and of their changing workplace conditions. *Journal of Teaching in Physical Education* 18(4):417-427.

Engelman, M. E., and J. R. Morrow, Jr. 1991. Reliability and skinfold correlates for traditional and modified pull-ups in children grades 3-5. *Research Quarterly for Exercise and Sport* 62(1):88-91.

Ennis, C. D. 2007. Defining learning as conceptual change in physical education and physical activity settings. *Research Quarterly for Exercise and Sport* 78(3):138-150.

Erbaugh, S. J. 1990. Reliability of physical fitness tests administered to young children. *Perceptual and Motor Skills* 71(3, Pt. 2):1123-1128.

Ernst, M. P., C. B. Corbin, A. Beighle, and R. P. Pangrazi. 2006. Appropriate and inappropriate uses of Fitnessgram: A commentary. *Journal of Physical Activity and Health* 3(Suppl. 2):S90-S100.

España-Romero, V., E. G. Artero, D. Jimenez-Pavon, M. Cuenca-Garcia, F. B. Ortega, J. Castro-Piñero, M. Sjostrom, M. J. Castillo-Garzon, and J. R. Ruiz. 2010. Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine* 31(7):490-497.

Flegal, K. M., M. D. Carroll, R. J. Kuczmarski, and C. L. Johnson. 1998. Overweight and obesity in the United States: Prevalence and trends, 1960-1994. *International Journal of Obesity* 22(1):39-47.

Fox, K. R. 1988. The self-esteem complex and youth fitness. *Quest* 40(3):230-246.

Freedson, P. S., K. J. Cureton, and G. W. Heath. 2000. Status of field-based fitness testing in children and youth. *Preventive Medicine* 31(2):S77-S85.

Gordon-Larsen, P., R. G. McMurray, and B. M. Popkin. 2000. Determinants of adolescent physical activity and inactivity patterns. *Pediatrics* 105(6):E83.

Harris, J., and L. Cale. 2007. Children's fitness testing: A feasibility study. *Health Education Journal* 66(2):153-172.

Hasselstrom, H., S. E. Hansen, K. Froberg, and L. B. Andersen. 2002. Physical fitness and physical activity during adolescence as predictors of cardiovascular disease risk in young adulthood. Danish Youth and Sports Study. An eight-year follow-up study. *International Journal of Sports Medicine* 23:S27-S31.

Hillman, C. H., S. M. Buck, J. R. Themanson, M. B. Pontifex, and D. M. Castelli. 2009. Aerobic fitness and cognitive development: Event-related brain potential and task performance indices of executive control in preadolescent children. *Developmental Psychology* 45(1):114-129.

Janz, K. F., J. D. Dawson, and L. T. Mahoney. 2002. Increases in physical fitness during childhood improve cardiovascular health during adolescence: The Muscatine Study. *International Journal of Sports Medicine* 23:S15-S21.

Kamijo, K., M. B. Pontifex, K. C. O'Leary, M. R. Scudder, C. T. Wu, D. M. Castelli, and C. H. Hillman. 2011. The effects of an afterschool physical activity program on working memory in preadolescent children. *Developmental Science* 14(5):1046-1058.

Keating, X. D. 2003. The current often implemented fitness tests in physical in physical education programs: Problems and future directions. *Quest* 55:141-160.

Keating, X. F. D., L. Harrison, L. Chen, P. Xiang, D. Lambdin, B. Dauenhauer, W. Rotich, and J. C. Piñero. 2009. An analysis of research on student health-related fitness knowledge in K-16 physical education programs. *Journal of Teaching in Physical Education* 28(3):333-349.

Kibbe, D. L., J. Hackett, M. Hurley, A. McFarland, K. G. Schubert, A. Schultz, and S. Harris. 2011. Ten years of take 10!®: Integrating physical activity with academic concepts in elementary school classrooms. *Preventive Medicine* 52:S43-S50.

Kollath, J. A., M. J. Safrit, W. Zhu, and L. G. Gao. 1991. Measurement errors in modified pull-ups testing. *Research Quarterly for Exercise and Sport* 62(4):432-435.

Kulinna, P. H. 2004. Physical activity and HRF knowledge: How much 1-6 grade students know. *International Journal of Physical Education* 41(3):111-121.

Kulinna, P. H., and S. Silverman. 2000. Teachers' attitudes toward teaching physical activity and fitness. *Research Quarterly for Exercise and Sport* 71(1):80-84.

Lefevre, J., R. Philippaerts, K. Delvaux, M. Thomis, A. L. Claessens, R. Lysens, R. Renson, B. Vanden Eynde, B. Vanreusel, and G. Beunen. 2002. Relation between cardiovascular risk factors at adult age, and physical activity during youth and adulthood: The Leuven Longitudinal Study on Lifestyle, Fitness and Health. *International Journal of Sports Medicine* 23:S32-S38.

Martin, S. B., A. Ede, J. R. Morrow, Jr., and A. W. Jackson. 2010. Statewide physical fitness testing: Perspectives from the gym. *Research Quarterly for Exercise and Sport* 81(Suppl. 3):S31-S41.

Meredith, M. D., and G. J. Welk, eds. 2010. *Fitnessgram/Activitygram test administration manual*. 4th ed. Champaign, IL: Human Kinetics Publishers.

Miech, R. A., S. K. Kumanyika, N. Stettler, B. G. Link, J. C. Phelan, and V. W. Chang. 2006. Trends in the association of poverty with overweight among US adolescents, 1971-2004. *Journal of the American Medical Association* 295(20):2385-2393.

Mitchell, M., D. Castelli, and S. Strainer. 2003. Student performance data, school attributes, and relationships. *Journal of Teaching in Physical Education* 22(5):494-511.

Morrow, J. R., and P. S. Freedson. 1994. Relationship between habitual physical activity and aerobic fitness in adolescents. *Pediatric Exercise Science* 6(4):315-329.

