Abstract

Title of Dissertation: DEVELOPMENT AND EVALUATION OF NEW ACCELEROMETER CUT POINTS FOR ADOLESCENT GIRLS

JoAnn Kuo

Dissertation directed by: Professor Deborah Rohm Young

Epidemiology and Biostatistics, Kinesiology

Current negative trends in adolescent disease risk factors (e.g., overweight) may be related to physical activity. To study these relationships using accelerometers, how to estimate physical activity from accelerometer counts must be better understood. **PURPOSES**: (1) To develop new accelerometer cut points for estimating physical activity using disease risk factors as criteria. (2) To evaluate how estimates of physical activity using these newly developed cut points agree with comparison measures (i.e., a previously suggested cut point and self-report physical activity recall questionnaires). **METHODS**: National Health and Nutrition Examination Survey (NHANES) and Trial of Activity for Adolescent Girls (TAAG) data were examined. New cut points were

developed using iterative correlations and signal detection and receiver operating characteristic (ROC) curves. To identify new cut points, potential cut points were identified in a development sample and validated in an evaluation sample. Agreement between new cut points and comparison measures was examined using concordance correlation coefficients, Bland-Altman plots, McNemar's tests, and proportions of agreement. **RESULTS**: Using the correlation method, two new combinations of light, moderate, and vigorous intensity cut points were identified in NHANES (1900, 4300, and 10000 counts/min and 1900, 4000, and 5000 counts/min) and two in TAAG (1450, 1950, and 2450 counts/30 sec and 1050, 1550, and 2050 counts/30 sec). Using the signal detection/ROC curve method, eleven new cut points were identified in NHANES (ranging from 100 to 2300 counts/min) and three in TAAG (ranging from 100 to 200 counts/min). Concordance correlation coefficients for minutes of activity with a previously suggested cut point tended to be stronger (≥ 0.60) with higher cut points (≥ 2300 count/min), while those with questionnaires were less than 0.10 or the 95% confidence intervals included zero. One new cut point (1800 counts/min) was similar (p = 0.6) to a comparison measure for classifying meeting recommendations.
CONCLUSIONS: Some cut points may be more strongly associated with disease risk factors than previously suggested cut points developed using oxygen consumption, but associations are not strong. The new cut points and comparison measures may be measuring different aspects of physical activity, as they were in poor agreement.

DEVELOPMENT AND EVALUATION OF NEW ACCELEROMETER CUT POINTS FOR ADOLESCENT GIRLS

by

JoAnn Kuo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Deborah Rohm Young, Chair
Associate Professor Olivia Carter-Pokras
Professor James Hagberg
Professor Bradley Hatfield
Professor Margarita Treuth

Development and Evaluation of New Accelerometer Cut Points for Adolescent Girls

List of Tables

List of Figures

Chapter 1: Dissertation Introduction

BACKGROUND

The prevalence of overweight and obesity increased among children and adolescents from NHANES II (1976-1980) to NHANES III (1988-1994) to NHANES 1999-2000 (Ogden, Flegal, Carroll, & Johnson, 2002). It showed no signs of changing through 2005-2006 (Ogden, Carroll, & Flegal, 2008). This trend may have negative health consequences. Overweight and obese adolescents are at increased risk for insulin resistance and diabetes, hypertension, dyslipidemia, and poor cardiorespiratory fitness, as well as obesity in adulthood (Boreham & Riddoch, 2001). The overweight and obesity trend may be related to inadequate physical activity. The Physical Activity Guidelines Advisory Committee (2008) found that youth physical activity is associated with body composition, as well as cardiovascular and metabolic health and cardiorespiratory fitness. Accurate measures of physical activity are necessary to effectively study physical activity and related disease risk factors. Increased understanding of these relationships is necessary to reverse the current trends in obesity, physical activity, and related health outcomes.

Physical activity assessment

Physical activity is any bodily movement produced by skeletal muscle that results in energy expenditure (Caspersen, Powell, & Christenson, 1985). In addition to determining the relationship between physical activity and health, physical activity is measured for a variety of reasons including to determine levels of physical activity, to understand determinants of physical activity, and to evaluate intervention effectiveness.

The ability to study these relationships of physical activity depends on having practical measures that are valid and reliable.

Energy expenditure is a physiological consequence of physical activity. Hence, several measures of physical activity are measures of energy expenditure. The most accurate measures of energy expenditure are direct and indirect calorimetry. While calorimetry is highly accurate at measuring energy expenditure, it is too expensive and not easily feasible to be used in large studies. Calorimetry methods also alter or inhibit regular activity patterns and cannot be used in free-living situations. Doubly labeled water (DLW) can be used to measure accurately measure total energy expenditure in free-living situations. However, patterns of physical activity (i.e., frequency, intensity, duration) cannot be assessed.

Physical activity is comprised of several dimensions, including frequency (how often), intensity (how hard), duration (how long), and type (what is done). Different combinations of these components may expend the same amount of energy yet have different health and physiological effects. For example, body weight is changed when there is a difference between energy intake and energy expenditure. Hence, assessing total energy expenditure may be sufficient. However, physical activity that causes weight change may not affect cardiovascular fitness, which requires physiological adaptations. Which physical activity dimensions are of interest depends on the study purpose. The exact pattern of physical activity needed to receive health benefits and how that may differ by population subgroups and by the outcome of interest has not been well elucidated (Physical Activity Guidelines Advisory Committee, 2008).

Self-report recall questionnaires have often been used in large, epidemiologic studies to assess physical activity because they are relatively inexpensive and can efficiently collect large amounts of detailed information. However, information from recall questionnaires may not be accurate. Respondents, particularly children and adolescents, may not be able to accurately recall details of their physical activity episodes. For example, they may not recall having a special school assembly instead of PE and report being physically active in PE that day. They may perceive the details of their activity differently than intended by the researcher. For example, they may report an activity as being vigorous that the researcher would consider moderate. Furthermore, respondents may alter their responses, consciously or subconsciously, because they want to appear more physically active. Hence, information collected using recall questionnaires is subject to the cognitive ability of respondents as well as social desirability bias.

Accelerometers

Accelerometers address some of the weaknesses of DLW and recall questionnaires for assessing physical activity in field studies. Rather than measuring the energy expenditure of physical activity, accelerometers detect bodily movement. The rationale is that the acceleration of the body is directly proportional to the muscular forces that cause it; thus, acceleration is related to energy expenditure. Actigraph (formerly known as Computer Science and Applications (CSA) and Manufacturing Technology, Inc. (MTI)) model 7164 utilizes a mechanical lever that measures changes in acceleration. To reduce non-activity artifacts, the acceleration signal passes through an analog band-pass filter. Actigraph 7164 contains an 8-bit solid-state analog-to-digital converter that digitizes the filtered signal to counts. It stores these counts data at user-specified epochs (ActiGraph, 2009). Hence, information about the frequency, duration, and intensity of physical activity can be objectively collected in the field by accelerometers. However, research is ongoing to determine how best to translate these counts into useful physical activity information.

One method for quantifying physical activity from accelerometer data is to classify the intensity of each epoch using cut points. Cut points are often determined using oxygen consumption as the criterion because oxygen consumption is related to energy expenditure. Treuth et al. (2004) and Puyau et al. (2002) have developed cut points for light, moderate, and vigorous intensity physical activity for young people for the Actigraph 7164 using oxygen consumption as their criterion. That is, oxygen consumption and accelerometer counts were simultaneously recorded while participants performed activities of various prescribed intensities. The sample in the Puyau study included 12 girls and 14 boys, 6 to 16 years old. Oxygen consumption was measured in a room calorimeter. Participants in the Treuth study were 74 eighth grade girls. Oxygen consumption was measured using a portable, breath-by-breath metabolic unit (Cosmed K4b2, Rome, Italy). In both studies, participants performed structured activities including resting and sedentary, light, moderate, and vigorous activities. Cut points were identified that best classified different intensity levels of activity (e.g., moderate). Puyau et al. used linear regression of energy expenditure on activity counts to identify cut points. Treuth et al. used a linear mixed model and examined false positive and negative classifications. Using accelerometer cut points such as these, the intensity of each epoch can be classified. That is, epochs with an accelerometer count at or above that cut point are classified as that physical activity intensity. Since the intensity of each epoch is identified, the frequency, duration, and intensity of physical activity during the measurement period can be assessed with custom designed software or statistical packages.

Despite the similarities of the Treuth and Puyau studies, two different sets of cut points were developed. Treuth cut points were 101, 3000, and 5201 counts/min and Puyau cut points were 800, 3200, and 8200 counts/min for light, moderate, and vigorous activity, respectively. Depending on which cut point is used, one cut point might classify an epoch as being active while another might classify the same epoch as not being active (e.g., 200 counts/min). Furthermore, the cut points may not be appropriate for field studies if structured activities performed in these studies were not performed the same way they would be in real-world situations. While these cut points reflect oxygen consumption, they may not reflect cut points that best classify the relationship between physical activity and disease risk factors, especially in field studies.

Rationale for focusing on adolescent girls

This dissertation focused on adolescent girls. While children tend to be the most active age group, physical activity levels tend to decline during adolescence (Kimm et al., 2002; McMurray, Harrell, Bangdiwala, & Hu, 2003; Nader, Bradley, Houts, McRitchie, & O'Brien, 2008), and adolescent girls tend to be less active than adolescent boys (Caspersen, Pereira, & Curran, 2000; McMurray et al., 2003; Nader et al., 2008). Among youth followed from 9 to 15 years of age, mean minutes of physical activity fell from approximately 3 hours a day to less than 1 hour a day (Nader et al., 2008). Although the rates of decrease were similar between girls and boys, girls were less active than boys.

Hence, adolescent girls are an important target for physical activity research. In order to more effectively study this population, practical and accurate measures of physical activity must be developed. Accelerometers may meet this need, but more needs to be understood about how to interpret the data accelerometers generate to produce meaningful estimates of physical activity (Ward, Brown Rodgers, & Vaughn, 2005).

Dissertation purposes

The purposes of this dissertation were to develop new cut points for estimating physical activity from accelerometer data among free-living adolescent girls using cardiovascular and metabolic disease risk factors as criteria (e.g., body mass index) and evaluate these new cut points. This dissertation addressed two research questions:

1. What accelerometer cut point values can be identified that produce physical activity estimates that are more strongly associated with selected disease risk factors than previously suggested cut points?
2. How well do physical activity estimates produced using new accelerometer cut points agree with those using previously developed cut points and self-report physical activity recall questionnaires?

This dissertation contributes to the literature by employing two methods to develop accelerometer cut points and examining the convergent validity (i.e., the extent to which different instruments intended to measure the same construct agree) of these new cut points with self-report physical activity recall questionnaires and previously suggested cut points. Previous studies have used oxygen consumption as criterion for developing cut points because it is related to energy expenditure, an outcome of physical activity. This study used disease risk factors that may be related to adolescent physical activity as its criteria for developing cut points. Using disease risk factors as criteria may develop new cut points that are more clinically relevant because they are based on outcomes directly related to health and disease. Studies examining the agreement between self-report and objective physical activity measures have typically used Pearson's, Spearman's, or intraclass correlations. This study uses the concordance correlation coefficient (Lin, 1989), which may be more useful or appropriate for understanding the convergent validity of the new cut points with self-report recall questionnaires and previously suggested cut points because it was developed to evaluate agreement.

METHODS

This study used data from the Trial of Activity for Adolescent Girls (TAAG) and the 2003-2004 National Health and Nutrition Examination Survey (NHANES). These two datasets included accelerometer and disease risk factor data from large numbers of free-living adolescent girls from different locations across the United States, and both studies utilized the same accelerometer: Actigraph, model 7164. In addition to accelerometers, both studies also included a self-report physical activity recall questionnaire (i.e., TAAG 3-day physical activity recall and NHANES physical activity questionnaire).

TAAG collected data from sixth and eighth grade girls at six schools at each of the six field centers. Baseline measures were conducted with sixth grade girls in 2003. Follow-up measures were conducted with eighth grade girls in the same schools two years later. Additionally, eighth grade girls in the same schools were measured in 2006. Because many of the sixth grade girls were also measured in eighth grade, this dissertation restricted the study population to eighth grade TAAG girls. One-hundred twenty girls from each school were randomly selected from each eighth grade cohort to wear accelerometers. TAAG disease risk factor data included body mass index (BMI), percent body fat, and cardiorespiratory fitness.

NHANES collected data from about 5,000 people of all ages in households each year (Centers for Disease Control and Prevention, 2008). This dissertation restricted the study population to adolescent girls in NHANES to keep it comparable to the TAAG population. Data collection methods differed slightly between 12 to 15 year olds and 16 to 19 year olds (e.g., proxy answered race/ethnicity questions for persons younger than 16 years) (Centers for Disease Control and Prevention, 2008), so the study population was limited to 12 to 15 year old girls in NHANES. NHANES disease risk factor data included BMI, percent body fat, waist circumference, total cholesterol, high-density lipoprotein cholesterol (HDL-C), and blood pressure. NHANES collected fasting blood work from a subsample to assess glucose, triglycerides, and low-density lipoprotein cholesterol (LDL-C) levels (Centers for Disease Control and Prevention, 2008); sample sizes were too small to analyze these data. Additionally, NHANES fitness data were not analyzed because they were not available at the time of analysis.

Variables created

To answer the above research questions, physical activity and cardiovascular and metabolic disease risk factor variables were created in each dataset. Physical activity estimates included a) time spent in light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous intensity physical activity and b) prevalence of meeting physical activity recommendations. Adolescent physical activity recommendations used included

Strong et al. (at least 60 minutes daily) (2005), 2005 Dietary Guidelines for Americans (at least 60 minutes on most, preferably all, days) (U.S. Department of Health and Human Services & U.S. Department of Agriculture, 2005), and Healthy People 2010 Objective 22.6 (at least 30 minutes on 5 or more days per week) (U.S. Department of Health and Human Services, 2000). Continuous values for disease risk factors were categorized based on current recommendations or proposed cut points (Barlow & the Expert Committee, 2007; Lohman & Falls, 2001; de Ferranti et al., 2004; American Academy of Pediatrics Committee on Nutrition, 1998).

Although using sample weights is suggested when participants do not have equal probability of being selected, such as with stratification or clustering, they were not used in these analyses because the purpose was not to estimate population parameters or their associated standard errors. However, one suggested approach is to use unweighted analyses and control for variables used in the sampling design (Korn & Graubard, 1991). Analyses were done overall and within race/ethnicity, age, and weight status groups, where appropriate, to control for potential confounding factors.

Developing cut points

To address the first research question, in both the TAAG and NHANES sample, 75% of the participants were randomly assigned to the development sample and 25% were randomly assigned to the evaluation sample. Potential accelerometer cut points were identified in the development sample based on the relationships between physical activity and selected disease risk factors. New cut points were identified by validating the potential cut points in the evaluation sample.

Two methods were used to develop new cut points. One was an iterative process based on correlations (Chapter 2). Using previously suggested cut points (i.e., (Puyau et al., 2002; Treuth et al., 2004) as a starting point, new physical activity estimates were generated using multiple values at relatively large intervals. Rank correlations of these new physical activity estimates with each continuous disease risk factor were determined. Cut points that maximized the correlations were the new starting points for subsequent iterations. The process was repeated using smaller intervals until the final cut point was determined.

The other method for developing new cut points used signal detection and receiver operating characteristic (ROC) curves (Chapter 3). Each participant was classified as high or low risk for each disease risk factor. At given accelerometer count increments, each participant was classified as meeting or not meeting physical activity recommendations. The cut points that minimized misclassification (e.g., not meeting recommendations and low risk) were identified. ROC curves (i.e., plot true positive rate against false positive rate for different possible cut points) were used to examine how well the objective monitor separated the sample into high and low risk.

Evaluating cut points

Only new cut points that were consistently more strongly associated with disease risk than previously suggested cut points in both the development and evaluation sample were evaluated in the second research question (Chapter 4). Individual-level agreement was examined between the new accelerometer cut points and previously suggested accelerometer cut points (i.e., (Puyau et al., 2002; Treuth et al., 2004) and self-report physical activity recall questionnaires. Concordance correlation coefficients and Bland-

Altman plots were used to evaluate the degree to which estimated minutes of physical activity agreed. McNemar's tests and overall and specific proportions of agreement were used to evaluate whether similar percentages of participants were classified as meeting and not meeting physical activity recommendations.

DISSERTATION ORGANIZATION

This dissertation is organized in 5 chapters. The current chapter expands on the background presented in the manuscripts in the subsequent chapters and describes the overall dissertation. Three manuscripts are included in this dissertation. The manuscripts in chapters 2 through 4 include participant characteristics data. These data were limited to participants with sufficient accelerometer data for analysis. Additional participant characteristic data including all age-eligible NHANES and TAAG participants are presented in Appendices A and B.

Chapter 2 presents the manuscript that examined using the iterative correlation method to develop new cut points (Research Question 1). Data in this manuscript were limited to the new cut points identified. Additional correlation data from the development and evaluation samples for all girls and population subgroups for all disease risk factors examined are located in Appendices C and D.

The manuscript that investigated using the signal detection and ROC curve method to develop new cut points (Research Question 1) comprises Chapter 3. Only misclassification and area under the ROC curve data for the final new cut points were presented in this manuscript. Comparisons with the previously suggested cut points were restricted to the light-to-vigorous cut points because moderate-to-vigorous and vigorous cut points were not distinguishing (i.e., previously suggested moderate-to-vigorous and vigorous cut points always had similar or higher misclassification rates than the new cut points). Additional signal detection and ROC curve data for all girls and for all population subgroups for all physical activity recommendations, all intensities of previously suggested cut points, and all disease risk factors examined in the development and evaluation samples can be found in Appendices E and F.

Chapter 4 contains the evaluation study manuscript (Research Question 2). Agreement for the new moderate-to-vigorous cut points from the iterative correlation method and the new cut points from the signal detection/ROC curve method are discussed. Only agreement with the moderate-to-vigorous physical activity from the self-report recall questionnaires and the Treuth moderate-to-vigorous cut point are presented. The Treuth cut point was used, rather than the Puyau cut point, because it was developed in a similar population (i.e., 13 to 14 year old girls). Furthermore, agreement for meeting physical activity recommendations is limited to the Dietary Guideline accumulated time recommendation. Dietary Guidelines was used because it recommended 60 minutes of activity, similar to the Strong recommendation, but activity did not have to be performed daily (i.e., performed on most days of the week), similar to the Health People recommendation. Accumulated time was used because it was more inclusive than the frequency and duration recommendation. Additional agreement data for minutes of activity for all the new cut points for all intensities with the self-report physical activity recall questionnaires and the Treuth and Puyau cut points for all the physical activity recommendations examined are presented in Appendices G through L. Appendices M through O present additional data for meeting recommendations using the self-report physical activity recall questionnaires and the Treuth and Puyau cut points for all the physical activity recommendations examined.

Chapter 5 discusses conclusions from the three manuscripts, including the strengths and weaknesses of this dissertation. It also presents some weaknesses of accelerometers as well as issues to consider for future studies.

Chapter 2: Development of Accelerometer Cut Points for Adolescent Girls Using an Iterative Correlation Method with Body Composition and Cardiorespiratory Fitness

ABSTRACT

Physical activity tends to decrease during adolescence, with girls being less active than boys. Trends in adolescent overweight and fitness may be related to physical activity. To effectively study these relationships with accelerometers, the data they generate must be better understood. PURPOSE: To identify accelerometer cut points using body composition and cardiorespiratory fitness as criteria. METHODS: Accelerometer data from adolescent girls in the National Health and Nutrition Examination Survey (NHANES) and Trial of Activity for Adolescent Girls (TAAG) were examined. New cut point combinations that maximized rank correlations were identified using an iterative process. Potential cut point combinations were identified in a development sample. New cut point combinations were identified by validating the potential cut point combinations in an evaluation sample. RESULTS: New cut point combinations were identified in NHANES for BMI percentile (1900, 4300, and 10000 counts/min) and waist circumference percentile (1900, 4000, and 5000 counts/min) and in TAAG for percent body fat (1450, 1950, and 2450 counts/30 sec) and cardiorespiratory fitness (1050, 1550, and 2050 counts/30 sec). Average correlations were less than 0.3 in magnitude, and intensity-specific correlations were less than 0.4. The strongest correlations tended to be observed for moderate and/or vigorous physical activity. CONCLUSIONS: Previously suggested cut points developed using oxygen consumption may not best reflect the intensity of physical activity needed for health

benefits. Physical activity measured by accelerometers and body composition and cardiorespiratory fitness are weakly correlated.