Morrow, J. R., Jr., S. B. Martin, and A. W. Jackson. 2010. Reliability and validity of the Fitnessgram: Quality of teacher-collected health-related fitness surveillance data. *Research Quarterly for Exercise and Sport* 81(Suppl. 3):S24-S30.

NASPE (National Association for Sport and Physical Education). 2004. *Moving into the future: National physical education standards*. 2nd ed. Reston, VA: McGraw-Hill.

NASPE. 2009a. *Appropriate instructional practice guidelines for elementary school physical education: A position statement*. 3rd ed. http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5287-207931.elementaryapproprac.pdf (accessed June 26, 2012).

NASPE. 2009b. *Appropriate instructional practice guidelines for high school physical education: A position statement*. 3rd ed. http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5288-573262.hsapproprac.pdf (accessed June 26, 2012).

NASPE. 2009c. *Appropriate instructional practice guidelines for middle school physical education: A position statement*. 3rd ed. http://www.cahperd.org/cms-assets/documents/toolkit/naspe_apprroprac/5289-666992.msapproprac.pdf (accessed June 26, 2012).

NASPE. 2010. *Shape of the nation report: Status of physical education in the USA*. http://www.aahperd.org/naspe/publications/Shapeofthenation.cfm (accessed June 26, 2012).

Ogden, C. L., M. D. Carroll, L. R. Curtin, M. M. Lamb, and K. M. Flegal. 2010. Prevalence of high body mass index in U.S. children and adolescents, 2007-2008. *Journal of the American Medical Association* 303(3):242-249.

O'Sullivan, M., and D. Deglau. 2006. Chapter 7: Principles of professional development. *Journal of Teaching in Physical Education* 25(4):441-449.

Parish, L. E., and D. C. Treasure. 2003. Physical activity and situational motivation in physical education: Influence of the motivational climate and perceived ability. *Research Quarterly for Exercise and Sport* 74(2):173-182.

Pate, R. R., M. Dowda, and J. G. Ross. 1990. Associations between physical activity and physical fitness in American children. *American Journal of Diseases in Children* 144(10):1123-1129.

Pate, R. R., M. Pratt, S. N. Blair, W. L. Haskell, C. A. Macera, C. Bouchard, D. Buchner, W. Ettinger, G. W. Heath, A. C. King, A. Kriska, A. S. Leon, B. H. Marcus, J. Morris, R. S. Paffenbarger, Jr., K. Patrick, M. L. Pollock, J. M. Rippe, J. Sallis, and J. H. Wilmore. 1995. Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of the American Medical Association* 273(5):402-407.

Pate, R. R., C. Y. Wang, M. Dowda, S. W. Farrell, and J. R. O'Neill. 2006. Cardiorespiratory fitness levels among US youth 12 to 19 years of age: Findings from the 1999-2002 National Health and Nutrition Examination Survey. *Archives of Pediatrics and Adolescent Medicine* 160(10):1005-1012.

Pate, R. R., D. S. Ward, J. R. O'Neill, and M. Dowda. 2007. Enrollment in physical education is associated with overall physical activity in adolescent girls. *Research Quarterly for Exercise and Sport* 78(4):265-270.

Pate, R. R., R. P. Saunders, J. R. O'Neill, and M. Dowda. 2011. Overcoming barriers to physical activity: Helping youth be more active. *ACMS Health and Fitness Journal* 15(1):7-12.

Payne, V. G., and J. R. Morrow. 1993. Exercise and VO_2max in children: A metaanalysis. *Research Quarterly for Exercise and Sport* 64(3):305-313.

Physical Activity Guidelines Advisory Committee. 2008. *Physical Activity Guidelines Advisory Committee report*. Washington, DC: U.S. Department of Health and Human Services.

Placek, J. H., L. L. Griffin, P. Dodds, C. Raymond, F. Tremino, and A. James. 2001. Middle school students' conceptions of fitness: The long road to a healthy lifestyle. *Journal of Teaching in Physical Education* 20(4):314-323.

Plowman, S. A. 2005. Physical activity and physical fitness: Weighing the relative importance of each. *Journal of Physical Activity and Health* 2:143-158.

Plowman, S. A. 2008. Muscular strength, endurance, and flexibility assessments. In *Fitnessgram/Activitygram reference guide*, edited by G. J. Welk and M. D. Meredith. Dallas, TX: Cooper Institute. Pp. 11.11-11.40.

Rasberry, C. N., S. M. Lee, L. Robin, B. A. Laris, L. A. Russell, K. K. Coyle, and A. J. Nihiser. 2011. The association between school-based physical activity, including physical education, and academic performance: A systematic review of the literature. *Preventive Medicine* 52(Suppl. 1):S10-S20.

Rosas, S., J. Case, and L. Tholstrup. 2009. A retrospective examination of the relationship between implementation quality of the coordinated school health program model and school-level academic indicators over time. *Journal of School Health* 79(3):108-115.

Safrit, M. J. 1990. The validity and reliability of fitness tests for children: A review. *Pediatric Exercise Science* 2:9-28.

Safrit, M. J. 1995. *Complete guide to youth fitness testing*. Champaign, IL: Human Kinetics Publishers.

Saint Romain, B., and M. T. Mahar. 2001. Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in Physical Education and Exercise Science* 5(2):67-80.

Simons, J., S. Dewitte, and W. Lens. 2003. "Don't do it for me. Do it for yourself!": Stressing the personal relevance enhances motivation in physical education. *Journal of Sport and Exercise Psychology* 25(2):145-160.

Standage, M., J. L. Duda, and N. Ntoumanis. 2003. Predicting motivational regulations in physical education: The interplay between dispositional goal orientations, motivational climate and perceived competence. *Journal of Sports Sciences* 21(8):631-647.