INTRODUCTION

The prevalence of overweight and obesity increased among children and adolescents from NHANES II (1976-1980) to NHANES III (1988-1994) to NHANES 1999-2000 (Ogden et al., 2002). It showed no signs of changing through 2005-2006 (Ogden et al., 2008). Adolescent cardiorespiratory fitness may also be decreasing, particularly among girls (Malina, 2007). These trends may have negative health consequences. Adolescent body composition and cardiorespiratory fitness have been associated with cardiovascular disease risk factors in adolescence (Eisenmann, Welk, Wickel, & Blair, 2007; Ondrak, McMurray, Bangdiwala, & Harrell, 2007; Andersen et al., 2008) and found to track moderately into adulthood (Eisenmann, Wickel, Welk, & Blair, 2005). These trends may be related to physical activity. The Physical Activity Guidelines Advisory Committee (2008) found that youth physical activity is associated with body composition, cardiorespiratory fitness, as well as cardiovascular and metabolic health.

The exact pattern of physical activity needed to receive health benefits and how that may differ by population subgroups and by the outcome of interest has not been well elucidated (Physical Activity Guidelines Advisory Committee, 2008). Adolescents are of particular interest as this is a time when physical activity tends to decrease (Kimm et al., 2002; McMurray et al., 2003; Nader et al., 2008). Furthermore, adolescent girls tend to be less active than adolescent boys (Caspersen et al., 2000; McMurray et al., 2003; Nader et al., 2008). Increased understanding of determinants of physical activity and related disease risk factors is necessary to reverse the current trends in obesity, physical activity, and related outcomes.

Accurate measures of physical activity are necessary to effectively study the relationship between physical activity and cardiorespiratory fitness and body composition. Self-report measures are often used to assess physical activity because they are inexpensive and can efficiently collect large amounts of detailed information. However, the information collected may be subject to biases depending on participants' inability to accurately recall or report their activity or their tendency to give more socially desirable answers. Accelerometers can objectively record the frequency, duration, and intensity of activity, thereby overcoming these weaknesses of self-report measures. However, it is still not clear how to interpret the data accelerometers produce (Ward et al., 2005).

Accelerometers record a count for each user-defined epoch that is proportional to the movement of activity. One method for quantifying physical activity is to classify the intensity of each epoch using cut points. The cut points are often determined using oxygen consumption as the criterion because oxygen consumption is an outcome of energy expenditure. For example, two previous studies have calibrated the Actigraph, model 7164 (formerly known as Computer Science and Applications (CSA) and Manufacturing Technology, Inc. (MTI)) against oxygen consumption among children or adolescents (Puyau et al., 2002; Treuth et al., 2004). The sample in the study by Puyau et al. included 12 girls and 14 boys, 6 to 16 years old. Oxygen consumption was measured in a room calorimeter. Participants in the Treuth et al. study were 74 eighth grade girls. Oxygen consumption was measured using a portable, breath-by-breath metabolic unit

(Cosmed K4b2, Rome, Italy). In both studies, oxygen consumption and accelerometer counts were simultaneously recorded while participants performed structured activities including resting, sedentary activities, light activities, moderate activities, and vigorous activities. Then cut points were identified that best classified different intensity levels of activity (e.g., moderate). Epochs with accelerometer counts at or above that cut point can be classified as that intensity.

The structured activities performed in calibration studies may not be performed the same way they would be in real-world situations. Furthermore, while using oxygen consumption reflects the energy expenditure associated with physical activity, they may not reflect other aspects of physical activity that may be related to disease risk. Cut points developed using disease risk factors as the criteria rather than oxygen consumption may be more appropriate to use when disease risk is the outcome of interest. Hence, the purpose of this study was to use data from two large-scale field studies that included adolescent girls to identify new cut points that are more strongly associated with measures of cardiorespiratory fitness and body composition than previously suggested cut points that used oxygen consumption as the criterion.

METHODS

Data from the National Health and Nutrition Examination Survey (NHANES) and Trial of Activity for Adolescent Girls (TAAG) were used to develop new accelerometer cut points for estimating physical activity in adolescent girls. Both of these datasets included accelerometer data from large numbers of adolescent girls from different locations across the United States.

Study design

This study used data from NHANES 2003-04. Each year, NHANES randomly selected about 5,000 people from fifteen different locations using a complex stratified, multi-stage probability sampling design to obtain a representative sample population. It oversampled Blacks, Mexican-Americans, adolescents, older people, and pregnant women. Household interviewers identified and enrolled survey participants, conducted household interviews, and appointed study participants for the mobile examination center (MEC) exam (Centers for Disease Control and Prevention, 2003). A proxy answered race/ethnicity questions for persons less than 16 years old during the interviewer-administered Sample Person Questionnaire. Race/ethnicity variables were derived by combining responses to questions on race and Hispanic origin. Participants could mark all that applied. Race categories included non-Hispanic White, non-Hispanic Black, Mexican American, other Hispanic, and other race – including multi-racial (Centers for Disease Control and Prevention, 2009). In the MEC, participants had their body measurements taken and received an accelerometer to monitor their physical activity (Centers for Disease Control and Prevention, 2008).

TAAG was a group-randomized multi-center physical activity trial for middle school girls. In Spring 2003, baseline measures were conducted with sixth grade girls attending six schools at each of the six field centers (University of Maryland, University of South Carolina, University of Minnesota, Tulane University, University of Arizona, and San Diego State University). Follow-up measures were conducted with eighth grade girls in the same 36 schools in Spring, 2005. Additionally, eighth grade girls in 34 of the schools were measured in Spring, 2006; two schools in New Orleans were closed due to

Hurricane Katrina. This study used data from the two eighth grade cohorts. All eighth grade girls in the participating schools were eligible. They were excluded if they could not read and understand questions written in English. Participants indicated their race/ethnicity on a checklist including White, Black or African American, Hispanic, Asian, Pacific Islander, American Indian or Alaska Native, or Other. They could mark all that applied. In each cohort, 120 girls from each school were randomly selected to wear an accelerometer. In 2005, cardiorespiratory fitness was measured using a submaximal cycle ergometer test with a subsample of 40 girls from each school. They were excluded from the fitness test if they were taking a contra-indicated medication.

Measurement

Physical activity. NHANES and TAAG used the ActiGraph accelerometer, model 7164, as an objective measure of physical activity. Actigraph is a uniaxial accelerometer that detects and records acceleration of movement, especially locomotor-type (e.g., walking, jogging) activities, at user-specified intervals. Girls wore it on an elasticized belt on their right hip, except when it might get wet (e.g., swimming or bathing) or while sleeping. NHANES participants wore the accelerometer for 7 full days. It started recording data in 1-minute epochs at 12:01 a.m. the day after the girl received the monitor. Participants returned the accelerometer by mail in postage-paid, padded envelopes (Centers for Disease Control and Prevention, 2006). TAAG participants wore the accelerometer for 6 or 7 complete days. It started recording data in 30-second epochs at 5:00 a.m. the day after the girl received the monitor. Participants returned their accelerometer one week later. If TAAG girls did not wear the monitor for at least 10 hours on 1 day or the monitor malfunctioned, they were asked to wear the accelerometer again for one more week.

In the current study, accelerometer non-wear time was defined as 60 minutes or more of consecutive zero counts. At least 10 hours of wear was required for a day to be valid and at least two valid weekdays and one valid weekend day was required for a participant to be included in the analysis. Counts of 24000/min (12000/30 sec) or greater were considered extreme values. They were included in the wear time as non-zero values, but they were not counted as physical activity time.

Minutes of light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous physical activity were determined for different combinations of cut points, including cut points previously developed in adolescents using oxygen consumption as the criterion by Treuth et al. (2004) (i.e., 101, 3000, and 5201 counts/min) and Puyau et al. (2002) (i.e., 800, 3200, and 8200 counts/min). The intensity of each epoch was classified based on its accelerometer count relative to the cut points.

Body mass index (BMI) percentile. Height and weight were measured to the nearest 0.1 cm and 0.1 kg, respectively. They were measured once in NHANES (Centers for Disease Control and Prevention, 2004a). TAAG used the average of two measurements. BMI was calculated as weight (kg) divided by height (m) squared. Age-specific BMI percentiles were determined using the 2000 CDC growth chart tables for girls (Centers for Disease Control and Prevention, 2000).

Percent body fat. Triceps skinfold was measured at the posterior midline of the right upper arm. NHANES used Holtain skinfold calipers and measured triceps skinfold once to the nearest 0.1 mm (Centers for Disease Control and Prevention, 2004a). TAAG used Lange skinfold calipers and used the average of three measurements recorded to the nearest 1 mm. Percent fat was estimated with an equation including race/ethnicity contrast, age, height, weight, and triceps skinfold [percent body fat mass = –23.393 + 2.269(BMI [kg/m^2]) + 1.943(triceps skinfold [mm]) – 2.995(race/ethnicity) – 0.524(age [yr]) – 0.058 (BMI [kg/m^2])(triceps skinfold [mm])] (Lohman et al., 2006). Race/ethnicity was 1 if non-Hispanic black and 0 if otherwise.

Waist circumference percentile. Waist circumference was measured in NHANES to the nearest 1 mm just above the uppermost lateral border of the right ilium (Centers for Disease Control and Prevention, 2004a). Age- and ethnicity-specific waist circumference percentiles were determined using values for girls from Fernandez et al. (2004).

Cardiorespiratory fitness. TAAG used physical work capacity (PWC) 170, a submaximal cycle ergometer test, to predict power output (watts/kg) at 170 beats per minute (bpm) as an estimate of cardiorespiratory fitness. Participants exercised on a Monark model 818 mechanically-braked cycle ergometer and wore a Polar heart rate monitor to measure heart rate during the test. The PWC-170 protocol consisted of up to four two-minute stages. Participants pedaled at a cadence of 60 revolutions per minute. The initial workload was based on the participant's weight. Resistance during the subsequent stages was increased based on the participant's average heart rate during the last 10 seconds of each stage. The test was ended if the average heart rate during the last 10 seconds of the stage exceeded 165 bpm or the participant exhibited physical or verbal signs or symptoms of exercise intolerance.

Identifying new cut point combinations

New combinations of light, moderate, and vigorous cut points were identified using an iterative process based on correlations. Potential new cut points were developed using a random sample of 75% of participants from each dataset, and a validation study of the potential new cut points was conducted with the remaining 25%. Potential cut points were identified in the development sample based on the rank correlations between physical activity and body composition variables. Additionally in TAAG, the correlation between physical activity and estimated cardiorespiratory fitness was used. New cut points were identified based on the rank correlations of the potential cut points in the evaluation sample. Unweighted analyses were used. However, correlations were examined overall and within 3 race/ethnicity groups in NHANES (i.e., White, Hispanic, and Black), within 4 race/ethnicity groups in TAAG (i.e., White, Hispanic, Black, and Asian), and within 2 age groups in NHANES (i.e., 12-13 and 14-15 years), to control for potential confounding factors and variables used in the sampling designs (Korn et al., 1991). Additionally, cardiorespiratory fitness was examined within weight status groups (i.e., under or normal weight and at risk for overweight or overweight) in TAAG.

The initial cut points tested were based on cut points previously suggested by Treuth et al. (2004) and Puyau et al. (2002). For NHANES accelerometer data, the initial light cut points tested started at 100 counts/min and went up to 1000 counts/min in increments of 100 counts/min, the initial moderate cut points started at 2000 counts/min and went up to 4000 counts/min in increments of 200 counts/min, and the initial vigorous cut points started at 3200 counts/min and went up to 9200 counts/min in increments of 500 counts/min. The NHANES cut points within a test combination had to differ by at least 1000 counts/min, which was equivalent to almost 1 MET according to the equation MET = 2.01 + 0.000856 (counts/min) (Treuth et al., 2004). The initial TAAG cut point numerators were half the value of the NHANES cut points because TAAG accelerometer data were collected in 30-second epochs and NHANES accelerometer data were collected in 1-minute epochs (e.g., 50 counts/30 sec in TAAG vs. 100 counts/min in NHANES).

For each combination of cut points, estimates of the average daily minutes of physical activity at light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous intensity were generated. Minutes of physical activity per day were normalized to a 12-hour day. For each intensity level, the average daily physical activity time was defined as the total normalized number of minutes of physical activity divided by the total number of days of valid accelerometer data. Then, Spearman rank correlations were determined in the development sample between time spent in different intensities of physical activity and body composition and cardiorespiratory fitness variables.

The test combination of cut points that maximized the average correlation (i.e., average of the correlations for light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous physical activity) was selected as the starting point for the subsequent iteration if two conditions were met. First, in order to only develop new cut point combinations when it was probable that physical activity and body composition or fitness were associated, the average correlation needed to have a magnitude greater than 0.100. Second, the maximum average correlation using the test combination had to be stronger than that using either the Treuth or the Puyau cut points. If both conditions were met, the process was repeated in the development sample using the selected test combination as the midpoints of the subsequent light, moderate, and vigorous cut points to test.

Eleven values for each intensity cut point at given intervals were tested – the cut point from the selected test combination and five cut points above and below. The second iteration tested cut points at intervals of 200 counts/min or 100 counts/30 sec. The third iteration tested cut points at intervals of 100 counts/min or 50 counts/30 sec. The smallest intervals examined were 100 counts/min and 50 counts/30 sec because they corresponded to approximately 0.1 MET according to the Treuth et al. (2004) equation. If the light cut point was too low to have five values below the cut point (e.g., 100 counts/min), additional values above the cut point were tested (e.g., from 100 to 1100 counts/min by 100 counts/min).

To be tested in the evaluation sample, the average correlation of the potential cut point combination in the final iteration in the development sample needed to be at least 0.050 stronger than either the Treuth or the Puyau average correlations. This was done in order to only validate potential cut point combinations when it was probable that it was more strongly correlated than previously suggested cut points. Spearman correlations using the potential cut point combinations and the Treuth and Puyau cut points were determined in the evaluation sample. Potential cut point combinations with an average correlation with a magnitude greater than 0.100 and at least 0.050 stronger than either the Treuth or the Puyau average correlations in the evaluation sample were identified as new cut point combinations.

RESULTS

Participant characteristics for the NHANES and TAAG datasets overall and within the development and evaluation samples are presented in Table 2.1. There were no significant differences between the development and evaluation samples (all $p > 0.10$).

Table 2.1. **NHANES and TAAG participant characteristics by sample**

Dataset/Characteristic	Full Sample N	Full Sample Mean (SD)	Development Sample N	Development Sample Mean (SE)	Evaluation Sample N	Evaluation Sample Mean (SE)	p-value[1]
NHANES							
Age (year)	332	14.0 (1.12)	249	14.0 (0.07)	83	14.1 (0.12)	0.12
BMI percentile[2]	327	66.5 (26.87)	245	67.0 (1.72)	82	65.0 (2.99)	0.55
WC percentile[3]	318	60.8 (24.20)	238	61.0 (1.56)	80	60.0 (2.77)	0.73
Percent body fat	308	28.7 (8.34)	227	28.7 (0.55)	81	28.9 (0.96)	0.83
TAAG							
Age (year)	4687	14.0 (0.48)	3516	14.0 (0.01)	1171	13.9 (0.01)	0.32
BMI percentile	4687	66.8 (27.43)	3516	67.1 (0.46)	1171	65.7 (0.82)	0.12
Percent body fat	4686	31.4 (8.46)	3516	31.5 (0.14)	1170	31.2 (0.25)	0.28
Cardiorespiratory fitness[4]	692	11.9 (3.77)	503	12.0 (0.17)	189	11.8 (0.27)	0.65

NHANES: National Health and Nutrition Examination Survey; TAAG: Trial of Activity for Adolescent Girls; BMI: body mass index; WC: waist circumference

1 P-values are from t-tests.

2 BMI percentiles are age- and sex-specific (CDC, 2000).

3 WC percentiles are age-, sex-, and ethnicity-specific (Fernandez et al, 2004).

4 Cardiorespiratory fitness is defined as the estimated power output (watts/kg body weight) at a heart rate of 170 beats per minute predicted from a multi-stage cycle ergometry test.

NHANES received accelerometer data from 451 girls aged 12 to 15 years, and 333 (74%) of those girls had sufficient data for analysis. Girls that were not included tended to have higher BMI percentile, waist circumference percentile, and percent body fat, though the differences were not significant (p = 0.05 to 0.08, data not shown). Thirty-eight percent of the 333 girls who were included in the analysis were Hispanic, 33% were Black, and 26% were White (data not shown). TAAG received accelerometer data from 7397 eighth grade girls, and 4696 (63%) of those girls had sufficient data for analysis. Girls that were not included tended to be older, have higher percent body fat, and be Black ($p < 0.0001$, data not shown). Nearly half of the 4696 TAAG girls were White (47%), 22% were Hispanic, 18% were Black, and 6% were Asian (data not shown). The average age in both datasets was 14.0 years. The average BMI percentiles (67th percentile) were similar in NHANES and TAAG. The average percent body fat was 28.7% in NHANES and 31.4% in TAAG. In NHANES, the average waist circumference percentile was the 61st percentile. In TAAG, the average estimated power output at a heart rate of 170 beats per minute was 12 watts/kg body weight.

In the NHANES development sample, 7 cut point combinations had an average correlation of at least 0.100 in the expected direction. Of those, 4 had average correlations of at least 0.050 better than one or both of the previously suggested cut point combinations. These 4 potential cut point combinations were identified among 12-13 year old girls for BMI percentile, waist circumference percentile, and percent body fat; and among Hispanic girls for percent body fat. In the TAAG development sample, 9 cut point combinations had an average correlation of at least 0.100 in the expected direction. Of those, 2 had average correlations of at least 0.050 better than one or both of the previously suggested cut point combinations. These 2 potential cut point combinations were identified among Black girls for percent body fat and cardiorespiratory fitness.

New cut points in NHANES

The correlations in the development and evaluation samples for the previously suggested and new cut point combinations identified in the NHANES dataset are presented in Table 2.2. The new cut point combinations were identified among 12-13 year old girls in NHANES for BMI percentile (1900, 4300, and 10000 counts/min) and waist circumference percentile (1900, 4000, and 5000 counts/min). The new light-to-vigorous cut points were 1100 to 1800 counts/min higher than the corresponding Treuth and Puyau cut points, and the new moderate-to-vigorous cut points were 800 to 1300 counts/min higher. The new vigorous cut point for BMI percentile was 1800 to 4800 counts/min higher than the previously suggested cut points, while the cut point for waist circumference percentile was 200 to 3200 counts/min lower.

The average correlations in NHANES between physical activity and BMI percentile and waist circumference percentile were stronger using the new cut point combinations compared with the Puyau cut points in both the development and evaluation samples. However, the average correlations were only stronger than the Treuth cut points in the development sample. In the development sample, the average correlation between physical activity and BMI percentile using the new cut point combination (−0.175) was 0.066 stronger than using the Puyau cut points and 0.122 stronger than using the Treuth cut points. However, in the evaluation sample, the average correlation with BMI percentile using the new cut point combination (−0.264) was 0.089 stronger than using the Puyau cut points but only 0.019 stronger than using the Treuth cut

Table 2.2. Spearman's rank correlations between body composition and average daily minutes spent in light, moderate, and/or vigorous intensity physical activity determined using Treuth, Puyau, and new cut points (counts/min) by sample among 12-13 year old girls in NHANES

	Physical Activity Intensity					
Body composition/Sample/Cut point	Light	Moderate	Vigorous	Light to Vigorous	Moderate to Vigorous	Average Correlation
BMI percentile						
Development sample (N = 126)						
Treuth[1]	0.104	−0.132	−0.160	0.076	−0.155	−0.053
Puyau[2]	−0.029	−0.146	−0.140	−0.064	−0.166	−0.109
New (1900, 4300, 10000)	−0.145	−0.188	−0.191	−0.159	−0.191	−0.175
Evaluation sample (N = 44)						
Treuth	−0.145	−0.260	−0.358	−0.162	−0.299	−0.245
Puyau	−0.107	−0.249	−0.129	−0.132	−0.259	−0.175
New (1900, 4300, 10000)	−0.254	−0.315	−0.156	−0.274	−0.324	−0.264
WC percentile						
Development sample (N = 123)						
Treuth	0.180	−0.115	−0.107	0.161	−0.134	−0.003
Puyau	−0.007	−0.124	−0.066	−0.036	−0.144	−0.075
New (1900, 4000, 5000)	−0.122	−0.166	−0.127	−0.140	−0.172	−0.145
Evaluation sample (N = 44)						
Treuth	−0.136	−0.217	−0.265	−0.142	−0.244	−0.201
Puyau	−0.060	−0.214	−0.044	−0.078	−0.202	−0.120
New (1900, 4000, 5000)	−0.202	−0.199	−0.280	−0.221	−0.238	−0.228

NHANES: National Health and Nutrition Examination Survey; BMI: body mass index; WC: waist circumference

1 Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 101-2999, 3000-5200, 5201, 101, and 3000 counts/min, respectively.

2 Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800-3199, 3200-8199, 8200, 800, and 3200 counts/min, respectively.

points. In the NHANES development sample, the average correlation between physical activity and waist circumference percentile using the new cut point combination (−0.145) was 0.070 stronger than using the Puyau cut points and 0.142 stronger than using the Treuth cut points. However, in the evaluation sample, the average correlation with waist circumference percentile using the new cut point combination (−0.228) was 0.108 stronger than using the Puyau cut points but only 0.027 stronger than using the Treuth cut points.

The strongest correlations in NHANES tended to be observed for higher intensity physical activity. The strongest correlations with BMI percentile were observed for moderate and moderate-to-vigorous intensity physical activity in both the development (−0.188 to −0.191) and evaluation (−0.315 to −0.324) samples. The strongest correlations with waist circumference percentile were observed for moderate and moderate-to-vigorous physical activity in the development sample (−0.166 to −0.172) but for vigorous activity in the evaluation sample (−0.280). Although the average correlations of the new cut point combinations were slightly stronger that the Treuth cut points in the evaluation sample, three intensity-specific correlations were lower for the new cut point than the corresponding Treuth cut point: vigorous intensity for BMI percentile (−0.156 vs. −0.358), moderate intensity for waist circumference percentile (−0.199 vs. −0.217), and moderate-to-vigorous intensity for waist circumference percentile (−0.238 vs. −0.244).

Two of the potential cut point combinations identified in the development sample were not identified as new cut points based on their performance in the evaluation sample. In the NHANES evaluation sample, the average correlation among Hispanic girls for percent body fat was less than 0.100 in magnitude (data not shown). The average correlation among 12-13 year old girls for percent body fat was greater than 0.100 in magnitude in the NHANES evaluation sample but less than 0.050 better than the average correlations for the Treuth and Puyau cut points (data not shown).

New cut points in TAAG

Table 2.3 presents the correlations in the development and evaluation samples for the previously suggested and new cut point combinations identified in the TAAG dataset. Both of the potential cut point combinations identified in the development sample were identified as new cut points in the evaluation sample. The new cut point combinations were identified among Black girls for percent body fat (1450, 1950, and 2450 counts/30 sec) and cardiorespiratory fitness (1050, 1550, and 2050 counts/30 sec). The new light-to-vigorous cut points were 650 to 1400 counts/30 sec higher than the corresponding Treuth and Puyau cut points. The new moderate-to-vigorous cut point for percent body fat was 350 to 450 counts/30 sec higher than the previously suggested cut points, while the cut point for cardiorespiratory fitness was 50 counts/30 sec higher than the Treuth cut point and 50 counts/30 sec lower than the Puyau cut point. The new vigorous cut points were 150 to 2050 counts/30 sec lower than the corresponding Treuth and Puyau cut points.

Among Black girls in TAAG, the average correlations between physical activity and percent body fat were stronger using the new cut point combinations compared with the Treuth and Puyau cut points in both the development and evaluation samples. In the development sample, the average correlation between physical activity and percent body fat using the new cut point combination (−0.106) was 0.062 stronger than using the

Table 2.3. **Spearman's rank correlations between body composition and physical fitness variables and average daily minutes spent in light, moderate, and/or vigorous intensity physical activity determined using Treuth, Puyau, and new cut points (counts/30 sec) by sample among Black girls in TAAG**

	Physical Activity Intensity					
Variable/Sample/Cut point	**Light**	**Moderate**	**Vigorous**	**Light to Vigorous**	**Moderate to Vigorous**	**Average Correlation**
Percent body fat						
Development sample (N = 649)						
Treuth[1]	0.041	–0.090	–0.115	0.029	–0.105	–0.048
Puyau[2]	0.069	–0.114	–0.095	0.032	–0.115	–0.044
New (1450, 1950, 2450)	–0.056	–0.123	–0.123	–0.099	–0.130	–0.106
Evaluation sample (N = 217)						
Treuth	–0.066	–0.269	–0.238	–0.103	–0.276	–0.191
Puyau	–0.060	–0.276	–0.169	–0.111	–0.278	–0.179
New (1450, 1950, 2450)	–0.244	–0.299	–0.257	–0.274	–0.288	–0.272
Cardiorespiratory fitness[3]						
Development sample (N = 110)						
Treuth	0.006	0.315	0.227	0.040	0.291	0.176
Puyau	0.108	0.306	0.148	0.165	0.288	0.203
New (1050, 1550, 2050)	0.214	0.327	0.260	0.273	0.297	0.274
Evaluation sample (N = 46)						
Treuth	0.217	0.327	0.230	0.231	0.331	0.267
Puyau	0.155	0.298	0.258	0.201	0.314	0.245
New (1050, 1550, 2050)	0.245	0.365	0.264	0.281	0.332	0.297

TAAG: Trial of Activity for Adolescent Girls

1 Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 51-1499, 1500-2600, 2601, 51, and 1500 counts/30 sec, respectively.

2 Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800, 3200, and 8200 counts/min, respectively. The cut points used were 400-1599, 1600-4099, and 4100, 400, and 1600 counts/30 seconds, respectively.

3 Cardiorespiratory fitness was defined as the estimated power output (watts/kg body weight) at a heart rate of 170 beats per minute predicted from a multi-stage cycle ergometry test.

Puyau cut points and 0.058 stronger than using the Treuth cut points. In the evaluation sample, the average correlation using the new cut point combination (−0.272) was 0.093 stronger than using the Puyau cut points and 0.081 stronger than using the Treuth cut points. The strongest correlations with percent body fat were observed for moderate and moderate-to-vigorous intensity physical activity in both the TAAG development (−0.123 to −0.130) and evaluation samples (−0.288 to −0.299). The average correlations among Black girls in TAAG between physical activity and cardiorespiratory fitness were stronger using the new cut point combinations compared with the Puyau cut points in both the development and evaluation samples. However, the average correlations were only stronger than the Treuth cut points in the development sample. In the development sample, the average correlation between physical activity and cardiorespiratory fitness using the new cut point combination (0.274) was 0.071 stronger than using the Puyau cut points and 0.098 stronger than using the Treuth cut points. However, in the evaluation sample, the average correlation using the new cut point combination (0.297) was 0.052 stronger than using the Puyau cut points but only 0.030 stronger than using the Treuth cut points. The strongest correlations with cardiorespiratory fitness were observed for moderate intensity physical activity in both the development (0.327) and evaluation samples (0.365).

DISCUSSION

Results from this study indicate that several combinations of cut points exist that may more strongly associate minutes of physical activity with body composition and cardiorespiratory fitness compared with previously suggested cut points. The previously suggested cut points used in this study were developed using oxygen consumption as the criterion (Puyau et al., 2002; Treuth et al., 2004) and thereby reflect energy expenditure. The new cut points may be more clinically relevant as they used disease risk factors, including body composition and cardiorespiratory fitness, as criteria. Hence, these cut points may be more appropriate to use than oxygen consumption cut points in studies with adolescent girls where the relationship between physical activity and obesity or fitness is of interest. However, the average correlations using the new cut points were only about 0.1 stronger, which may not represent a meaningful improvement.

Compared with the previously suggested cut points of comparable intensity, the new light and moderate cut points tended to be higher. Hence, the minimum intensity threshold for receiving health benefits from light and moderate physical activity may be higher than the previously suggested cut points. In contrast, the new vigorous cut points tended to be lower than the previously suggested vigorous cut points. Hence, beneficial effects of vigorous physical activity may be observed at a lower intensity than the previously suggested cut points. Furthermore, the range of intensities for moderate activity may be narrower.

Even though four new combinations of cut points were identified that were more strongly correlated with body composition and cardiorespiratory fitness than previously suggested cut points, the correlations were weak, with average correlations less than 0.3 and intensity-specific correlations less than 0.4. Other factors, such as diet or genetics, may account for some of the variation. However, these factors were not examined in this study. The strongest intensity-specific correlations tended to be observed for moderate and moderate-to-vigorous cut points. This agrees with the current physical activity recommendations, which emphasize performing activities of at least moderate intensity

(Strong et al., 2005; U.S. Department of Health and Human Services et al., 2005; U.S. Department of Health and Human Services, 2000; Physical Activity Guidelines Advisory Committee, 2008).

It is unclear why new cut points were identified for BMI percentile in NHANES but not TAAG or why the opposite was the case for percent body fat. It may be due to the different racial/ethnic distributions. NHANES had a larger proportion of Black and Hispanic girls. Moreover, TAAG included a considerable number of Asian girls. Another reason may be age differences in the samples. TAAG girls were primarily 14 years old, while NHANES girls encompassed a broader range of ages. Hence, new cut points may have only been identified in NHANES among 12-13 year old girls because the relationship between physical activity and body composition may have been affected by level of physical maturity. That is, the relationship observed among the younger age group in NHANES differed from the older age group in NHANES and in TAAG. Furthermore, new cut points may have only been identified in TAAG among Black girls because there was an interaction effect between age and race that attenuated the relationship between physical activity and the examined disease risk factors in NHANES. That is, the relationship observed in the limited age range in TAAG may not have been observed in NHANES because of the larger proportion of younger girls. Some race and age effects for body composition have been observed in the National Heart, Lung, and Blood Institute (NHLBI) Growth and Health Study (NGHS), which followed 9 and 10 year old girls for 10 years. BMI and sums of triceps and subscapular skinfolds were higher for Black than White girls at all ages, and the differences increased with increasing age (Kimm et al., 2001; Morrison et al., 2001). However, while percent body fat measured by bioelectrical impedance increased for Black girls at all ages, it did not increase in White girls until 14 years. As such, from 9 to 11 years, White girls had higher percent body fat than Black girls; they were similar at 12 and 13 years; then from 14 years and older, Black girls had higher values than White girls (Morrison et al., 2001). These different patterns in body composition measures may have affected the ability to identify new cut points in this study.

In order to limit the cut points tested to a reasonable number, this method used previously developed cut points as a starting point. Hence, one limitation to this method was that it required a prior calibration study. One benefit of this was that it gave the new cut points a frame of reference. That is, because cut points developed using oxygen consumption as the criterion were used as a starting point, intensity levels could be assigned to the new cut points. Still, this methodology allowed for the new cut points to move away from the starting point of the previously suggested cut points. For example, the new light cut points were more than 1000 counts/min above the previously suggested light cut points.

One strength of this study was the data used were from two large field studies. Hence, data were obtained from a substantial number of adolescent girls performing free-living physical activities. However, because the data were not collected to calibrate accelerometer data, a large proportion of participants lacked sufficient accelerometer data and were excluded from analysis. Furthermore, because these data were collected from the general population of adolescent girls, the samples comprised low-risk populations (e.g., mean BMI percentile was normal weight). Restricting the samples to adolescent girls targeted a population at risk for low physical activity and controlled for age and gender. However, participants may not have been exposed to their level of physical activity long enough to observe an influence on disease risk factors. These methods may have worked better with populations with higher disease risk (e.g., older adults, overweight). Another weakness of this study was that the data were cross-sectional. Hence, a temporal relationship between physical activity and risk factors could not be established. Furthermore, the one week of accelerometer data may not be representative of regular physical activity patterns, thereby making it difficult to detect a relationship between regular physical activity and disease risk factors.

In summary, this study utilized body composition and cardiorespiratory fitness as the criteria to identify new accelerometer cut points for assessing physical activity. The results suggest that the previously suggested cut points developed using oxygen consumption may not best reflect the intensity of physical activity needed for similar health benefits. The new cut points identified indicate that health benefits from light and moderate physical activity may be conferred at a higher intensity than the previously suggested cut points and those from vigorous activity at a lower intensity. Furthermore, they support current recommendations that encourage engaging in moderate and moderate-to-vigorous physical activity. The actual intensity of physical activity needed for health benefits may differ by disease risk factor and population subgroup. Moreover, there may be interactions, such as between race and age, that affect the relationship between physical activity and body composition and cardiorespiratory fitness.

REFERENCES

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

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med.Sci.Sports Exerc., 32,* 1601-1609.

Centers for Disease Control and Prevention, National Center for Health Statistics (2008). 2003-2004 National Health and Nutrition Examination Survey: survey operations manuals, brochures, consent documents. http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/current_nhanes_03_04.htm [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2006). National Health and Nutrition Examination Survey 2003-2004 documentation, codebook, and frequencies: exam component – physical activity monitor examination data. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/paxraw_c.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2003). National Health and Nutrition Examination Survey overview. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/overviewbrochure_0304.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2004). National Health and Nutrition Examination Survey anthropometry procedures manual. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/BM.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2000). CDC growth charts: United States. http://www.cdc.gov/growthcharts/ [On-line].

Eisenmann, J. C., Welk, G. J., Wickel, E. E., & Blair, S. N. (2007). Combined influence of cardiorespiratory fitness and body mass index on cardiovascular disease risk factors among 8-18 year old youth: the Aerobics Center Longitudinal Study. *Int.J.Pediatr.Obes., 2,* 66-72.

Eisenmann, J. C., Wickel, E. E., Welk, G. J., & Blair, S. N. (2005). Relationship between adolescent fitness and fatness and cardiovascular disease risk factors in adulthood: the Aerobics Center Longitudinal Study (ACLS). *Am.Heart J., 149,* 46-53.

Fernandez, J. R., Redden, D. T., Pietrobelli, A., & Allison, D. B. (2004). Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J.Pediatr., 145,* 439-444.

Kimm, S. Y., Glynn, N. W., Kriska, A. M., Barton, B. A., Kronsberg, S. S., Daniels, S. R. et al. (2002). Decline in physical activity in black girls and white girls during adolescence. *N.Engl.J.Med., 347,* 709-715.

Kimm, S. Y. S., Barton, B. A., Obarzanek, E., McMahon, R. P., Sabry, Z. I., Waclawiw, M. A. et al. (2001). Racial divergence in adiposity during adolescence: the NHLBI Growth and Health Study. *Pediatrics, 107,* e34.

Korn, E. L. & Graubard, B. I. (1991). Epidemiologic studies using surveys: accounting for the sampling design. *Am.J.Public Health, 81,* 1166-1173.

Lohman, T. G., Ring, K., Schmitz, K. H., Treuth, M. S., Loftin, M., Yang, S. et al. (2006). Associations of body size and composition with physical activity in adolescent girls. *Med.Sci.Sports Exerc., 38,* 1175-1181.

Malina, R. M. (2007). Physical fitness of children and adolescents in the United States: status and secular change. *Med.Sports Sci., 50,* 67-90.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. I., & Hu, J. (2003). Tracking of physical activity and aerobic power from childhood through adolescence. *Med.Sci.Sports Exerc., 35,* 1914-1922.

Morrison, J. A., Barton, B. A., Obarzanek, E., Crawford, P. B., Guo, S. S., Schreiber, G. B. et al. (2001). Racial differences in the sums of skinfolds and percentage of body fat estimated from impedance in black and white girls, 9 to 19 years of age: the National Heart, Lung, and Blood Institute Growth and Health Study. *Obes.Res., 9,* 297-305.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA, 300,* 295-305.

Ogden, C. L., Carroll, M. D., & Flegal, K. M. (2008). High body mass index for age among US children and adolescents, 2003-2006. *JAMA, 299,* 2401-2405.

Ogden, C. L., Flegal, K. M., Carroll, M. D., & Johnson, C. L. (2002). Prevalence and trends in overweight among US children and adolescents, 1999-2000. *JAMA, 288,* 1728-1732.

Ondrak, K. S., McMurray, R. G., Bangdiwala, S. I., & Harrell, J. S. (2007). Influence of aerobic power and percent body fat on cardiovascular disease risk in youth. *J.Adolesc.Health, 41,* 146-152.

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

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obes.Res., 10,* 150-157.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B. et al. (2005). Evidence based physical activity for school-age youth. *J.Pediatr., 146,* 732-737.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J. et al. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med.Sci.Sports Exerc., 36,* 1259-1266.

U.S. Department of Health and Human Services (2000). *Healthy People 2010: understanding and improving health and objectives for improving health* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services & U.S. Department of Agriculture (2005). *Dietary guidelines for Americans* http://www.health.gov/dietaryguidelines/dga2005/document/ [On-line].

Ward, D. S., Brown Rodgers, A., & Vaughn, A. (2005). Objective monitoring of physical activity: closing the gaps in the science of accelerometry. *Med.Sci.Sports Exerc.*, *37*.

Chapter 3: Development of Accelerometer Cut Points for Adolescent Girls Using Signal Detection and Receiver Operating Characteristic Curves

ABSTRACT

Accelerometers objectively measure physical activity by recording counts per epoch. It is still unclear how to convert accelerometer counts into minutes of physical activity. PURPOSE: To use signal detection and receiver operating characteristic (ROC) curves to identify accelerometer cut points that are better at minimizing misclassification of cardiovascular and metabolic disease risk factors (signals) than previously suggested cut points by whether physical activity recommendations were met (detector). METHODS: Accelerometer data from adolescent girls in the National Health and Nutrition Examination Survey (NHANES) and Trial of Activity for Adolescent Girls (TAAG) were examined overall and within population subgroups. Potential cut points that minimized misclassification within a population subgroup for which the area under the ROC curve was greater than 0.5 were identified in a development sample. New cut points were identified by validating the potential cut points in an evaluation sample. RESULTS: The 11 new cut points identified in NHANES (ranging from 100 to 2300 counts/min) and the 3 new cut points identified in TAAG (ranging from 50 to 100 counts/30 sec) were lower than previously suggested moderate cut points developed using oxygen consumption. The mean misclassification was approximately 32.5% in NHANES and 19.5% in TAAG. The mean area under the ROC curve was approximately 0.57 in NHANES and 0.52 in TAAG. CONCLUSIONS: Previously suggested cut points developed using oxygen consumption may not best classify markers of disease risk. Cut

points may differ by disease risk factor and population subgroup. Accelerometers may not be good at classifying disease risk.

INTRODUCTION

Physical activity has been associated with several health benefits including higher physical fitness, healthier body composition, and more favorable cardiovascular and metabolic disease risk profiles (U.S. Department of Health and Human Services, 1996; Physical Activity Guidelines Advisory Committee, 2008). However, the exact pattern of physical activity needed to receive health benefits is not well elucidated. Moreover, it may differ by population subgroups and by the outcome of interest (Physical Activity Guidelines Advisory Committee, 2008). The inability to define this pattern of physical ability is due in part to the lack of ability to easily and accurately measure this information. Most population-based research has relied on self-report measures for collecting physical activity data because they tend to be lower cost and can be used to collect considerable detailed information. However, the information collected may be subject to biases depending on the participants' ability to accurately recall or report their activity or their tendency to give more socially desirable answers.

Recent large scale studies (e.g., National Health and Nutrition Examination Survey (NHANES) and Trial of Activity for Adolescent Girls (TAAG)) have employed accelerometers to assess physical activity, which addresses the aforementioned weaknesses of self-report measures. Accelerometers can objectively measure and record the frequency, intensity, and duration of physical activity, reflected by accelerometer counts per epoch (e.g., minute). However, it is still unclear how to interpret accelerometer data (e.g., how to convert counts into minutes of physical activity).