Stewart, S., and M. Mitchell. 2003. Instructional variables and student knowledge and conceptions of fitness. *Journal of Teaching in Physical Education* 22:533-551.

Stroot, S. A., C. Collier, M. O'Sullivan, and K. England. 1994. Contextual hoops and hurdles: Workplace conditions in secondary physical-education. *Journal of Teaching in Physical Education* 13(4):342-360.

Tremblay, M., and M. Lloyd. 2010. Physical literacy measurement: The missing piece. *Physical and Health Education Journal* 76(1):26-30.

Twisk, J. W. R., H. C. G. Kemper, and W. van Mechelen. 2002. The relationship between physical fitness and physical activity during adolescence and cardiovascular disease risk factors at adult age. The Amsterdam Growth and Health Longitudinal Study. *International Journal of Sports Medicine* 23:S8-S14.

Welk, G. J., A. W. Jackson, J. R. Morrow, Jr., W. H. Haskell, M. D. Meredith, and K. H. Cooper. 2010. The association of health-related fitness with indicators of academic performance in Texas schools. *Research Quarterly for Exercise and Sport* 81(Suppl. 3):S16-S23.

Whitehead, J. R., and C. B. Corbin. 1991. Youth fitness testing: The effect of percentile-based evaluative feedback on intrinsic motivation. *Research Quarterly for Exercise and Sport* 62(2):225-231.

Winnick, J. P., and F. X. Short. 1999. *The Brockport physical fitness test manual*. Champaign, IL: Human Kinetics Publishers.

10

Future Needs

KEY MESSAGES

The committee's review of the scientific literature revealed that studies on fitness measures for youth often were not designed to answer questions related to understanding the relationships between fitness measures and health across all ages, genders, and racial/ethnic populations. Nonetheless, there is sufficient evidence suggesting relationships between specific test items and health outcomes in youth. More research in priority areas is warranted so that those relationships can be confirmed and applied to the development of batteries of health-related fitness tests for youth. Priority areas for research relate to understanding the relationship between fitness measures and health, including the effect of modifying factors, and to developing cut-points (cutoff scores) for interpreting performance on fitness tests.

The review of the literature conducted for this study revealed a number of associations between fitness tests and health markers or risk factors. A key revelation, however, came from the review of the study designs themselves. Many of the studies reviewed were performed in the school environment. Conducting such studies will not, in principle, introduce biases as long as the investigators are careful to adhere to Recommendation 10-1 below with regard to well-designed studies. Of greater concern, however, is that, perhaps as a result of the recent emphasis on physical activity, many of the studies reviewed were not designed to assess the independent association between performance on a fitness test and a health marker or risk factor. This is a fundamental but crucial point that determined the value of studies for the committee's work. Studies also would have been more valuable had they accounted for the effects of various modifying factors. Although not included in the Centers for Disease Control and Prevention's (CDC's) search strategy, studies predicting health outcomes in adulthood would be valuable as well. For example, selected measures of strength and power track both during adolescence and from adolescence into adulthood. For many test items, however, it remains to be determined whether fitness performance in youth is predictive of health outcomes in later life. In sum, the literature review revealed many gaps in our understanding of the relationship between fitness measures and health. The committee offers the following recommendations for designing and conducting research to address these gaps.

RECOMMENDATION 10-1. Well-designed research studies aimed at advancing understanding of the associations between fitness components and health in youth should be undertaken. Researchers should ensure that the interventions studied are both specific and sufficient (i.e., appropriate dosage and duration) to induce a change in fitness. In addition, studies should be designed so that the effect of potential confounders (e.g., nutrition, physical activity, demographic variables, maturity status) and the potential for adverse events can be analyzed.

RECOMMENDATION 10-2. Longitudinal studies should be conducted to provide empirical evidence concerning how health markers related to fitness track from youth into adulthood.

RECOMMENDATION 10-3. Randomized controlled trials and longitudinal studies should be undertaken to understand the following issues regarding the relationships between (1) specific fitness tests and health, and (2) fitness components and health:

- **Studies should explore the relationship between body composition measures and physical fitness tests and the potential interactions among body composition, fitness, and health in youth.**
- **Studies should examine the relationship between changes in cardiorespiratory endurance as measured by field tests, including the shuttle run and timed and distance runs, and subsequent changes in health risk factors in youth beyond weight status and cardiometabolic risk factors. Examples include bone health and neurocognitive function and behavior.**

- **Studies should address the relationship between specific musculoskeletal fitness test items and health markers in youth. Priority should be given to test items for which there is growing evidence, such as the handgrip strength or standing long jump test, or others that are promising. Since musculoskeletal fitness is a multivariate construct, the studies should be designed so that a variety of tests are conducted.**
- **Studies should investigate the relationship between specific flexibility test items (e.g., sit-and-reach and its modifications), either by themselves or in combination with musculoskeletal fitness test items, and potential health markers (e.g., back pain, posture, injury prevention). Such studies should include stretching interventions specifically designed to produce changes in joint-specific flexibility. Since flexibility is a multivariate construct, the studies should be designed so that a variety of tests are conducted. Researchers should investigate the development and validation of a general marker of musculoskeletal systemic flexibility and its relationship to health markers and risk factors.**
- **Studies should examine the potential effects of modifying factors (i.e., age, gender, race/ethnicity, body composition, maturity status, training status/practice, motor skill, socioeconomic factors) on fitness components and on the relationship between a change in a health-related fitness component and health markers in specific populations.**

RECOMMENDATION 10-4. Developers of national surveys of health-related physical fitness in youth should consider the inclusion of measures of cardiometabolic health, bone health, and neurocognitive function. The collection of fitness and health data in the same individuals would allow investigators to further confirm whether direct relationships between specific test items and health markers and risk factors exist.