Two previous studies have calibrated the Actigraph (formerly known as Computer Science and Applications (CSA) and Manufacturing Technology, Inc. (MTI)) model 7164 against oxygen consumption among adolescents. The sample in the study by Puyau et al. (2002) included 12 girls and 14 boys, 6 to 16 years old. Oxygen consumption was measured in a room calorimeter. Participants in the study by Treuth et al. (2004) were 74 eighth grade girls. Oxygen consumption was measured using a portable, breath-by-breath metabolic unit (Cosmed K4b2, Rome, Italy). In both studies, participants performed structured activities including resting, sedentary activities, light activities, moderate activities, and vigorous activities. Puyau et al. (2002) used linear regression of energy expenditure on activity counts to identify the cut points for different intensity levels. Treuth et al. (2004) used a linear mixed model and examined false positive and negative classifications to identify cut points.

Signal detection and receiver operating characteristic (ROC) curves have been used to examine the association between the presence and absence of a signal and a detector's ability to detect the difference (Kraemer, 1988). In medical decision making, they have been used to evaluate the performance of diagnostic tests. Evenson et al. (2008) recently used ROC curves to identify accelerometer cut points for sedentary, light, moderate, and vigorous activity in children 5 to 8 years old. Participants performed controlled bouts of activities ranging from resting to running 4 mph while wearing two accelerometers (ActiGraph and Actical). Cut points that gave equal weight to (i.e., maximized) sensitivity and specificity for sedentary, moderate, and vigorous activity were identified. The area under the ROC curve was also used to evaluate how accurately each accelerometer identified the activity intensity.

In the current study, signal detection and ROC curves were used with the accelerometer as the detector. Several cardiovascular and metabolic disease risk factors were used as signals rather than controlled bouts of activity. The structured activities performed in the aforementioned calibration studies may not be performed the same way they would be in real-world situations. Furthermore, while these cut points reflect energy expenditure associated with physical activity because they were developed using oxygen consumption, they may not reflect cut points that best classify the relationship between physical activity and disease risk.

The purpose of this study was to use signal detection and ROC curves to identify cut points that minimized misclassification of selected cardiovascular and metabolic disease risk factors (signals) better than previously suggested cut points by whether physical activity recommendations were met, as measured by accelerometers (detector) in two large field studies. This study focused on adolescent girls because they are at particular risk of low physical activity levels. Physical activity levels tend to decline during adolescence (Kimm et al., 2002; McMurray et al., 2003; Nader et al., 2008), and adolescent girls tend to be less active than adolescent boys (Caspersen et al., 2000; McMurray et al., 2003; Nader et al., 2008).

METHODS

This study used data from two studies that collected accelerometer data from large numbers of adolescent girls from different locations across the United States: the 2003-2004 NHANES and TAAG. Potential new cut points were developed using a random sample of 75% of participants from each study. A validation study of the potential new cut points were conducted with the remaining 25%. Unweighted analyses were used.

However, analyses were done for each risk factor overall and within 3 race/ethnicity groups in NHANES (i.e., White, Hispanic, and Black), 4 race/ethnicity groups in TAAG (i.e., White, Hispanic, Black, and Asian), and 2 age groups in NHANES (i.e., 12-13 and 14-15 years), to control for potential confounding factors and variables used in the sampling designs (Korn et al., 1991). Additionally, total cholesterol and high density lipoprotein (HDL) cholesterol were examined within weight status groups in NHANES. Overweight was only examined for all girls (i.e., not by population subgroups) in NHANES due to limited sample size.

Study designs

NHANES used a complex stratified, multi-stage probability sampling design to obtain a representative sample population and over-sampled Blacks, Mexican-Americans, adolescents, older people, and pregnant women. Each year, NHANES randomly selected about 5,000 people of all ages in households from fifteen different locations. Household interviewers identified and enrolled survey participants, conducted household interviews, and appointed study participants for the mobile examination center (MEC) exam (Centers for Disease Control and Prevention, 2003). A proxy answered race/ethnicity questions for persons less than 16 years old during the interviewer-administered Sample Person Questionnaire. Race/ethnicity variables were derived by combining responses to questions on race and Hispanic origin. Participants could mark all that applied. Race categories included non-Hispanic White, non-Hispanic Black, Mexican American, other Hispanic, and other race – including multi-racial (Centers for Disease Control and Prevention, 2009). Body measurements and blood sample draws were conducted in the MEC by health professionals (Centers for Disease Control and Prevention, 2004a;

Centers for Disease Control and Prevention, 2004b). Additionally, participants received an accelerometer at the MEC visit to assess physical activity.

TAAG was a group-randomized, multi-center trial of a physical activity intervention. Data were collected at six schools recruited at each of the six field centers (University of Maryland, University of South Carolina, University of Minnesota, Tulane University, University of Arizona, and San Diego State University). Baseline measures were conducted with sixth grade girls attending the participating schools in Spring, 2003. Follow-up measures were conducted with eighth grade girls in the same schools in Spring, 2005. Additionally, eighth grade girls in 34 of the same schools were measured in Spring, 2006; two schools in New Orleans were closed due to Hurricane Katrina. This study used data from the two eighth grade cohorts. All eighth grade girls in the TAAG schools were eligible to participate. They were excluded if they could not read and understand questions written in English. One-hundred twenty girls from each school were randomly selected from each eighth grade cohort to wear accelerometers. Data were collected over three visits at school. At one visit, participants had their height, weight, and triceps skinfold measured and received their accelerometer. One week later, participants returned their accelerometer. Participants completed a student questionnaire, which included demographic questions, either before or after wearing the accelerometer. Participants indicated their race/ethnicity on a checklist including White, Black or African American, Hispanic, Asian, Pacific Islander, American Indian or Alaska Native, or Other. They could mark all that applied.

<u>Measurement</u>

Accelerometer. The objective physical activity measure in NHANES and TAAG was the ActiGraph accelerometer, model 7164. It is a uniaxial accelerometer that detects and records acceleration of movement, especially locomotor-type (e.g., walking, jogging) activities, at user-specified intervals. Girls wore the accelerometer on an elasticized belt on their right hip. They were told to keep the monitor dry (i.e., remove it before swimming or bathing) and to remove it at bedtime. NHANES participants wore the accelerometer for 7 full days. It started recording data in 1-minute epochs at 12:01 a.m. the day after the girl was given the monitor at the examination. Participants returned the accelerometer by mail in postage-paid, padded envelopes (Centers for Disease Control and Prevention, 2006). TAAG participants wore the accelerometer for 6 or 7 complete days. It started recording data in 30-second epochs at 5:00 a.m. the day after the accelerometer was given to the girl. Participants returned their accelerometer one week later. If participants did not wear the monitor for at least 10 hours on 1 day or the monitor malfunctioned, they were asked to wear the accelerometer again for one more week.

In the current study, accelerometer non-wear time was defined as 60 minutes or more of consecutive zero counts. At least 10 hours of wear was required for a day to be valid and at least two valid weekdays and one valid weekend day were required for a participant to be included in the analysis. Counts of 24000/min (12000/30 sec) or greater were considered extreme values. They were included in the wear time as non-zero values, but they were not counted as physical activity time.

Physical activity recommendations. Three physical activity recommendations for adolescents were examined. Strong et al. (2005) recommend that school-age youth participate in 60 minutes or more of moderate to vigorous physical activity daily. The 2005 Dietary Guidelines for Americans (U.S. Department of Health and Human Services et al., 2005) recommends that children and adolescents engage in at least 60 minutes of moderate intensity physical activity on most, preferably all, days of the week. Healthy People 2010 Objective 22.6 (U.S. Department of Health and Human Services, 2000) recommends that adolescents engage in moderate physical activity for at least 30 minutes on 5 or more days per week. These three were examined because each has a different combination of frequency and duration recommended, which may affect the cut points identified to detect elevated disease risk.

Whether recommendations were met were examined for accelerometer cut points at 100 counts/min increments up to 10500 counts/min in NHANES because 100 counts/min corresponds to approximately 0.1 MET according to the Treuth (2004) equation, MET = 2.01 + 0.000856 (counts/min). In TAAG, 50 counts/30 sec increments up to 5250 counts/30 sec were used because NHANES used 1-minute epochs and TAAG used 30-second epochs. Epochs with accelerometer counts at or above the cut point were counted to determine the number of minutes of physical activity per day. Because it is not clear whether it is necessary to meet the frequency and the duration recommended or it is adequate to accumulate the recommended time, meeting each recommendation was examined both ways. A participant was defined as meeting a frequency and duration recommendation if both the frequency and the duration recommendations were met or exceeded (e.g., at least 30 minutes on at least 5 days to meet the Healthy People recommendation). A participant was defined as meeting an accumulated time recommendation if the product of the frequency and duration recommendations was met or exceeded (e.g., at least 150 minutes a week to meet Healthy People recommendation).

Cardiovascular and metabolic disease risk factors. For the signals, continuous values for disease risk factors were categorized based on current recommendation or proposed cut points. The signals detected included at risk for overweight or overweight, overweight, high body fat, moderate or high central adiposity, borderline or high total cholesterol, and borderline or low HDL cholesterol.

Height and weight were measured to the nearest 0.1 cm and 0.1 kg, respectively. They were measured once in NHANES (Centers for Disease Control and Prevention, 2004a). The average of two measurements was used in TAAG. Body mass index (BMI) was calculated as weight (kg) divided by height (m) squared. Age-specific BMI percentiles were determined using the 2000 CDC growth chart tables for girls (Centers for Disease Control and Prevention, 2000). Weight status was classified based on the CDC weight status categories definitions (Barlow et al., 2007). Participants with a BMI less than the 85^{th} percentile were classified as under or normal weight, greater than or equal to the 85^{th} percentile but less than the 95^{th} percentile as at risk for overweight, and greater than or equal to the 95^{th} percentile as overweight.

Triceps skinfold was measured at the posterior midline of the right upper arm. NHANES used Holtain skinfold calipers and measured triceps skinfold once to the nearest 0.1 mm (Centers for Disease Control and Prevention, 2004a). TAAG used Lange skinfold calipers and used the average of three measurements recorded to the nearest 1 mm. Percent fat was estimated with an equation including race/ethnicity contrast, age, height, weight, and triceps skinfold [percent body fat mass = –23.393 + 2.269(BMI [kg/m^2]) + 1.943(triceps skinfold [mm]) – 2.995(race/ethnicity) – 0.524(age [yr]) – 0.058 (BMI [kg/m^2])(triceps skinfold [mm])] (Lohman et al., 2006). Race/ethnicity was 1 if non-Hispanic black and 0 if otherwise. Estimated percent body fat from skinfold measurements were classified using standards for Healthy Fitness Zone from FITNESSGRAM (Lohman et al., 2001). Participants with an estimated percent body fat less than 32% were classified as low or moderate fat. Those with greater than 32% were classified as high fat.

Waist circumference was measured in NHANES to the nearest 1 mm just above the uppermost lateral border of the right ilium (Centers for Disease Control and Prevention, 2004a). Age-, sex-, and ethnicity-specific waist circumference percentiles were determined using values from Fernandez et al. (2004). A waist circumference cutoff value of the 75th percentile has been suggested as a component of metabolic syndrome (de Ferranti et al., 2004). Participants with a waist circumference less than the 75th percentile were classified as having low central adiposity, and those at or above the 75th percentile were classified as having moderate or high central adiposity.

Total cholesterol and high-density lipoprotein (HDL) cholesterol were determined in NHANES participants who had their blood drawn after fasting at least 8 hours. Serum samples were frozen and shipped to Johns Hopkins University for lipids analysis using standard procedures (Centers for Disease Control and Prevention, 2004b). Participants with total cholesterol less than 170 mg/dl were classified as acceptable and 170 mg/dl or greater were classified as borderline or high, based on the American Academy of Pediatrics Committee on Nutrition (1998). In a study of adolescent metabolic syndrome, low HDL cholesterol was defined as less than 50 mg/dl (de Ferranti et al., 2004), based on the percentiles corresponding to the cut points for adults (i.e., 40 mg/dl). Participants with HDL cholesterol greater than 50 mg/dl were classified as acceptable and 50 mg/dl or less were classified as borderline or low.

Signal detection and Receiver Operating Characteristic (ROC) curves

Whether physical activity recommendations were met at incremental cut points was used to detect elevated disease risk (signal). A positive test for detecting elevated disease risk (e.g., high body fat) was defined as not meeting a given physical activity recommendation. Hence, a true positive was defined as not meeting a recommendation and having elevated risk; a false positive was defined as not meeting a recommendation and having lower risk; and a false negative was defined as meeting a recommendation and having elevated risk.

The proportions of true and false positives and misclassification (i.e., false positives and false negatives) at each incremental cut point were determined. The ROC curve is a plot of the true positive rate against the false positive rate at each of the incremental cut points. The area under the ROC curve was approximated using the trapezoidal rule. The larger the area under the curve, the more accurate the detector is. An area of 0.5 or less represents a “useless” detector because for every increase in true positives, at least as much of an increase is observed in false positives.

The highest cut points that minimized misclassification in the development sample for each of the population-risk factor combinations were identified. The percent misclassification for each of these potential cut points was compared with cut points previously developed in adolescents using oxygen consumption as the criterion (Puyau et al., 2002; Treuth et al., 2004). If misclassification using potential cut points was more than 5.0% lower than that using either of the previously suggested light-to-vigorous (light) cut points, the area under the ROC curve was examined. If the area under the ROC curve for that population-risk factor combination was greater than 0.50, then the potential cut point was tested in the evaluation sample. Potential cut points that met both criteria in the evaluation sample were identified as new cut points. The percent misclassification of potential cut points was compared with previously suggested light, moderate, and vigorous cut points. However, only comparisons with the light cut points are presented because the light cut points were better at minimizing misclassification than the higher intensity cut points (i.e., previously suggested moderate and vigorous cut points always had higher misclassification than the new cut points).

RESULTS

Table 3.1 presents participant characteristics for the NHANES and TAAG datasets overall and within the development and evaluation samples. There were no significant differences between the development and evaluation samples ($p > 0.10$). NHANES received accelerometer data from 451 girls aged 12 to 15 years, and 333 of those girls had sufficient data for analysis. Girls that were not included tended to have higher BMI percentile, waist circumference percentile, and percent body fat, though the differences were not significant ($p = 0.05$ to 0.08, data not shown). Of 333 girls included, 51% were 12-13 years old and 49% were 14-15 years old. The majority of NHANES girls were Hispanic (38%) followed by Black (33%) and White (26%). One-third of the girls in NHANES were at risk for overweight or overweight, and 15% were overweight. Similar prevalence rates were observed for the other risk factors examined: moderate or

Table 3.1. NHANES and TAAG participant characteristics by sample

	n (%)			
Dataset/Characteristic	Total	Development Sample	Evaluation Sample	p-value[1]
NHANES, all girls	**333 (100.0)**	**250 (100.0)**	**83 (100.0)**	
Age (year)				
12	77 (23.1)	62 (24.8)	15 (18.1)	0.30
13	94 (28.2)	65 (26.0)	29 (34.9)	
14	92 (27.6)	72 (28.8)	20 (24.1)	
15	70 (21.0)	51 (20.4)	19 (22.9)	
Race/Ethnicity				
Hispanic	126 (37.8)	97 (38.8)	29 (34.9)	0.73
Black, non-Hispanic	110 (33.0)	79 (31.6)	31 (37.3)	
White, non-Hispanic	85 (25.5)	64 (25.6)	21 (25.3)	
Other and multiple race	12 (3.6)	10 (4.0)	2 (2.4)	
Weight Status				
UW or NW (<85th BMI %ile[2])	218 (66.7)	165 (67.3)	53 (64.6)	0.65
AR or OW (≥85th BMI %ile)	109 (33.3)	80 (32.7)	29 (35.4)	
UW, NW, or AR (<95th BMI %ile)	278 (85.0)	206 (84.1)	72 (87.8)	0.41
OW (≥95th BMI %ile)	49 (15.0)	39 (15.9)	10 (12.2)	
Central Adiposity				
Low (<75th WC %ile[3])	207 (65.1)	160 (67.2)	47 (58.8)	0.17
Moderate or High (≥75th WC %ile)	111 (34.9)	78 (32.8)	33 (41.3)	
Percent Body Fat				
Low or Moderate (≤32%)	193 (62.7)	143 (63.0)	50 (61.7)	0.84
High (>32%)	115 (37.3)	84 (37.0)	31 (38.3)	
Total Cholesterol				
Acceptable (<170 mg/dl)	209 (67.9)	156 (67.0)	53 (70.7)	0.55
Borderline or High (≥170 mg/dl)	99 (32.1)	77 (33.0)	22 (29.3)	
HDL Cholesterol				
Acceptable (>50 mg/dl)	191 (62.0)	140 (60.1)	51 (68.0)	0.22
Borderline or Low (≤50 mg/dl)	117 (38.0)	93 (39.9)	24 (32.0)	

Dataset/Characteristic	n (%) Total	Development Sample	Evaluation Sample	p-value[1]
TAAG, all girls	**4696 (100.0)**	**3522 (100.0)**	**1174 (100.0)**	
Age (year)				
12	9 (0.2)	5 (0.1)	4 (0.3)	0.21
13	2743 (58.5)	2043 (58.1)	700 (59.8)	
14	1785 (38.1)	1352 (38.5)	433 (37.0)	
15	131 (2.8)	104 (3.0)	27 (2.3)	
16	19 (0.4)	12 (0.3)	7 (0.6)	
Race/Ethnicity				
White, non-Hispanic	2190 (46.7)	1649 (46.9)	541 (46.2)	0.65
Hispanic	1017 (21.7)	770 (21.9)	247 (21.1)	
Black, non-Hispanic	866 (18.5)	649 (18.5)	217 (18.5)	
Asian	286 (6.1)	204 (5.8)	82 (7.0)	
Other and multiple race	327 (7.0)	244 (6.9)	83 (7.1)	
Weight Status				
UW or NW (<85th BMI %ile)	3044 (65.0)	2279 (64.8)	765 (65.3)	0.75
AR or OW (≥85th BMI %ile)	1643 (35.1)	1237 (35.2)	406 (34.7)	
UW, NW, or AR (<95th BMI %ile)	3867 (82.5)	2892(82.3)	975 (83.3)	0.43
OW (≥95th BMI %ile)	820 (17.5)	624 (17.7)	196 (16.7)	
Percent Body Fat				
Low or Moderate (≤32%)	2512 (53.6)	1880 (53.5)	632 (54.0)	0.77
High (>32%)	2175 (46.4)	1636 (46.5)	539 (46.0)	

NHANES: National Health and Nutrition Examination Survey; TAAG: Trial of Activity for Adolescent Girls; BMI: body mass index; UW: underweight; NW: normal weight; AR: at risk for overweight; OW: overweight; WC: waist circumference; HDL: high density lipoprotein

[1] P-values are from chi-squared tests, except if the expected value for any one cell is less than 5. Then p-values are from Fisher's exact tests.

[2] BMI percentiles are age- and sex-specific (CDC, 2000).

[3] WC percentiles are age-, sex-, and ethnicity-specific (Fernandez et al., 2004).

high central adiposity (35%); high body fat (36%); borderline or high total cholesterol (32%); and borderline or low HDL-cholesterol (38%). TAAG received accelerometer data from 7397 eighth grade girls, and 4696 (63%) of those girls had sufficient data for analysis. Girls that were not included tended to be older, have higher percent body fat, and be Black ($p < 0.0001$, data not shown). Most of the 4696 girls in TAAG (97%) were 13-14 years old. The majority of girls were White (47%) followed by Hispanic (22%), Black (18%), and Asian (6%). Among TAAG girls, 35% were at risk for overweight or overweight, 17% were overweight, and 46% had high body fat.

Potential cut points in development samples

The number of potential cut points identified in the development samples by risk factor and population group is shown in Table 3.2. In the NHANES development sample, 86 cut points with misclassification rates more than 5% lower than at least one of the previously suggested light cut points were identified. Of these, 68 were for population group-risk factor combinations with areas under the ROC curve that were not "useless," resulting in 7 potential cut points for at risk for overweight, 2 for overweight, 13 for moderate or high central adiposity, 10 for high body fatness, 9 for borderline or high total cholesterol, and 27 for borderline or low HDL cholesterol. No potential cut points were identified for at risk for overweight among 12-13 year old or White girls, for moderate or high central adiposity or high body fatness among 14-15 year old or Black girls, or for borderline or high total cholesterol among Hispanic girls.

In the TAAG development sample, 32 cut points that had misclassification rates more than 5% lower than at least one of the previously suggested light cut points were identified. Of these, 11 were for population group-risk factor combinations that had areas

Table 3.2. Number of potential cut points identified in the development sample by population subgroup, disease risk factor, and dataset

Dataset/Risk Factor	Total	All girls	12-13 y	14-15 y	White	Hispanic	Black	Asian	UW or NW	AR or OW
NHANES	**68**	**11**	**10**	**5**	**15**	**10**	**6**	**--**	**4**	**7**
AR or OW	7	2	0	3	0	1	1	--	--	--
OW	2	2	--	--	--	--	--	--	--	--
Moderate or High CA	13	3	1	0	6	3	0	--	--	--
High BF	10	1	3	0	5	1	0	--	--	--
Borderline or High TC	9	1	1	1	1	0	2	--	2	1
Borderline or Low HDL-C	27	2	5	1	3	5	3	--	2	6
TAAG	**11**	**2**	**--**	**--**	**6**	**2**	**1**	**0**	**--**	**--**
AR or OW	2	0	--	--	2	0	0	0	--	--
OW	6	2	--	--	3	0	1	0	--	--
High BF	3	0	--	--	1	2	0	0	--	--

NHANES: National Health and Nutrition Examination Survey; TAAG: Trial of Activity for Adolescent Girls; UW: underweight; NW: normal weight; AR: at risk for overweight; OW: overweight; CA: central adiposity; BF: body fatness; TC: total cholesterol; HDL-C: high density lipoprotein cholesterol
--: cut points not developed for population subgroup-risk factor combination

under the ROC curve that were not “useless,” resulting in 2 potential cut points for at risk for overweight, 6 for overweight, and 3 for high body fatness. No potential cut points were identified for at risk for overweight among all, Hispanic, or Asian girls; for overweight among Hispanic or Asian girls; or for high body fatness among all, Black, or Asian girls.

New cut points in NHANES

Table 3.3 presents the new cut points, which had misclassification rates more than 5% lower than at least one of the previously suggested light cut points and areas under the ROC curve that were not “useless” in both the NHANES development and evaluation samples. In the NHANES evaluation sample, 25 of the 86 potential cut points had lower misclassification rates than at least one of the previously suggested cut points. Of these, 11 were for population group-risk factor combinations with areas under the ROC curve that were not “useless.” The 11 new cut points in NHANES included 100, 400, 700, 1200, 1800, 2300 counts/min.

Two new cut points (1800 and 2300 counts/min for moderate or high central adiposity among White girls) had lower misclassification rates than both of the previously suggested cut points in both samples. In the development sample, the misclassification rate was 35% for 1800 counts/min and 37% for 2300 counts/min, while the rates for the Treuth and Puyau cut points were 49%. A similar pattern (38% vs. 52%) was observed in the evaluation sample. Two new cut points (1200 counts/min for moderate or high central adiposity among White girls and 700 counts/min for borderline or low HDL cholesterol among White girls) had misclassification rates equal to the Puyau cut point in the evaluation sample. The misclassification rate was 38% for 1200

Table 3.3. Physical activity recommendation used to identify new cut points and misclassification rate for new cut points and Treuth and Puyau light cut points (counts/min) and area under the ROC curve by sample for disease risk factors among population subgroups with new cut points in NHANES

Risk Factor/ Subgroup	PA Recom[1]	New cut point	Development Sample: Misclassified [n (%)] New	Treuth[2]	Puyau[2]	ROC Curve Area[3]	Evaluation Sample: Misclassified [n (%)] New	Treuth	Puyau	ROC Curve Area
AR or OW										
All girls	ST–FD	**100**	79 (32.2)	79 (32.2)	126 (51.4)	0.53	29 (35.4)	29 (35.4)	41 (50.0)	0.55
14–15 y	ST–FD	**100**	38 (31.9)	38 (31.9)	60 (50.4)	0.58	14 (36.8)	14 (36.8)	16 (42.1)	0.60
OW										
All girls	ST–AT	**400**	38 (15.5)	39 (15.9)	64 (26.1)	0.55	10 (12.2)	10 (12.2)	16 (19.5)	0.51
Moderate or high CA										
White	ST–AT	**1200**	22 (34.9)	31 (49.2)	30 (47.6)	0.61	8 (38.1)	11 (52.4)	8 (38.1)	0.56
White	DG–AT	**1800**	22 (34.9)	31 (49.2)	31 (49.2)	0.61	8 (38.1)	11 (52.4)	11 (52.4)	0.60
White	HP–AT	**2300**	23 (36.5)	31 (49.2)	31 (49.2)	0.63	8 (38.1)	11 (52.4)	11 (52.4)	0.60
High BF										
All girls	ST–FD	**100**	86 (37.9)	86 (37.9)	114 (50.2)	0.52	31 (38.3)	31 (38.3)	39 (48.2)	0.55
Borderline or high TC										
12–13 y	ST–FD	**400**	32 (27.4)	33 (28.2)	56 (47.9)	0.56	9 (22.0)	9 (22.0)	24 (58.5)	0.51
White	ST–FD	**400**	21 (35.0)	25 (41.7)	30 (50.0)	0.57	6 (33.3)	6 (33.3)	10 (55.6)	0.56

Risk Factor/ Subgroup	PA Recom[1]	New cut point	Development Sample: Misclassified [n (%)] New	Treuth[2]	Puyau[2]	ROC Curve Area[3]	Evaluation Sample: Misclassified [n (%)] New	Treuth	Puyau	ROC Curve Area
Borderline or low HDL-C										
14–15 y	ST–FD	**100**	47 (40.5)	47 (40.5)	53 (45.7)	0.57	9 (26.5)	9 (26.5)	17 (50.0)	0.64
White	DG–FD	**700**	18 (30.0)	21 (35.0)	22 (36.7)	0.60	7 (38.9)	9 (50.0)	7 (38.9)	0.51

NHANES: National Health and Nutrition Examination Survey; AR: at risk for overweight; OW: overweight; CA: central adiposity; BF: body fatness; TC: total cholesterol; HDL-C: high-density lipoprotein cholesterol

1 Physical activity recommendations (PA recom) included Strong et al. (ST) (2005); 2005 Dietary Guidelines (DG) (USDHHS et al., 2005); and Healthy People 2010 (HP) (USDHHS, 2000). Each recommendation was evaluated in terms of frequency and duration (FD) and accumulated time (AT).

2 Treuth et al. (2004) and Puyau et al. (2002) cut points for light intensity physical activity were 101 and 800 counts/min, respectively.

3 Receiver operating characteristic (ROC) curve is a plot of the true positive and false positive rate for each cut point tested. Area under the curve was approximated as the integral using the trapezoidal rule.

counts/min compared with 52% for the Treuth cut point and 38% for the Puyau cut point. A similar pattern was observed for 700 counts/min (39% vs. 50% and 39%). The other seven new cut points had lower misclassification rates than the Puyau but not the Treuth cut points in both samples.

The mean misclassification rate was 32.4% in the NHANES development sample and 32.5% in the evaluation sample using the new cut points; 37.4% in both samples using the Treuth cut point, and 45.9% in the development sample and 46.0% in the evaluation sample using the Puyau cut point (data not shown). The mean area under the curve was 0.58 in the development sample and 0.56 in the evaluation sample (data not shown). Most of the new cut points were developed using the Strong frequency and duration physical activity recommendation (6 cut points), followed by 2 using the Strong accumulated time recommendation, and 1 each using the Dietary Guidelines and Healthy People accumulated time recommendations.

New cut points in TAAG

Table 3.4 presents the misclassification rates and areas under the ROC curves in the TAAG development and evaluation samples for the new cut points. In the TAAG evaluation sample, 8 of the 11 potential cut points had lower misclassification rates than at least one of the previously suggested light cut points. Of these, 3 were for population group-risk factor combinations with areas under the ROC curve that were not "useless." The 3 new cut points in TAAG included 50 and 100 counts/30 sec. All three new cut points had lower misclassification rates than the Puyau cut point but equivalent to the Treuth cut point in both samples.

Table 3.4 Physical activity recommendation used to identify new cut points and misclassification rate for new cut points and Treuth and Puyau light cut points (counts/30 sec) and area under the ROC curve by sample for overweight among population subgroups with new cut points in TAAG

			Development Sample				Evaluation Sample			
			Misclassified [n (%)]				Misclassified [n (%)]			
Subgroup	PA Recom[1]	New cut point	New	Treuth[2]	Puyau[2]	ROC Curve Area[3]	New	Treuth	Puyau	ROC Curve Area
All girls	ST–AT	**50**	647 (18.4)	647 (18.4)	872 (24.8)	0.52	205 (17.5)	205 (17.5)	302 (25.8)	0.51
White	ST–AT	**50**	204 (12.4)	204 (12.4)	317 (19.2)	0.54	69 (12.8)	69 (12.8)	114 (21.1)	0.52
Black	DG–FD	**100**	177 (27.3)	177 (27.3)	220 (33.9)	0.51	61 (28.1)	61 (28.1)	86 (39.6)	0.53

TAAG: Trial of Activity for Adolescent Girls;

[1] Physical activity recommendations (PA recom) included Strong et al. (ST) (2005); 2005 Dietary Guidelines (DG) (USDHHS et al., 2005); and Healthy People 2010 (HP) (USDHHS, 2000). Each recommendation was evaluated in terms of frequency and duration (FD) and accumulated time (AT).

[2] Cut points for light intensity physical activity from Treuth et al. (2004) and Puyau et al. (2002) were 51 and 400 counts/30 sec, respectively.

[3] Receiver operating characteristic (ROC) curve is a plot of the true positive and false positive rate for each cut point tested. Area under the curve was approximated as the integral using the trapezoidal rule.

The mean misclassification rate was 19.4% in the development sample and 19.5% in the evaluation sample using the new cut points and the Treuth cut point (data not shown). Using the Puyau cut point, the mean misclassification rate was 26.0% in the development sample and 28.8% in the evaluation sample (data not shown). The mean area under the ROC curve was 0.52 in both samples (data not shown). Two of the new cut points were developed using the Strong accumulated time recommendation, and one was developed using the Dietary Guidelines frequency and duration recommendation.

DISCUSSION

The previously suggested cut points used in this study were developed using oxygen consumption as the criterion (Puyau et al., 2002; Treuth et al., 2004), thereby reflecting the energy expenditure associated with physical activity. Results from this study indicate that several cut points exist that may better predict the status of selected cardiovascular and metabolic disease risk factors compared with previously suggested cut points, particularly the Puyau cut point. These new cut points may be more clinically relevant as they use disease risk factors, including body composition and lipids and lipoproteins, as criteria. Hence, these cut points may be more appropriate to use than oxygen consumption cut points in studies with adolescent girls where the interest is in the relationship between physical activity and disease risk.

The new cut points identified tended to be in the range of light intensity activities, as defined by the previously suggested cut points. Furthermore, the new cut points always performed at least as well, if not better than the previously suggested moderate or vigorous cut points. This suggests that light intensity activities might be sufficient to receive health benefits. Similarly, studies using data from the Australian Diabetes, Obesity, and Lifestyle Study have found that light intensity time was independently associated with waist circumference, clustered metabolic risk, and 2-hour post-challenge plasma glucose in adults without diagnosed diabetes (Healy et al., 2007; Healy et al., 2008). This may be helpful for public health interventions. Light activities might be more acceptable to girls because they may address some of the perceived barriers to physical activity, including injury, physical discomfort (e.g., sweating, tiredness, breathing harder), and personal appearance (e.g., ruining hair style) (Allison, Dwyer, & Makin, 1999; Dunton & Schneider, 2006; Grieser et al., 2006).

Although the new cut points reflected lower counts than the previously suggested cut points, there was still some variability across population subgroups and outcomes. Hence, the physical activity intensity needed to confer health benefits might vary depending on the outcome of interest. In this study, higher cut points were identified for central adiposity and HDL cholesterol than other risk factors. Similarly, waist circumference was associated with vigorous but not light or moderate physical activity among Spanish adolescents (Moliner-Urdiales et al., 2009). Hence, higher intensity activity may be required to reduce waist circumference and raise HDL cholesterol levels. New cut points were not identified for each population-risk factor subgroup. One reason may be that factors that differ between the population subgroups, such as physical maturity, environment, or genetics, may moderate or mediate the relationship between physical activity and disease risk factors. Hence, cut points might also depend on the population of interest. Therefore, future studies and interventions that use accelerometer cut points specific to the outcome and population of interest might observe stronger relationships and the relationships might be better understood.

While several new cut points were identified, accelerometers may not be good at classifying cardiovascular and metabolic disease risk. In this study, the majority of relationships tested had low areas under the curve and substantial misclassification. Other factors, such as diet or genetics, that were not examined in this study may account for some of the misclassification. This may be due in part to the fact that it was not a high risk population. Restricting the sample to adolescent girls in the general population targeted a population at risk for low physical activity and controlled for age and gender. However, participants may not have been exposed to their level of physical activity long enough to observe an influence on disease risk factors. Even though approximately one-third of the sample was at elevated risk for the factors examined, these methods may have worked better with higher risk populations (e.g., older adults, overweight).

Another strength of this study was that it used data from two large field studies. Hence, the activities performed were free-living. However, the data were not collected to calibrate accelerometer data. One drawback to this was that a large proportion of participants were excluded from analysis because they did not have sufficient accelerometer data, which is difficult to assure in a large field study. Another weakness of this study was that it used cross-sectional data. The assumption is that physical activity influences cardiovascular and metabolic disease risk factors. However, with no temporal relationship between the measures, this could not be evaluated. Furthermore, the accelerometer data only represented one week of activity, which may not represent habitual physical activity patterns.

This study used a unique approach to identifying accelerometer cut points. Utilizing this approach allowed new cut points to be identified using more readily available criterion data (i.e., disease risk factors) rather than oxygen consumption. However, this assumes that the cut points used to classify disease risk represented appropriate signals. Furthermore, it was assumed that the physical activity recommendations were appropriate definitions for the detector. To account for some of the uncertainty about the recommendations, multiple recommendations were used and they were each interpreted two ways.

In summary, this study demonstrated the use of signal detection and ROC curves in identifying new accelerometer cut points for assessing physical activity. The new cut points identified suggest that light intensity may be sufficient to obtain similar health benefits and the actual intensity may differ by disease risk factor and population subgroup. While accelerometers did not identify disease risk well, as made evident by ROC curves, new cut points were identified. This suggests that cut points developed using oxygen consumption may not best reflect the clinically relevant aspects of physical activity measured by accelerometers.

REFERENCES

Allison, K. R., Dwyer, J. J., & Makin, S. (1999). Perceived barriers to physical activity among high school students. *Prev.Med., 28,* 608-615.

American Academy of Pediatrics Committee on Nutrition (1998). Cholesterol in childhood. *Pediatrics, 101,* 141-147.

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

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med.Sci.Sports Exerc., 32,* 1601-1609.

Centers for Disease Control and Prevention, National Center for Health Statistics (2006). National Health and Nutrition Examination Survey 2003-2004 documentation, codebook, and frequencies: exam component physical activity monitor examination data. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/paxraw_c.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2009). National Health and Nutrition Examination Survey 2003-2004 documentation, codebook, and frequencies: demographic variables and sample weights. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/demo_c.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2003). National Health and Nutrition Examination Survey overview. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/overviewbrochure_0304.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2004a). National Health and Nutrition Examination Survey anthropometry procedures manual. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/BM.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2004b). National Health and Nutrition Examination Survey laboratory procedures manual. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/lab_pm.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2000). CDC growth charts: United States. http://www.cdc.gov/growthcharts/ [On-line].

de Ferranti, S. D., Gauvreau, K., Ludwig, D. S., Neufeld, E. J., Newburger, J. W., & Rifai, N. (2004). Prevalence of the metabolic syndrome in American adolescents: findings from the third National Health and Nutrition Examination Survey. *Circulation, 110*, 2494-2497.

Dunton, G. F. & Schneider, M. (2006). Perceived barriers to walking for physical activity. *Prev.Chronic Dis., 3,* A116.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *J.Sports Sci.*, 1-9.

Fernandez, J. R., Redden, D. T., Pietrobelli, A., & Allison, D. B. (2004). Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J.Pediatr., 145,* 439-444.

Grieser, M., Vu, M. B., Bedimo-Rung, A. L., Neumark-Sztainer, D., Moody, J., Young, D. R. et al. (2006). Physical activity attitudes, preferences, and practices in African American, Hispanic, and Caucasian girls. *Health Educ.Behav., 33,* 40-51.

Healy, G. N., Dunstan, D. W., Salmon, J., Cerin, E., Shaw, J. E., Zimmet, P. Z. et al. (2007). Objectively measured light intensity physical activity is independently associated with 2-h plasma glucose. *Diabetes Care, 30,* 1384-1389.

Healy, G. N., Wijndaele, K., Dunstan, D. W., Shaw, J. E., Salmon, J., Zimmet, P. Z. et al. (2008). Objectively measured sedentary time, physical activity, and metabolic risk. *Diabetes Care, 31,* 369-371.

Kimm, S. Y., Glynn, N. W., Kriska, A. M., Barton, B. A., Kronsberg, S. S., Daniels, S. R. et al. (2002). Decline in physical activity in black girls and white girls during adolescence. *N.Engl.J.Med., 347,* 709-715.

Korn, E. L. & Graubard, B. I. (1991). Epidemiologic studies using surveys: accounting for the sampling design. *Am.J.Public Health, 81,* 1166-1173.

Kraemer, H. C. (1988). Assessment of 2 x 2 associations: generalization of signal-detection methodology. *Am.Statistician, 42,* 37-49.

Lohman, T. G. & Falls, H. B. (2001). Body composition assessment. In G.J. Welk, J. R. Morrow, & H. B. Falls (Eds.), *FITNESSGRAM reference guide* (pp. 87-92). Dallas: The Cooper Institute for Aerobics Research.

Lohman, T. G., Ring, K., Schmitz, K. H., Treuth, M. S., Loftin, M., Yang, S. et al. (2006). Associations of body size and composition with physical activity in adolescent girls. *Med.Sci.Sports Exerc., 38,* 1175-1181.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. I., & Hu, J. (2003). Tracking of physical activity and aerobic power from childhood through adolescence. *Med.Sci.Sports Exerc., 35,* 1914-1922.

Moliner-Urdiales, D., Ruiz, J. R., Ortega, F. B., Rey-Lopez, J. P., Vicente-Rodriguez, G., España-Romero, V. et al. (2009). Association of objectively assessed physical activity with total and central body fat in Spanish adolescents; The HELENA Study. *Int.J.Obes., advance online publication,* doi:10.1038.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA, 300,* 295-305.

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

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obes.Res., 10,* 150-157.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B. et al. (2005). Evidence based physical activity for school-age youth. *J.Pediatr., 146*, 732-737.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J. et al. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med.Sci.Sports Exerc., 36,* 1259-1266.

U.S. Department of Health and Human Services (1996). *Physical activity and health: a report of the Surgeon General* Atlanta, GA: U.S. Department of Health and

Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.

U.S. Department of Health and Human Services (2000). *Healthy People 2010: understanding and improving health and objectives for improving health* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services & U.S. Department of Agriculture (2005). *Dietary guidelines for Americans*. http://www.health.gov/dietaryguidelines/dga2005/document/ [On-line].