RECOMMENDATION 10-5. When an association between a fitness test and a health marker is confirmed, research should be conducted to establish and validate health-related cut-points for that test. For example, given the association of skinfold measures with health markers, large national studies should be conducted to establish health-related cut-points for skinfold measures in youth.

As noted in the above recommendations, the field tests identified in this report are recommended for use in future research. The committee, however, acknowledges the need for continued research designed to identify the best health-related field tests. The committee has not recommended some fitness test items simply because they have not been studied well enough to justify their inclusion. It is not the committee's intent to eliminate from future consideration or from a research agenda those test items that currently do not meet the level of evidence necessary for inclusion in a battery of fitness tests for youth.

Appendix A

Agenda

Committee on Fitness Measures and Health Outcomes in Youth
November 15-16, 2011
The Keck Center
500 Fifth Street, NW, Washington, DC 20001

Meeting Goals

- To gather information about the relationship between fitness measures and health outcomes
- To learn lessons from implementation of fitness batteries in the field in the United States and other countries

Tuesday, November 15, 2011: Day 1

1:00 PM Welcome and Introductions
Russell Pate, Committee Chair

SESSION 1: ISSUES ASSOCIATED WITH MEASURING FITNESS IN YOUTH

Objective: To develop a shared understanding of specific issues and considerations related to fitness components and the interpretation of test performance results.

Moderator: Russell Pate

1:10 Associations Between Strength and Flexibility and Health
Sharon Plowman, Northern Illinois University

1:40 Fitness Measures and Metabolic Health
Lars Anderson, University of Southern Denmark (by phone)

2:10 Body Composition: Methods and Its Use in a Fitness Test
Tim Lohman, University of Arizona

2:40 Setting Cut-Points: Approaches and Issues
Greg Welk, Iowa State University

3:10 Discussion

3:40 Break

SESSION 2: CONSIDERATIONS FOR THE FIELD-BASED APPLICATION OF FITNESS MEASURES

Objective: To achieve greater awareness of the issues and considerations associated with developing, implementing, and evaluating a fitness test battery based on the experiences of those working with test batteries that are currently being used around the world.

Moderator: Russell Pate

3:50 Implementation of Fitness Batteries in Various Countries
U.S. Fitnessgram
Greg Welk, Iowa State University

EUROFIT
Colin Boreham, University College Dublin
Mark Tremblay, University of Ottawa, Canada

4:35 Panel Discussion

5:30 Experiences in Implementing Fitness Testing in Urban School Settings
Lori Benson, YMCA of Greater New York

6:00 Adjourn

Wednesday, November 16, 2011: Day 2

1:00 PM	Welcome and Introductions *Russell Pate, Committee Chair*

SESSION 3: FITNESS, BRAIN FUNCTION, AND ACADEMIC ACHIEVEMENT

1:10	Fitness and Brain Function *Charles H. Hillman,* *University of Illinois at Urbana-Champaign*
1:30	Fitness and Academic Achievement *Joseph Donnelly, University of Kansas*
1:50	Questions and Answers
2:00	Adjourn

Appendix B

Glossary

Adolescence: These years, from puberty to adulthood, may be roughly divided into three stages: early adolescence, generally ages 12 and 13; middle adolescence, ages 14 to 16; and late adolescence, ages 17 to 21. In addition to physiological growth, seven key intellectual, psychological, and social developmental tasks take place during these years. The fundamental purpose of these tasks is to form one's own identity and to prepare for adulthood.[1]

Adverse events: In the context of this report, any unexpected, damaging effect that occurs as the result of a performing a fitness test, such as an injury or physical pain.

Body composition: The components that make up body weight, including fat, muscle, and bone content. The committee defined body composition operationally as a component of fitness, a marker of health, and a modifier of fitness for the purposes of this report.

Cardiorespiratory endurance: The ability to perform large-muscle, whole-body exercise at moderate to high intensities for extended periods of time (Saltin, 1973).

[1] Available at http://www.healthychildren.org/English/ages-stages/teen/pages/Stages-of-Adolescence.aspx (accessed August 23, 2012).

Childhood: The period in human development that extends from birth until the onset of puberty.[2]

Concurrent validity: A type of measurement validity; a form of criterion-related validity; the degree to which the outcomes of one test correlate with outcomes on a criterion test when both tests are given at relatively the same time (Portney and Watkins, 2008).

Criterion-referenced standards: Evaluation standards used to interpret physical fitness test scores and provide information about a participant's health status. They are considered the most accurate measure of a construct and are used to validate field-based measures.

Cut-point (cutoff score): A test score that represents the minimum level of performance that must be achieved for a participant to be said to be at reduced risk or fit/healthy. Cut-points are critical to interpretation of the results of health-related fitness tests since they serve as a way to distinguish individuals and populations that may be at risk of poor health outcomes from those that are not.

Disability: Any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being. For the purposes of this report, this term should be construed in the broadest sense, covering impairments (i.e., a problem in body function or structure), activity limitations (i.e., a difficulty encountered by an individual in executing a task or action), and participation restrictions (i.e., a problem experienced by an individual in involvement in life situations). Adapted from the World Health Organization definition.[3]

Feasibility: In this report, the degree to which a fitness test can be administered and interpreted with ease. It is defined by criteria such as ease of administration; burden on participants and administrators; privacy and safety; equipment and space; complexity; and suitability for all socioeconomic levels, education levels, and ages.

Fitnessgram: A health-related fitness test and reporting program introduced by the Cooper Institute in 1988 (Cooper Institute, 2010).

---

[2]Available at http://medical-dictionary.thefreedictionary.com/childhood (accessed August 25, 2012).

[3]Available at http://www.who.int/topics/disabilities/en/ (accessed August 8, 2012).

Flexibility: The intrinsic property of body tissues that determines the range of motion achievable without injury at a joint or group of joints (adapted from Holt et al., 1996).

Health marker: An indicator of a particular health or disease state within an organism.