Chapter 4: Evaluation of Convergent Validity of New Accelerometer Cut Points for Adolescent Girls with Recall Questionnaires and a Previously Suggested Cut Point

ABSTRACT

Accelerometer cut points have been used to classify the intensity of activity of a given epoch. However, different cut points might not classify the same epoch as the same intensity. PURPOSE: To examine how estimates of physical activity using newly developed cut points agree with a cut point previously suggested by Treuth (2004) and two self-report physical activity questionnaires. METHODS: The new cut points had been developed using iterative correlations and signal detection and receiver operating characteristic (ROC) curves. Minutes of physical activity and individuals meeting recommendations for each cut point and questionnaire were determined in the 2003-2004 National Health and Nutrition Examination Survey (NHANES) and the Trial of Activity for Adolescent Girls (TAAG). Agreement was examined using concordance correlation coefficients, McNemar's tests, and proportions of agreement. RESULTS: Concordance correlation coefficients for minutes of activity with the previously suggested cut point tended to be stronger (≥ 0.6) with higher cut points (≥ 2300 count/min), while those with questionnaires were less than 0.10 or the 95% confidence intervals included zero. Some proportions of agreement for meeting recommendations were moderate (> 0.6) with the questionnaires and were high (> 0.9) with the previously suggested cut point, but only one new cut point was not significantly different (1800 counts/min and NHANES questionnaire, $p = 0.6$). CONCLUSIONS: The new cut points tended to be in poor

agreement with the comparison measures, especially the questionnaires. They are likely measuring different aspects of physical activity.

INTRODUCTION

Accurate physical activity measures are needed to better understand the relationship of physical activity with health outcomes, determinants of physical activity, effects of physical activity interventions, as well as the prevalence of physical activity. However, physical activity is difficult to accurately assess.

Self-report questionnaires are often used in large, epidemiologic studies to assess physical activity because they are relatively inexpensive and can efficiently collect large amounts of detailed information. However, information from self-report questionnaires may not be accurate due to cognitive abilities and social desirability bias (Baranowski, 1988; Montoye, Kemper, Saris, & Washburn, 1996). Physical activity dimensions of interest may include frequency, intensity, duration, type, and context. Respondents, particularly children and adolescents, may not be able to accurately recall or report all the relevant details of their physical activity. They may also perceive the details of their activity differently than intended by the researcher. For example, they may report an activity as being vigorous that the researcher would consider moderate. Furthermore, respondents may alter their responses because they want, consciously or subconsciously, to appear more physically active.

Accelerometers may help improve the accuracy of physical activity assessment as they can overcome weaknesses of self-report questionnaires. Accelerometers record the intensity of movement as counts per user-specified epochs (e.g., per minute). Hence, they objectively collect data on the frequency, intensity, and duration of physical activity.

However, accelerometer counts have no biological meaning, and it is still not clear how best to translate accelerometer data into physical activity estimates (Chen & Bassett, Jr., 2005). Cut points have been used to classify the intensity of activity during a given epoch, so the frequency and duration of activity at different intensities can be estimated.

Treuth et al. (2004) and Puyau et al. (2002) have developed cut points for light, moderate, and vigorous intensity for young people using the Actigraph 7164 with oxygen consumption as their criterion. Participants performed structured activities while oxygen consumption and accelerometer counts were simultaneously recorded. Despite the similarities of these studies, two different sets of cut points were developed. Treuth cut points were 101, 3000, and 5201 counts/min, and Puyau cut points were 800, 3200, and 8200 counts/min. Depending on which cut point is used, one cut point might classify an epoch as being active while another might classify the same epoch as not being active. Hence, it is important to understand how using different accelerometer cut points might affect physical activity estimates.

While oxygen consumption reflects the energy expenditure associated with physical activity, using disease risk factors to develop cut points may be more appropriate where the relationship between physical activity and health and disease is of interest. New accelerometer cut points for adolescent girls were developed in two studies using disease risk factors as the criterion measures (Chapters 2 and 3). Physical activity has been associated with several health benefits including higher physical fitness, healthier body composition, and more favorable cardiovascular and metabolic disease risk profiles (U.S. Department of Health and Human Services, 1996; Physical Activity Guidelines Advisory Committee, 2008). The two studies to develop cut points and the current study focused on adolescent girls as physical activity levels tend to decline during adolescence (Kimm et al., 2002; McMurray et al., 2003; Nader et al., 2008) and adolescent girls tend to be less active than adolescent boys (Caspersen et al., 2000; McMurray et al., 2003; Nader et al., 2008). The purpose of this study was to understand how well estimates of moderate-to-vigorous physical activity using these new cut points agree with a previously suggested moderate-to-vigorous cut point and two self-report physical activity questionnaires.

METHODS

This study examined agreement using data from two studies that collected accelerometer data from large numbers of adolescent girls from different locations across the United States: the 2003-2004 National Health and Nutrition Examination Survey (NHANES) and the Trial of Activity for Adolescent Girls (TAAG). NHANES is a continuous, annual survey that collected information about health and diet from a nationally representative sample of the U.S. noninstitutionalized household population. TAAG is a group-randomized, multi-center trial of a school- and community-linked intervention to decrease the decline in physical activity of middle school girls. Baseline measures were conducted with sixth grade girls in 2003. Follow-up measures were conducted with eighth grade girls in 2005. The sustainability of the intervention was determined by measuring eighth grade girls in 2006. The new accelerometer cut points examined in this study were identified based on the relationships between physical activity and selected cardiovascular and metabolic disease risk factors among 333 girls 12 to 15 years old in the 2003-2004 NHANES and 4696 eighth grade girls in TAAG.

Developing new cut points

Two methods used to develop new cut points are described briefly here and in detail elsewhere. One method was an iterative process based on correlations (Chapter 2). Using previously suggested cut points (Treuth et al., 2004; Puyau et al., 2002) as a starting point, physical activity estimates were generated using multiple values above and below each cut point at relatively large intervals. Rank correlations of these new physical activity estimates with each continuous disease risk factor were determined. Cut points that maximized the correlations were the new starting points for subsequent iterations. The process was repeated using smaller intervals until final cut points were determined.

The other method for developing new cut points used signal detection and receiver operating characteristic (ROC) curves (Chapter 3). Each participant was classified as high or low risk for each disease risk factor. At given accelerometer count increments (e.g., 100 counts/min), each participant was classified as meeting or not meeting physical activity recommendations. Three recommendations were used: Strong et al. (2005), 2005 Dietary Guidelines for Americans (U.S. Department of Health and Human Services et al., 2005), and Healthy People 2010 Objective 22.6 (U.S. Department of Health and Human Services, 2000). Cut points that minimized misclassification (e.g., not meeting recommendations and low risk) were identified. ROC curves (i.e., plots of the true positive rate against the false positive rate for different possible cut points) were used to examine how well the accelerometer separated the sample into high and low risk.

Only cut points that were consistently (i.e., in the development and evaluation sample) more strongly associated with disease risk than the previously suggested cut points (Treuth et al., 2004; Puyau et al., 2002) were considered new cut points and were evaluated for agreement in this study. New cut points were developed in each dataset overall and within race/ethnicity and age groups to control for potential confounding factors and variables used in the sampling designs (Korn et al., 1991). Agreement was evaluated within the same population subgroup the cut point was developed. Of the 333 girls in NHANES, 171 (51%) were 12-13 years old, 162 (49%) were 14-15 years old, and 85 (26%) were White. Of the 4696 eighth grade girls in TAAG, 2190 (47%) were White and 866 (18%) were Black.

Determining physical activity variables

Minutes of physical activity. Average daily time spent in physical activity was defined as the total time spent doing moderate-to-vigorous activities divided by the total number of days. For accelerometer data, each epoch measured that had a count equal to or above a given cut point was classified as moderate-to-vigorous for that cut point. For questionnaire data, physical activity intensity was classified based on assigned MET values and respondent perceptions. Moderate-to-vigorous physical activities were defined as those assigned a value of 3 METS or more.

The physically active time from questionnaires was determined differently due to differences in methodology. In the NHANES physical activity questionnaire, participants reported the number of times they performed a physical activity for at least 10 minutes in the past month and the average number of minutes they did it each time. In the TAAG 3-day physical activity recall (3DPAR), participants reported the main activity performed by 30-minute blocks of time for the previous three days. However, they could have performed additional activities or taken breaks from the activity. Hence, to take into account those departures from the reported activity, each block of time was counted as 10 minutes of activity. For accelerometer and questionnaire data, minutes of physical activity per day were normalized to a 12-hour day. The total normalized number of minutes of moderate-to-vigorous physical activity was divided by the total time frame of the questionnaire or the total number of days of valid accelerometer data to calculate the average daily time spent in moderate-to-vigorous physical activity.

Meeting physical activity recommendations. Participants were classified as meeting the Dietary Guidelines accumulated time recommendation if the average daily time was greater than 60 minutes on 4 days per week (i.e., 60*4/7). The Dietary Guidelines recommendation was used because it recommended 60 minutes of activity, similar to the Strong recommendation, but activity did not have to be performed daily (i.e., performed on most days of the week), similar to the Health People recommendation.

Statistical analysis

The degree to which minutes of physical activity estimated using the new accelerometer cut points agreed at the individual level with those estimated using self-report questionnaires and the Treuth (2004) moderate-to-vigorous cut point were examined using the concordance correlation coefficient (Lin, 1989) and Bland-Altman plots (1986). The Treuth cut point was used as a comparison measure because it was developed in a similar population (i.e., 13 to 14 year old girls). Proportions of agreement and McNemar's tests were used to examine agreement for meeting the Dietary Guidelines accumulated time recommendation. Because this study was interested agreement at the individual level, rather than at the population level, unweighted analyses were used.

Minutes of physical activity. Concordance correlation coefficients were used to examine the agreement between the new accelerometer cut points and comparison measures (i.e., the Treuth cut point and self-report questionnaires) for measuring minutes of moderate-to-vigorous physical activity. The concordance correlation coefficient is the correlation between two measures that fall on the 45° line through the origin (Lin, 1989). If the estimates were exactly the same, a plot of the new cut point against the comparison measure would be a line through the origin with a slope of 1. The concordance correlation coefficient was calculated for this study as:

$$r_c = \frac{2 \times S_{\text{new cut point, comparison}}}{S^2_{\text{new cut point}} + S^2_{\text{comparison}} + (\text{mean}_{\text{new cut point}} - \text{mean}_{\text{comparison}})^2}$$

where $S_{\text{new cut point, comparison}}$ is the covariance between minutes of physical activity using the new cut point and using the comparison measure, $S^2_{\text{new cut point}}$ and $S^2_{\text{comparison}}$ are the variances, and $\text{mean}_{\text{new cut point}}$ and $\text{mean}_{\text{comparison}}$ are the means.

Bland-Altman plots were also used to graphically examine patterns in the individual differences between minutes of physical activity estimated using the new cut points and comparison measures. The differences between the minutes of physical activity estimated using the comparison measures and the new cut points (i.e., the new cut point minus the comparison measure estimates) were plotted on the vertical axis, and the average minutes (i.e., the sum of the new cut point and the comparison measure estimates divided by 2) were plotted on the horizontal axis. Patterns were characterized relative to increasing average minutes of activity (i.e., moving from left to right on the horizontal axis) as: a) tending to get more positive; b) tending to get more negative; c) increasing in variation; and d) staying relatively constant. Examples of each pattern are presented in Figure 4.1. The mean of the differences and the limits of agreement (i.e., mean ± 1.96SD) were also determined to examine bias. Negative differences mean the new cut point estimates are less than the comparison measure, and positive differences mean the new cut point estimates are greater than the comparison measure.

Meeting physical activity recommendation. McNemar's test was used to examine whether similar percentages of participants were classified as meeting and not meeting the Dietary Guidelines accumulated time physical activity recommendation using the new cut points and the comparison measures. The proportions of agreement overall and specifically for meeting and not meeting the Dietary Guidelines accumulated time recommendation (i.e., proportion of both measures producing the same classification among those classified as such by either measure) were also determined.

RESULTS

New cut points identified

The new moderate-to-vigorous cut points identified using the correlation method were higher than cut points identified using signal detection and ROC curves. In NHANES, the moderate-to-vigorous cut points from the correlation method included 4000 and 4300 counts/min, while the cut points from the signal detection and ROC curve method included 100, 400, 700, 1200, 1800, and 2300 counts/min. The correlation method only identified new cut points among 12-13 year old girls in NHANES. The four highest new cut points from signal detection and ROC curves (i.e., ≥700 counts/min) were all identified among White girls. In TAAG, the moderate-to-vigorous cut points from the correlation method included 1550 and 1950 counts/30 sec, while the cut points

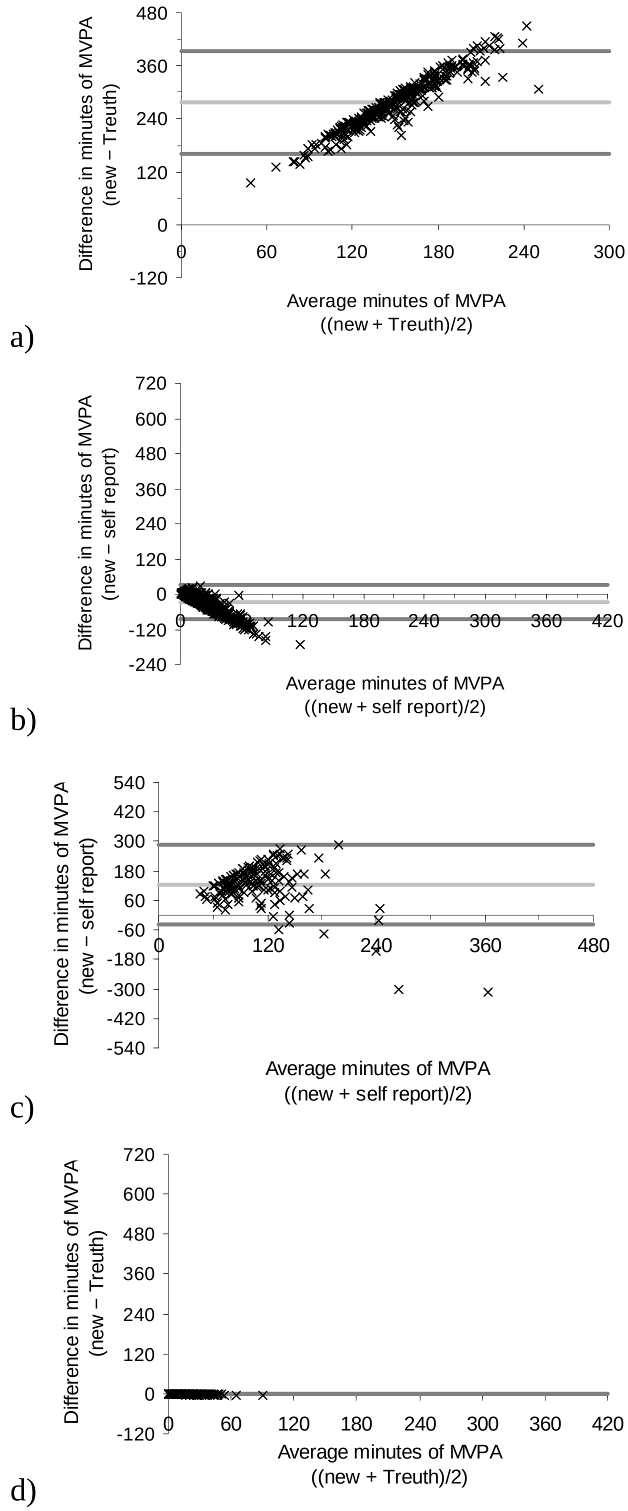


Figure 4.1. Examples of patterns from Bland-Altman plots: a) differences getting more positive, b) differences getting more negative, c) differences increasing in variation, and d) differences staying relatively constant

from the signal detection and ROC curve method included 50 and 100 counts/30 sec. The correlation method only identified new cut points among Black girls in TAAG.

<u>Agreement with NHANES questionnaire in NHANES</u>

Agreement for minutes of physical activity. Table 4.1 presents results from concordance correlations and Bland-Altman plots with the NHANES questionnaire. The new accelerometer cut points and NHANES questionnaire were not in agreement about the estimated average daily minutes of moderate-to-vigorous physical activity for any of the population subgroups. All of the 95% confidence intervals for the concordance correlations included zero.

The smallest average difference in estimated minutes of activity between the NHANES questionnaire and the new cut points was for 1200 counts/min. The 1200 counts/min cut point averaged 9 minutes more activity than the NHANES questionnaire. The four highest new cut points of 1800, 2300, 4000, and 4300 counts/min averaged 17 to 42 minutes less activity than the NHANES questionnaire. The greatest average differences were observed for the three lowest new cut points of 100, 400, and 700 counts/min. These three cut points averaged approximately 1 to 4 hours more activity than the NHANES questionnaire.

In the Bland-Altman plots, variation in the differences in minutes of activity between the NHANES questionnaire and the four lowest new cut points of 100, 400, 700, and 1200 counts/min tended to increase with higher average minutes of activity, while differences tended to become more negative with higher average minutes of activity for the four highest new cut points. For all cut points, most of the differences observed

Table 4.1. Concordance correlation coefficient (r_c), mean difference in minutes of moderate-to-vigorous physical activity, limits of agreement, and Bland-Altman plot pattern between minutes of physical activity measured by new accelerometer cut points (counts/min) and comparison measures in NHANES by population subgroup

Comparison[1] / Subgroup	Cut point	r_c (95% CI)	Mean difference[2]	Limits of agreement[3]	Plot pattern[4]
Questionnaire					
All girls	100	0.01 (-0.006, 0.020)	244.9	76.8 - 413.0	〈
	400	0.02 (-0.013, 0.050)	118.7	−29.4 - 266.7	〈
12-13 y	400	0.01 (-0.039, 0.051)	125.2	−37.4 - 287.9	〈
	4000	0.00 (-0.022, 0.031)	−41.3	−176.0 - 93.5	↓
	4300	0.00 (-0.019, 0.027)	−42.4	−176.9 - 92.1	↓
14-15 y	100	0.01 (-0.007, 0.028)	234.1	81.8 - 386.5	〈
White	400	0.05 (-0.026, 0.118)	114.6	−42.5 - 271.6	〈
	700	0.10 (-0.022, 0.214)	56.3	−90.5 - 203.1	〈
	1200	0.13 (-0.011, 0.263)	8.7	−132.2 - 149.6	〈
	1800	0.09 (-0.011, 0.192)	−17.2	−156.6 - 122.3	↓
	2300	0.05 (-0.027, 0.132)	−28.2	−168.9 - 112.4	↓
Treuth					
Overall	100	0.00 (0.003, 0.006)	275.9	159.5 - 392.3	↑
	400	0.02 (0.013, 0.022)	149.6	67.3 - 231.9	↑
12-13 years	400	0.02 (0.013, 0.027)	159.5	73.5 - 245.5	↑
	4000	0.71 (0.656, 0.755)	−7.1	−16.3 - 2.1	↔
	4300	0.60 (0.543, 0.657)	−8.3	−19.3 - 2.8	↔
14-15 years	100	0.00 (0.000, 0.005)	261.7	149.3 - 374.0	↑
White	400	0.02 (0.012, 0.034)	151.8	69.2 - 234.3	↑
	700	0.06 (0.034, 0.077)	93.9	33.7 - 154.1	↑
	1200	0.17 (0.119, 0.222)	46.4	9.3 - 83.5	↑
	1800	0.45 (0.370, 0.540)	20.4	1.4 - 39.3	↔
	2300	0.77 (0.713, 0.833)	9.1	−0.8 - 19.1	↔

[1] Comparison measures were moderate-to-vigorous physical activity from the NHANES questionnaire (≥ 3 METS) and the Treuth (2004) accelerometer cut point (≥ 3000 counts/min).

[2] Difference in minutes was calculated as estimated minutes using the new cut point minus those using the comparison measure.

[3] Limits of agreement were calculated as the mean ± 1.96×SD.

[4] Patterns from the Bland-Altman plots were differences tend to get more positive (↑), more negative (↓), increase in variation (〈), or stay relatively constant (↔) with increasing average minutes of activity.

outside the limits of agreement were below the lower limits (data not shown). The limits of agreement were 124 to 168 minutes away from the mean.

Agreement for meeting physical activity recommendation. The proportions of agreement and results from McNemar's test for the Dietary Guidelines accumulated time recommendation for NHANES are presented in Table 4.2. The greatest proportions of overall agreement (0.62) with the NHANES questionnaire were for the new cut points of 1200 and 1800 counts/min. The new cut point of 1200 counts/min also had the greatest proportion of positive agreement (0.69) with the NHANES questionnaire. The two highest new cut points of 4000 and 4300 counts/min had the greatest proportion of negative agreement (0.72) but zero positive agreement with the NHANES questionnaire. The three lowest new cut points of 100, 400, and 700 counts/min had zero negative agreement with the NHANES questionnaire, as these cut points classified all girls as meeting the recommendation. The percentage of participants classified as meeting and not meeting the recommendation was not significantly different from the NHANES questionnaire for one cut point: 1800 counts/min ($p = 0.59$).