Health risk factor: A characteristic statistically associated with, although not necessarily causally related to, an increased risk of morbidity or mortality.[4] In this report, it refers to markers (characteristics) associated with a disease or illness that increase the chances of contracting that disease or illness.

Hypokinetic-related diseases: A term coined by Kraus and Raab, who described "somatic or mental derangements" that are "caused by insufficient motion" (Kraus and Raab, 1961, p. 8). In this report, hypokinetic-related diseases are health problems or illnesses that are caused partly by the lack of regular physical activity (Corbin and Lindsey, 2007, p. 320).

Metabolic health: Freedom from diseases or conditions associated with metabolic risk factors (metabolic syndrome); the sum of all cellular processes that provide the human body with the ability to function optimally and resist disease.

Modifying factors: As related to physical fitness tests, those factors that can independently affect an individual's level of fitness. They include both factors that are measurable in the field (e.g., gender, race, ethnicity, maturity) and those that are not (e.g., heredity, practice level, skill level).

Motor skill: The ability to perform complex muscle-and-nerve acts that produce movement. Fine motor skills are small movements such as writing and tying shoes; gross motor skills are large movements such as walking and kicking.

Muscle endurance: The ability of a muscle or group of muscles to perform repeated contractions against a constant external load for an extended period of time (Kell et al., 2001). The constant load can be either an absolute external resistance, which provides a measure of absolute endurance, or a relative load based on an individual's maximal strength, which provides a measure of relative endurance.

---

[4]Available at http://dictionary.webmd.com/terms/risk-factor (accessed August 25, 2012).

Muscle power: A physiological construct reflecting the rate at which work is performed (Knuttgen and Kraemer, 1987). It is derived from the product of the force production of a muscle or group of muscles and the velocity of the muscle contraction during a single- or multijoint action (Sale and Norman, 1982).

Muscle strength: The ability of skeletal muscle (single or group) to produce measureable force, torque, or moment about a single or multiple joints, typically during a single maximal voluntary contraction and under a defined set of controlled conditions, which include specificity of movement pattern, muscle contraction type (concentric, isometric, or eccentric), and contraction velocity (Farpour-Lambert and Blimkie, 2008; Kell et al., 2001; Sale and Norman, 1982).

Musculoskeletal fitness: A theoretical construct reflecting the integrated function of an individual's muscle strength, endurance, and power to enable the performance of work against one's own body or an external resistance.

Physical fitness: A set of attributes that people have or achieve relating to their ability to perform physical activity (HHS, 1996).

Reliability: The dependability of test scores, their freedom from error, and their reproducibility in repeated trials on the same individual. A reliable test will have little test-retest, intratester, and intertester variability.

Validity: The extent to which a test measures what it is designed to measure; the degree to which evidence supports the interpretation of test scores (Eignor, 2001).

VO_2max: Maximal oxygen consumption—the maximum capacity of an individual's body to transport and use oxygen during incremental exercise. It is considered to be the criterion-referenced standard for the fitness component of cardiorespiratory endurance.[5]

Youth: For this report, a period of human development that includes ages 5-18.

[5]Available at http://en.wikipedia.org/wiki/VO2_max (accessed August 28, 2012).

REFERENCES

Corbin, C. B., and R. Lindsey. 2007. *Fitness for life*. 5th ed. Glenview, IL: Scott, Foresman and Company.

Eignor, D. R. 2001. Standards for the development and use of tests: The standards for educational and psychological testing. *European Journal of Psychological Assessment* 17(3):157-163.

Farpour-Lambert, N. J., and C. J. R. Blimkie. 2008. Muscle strength. In *Paediatric exercise science and medicine*. 2nd ed., edited by N. Armstrong and W. van Mechelen. Oxford, UK: Oxford University Press. Pp. 37-53.

HHS (U.S. Department of Health and Human Services). 1996. *Physical activity and health: A report of the Surgeon General*. Atlanta, GA: HHS, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

Holt, J., L. E. Holt, and T. W. Pelham. 1996. Flexibility redefined. In *Biomechanics in sports XIII,* edited by T. Bauer. Thunder Bay, Ontario: Lakehead University. Pp. 170-174.

Kell, R. T., G. Bell, and A. Quinney. 2001. Musculoskeletal fitness, health outcomes and quality of life. *Sports Medicine* 31(12):863-873.

Knuttgen, H. G., and W. J. Kraemer. 1987. Terminology and measurement in exercise performance. *Journal of Strength and Conditioning Research* 1(1):1-10.

Kraus, H., and W. Raab. 1961. *Hypokinetic disease*. Springfield, IL: Charles C. Thomas Publishers.

Meredith, M. D., and G. J. Welk, eds. 2010. *Fitnessgram/Activitygram test administration manual*. 4th ed. Champaign, IL: Human Kinetics Publishers.

Portney, L. G., and M. P. Watkins. 2008. *Foundations of clinical research: Applications to practice*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

Sale, D. G., and R. W. Norman. 1982. Testing strength and power. In *Physiological testing of the elite athlete*, edited by J. D. Macdougall, H. A. Wenger, and H. J. Green. Canada: Mutual Press. Pp. 7-37.

Saltin, B. 1973. Oxygen transport by the circulatory system during exercise in man. In *Limiting factors of physical performance*, edited by J. Keul. Stuttgart, Germany: Thieme Medical Publishers. Pp. 235-252.

Appendix C

Acronyms

1RM	one repetition maximum
AAAPE	American Association for the Advancement of Physical Education
AAHPER	American Alliance for Health, Physical Education, and Recreation
AAHPERD	American Alliance for Health, Physical Education, Recreation and Dance
AAU	Amateur Athletic Union
ACLS	Aerobics Center Longitudinal Study
ACSM	American College of Sports Medicine
AF	abdominal fat
AFEA	Australian Fitness Education Award
ALPHA	Assessing Levels of Physical Activity (test battery)
ANOVA	analysis of variance
BF	body fat
BIA	bioelectric impedance analysis
BMC	bone mineral content
BMD	bone mineral density
BMI	body mass index
BP	blood pressure or bench press
BRFSS	Behavioral Risk Factor Surveillance System

CAHPER	Canadian Association for Health, Physical Education and Recreation
CAPL	Canadian Assessment of Physical Literacy Test
CDC	Centers for Disease Control and Prevention
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CM	centimeter
CMJ	countermovement jump
CRP	*C-reactive protein*
CT	computed tomography
CVD	cardiovascular disease
DBP	diastolic blood pressure
DOMS	delayed onset muscle soreness
DXA	dual-energy X-ray absorptiometry
E	early velocity
E1	early diastolic peak myocardial velocity
EUROFIT	European Test of Physical Fitness
F	female
FFM	fat-free mass
FM	fat mass
FMS	functional movement screening
FN	femoral neck
HDL-C	high-density lipoprotein cholesterol
HOMA-IR	homeostatic model assessment-insulin resistance
IOM	Institute of Medicine
JHS	joint hypermobility syndrome
LDL-C	low-density lipoprotein cholesterol
LE	level of evidence
LJ	long jump
LM	lean mass
LP	leg press
LS	lumbar spine
M	male
MRI	magnetic resonance imaging
MTU	muscle-tendon unit

NAPFA	National Physical Fitness Award (Singapore)
NASPE	National Association for Sport and Physical Education
NHANES	National Health and Nutrition Examination Survey
NHES	National Health Examination Survey
NHIS	National Health Interview Survey
NIS	National Immunization Survey
PA	physical activity
PACER	progressive aerobic cardiovascular endurance run
PCFSN	President's Council on Physical Fitness, Sports and Nutrition
PCPF	President's Council on Physical Fitness
PCPFS	President's Council on Physical Fitness and Sports
PFI	Physical Fitness Index
PNF	proprioceptive neuromuscular facilitation
PQCT	peripheral quantitative tomography
PSDQ	Physical Self Description Questionnaire
PWC-170	Physical Working Capacity-170 test
ROC	receiver operating characteristic
ROM	range of motion
SBP	systolic blood pressure
SD	standard deviation
SLJ	standing long jump
TC	total cholesterol
TG	triglycerides
TFM	total fat mass
TT	treadmill time
VJ	vertical jump
VO_2max	maximal oxygen consumption
VO_2peak	peak oxygen consumption
W	watts
WC	waist circumference
YMCA	Young Men's Christian Association

Appendix D

Biographical Sketches of Committee Members

Russell R. Pate, Ph.D. (*Chair*), is professor of exercise science at the Norman J. Arnold School of Public Health, University of South Carolina, Columbia. Dr. Pate's research interests and expertise focus on physical activity measurement, determinants, and promotion in children and youth. He also directs a national postgraduate course aimed at developing research competencies related to physical activity and public health. Dr. Pate is involved in the Centers for Disease Control and Prevention–funded Prevention Research Center at the University of South Carolina. His research includes studies on preschoolers' physical activity levels and how schools can influence these levels, as well as multicenter trials on the promotion of physical activity among middle and high school-age girls. Dr. Pate was a member of the Physical Activity Guidelines Advisory Committee of the U.S. Department of Health and Human Services and served on the 2005 Dietary Guidelines Advisory Committee. He is a past president of both the American College of Sports Medicine and the National Coalition on Promoting Physical Activity. Dr. Pate served as a member of the Institute of Medicine Committee on Prevention of Obesity in Children and Youth and Committee on Progress in Preventing Obesity in Children and Youth, and is a current member of the Standing Committee on Childhood Obesity Prevention. He received a B.S. in physical education from Springfield College and an M.S. and Ph.D. in exercise physiology from the University of Oregon.

Cameron Blimkie, Ph.D., is professor in the Department of Kinesiology at McMaster University in Ontario, Canada. He is a specialist in pediatric exercise physiology, known internationally for his expertise in the neuromuscular and musculoskeletal adaptations of children to exercise. He recently completed a 6-year term as director of the Graduate Program in Kinesiology. Dr. Blimkie is former foundation chair of pediatric exercise science research at the New Children's Hospital in Sydney, Australia, and has held invited visiting professorships at the Catholic University of Leuven, Belgium, and the Department of Pediatrics, the University Hospital of Geneva, Switzerland. He is a former vice president of the Canadian Society for Exercise Physiology and served as a consultant to the Swiss Musculo-Skeletal Health Research Program. Dr. Blimkie has also been a member of the editorial board of the journal *Pediatric Exercise Science* for the past 20 years and is a longtime fellow of the American College of Sports Medicine (ACSM). He was an invited contributor to the Clinical Guidelines on Osteoporosis issued by the Osteoporosis Society of Canada, to the Bone Health section of the recent Guidelines for Youth Physical Activity sponsored by the Centers for Disease Control and Prevention, and to a recent consensus statement on Training Considerations for Young Athletes sponsored by the International Olympic Committee Medical Commission. Dr. Blimkie's research spans a spectrum of special clinical populations, including children with obesity, diabetes, and cystic fibrosis, as well as normal healthy children and elite young athletes. He received his B.A. in combined arts and physical education from McMaster University and his M.A. in physical education and Ph.D. in medical physiology from the University of Western Ontario.