<u>Agreement with Treuth cut point in NHANES</u>

Agreement for minutes of physical activity. As shown in Table 4.1, the strongest concordance correlation for estimated minutes of activity with the Treuth moderate-to-vigorous cut point was observed for the new cut point of 2300 counts/min ($r_c = 0.8$). Moderate to strong correlations were also observed for the new cut points of 1800, 4000, and 4300 counts/min ($r_c = 0.5$-0.7). The four lowest new cut points of 100, 400, 700, and 1200 counts/min were very weakly to weakly correlated ($r_c = 0.2$ or less) with the Treuth cut point.

Table 4.2. Proportions of agreement and p-values from McNemar's tests between using new accelerometer cut points (counts/min) and comparison measures for classification of meeting Dietary Guidelines accumulated time physical activity recommendation in NHANES by population subgroup

Cut point	Questionnaire[2]				Treuth[2]			
	Overall[1]	Met[1]	Not met[1]	p-value	Overall	Met	Not met	p-value
All girls								
100	0.392	0.563	0.000	+++	0.036	0.070	0.000	+++
400	0.392	0.563	0.000	+++	0.036	0.070	0.000	+++
12-13 y								
400	0.422	0.594	0.000	+++	0.041	0.079	0.000	+++
4000	0.565	0.000	0.722	<.0001	0.971	0.444	0.985	0.03
4300	0.565	0.000	0.722	<.0001	0.971	0.444	0.985	0.03
14-15 y								
100	0.359	0.529	0.000	+++	0.031	0.060	0.000	+++
White								
400	0.457	0.627	0.000	+++	0.035	0.068	0.000	+++
700	0.457	0.627	0.000	+++	0.035	0.068	0.000	+++
1200	0.617	0.687	0.508	<.0001	0.271	0.088	0.392	<.0001
1800	0.617	0.563	0.659	0.59	0.624	0.158	0.758	<.0001
2300	0.556	0.217	0.690	<.0001	0.918	0.462	0.955	0.0082

+++ Value not determined because all participants classified as meeting that recommendation using that cut point.

[1] Overall agreement was the proportion of both measures producing the same classification overall. Met/not met (positive/negative) agreement was the proportion of both measures producing met/not met classification among those classified as met/not met by either measure.

[2] Comparison measures were moderate-to-vigorous physical activity from the NHANES questionnaire (≥ 3 METS) and the Treuth (2004) accelerometer cut point (≥ 3000 counts/min).

The smallest average difference in estimated minutes of activity between the Treuth and the new cut points was for 4000 counts/min. The new cut point of 4000 counts/min averaged 7 minutes less activity than the Treuth cut point. Differences in activity for the new cut points immediately above and below the 4000 counts/min cut point (2300 and 4300 counts/min) were within 8 to 9 minutes of the Treuth cut point. The five lowest new cut points of 100, 400, 700, 1200, and 1800 counts/min all averaged more minutes of activity than the Treuth cut point, ranging from approximately 20 minutes more for 1800 counts/min to over 4 hours more for 100 counts/min.

In the Bland-Altman plots, there was little variation in the differences in activity between the Treuth cut point and the four highest new cut points (1800, 2300, 4000, and 4300 counts/min), while variation for the four lowest new cut points (100, 400, 700, and 1200 counts/min) tended to increase with higher average minutes of activity. The limits of agreement were 9 to 11 minutes away from the mean for the three highest new cut points, 19 to 37 minutes away for 1200 and 1800 counts/min, and 1 to 2 hours away for the three lowest new cut points. Differences outside the limits of agreement were above the upper limits for most of the new cut points (700, 1200, 1800, and 2300 counts/min) and below the lower limits for the two highest new cut points (data not shown). Differences were observed both above and below the limits of agreement for the 100 and 400 counts/min cut points, but most of these differences for the 400 counts/min cut point were above the upper limit (data not shown).

Agreement for meeting physical activity recommendation. The greatest proportions of overall (0.97) and negative (0.99) agreement for the Dietary Guidelines accumulated time recommendation with the Treuth moderate-to-vigorous cut point were for the two highest new cut points of 4000 and 4300 counts/min (Table 4.2). The new cut point of 2300 counts/min had the greatest proportion of positive agreement (0.46). The three lowest new cut points of 100, 400, and 700 counts/min had the smallest proportions of overall (0.03 to 0.04) and positive (0.06 to 0.08) agreement and zero negative agreement with the Treuth cut point. The percentages of participants classified as meeting and not meeting the recommendation were significantly different from the Treuth cut point for all new cut points tested. However, the difference for two cut points (4000 and 4300 counts/min) was not highly significant (p = 0.03).

Agreement with 3DPAR in TAAG

Agreement for minutes of physical activity. Table 4.3 presents results from concordance correlation and Bland-Altman plots for TAAG 3DPAR. The accelerometer and 3DPAR were in very weak agreement ($r_c < 0.1$) about the estimated average daily minutes of moderate-to-vigorous physical activity for any of the population subgroups.

The smallest average differences in estimated minutes of activity between the TAAG 3DPAR and the new cut points were for 1550 and 1950 counts/30 sec. These new cut points averaged 21 and 27 minutes less activity, respectively, than the TAAG 3DPAR. The lower new cut points of 50 and 100 counts/30 sec averaged 2.8 to 3.7 hours more activity than the TAAG 3DPAR.

In the Bland-Altman plots, variation in the differences of activity between the TAAG 3DPAR and the new cut points tended to become more positive with higher average minutes for 50 and 100 counts/30 sec and more negative for 1550 and 1950 counts/30 sec. The limits of agreement were 115 to 123 minutes away from the mean for the two lowest new cut points and 57 minutes away for the two highest new cut points.

Table 4.3. Concordance correlation coefficient (r_c), mean difference in minutes of moderate-to-vigorous physical activity, limits of agreement, and Bland-Altman plot pattern between minutes of physical activity measured by new accelerometer cut points (counts/30 sec) and comparison measures in TAAG by population subgroup

Comparison[1] / Subgroup	Cut point	r_c (95% CI)	Mean difference[2]	Limits of agreement[3]	Plot pattern[4]
3DPAR					
All girls	50	0.01 (0.007, 0.011)	220.2	97.4 - 343.1	↑
White	50	0.01 (0.007, 0.012)	222.0	101.3 - 342.8	↑
Black	100	0.01 (0.006, 0.019)	165.4	50.8 - 280.1	↑
	1550	0.07 (0.044, 0.093)	−21.2	−78.4 - 36.0	↓
	1950	0.04 (0.025, 0.053)	−26.9	−83.8 - 30.0	↓
Treuth					
All girls	50	0.01 (0.010, 0.011)	238.9	129.9 - 348.0	↑
White	50	0.01 (0.009, 0.012)	239.1	129.3 - 348.8	↑
Black	100	0.01 (0.013, 0.017)	185.7	88.7 - 282.8	↑
	1550	0.99 (0.990, 0.992)	−1.0	−2.3 - 0.2	↔
	1950	0.62 (0.592, 0.639)	−6.8	−14.7 - 1.1	↔

1 Comparison measures were moderate-to-vigorous physical activity from the TAAG 3-day Physical Activity Recall (3DPAR) (≥ 3 METS) and the Treuth (2004) accelerometer cut point (≥1500 counts/30 sec).

2 Difference in minutes was calculated as estimated minutes using the new cut point minus those using the comparison measure.

3 Limits of agreement were calculated as the mean ± 1.96×SD.

4 Patterns from the Bland-Altman plots were differences tend to get more positive (↑), more negative (↓), increase in variation (〈), or stay relatively constant (↔) with increasing average minutes of activity.

Differences were observed both above and below the limits of agreement for the 50 and 100 counts/30 sec cut points, while most of the differences observed outside the limits of agreement for 1550 and 1950 counts/30 sec were below the lower limits (data not shown).

Agreement for meeting physical activity recommendation. The proportions of agreement and results from McNemar's test for the Dietary Guidelines accumulated time recommendation in TAAG are presented in Table 4.4. The greatest proportions of overall agreement (0.61) with the TAAG 3DPAR were for the two highest new cut points of 1550 and 1950 counts/30 sec. These cut points also had the greatest proportions of negative agreement (0.75 and 0.76, respectively) with the TAAG 3DPAR. The lowest new cut point of 50 counts/30 sec had the greatest proportions of positive agreement with the TAAG 3DPAR among all girls (0.60) and White girls (0.61). The percentage of participants classified as meeting and not meeting the recommendation was significantly different from the TAAG 3DPAR for all new cut points (all $p < 0.0001$).

Agreement with Treuth cut point in TAAG

Agreement for minutes of physical activity. As shown in Table 4.3, the strongest concordance correlation for minutes of activity with the Treuth moderate-to-vigorous cut point was observed for the new cut point of 1550 counts/30 sec ($r_c \approx 1.0$). A strong correlation was also observed for the new cut point of 1950 counts/30 sec ($r_c = 0.6$). The two lowest new cut points of 50 and 100 counts/30 sec were very weakly correlated ($r_c = 0.01$) with the Treuth cut point.

The smallest average differences in estimated minutes of activity between the Treuth and the new cut points were for 1550 and 1950 counts/30 sec. The new cut point

Table 4.4. Proportions of agreement and p-values from McNemar's tests between using new accelerometer cut points (counts/30 sec) and comparison measures for classification of meeting Dietary Guidelines accumulated time physical activity recommendation in TAAG by population subgroup

Cut point	3DPAR				Treuth			
	Overall[1]	Met[1]	Not met[1]	p-value	Overall	Met	Not met	p-value
All girls								
50	0.432	0.601	0.015	<.0001	0.076	0.127	0.018	<.0001
White								
50	0.441	0.610	0.015	<.0001	0.093	0.158	0.016	<.0001
Black								
100	0.386	0.554	0.015	<.0001	0.047	0.066	0.028	<.0001
1550	0.612	0.073	0.754	<.0001	0.995	0.926	0.998	<.05
1950	0.611	0.012	0.758	<.0001	0.970	0.188	0.985	<.0001

[1] Overall agreement was the proportion of both measures producing the same classification overall. Met/not met (positive/negative) agreement was the proportion of both measures producing met/not met classification among those classified as met/not met by either measure.

[2] Comparison measures were moderate-to-vigorous physical activity from the TAAG 3-day Physical Activity Recall (3DPAR) (≥ 3 METS) and the Treuth (2004) accelerometer cut point (≥1500 counts/30 sec).

of 1550 counts/30 sec averaged 1 minute less activity than the Treuth cut point and 1950 counts/30 sec averaged 7 minutes less. The two lowest new cut points of 50 and 100 counts/30 sec averaged 3 to 4 hours more activity than the Treuth cut point.

In the Bland-Altman plots, variation in the differences in activity between the Treuth and the new cut points tended to become more positive with higher average minutes of activity for 50 and 100 counts/30 sec, but there was little variation in the differences for 1550 and 1950 counts/30 sec. The limits of agreement were 97 to 110 minutes away from the mean for the two lowest new cut points and 1 to 8 minutes away for the two highest new cut points. Differences were observed both above and below the limits of agreement for the 50 and 100 counts/30 sec cut points, while the differences observed outside the limits of agreement for 1550 and 1950 counts/30 sec were below the lower limits (data not shown).

Agreement for meeting physical activity recommendation. The greatest proportions of overall agreement (≥ 0.97) for the Dietary Guidelines accumulated time recommendation with the Treuth cut point were for the two highest new cut points of 1550 and 1950 counts/30 sec (Table 4.4). These cut points had the greatest proportions of negative agreement (> 0.98) with the Treuth cut point. The new cut point of 1550 counts/30 sec also had the greatest proportion of positive agreement (0.93) with the Treuth cut point. The difference in the percentage of participants classified as meeting and not meeting the recommendation was significantly different from the Treuth cut point for all of the new cut points tested. However, the difference for one cut point (1550 counts/30 sec) was not highly significant ($p = 0.05$).

DISCUSSION

This study examined the agreement of eleven new cut points identified using NHANES data and five new cut points identified using TAAG data. The new cut points identified using signal detection and ROC curves were lower than the moderate-to-vigorous cut points identified using correlations. One reason may be that the correlation method used previously suggested cut points as a starting point, testing cut points between 2000 and 4000 counts/min (1000 and 2000 counts/30 sec) whereas the signal detection/ROC curve method tested cut points between 100 and 10500 counts/min (50 to 5250 counts/30 sec). The correlation method did allow for movement away from the initial starting point with subsequent iterations, yet most of the new cut points still fell within the initial range tested. The correlation cut points may have been higher than the signal detection/ROC curve cut point because of the lower limits associated with each methodology. The cut points tested using signal detection and ROC curves could be as low as 100 counts/min (50 counts/30 sec), while the moderate-to-vigorous cut points tested using correlations could not be lower than 1100 counts/min (550 counts/30sec). This was because there needed to be at least 1000 counts/min (500 counts/30 sec) between the light-to-vigorous and moderate-to-vigorous cut points and the lower limit for the light-to-vigorous cut point was 100 counts/min (50 counts/30 sec). It is unlikely that the lower limits restricted the cut points from being more similar considering the cut points identified using the correlation method (4000 and 4300 counts/min and 1550 and 1950 counts/30 sec) were much higher than the lower limits.

The physical activity estimates using the new cut points were in poor agreement with the self-report measures in both datasets. For minutes of moderate-to-vigorous physical activity, all of the concordance correlation coefficients with the TAAG 3DPAR were less than 0.10 and all the 95% confidence intervals for all the concordance correlation coefficients with the NHANES questionnaire included zero. For recommendation classification, only one cut point was not significantly different from the NHANES questionnaire (1800 counts/min) and all cut points were significantly different from the TAAG 3DPAR. Troped et al. (2007) also found poor agreement between the Youth Risk Behavior Survey (YRBS) and the Actigraph, with Kappa coefficients ranging from -0.05 to 0.03 for meeting the Health People 2010 moderate physical activity recommendation among 6th and 7th grade students. However, other studies that have used the Actigraph to validate questionnaires with similar populations have reported stronger agreement. Pate et al. (2003) found Pearson correlations for moderate-to-vigorous physical activity ranging from 0.23 to 0.35 between the 3DPAR and the Actigraph among 8th and 9th grade girls. Welk et al. (2007) found a correlation of 0.76 between the Youth Media Campaign Longitudinal Survey (YMCLS) and the Actigraph for minutes of moderate-to-vigorous physical activity among 11 to 13 year old girls.

Agreement with the self-report questionnaire could not be tested for some new cut points identified with the signal detection/ROC curve method (100, 400, and 700 counts/min in NHANES) because these cut points classified everyone as meeting the Dietary Guidelines accumulated time physical activity recommendation. Although this was the only recommendation examined in the current study, the new cut points were identified using 3 different recommendations (i.e., Strong (2005), 2005 Dietary Guidelines for Americans (U.S. Department of Health and Human Services et al., 2005), and Health People 2010 Objective 22.6 (U.S. Department of Health and Human Services, 2000)) and each recommendation was examined in terms of frequency and duration in addition to accumulated time. Meeting the frequency and duration recommendations required meeting both components of the recommendation (e.g., at least 60 minutes on at least 4 days per week). Accumulated time recommendations could be met with more time on fewer days or less time on more days as long as the daily average time was met (e.g., 60 min * 4 days / 7 days/wk). The Dietary Guidelines accumulated time recommendation was one of six recommendations used to develop the new cut points, and it was not used to identify any of the cut points that classified everyone as meeting it. In other words, these new cut points that were identified using other recommendations were not useful for separating people into meeting or not meeting the Dietary Guidelines accumulated time recommendation. Hence, a cut point that minimized misclassification for one recommendation may not be useful if a different recommendation is used.

The closer the new cut points were to the Treuth cut point, the better the agreement between them. The cut points that were closer to the Treuth cut point tended to be identified using the correlation method, but one cut point (2300 counts/min) identified in NHANES using the signal detection/ROC method was also close. The new cut points from the correlation method were higher than the Treuth cut point in both datasets, but they were closer to the Treuth cut point in TAAG compared with in NHANES. This may be because the Treuth cut points were developed in a very similar population (i.e., 8th grade girls) in a preliminary study for TAAG (Treuth et al., 2004). Although closer cut points agreed better, small differences between cut points made notable differences in agreement for recommendation classification. For example, a difference of 50 counts/30 sec (i.e., 1550 counts/30 sec) was significantly different from the Treuth cut point and the positive agreement dropped to 0.93. One review study that examined the effect of different cut points on minutes of moderate-to-vigorous physical activity found that the mean minutes/day using the Puyau cut point was significantly lower that using the Treuth cut point (Reilly et al., 2008).

High overall agreement does not necessarily reflect both positive and negative agreement. For example, 1950 counts/30 sec had agreement of 0.97 with the Treuth cut point but the positive agreement was only 0.19. The lower cut points tended to have higher positive than negative agreement with the Treuth cut point and vice versa for the higher cut points. This may be an important consideration when selecting a cut point depending on whether one is more interested in having a sensitive or specific cut point.

This study examined the convergent validity of new cut points with the self-report questionnaires and a previously developed cut point. The methods used to develop these new cut points varied in several ways, including the values tested and the physical activity variables used. These differences may explain why they identified different new cut points.

The methods varied in the values tested. The signal detection/ROC curve method tested a wide range of values, independent of previously suggested cut points. In contrast, the correlation method tested a limited number of cut points; previously suggested cut points developed using oxygen consumption as criterion were used as starting points. Hence, new cut points from the correlation method have associated intensity levels, even though they may be very different from the starting points. However, it is unclear what intensity the new cut points from the signal detection/ROC curve method represent, especially if they are in between intensity levels of previous cut points. Hence, it may not be appropriate to compare them against measures of moderate-to-vigorous physical activity.

The physical activity variables differed between methods. The physical activity variable used in the correlation method was a continuous variable of minutes of physical activity, while the signal detection/ROC curve method used a dichotomous variable based on whether physical activity recommendations were met. Using a continuous variable may be inappropriate if there is a minimum threshold that needs to be met to observe benefits. Hence, it does not address the pattern (i.e., frequency, duration, intensity) of physical activity necessary to confer benefits. However, while the cut points from the signal detection/ROC curve method are associated with a pattern of physical activity, the cut points identified depend on the recommendation used and assume the recommendation is appropriate. Furthermore, using a dichotomous variable lacks detailed information for identifying the amount of difference between different physical activity levels (e.g., dose-response).

One strength of this study was the data used were from two large field studies. Hence, data were obtained from a substantial number of adolescent girls and real-world physical activities were assessed. However, because the data were not collected to calibrate or validate accelerometer data, a large proportion of participants were excluded from analysis because they did not have sufficient accelerometer data. Furthermore, because these data were collected from the general population of adolescent girls, the samples comprised low-risk populations. Restricting the samples to adolescent girls targeted a population at risk for low physical activity and controlled for age and gender. However, participants may not have been exposed to their level of physical activity long enough to observe an influence on disease risk factors. These methods may have worked better with higher disease risk populations (e.g., older adults, overweight). Another weakness of this study is that the data were cross-sectional. Hence, a temporal relationship between physical activity and risk factors cannot be examined. Regardless, the one week of accelerometer data collected may not represent regular physical activity patterns. Furthermore, the accelerometer and self-report data did not reflect the same time frames. Participants in NHANES completed the 30-day physical activity recall before they received the accelerometer. Most participants in TAAG completed the 3DPAR when they returned the accelerometer. Hence, three of the accelerometer days may correspond to the same days as the 3DPAR, but the NHANES questionnaire does not overlap at all with accelerometer timing. Furthermore, compared with the 3DPAR, the longer time frame of the NHANES questionnaire may better represent regular physical activity patterns, but it is more subject to reporting errors. Examining this level of agreement was beyond the scope of this study.