Darla Castelli, Ph.D., is an associate professor in the Department of Kinesiology and Health Education at the University of Texas at Austin. Formerly a health and physical education teacher, she is a past president of the Maine Association for Health, Physical Education, Recreation and Dance (AHPERD) and a 1995 teacher of the year. From a pedagogical perspective, Dr. Castelli studies the effects of physical activity and metabolic risk factors on cognitive performance in school-aged children. She has presented her work at congressional briefings in Washington, DC, in support of the Fit Kids Act. For her role in this research, she was named a 2006 young scholar by the International Association of Physical Education in Higher Education and a past-president's scholar by the 2007 Illinois Association for Health, Physical Education, Recreation and Dance. Dr. Castelli has published in *Preventive Medicine*, *Medicine and Science in Sport and Exercise, Research Quarterly for Exercise and Sport*, and *Developmental Psychology.* Her research has been funded by the National Institutes of Health, the American Dietetic Association, and the U.S. Department of Education. She obtained her M.S. from Northern Illinois University and her Ph.D. from the University of South Carolina.

Charles B. Corbin, Ph.D., is professor emeritus in the Department of Exercise and Wellness at Arizona State University. Dr. Corbin has published more than 200 professional, research, and popular articles and 90 books on fitness and wellness. He is a fellow emeritus and former president of the National Academy of Kinesiology. Among his awards are the Healthy American Fitness Leaders Award of the President's Council on Physical Fitness and Sports (PCPFS, now President's Council on Physical Fitness, Sports and Nutrition [PCFSN]) and National Jaycees; the American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) Honor Award; the Physical Fitness Council Honor Award; the National Association for Sport and Physical Education (NASPE) Hall of Fame; ACSM Cureton Lecturer; the Distinguished Service Award of the President's Council; and the Gulick Award (AAHPERD). He is a lifetime member of AAHPERD and a longtime member and fellow in ACSM. Dr. Corbin was named alliance scholar by AAHPERD and distinguished scholar by the National Association of Kinesiology and Physical Education in Higher Education (NAKPEHE). He served as editor of several periodicals, including *Quest* and the *PCFSN Research Digest*. He was the first chair of the Scientific Board of the PCFSN. He was named as a distinguished alumnus by the Department of Kinesiology and Community Health at the University of Illinois and was honored as centennial scholar by the University of New Mexico at the university's 100th anniversary. Dr. Corbin received his B.S. from the University of New Mexico, his M.S. from the University of Illinois, and his Ph.D. from the University of New Mexico.

Stephen R. Daniels, M.D., Ph.D., is professor and chair of the Department of Pediatrics at University of Colorado School of Medicine. He is also pediatrician-in-chief and L. Joseph Butterfield chair in pediatrics at The Children's Hospital. Dr. Daniels held numerous academic and clinical appointments at the University of Cincinnati College of Medicine and the Cincinnati Children's Hospital before joining the University of Colorado School of Medicine and The Children's Hospital. His area of expertise is preventive cardiology, and he has a longtime interest in the application of sophisticated epidemiologic and biostatistical methods to pediatric clinical research problems. The role of lifestyle factors, such as diet and physical activity, is central to many of his studies. Dr. Daniels has received numerous awards and honors throughout his career. He has been an active participant and leader in many national investigative committees and study sections, including the American Academy of Pediatrics' Committee on Nutrition; the American Heart Association's Council for Cardiovascular

Disease in the Young; and the International Pediatric Hypertension Association's executive board, which he recently chaired. He has also served as a frequent participant in grant review study sections and science panels of the National Institutes of Health. Dr. Daniels has served as associate editor for the *Journal of Pediatrics* since 1995. He is a peer reviewer for many other journals and is widely published in the medical literature. He is co-author of *Medical Epidemiology*, an introductory textbook for medical students, and co-author and editor of the book *Pediatric Prevention of Atherosclerotic Cardiovascular Disease*, published in 2006. He earned his M.D. from the University of Chicago in 1977, his M.P.H. from Harvard University in 1979, and his Ph.D. in epidemiology from the University of North Carolina in 1989. He completed his residency in pediatrics and his fellowship in pediatric cardiology at the Cincinnati Children's Hospital Medical Center in 1981 and 1984, respectively.

Harold W. Kohl III, Ph.D., is professor in the School of Public Health and Michael and Susan Dell Center for Healthy Living at the University of Texas Health Science Center-Houston, and professor in the Department of Kinesiology and Health Education at the University of Texas-Austin. Dr. Kohl is founder and director of the University of Texas Physical Activity Epidemiology Program, where he is responsible for student training, research, and community service related to physical activity and public health. Dr. Kohl presently is co-editor-in-chief of the *Journal of Physical Activity and Health*. His previous service includes directing physical activity epidemiology and surveillance projects in the Division of Nutrition, Physical Activity, and Obesity at the Centers for Disease Control and Prevention. Dr. Kohl's research focuses on the specific area of epidemiology related to physical inactivity and obesity, in adults but also in children. He has researched the development of a system for measuring the effectiveness of state physical education policies in schools, as well as historical changes in physical activity. Dr. Kohl also studies the effect of the built environment on physical activity and is currently researching a planned development that implements "smart growth" techniques that support physically active lifestyles. He received an M.S.P.H. from the University of South Carolina School of Public Health in epidemiology and biostatistics and a Ph.D. from the University of Texas Health Science Center-Houston School of Public Health in community health studies.

Robert M. Malina, Ph.D., is professor emeritus in the Department of Kinesiology and Health Education at the University of Texas at Austin and an adjunct research professor in the Department of Kinesiology at Tarleton State University. Dr. Malina is also a visiting professor at the Research Institute for Sport and Exercise Sciences at Liverpool John Moores University in the United Kingdom and at the University School of Physical Education in Wroclaw, Poland. He has extensive research experience in the field of physical growth and maturation of children and youth and has published about 700 articles and book chapters. He has received many honors, including the Alliance Scholar Award from AAHPERD, the Citation Award from ACSM, the Honor Award from the North American Society for Pediatric Exercise Medicine, the Distinguished Scholar Award from the North American Society for the Psychology of Sport and Physical Activity, the Hetherington Award from the American Academy of Kinesiology and Physical Education, and the Franz Boas Distinguished Achievement Award from the Human Biology Association. He is a foreign member of the Polish Academy of Sciences (Division II, Biological Sciences) and a member of many professional organizations, including AAHPERD, ACSM, the American Association of Physical Anthropologists, the Society for the Study of Human Biology, the Human Biology Association, the European Anthropological Association, the Society for Research in Child Development, the American Association for the Advancement of Science, the American Academy of Kinesiology and Physical Education, and the International Association of Sport Kinetics. He also serves on the editorial board of numerous journals, including *Annals of Human Biology*; *Journal of Sports Sciences*; *Clinical Journal of Sports Medicine*; *Pediatric Exercise Science*; *Ovidius University Annals, Series Physical Education and Sport*; *Italian Journal of Sport Sciences*; *International Journal of Sport and Health Science*; and *Indian Journal of Sport Sciences*. Dr. Malina earned a Ph.D. in physical education from the University of Wisconsin, Madison in 1963 and a second Ph.D. in anthropology from the University of Pennsylvania in 1968. He was awarded honorary doctorates from Catholic University (Belgium) in 1989, University School of Physical Education in Cracow (Poland) in 2001, University School of Physical Education in Wrocław (Poland) in 2006, and University of Coimbra (Portugal) in 2008.

Jennifer Sacheck, Ph.D., is associate professor of nutrition in the John Hancock Research Center on Physical Activity, Nutrition, and Obesity Prevention at the Friedman School of Nutrition Science and Policy and a scientist in the Antioxidants Research Laboratory at the Jean Mayer Human Nutrition Research Center on Aging at Tufts University. She is an active member of ACSM and the Obesity Society. Dr. Sacheck's research expertise lies at the intersection of diet, physical activity, and health. Past and current studies in youth have involved biological, anthropometric, dietary, physical activity, and fitness measurements in both school and community settings. Currently, she is examining the impact of school-based physical fitness results and obesity on cardiovascular disease risk factors in urban schoolchildren and the relationship between changes in school-based fitness and the remission of overweight/obesity. Dr. Sacheck has authored reports on the childhood obesity epidemic in New England for the Massachusetts Health Policy Forum and the Harvard Pilgrim Healthcare Foundation and has also led several evaluations of community-based programs that target disadvantaged overweight/obese youth. She received her B.S. in biology from Syracuse University; her M.S. in exercise science from the University of Massachusetts, Amherst; and her Ph.D. in nutritional science from Tufts University. She completed her postdoctoral training at Harvard Medical School.

David Stodden, Ph.D., is an associate professor in the Department of Health, Exercise, and Sport Sciences at Texas Tech University. Prior to his current position, he was an assistant professor in the School of Human Movement, Sport, and Leisure Studies at Bowling Green State University and a consultant and strength and conditioning coach for the Cleveland Indians organization. His research focuses on promoting the acquisition of fundamental motor skills and on the association of motor skill competence with physical activity, health-related physical fitness, perceived competence, and obesity across the life span. Dr. Stodden is a member of the North American Society for the Psychology of Sport and Physical Activity, AAHPERD, and the National Strength and Conditioning Association (NSCA). He is serving as academic support on a committee for the National Physical Activity Plan, Education Sector. He also is a certified strength and conditioning specialist (NSCA). Dr. Stodden received his B.S. in biology from Buena Vista University, his M.S. in exercise and sport science from Iowa State University, and his Ph.D. in motor behavior from Auburn University.

Melicia Whitt-Glover, Ph.D., is president and chief executive officer of Gramercy Research Group, LLC. She is also adjunct assistant professor in the Department of Applied Behavioral Science at the University of Kansas, adjunct faculty member in the Department of Nutrition at the University of North Carolina at Chapel Hill, and adjunct associate professor in the Department of Epidemiology and Prevention at Wake Forest University Health Sciences. She serves as a reviewer for journals that include *American Journal of Preventive Medicine, Annals of Behavioral Medicine, Ethnicity and Disease, International Journal of Obesity, Journal of Physical Activity and Health, Medicine and Science in Sports & Exercise, Research Quarterly for Exercise & Sport*, and *Sports Medicine*. Dr. Whitt-Glover serves on the editorial board of *American Journal of Health Behavior* and as a member of the African American Collaborative Obesity Research Network, ACSM, and the International Society of Behavioral Nutrition and Physical Activity. Her honors and awards include an ACSM fellowship and an African American Professors Program fellowship from the University of South Carolina. Dr. Whitt-Glover received her B.A. and M.A. in exercise physiology from University of North Carolina at Chapel Hill and her Ph.D. in epidemiology from the University of South Carolina.

Weimo Zhu, Ph.D., is professor in the Department of Kinesiology and Community Health, College of Applied Health Sciences, at the University of Illinois at Urbana-Champaign (UIUC). He is a member of several professional organizations, including AAHPERD, the American Educational Research Association, the Rasch Measurement Special Interest Group, the American Educational Research Association, the National Council on Measurement in Education, ACSM, and the American Statistical Association. He serves on the editorial boards of several journals, including *International Journal of Applied Sport Science*, *Measurement in Physical Education and Exercise Science*, *Journal of Applied Measurement*, *Journal of Physical Activity and Health*, and *International Education*. Honors and awards received by Dr. Zhu include an American Academy of Kinesiology and Physical Education fellowship; membership on the President's Council on Physical Fitness and Sports Science Board; faculty fellow at the Academy for Entrepreneurial Leadership, UIUC; Outstanding Leadership ACSM-UIUC Walking for Health Specialty Conference: October 13-15, 2005, from ACSM; and the Measurement Honor Award from the Measurement and Evaluation Council, AAHPERD. Dr. Zhu received his B.S. in physical education from Nanjing Normal University, China; his M.S. in exercise physiology from Shanghai Institute of Physical Education, China; and his Ph.D. in physical education from the University of Wisconsin, Madison.