In summary, new cut points were identified using two different methods in two separate studies. This study simultaneously evaluated the agreement of these new cut points against two comparison measures: self-report questionnaires and a previously developed cut point. The new cut points tended to be in poor agreement with comparison measures, particularly with the self-report measures, for both the continuous and dichotomous physical activity variables. From this study, it is not clear which, if any, are valid measures of physical activity as none of them can confidently be considered the gold standard. However, the lack of agreement with the comparison measures implies that the new cut points are different from the comparison measures and, thus, may be measuring a different aspect of physical activity.

REFERENCES

Baranowski, T. (1988). Validity and reliability of self report measures of physical activity: an information-processing perspective. *Res.Q.Exerc.Sport, 59,* 314-327.

Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1,* 307-310.

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med.Sci.Sports Exerc., 32,* 1601-1609.

Chen, K. Y. & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: current and future. *Med.Sci Sports Exerc., 37,* S490-S500.

Kimm, S. Y., Glynn, N. W., Kriska, A. M., Barton, B. A., Kronsberg, S. S., Daniels, S. R. et al. (2002). Decline in physical activity in black girls and white girls during adolescence. *N.Engl.J.Med., 347,* 709-715.

Korn, E. L. & Graubard, B. I. (1991). Epidemiologic studies using surveys: accounting for the sampling design. *Am.J.Public Health, 81,* 1166-1173.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45,* 255-268.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. I., & Hu, J. (2003). Tracking of physical activity and aerobic power from childhood through adolescence. *Med.Sci.Sports Exerc., 35*, 1914-1922.

Montoye, H. J., Kemper, H. C. G., Saris, W. H. M., & Washburn, R. A. (1996). Questionnaires and interviews. In *Measuring Physical Activity and Energy Expenditure* (pp. 42-71). Champaign, IL: Human Kinetics.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA, 300,* 295-305.

Pate, R. R., Ross, R., Dowda, M., Trost, S. G., & Sirard, J. R. (2003). Validation of a 3-day physical activity recall instrument in female youth. *Pediatr.Exerc.Sci., 15,* 257-265.

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

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obes.Res., 10*, 150-157.

Reilly, J. J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J. Y. (2008). Objective measurement of physical activity and sedentary behaviour: review with new data. *Arch.Dis.Child., 93*, 614-619.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B. et al. (2005). Evidence based physical activity for school-age youth. *J.Pediatr.*, *146*, 732-737.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J. et al. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med.Sci.Sports Exerc.*, *36*, 1259-1266.

Troped, P. J., Wiecha, J. L., Fragala, M. S., Matthews, C. E., Finkelstein, D. M., Kim, J. et al. (2007). Reliability and validity of YRBS physical activity items among middle school students. *Med.Sci.Sports Exerc.*, *39,* 416-425.

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

U.S. Department of Health and Human Services (2000). *Healthy People 2010: understanding and improving health and objectives for improving health* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services & U.S. Department of Agriculture (2005). *Dietary guidelines for Americans*. http://www.health.gov/dietaryguidelines/dga2005/document/ [On-line].

Welk, G. J., Wickel, E., Peterson, M., Heitzler, C. D., Fulton, J. E., & Potter, L. D. (2007). Reliability and validity of questions on the Youth Media Campaign llongitudinal Survey. *Med.Sci.Sports Exerc., 39,* 612-621.

Chapter 5: Dissertation Discussion

The ability to effectively study physical activity requires practical measures that are valid and reliable. Accelerometers may fulfill this need, but more research needs to be done to understand how the data accelerometers generate should be interpreted. One method for quantifying physical activity from accelerometer data is to classify the intensity of each epoch using cut points. Cut points are often determined using oxygen consumption as the criterion. This dissertation developed new cut points among free-living adolescent girls using an iterative correlation method and signal detection and ROC curves with cardiovascular and metabolic disease risk factors as criteria. Additionally, this dissertation examined the convergent validity of the new cut points with self-report recall questionnaires and a previously suggested cut point using concordance correlation coefficients, Bland-Altman plots, McNemar's tests, and proportions of agreement. Conclusions from this dissertation are discussed below, followed by strengths and weakness of the study, weaknesses of accelerometers, and issues that were beyond the scope of this dissertation to consider for future studies.

DISSERTATION CONCLUSIONS

New cut points are more strongly associated with disease risk factors. This dissertation demonstrated that cut points exist that are better associated with disease risk factors than previously suggested cut points that were developed using oxygen consumption as the criterion. Because these new cut points used disease risk factors as the criteria, they may be more clinically relevant. Hence, these cut points may be more appropriate to use than oxygen consumption cut points in studies with adolescent girls where the relationship between physical activity and risk factors, such as obesity, is of interest.

Health benefits may be achieved at intensities different from those previously suggested. Differences between the new cut points identified using the iterative correlation method and the previously suggested cut points (Puyau et al., 2002; Treuth et al., 2004) suggest that a different intensity of physical activity may be needed to receive health benefits. Compared with the previously suggested cut points of comparable intensity, the new light and moderate cut points tended to be higher. Hence, the minimum intensity threshold for receiving health benefits from light and moderate physical activity may be higher than the previously suggested cut points. In contrast, the new vigorous cut points tended to be lower than the previously suggested vigorous cut points. Hence, beneficial effects of vigorous physical activity may be observed at a lower intensity than the previously suggested cut points.

Light intensities may be sufficient to receive health benefits. Findings from this dissertation suggest that light intensity activities might be enough to receive health benefits. The new cut points identified using the signal detection/ROC curve method tended to be in the range of light intensity, as defined by the previously suggested cut points. Furthermore, the new cut points always performed at least as well, if not better than the previously suggested moderate or vigorous cut points. Associations between light physical activity and blood pressure, triglycerides, and insulin have been found in the European Youth Heart Study (Ekelund et al., 2007). If light activities are sufficient, this may be helpful for public health interventions as light activities might be more acceptable, especially to sedentary people. However, the strongest intensity-specific correlations tended to be observed for moderate and moderate-to-vigorous cut points. This agrees with the current physical activity recommendations, which emphasize performing moderate-to-vigorous intensity activities (Strong et al., 2005; U.S. Department of Health and Human Services et al., 2005; U.S. Department of Health and Human Services, 2000; Physical Activity Guidelines Advisory Committee, 2008).

Intensity of cut points differs by method of development. The intensity of the cut points identified may depend on the development method used. The new cut points identified using the signal detection/ROC curve method tended to be lower than the cut points identified using the iterative correlation method. Differences between the two development methods may explain why different cut points were identified. The signal detection/ROC curve method tested a wide range of values, independent of previously suggested cut points, while the correlation method tested a limited number of cut points, using previously suggested cut points that were developed using oxygen consumption as criterion as starting points. Additionally, the physical activity variable used in the correlation method was a continuous variable of minutes of physical activity, while the signal detection/ROC curve method used a dichotomous variable based on whether physical activity recommendations were met.

Accelerometers may not be good at classifying disease risk. While several new cut points were identified, accelerometers may not be good at classifying cardiovascular and metabolic disease risk. The majority of relationships tested had low areas under the ROC curve. The areas under the ROC curve for the new cut points ranged from 0.51 to 0.64. The correlations between minutes of physical activity with disease risk factors were weak. The average correlations for the new cut points were less than 0.3, and the intensity-specific correlations were less than 0.4. The new cut points also tended to have sizeable misclassifications, ranging from 12 to 41%. The relationship between physical activity measured by accelerometers and disease risk factors varied greatly. Many of the potential cut points identified in the development samples were no longer associated with disease risk factors in the evaluation samples.

New cut points did not agree well with comparison measures. Agreement between the new cut points with the self-report questionnaires and the previously suggested cut points tended to be poor. For minutes of moderate-to-vigorous physical activity, all of the concordance correlation coefficients with the TAAG 3DPAR were less than 0.10 and the 95% confidence intervals for all the concordance correlation coefficients with the NHANES questionnaire included zero. For recommendation classification, only one cut point was not significantly different from the NHANES questionnaire and all cut points were significantly different from the TAAG 3DPAR. The closer a new cut point was to the previously suggested cut point, the better the agreement between them. However, small differences between cut points made notable differences in agreement for recommendation classification. For example, 1550 counts/30 sec was significantly different from the Treuth cut point of 1500 counts/30 sec, and the positive agreement dropped to 0.93. High overall agreement does not necessarily reflect both positive and negative agreement. For example, 1950 counts/30 sec had an overall agreement of 0.97 with the Treuth cut point of 1500 counts/30 sec, but the positive agreement was only 0.19. This may be important to consider when selecting a cut point depending on whether higher sensitivity or specificity is desired.

DISSERTATION STRENGTHS AND WEAKNESSES

Several strengths and limitations of this dissertation should be acknowledged. One strength of this dissertation was that the data used were from two large field studies. Hence, data were obtained from a substantial number of adolescent girls and free-living physical activities were assessed. However, because the data were not collected to calibrate or validate accelerometer data, a large proportion of participants were excluded from analyses because they did not have sufficient accelerometer data. Furthermore, because these data were collected from general populations of adolescent girls, the samples comprised low-risk populations. Restricting the samples to adolescent girls targeted a population at risk for low physical activity and controlled for age and gender. However, these methods may have worked better with higher risk populations (e.g., older, overweight).

Data were examined by population subgroups, which may experience different effects of physical activity on disease risk factors. The new cut points varied some across population subgroups and risk factors. Similarly, new cut points were not identified for each population subgroup-risk factor combination. Hence, the physical activity intensity needed to confer health benefits might vary depending on the outcome and population of interest. Therefore, future studies and interventions that use accelerometer cut points specific to the outcome and population of interest might observe stronger relationships and physical activity and these relationships might be better understood.

Another weakness of this study is that the data were cross-sectional. Hence, a temporal relationship between physical activity and risk factors cannot be established. That is, whether the level of physical activity caused the risk factor level could not be determined. Moreover, the one week of accelerometer data may not represent regular physical activity patterns, which would make it difficult to find a relationship between physical activity and disease risk factors even if one existed.

WEAKNESSES OF ACCELEROMETERS

Accelerometers are often touted as the best available measure of physical activity for field studies. However, it is important to acknowledge some of its weaknesses.

Accelerometers misclassify some activities. Accelerometers are usually worn on the hip to measure locomotor activities, such as walking. These comprise the majority of physical activity for most people. However, accelerometers fail to detect other types of activities, such as cycling or weight lifting. Additionally, accelerometers are not able to pick up the additional work associated with moving uphill, up stairs, or on soft terrain. Furthermore, activities comprised predominately of upper body or arm movements are not captured well by an accelerometer worn on the hip. Hence, activities of different energy expenditures may have similar accelerometer counts or activities of different accelerometer counts may be of similar intensity.

Accelerometers can only capture activities that occur while they are worn. Accelerometers cannot measure physical activity if they are not worn. Hence, they are not practical to use long term (e.g., year). Furthermore, they cannot be used to assess past behavior. Accelerometers also miss activities if they are removed to perform the activity (e.g., swimming, contact sports). They would also miss activities if the participant forgets to wear it or chooses not to wear it (i.e., non-compliance).

Accelerometers do not measure all important aspects of physical activity. Accelerometers cannot assess some aspects of physical activity that may be of interest.

While frequency, intensity, and duration are dimensions of physical activity that can be assessed with accelerometers, type of activity cannot. Furthermore, accelerometers cannot assess the physical activity environment, such as the physical and social context.

ISSUES FOR FUTURE STUDIES

Several issues were not addressed in this dissertation that should be considered in future studies.

Identifying an appropriate gold standard. It was not unexpected that the new cut points were in poor agreement with the self-report recall questionnaires and the previously suggested cut points in this dissertation. This lack of agreement suggests that they are measuring different aspects of physical activity. However, the comparison measures may not be appropriate gold standards to evaluate validity. A valid measure measures what it is intended to measure (i.e., the truth). In calibration studies such as Treuth et al. (2004) and Puyau et al. (2002), the behaviors performed were controlled. Hence, the truth was known. The truth was not known in this dissertation. The questionnaires were used as comparison measures because they are accepted measures of physical activity administered to the same population. Previously suggested cut points were used as comparison measures because they were developed in similar populations. However, the observed agreement might be better with a more appropriate comparison measure. New technologies that incorporate multiple measures, such as accelerometry, physiologic measures, global positioning system (GPS), and cell phones, may be more appropriate.

Determining the best way to reduce accelerometer data. Masse et al. (2005) examined issues to consider when reducing accelerometer data including the number of consecutive zeros that signify non-wear time, the number of minutes of wear time required for a valid day, counts that are considered extreme values, the number of days needed for valid estimates, and defining bouts of physical activity. In this dissertation, less stringent criteria tended to be used (e.g., 60 minutes of consecutive zero counts as non-wear time rather than 20 minutes) because more stringent criteria tend to reduce the wear time and sample size (Masse et al., 2005). Additionally, the minimum length for bouts of activity, which may be important for observing health effects, was not examined in this dissertation. Using different criteria for reducing accelerometer data may have affected which cut points were identified.

Comparing individual- and group-level agreement. Individual-level agreement was poor in this dissertation. However, group-level agreement might be better if the individual differences average out for the group. Acceptable group-level agreement may be sufficient for some studies. For example, a prevalence study may be interested in the level of physical activity in the population and not be concerned with the physical activity level of individuals within that population. From this dissertation, it is not clear how well these cut points would perform at the group level.

Examining the effect of other important factors. This dissertation used the relationship between physical activity and disease risk factors to identify new cut points. The analyses controlled for demographic characteristics, such as age and race/ethnicity, by examining this relationship within population subgroups. However, they did not control for other factors, such as diet and genetics, that may also affect disease risk factors. For example, if a participant had a high fat diet, a poor lipid profile may be observed due to the diet rather than the physical activity level. Hence, future studies should consider other important factors in addition to demographic characteristics.

In conclusion, this dissertation used two methods to develop new accelerometer cut points among free-living adolescent girls. Cardiovascular and metabolic disease risk factors were used as criteria rather than oxygen consumption. The new cut points suggest that cut points exist that are better associated with disease risk factors than previously suggested cut points developed using oxygen consumption. However, the new cut points tended to be in poor agreement with comparison measures and accelerometers may not be good at classifying cardiovascular and metabolic disease risk. Future studies should explore other methods for identifying cut points as well as address issues, such as how to reduce accelerometer data and comparing individual- and group-level agreement, to better understand the data that accelerometers produce.

Appendix A. Decisions made regarding methods

Decision	**Rationale**
Define non-wear time as 60 minutes or more of consecutive zero counts	Sixty minutes was used to maximize the wear time and sample size (Masse et al., 2005).
Require at least 10 hours of wear for a day to be valid	In a review by Masse et al. (2005), 10 hours/day was the cut point used most often to determine whether a day was included in the analysis.
Require at least 2 valid weekdays and 1 valid weekend days to include in analyses	The minimum number of days used most often in analyses was 3 days in a review by Masse et al. (2005). Additionally, physical activity time tends to differ between weekdays and weekend days (Trost, Pate, Freedson, Sallis, & Taylor, 2000)

Decision	Rationale
Consider counts of 24000/min (12000 counts/30 sec) or greater as extreme values	TAAG chose a maximum possible value of 12,000 counts/30 sec after consulting with Russ Pate. Additionally, assuming a maximum possible MET value of 22 METs and solving the Treuth et al. (2004) MET prediction equation results in approximately 12000 counts/30 sec (Kim Ring, personal communication, 2/26/07).
Freedson et al. (1997) and Trost et al. (1998) cut points not used	These suggest prediction equations that include age (Freedson) and body mass (Trost) as variables to estimate METs (Freedson) and kcal/min (Trost), while Puyau et al. (2002) and Treuth et al. (2004) suggest cut points than define intensity levels.

Decision	Rationale
Use (1) average correlations greater than 0.100 in magnitude and 0.050 stronger than that using either the Treuth or the Puyau cut points and (2) misclassification more than 5.0% lower than that using either the Treuth or the Puyau cut points and area under the receiver operating characteristic (ROC) curve greater than 0.50 as criteria for new cut points	Correlations, misclassification percentages, and area under the ROC curves were determined in the development sample and examined. The criteria values for new cut points identified a practical number of new cut points and were considered reasonable criteria.
Use highest cut point when more than one cut point has same misclassification	Higher cut points err on the conservative side for estimating physical activity.
Define 30-minute blocks of physical activity reported using 3-day physical activity recall (3DPAR) as 10 minutes of activity	Physical activity estimates were determined using 10, 15, and 20 minutes. Ten minutes was determined to produce the most reasonable estimates.

Appendix B. Continuous participant characteristics by inclusion in analysis by dataset

	All girls		Included[1]				
			Yes		No		
Characteristic	N	Mean (SD)	N	Mean (SE)	N	Mean (SE)	p-value[2]
NHANES	**549**	**n/a**	**333**	**n/a**	**118**	**n/a**	**n/a**
Age (year)	537	14.0 (1.11)	332	14.0 (0.06)	205	14.0 (0.08)	0.74
BMI percentile[3]	531	68.2 (26.14)	327	66.5 (1.49)	204	71.0 (1.73)	0.05
Waist circumference percentile[3]	515	62.3 (23.75)	318	60.8 (1.36)	197	64.8 (1.63)	0.06
Percent body fat	498	29.3 (8.33)	308	28.7 (0.48)	190	30.1 (0.60)	0.08
SBP percentile[3]	521	34.2 (26.81)	323	33.2 (1.45)	198	35.7 (2.00)	0.31
DBP percentile[3]	470	36.8 (25.65)	291	37.4 (1.52)	179	35.8 (1.89)	0.49
Total cholesterol (mg/dl)	485	160.3 (28.22)	308	159.8 (1.61)	177	161.2 (2.11)	0.59
HDL-cholesterol (mg/dl)	485	55.3 (11.87)	308	54.7 (0.68)	177	56.2 (0.87)	0.19
TAAG	**7910**	**n/a**	**4696**	**n/a**	**3214**	**n/a**	**n/a**
Age (year)	7466	14.0 (0.51)	4687	14.0 (0.01)	2779	14.0 (0.01)	<0.0001
BMI percentile[3]	7465	66.7 (27.36)	4687	66.8 (0.40)	2778	66.5 (0.52)	0.67
Percent body fat	7463	31.4 (8.41)	4686	31.4 (0.12)	2777	31.4 (0.16)	0.98
Cardiorespiratory fitness[4]	1235	11.9 (3.77)	692	11.9 (0.14)	543	11.8 (0.16)	0.41

NHANES: National Health and Nutrition Examination Survey. TAAG: Trial of Activity for Adolescent Girls; BMI: body mass index; SBP: systolic blood pressure; DBP: diastolic blood pressure; HDL: high-density lipoprotein;

1 Girls were included in analyses if they had at least two valid (i.e., at least 10 hours of wear time) weekdays and one valid weekend day of accelerometer data.

2 P-values are from t-tests.

3 BMI percentiles are age- and sex-specific (CDC, 2000). Waist circumference percentiles are age-, sex-, and ethnicity-specific (Fernandez et al., 2004). Blood pressure percentiles are age-, sex-, and height-specific (National High Blood Pressure Education Program, 2005).

4 Cardiorespiratory fitness is defined as the estimated power output (watts/kg body weight) at a heart rate of 170 beats per minute predicted from a multi-stage cycle ergometry test.

Appendix C. Categorical participant characteristics by inclusion in analysis by dataset

Characteristic	n (%) All girls	Included[1] Yes	Included[1] No	p-value[2]
Total in NHANES	**549 (100.0)**	**333 (100.0)**	**118 (100.0)**	**n/a**
Age (year)				
12	132 (24.0)	77 (23.1)	55 (25.5)	0.74
13	150 (27.3)	94 (28.2)	56 (25.9)	
14	157 (28.6)	92 (27.6)	65 (30.1)	
15	110 (20.0)	70 (21.0)	40 (18.5)	
Race/Ethnicity				
White, non-Hispanic	140 (25.5)	85 (25.5)	55 (25.5)	0.21
Hispanic	192 (35.0)	126 (37.8)	66 (30.6)	
Black, non-Hispanic	192 (35.0)	110 (33.0)	82 (38.0)	
Other and multiple race	25 (4.6)	12 (3.6)	13 (6.0)	
BMI percentile[3]				
<85th percentile	348 (65.5)	218 (66.7)	130 (63.7)	0.49
≥85th percentile	183 (34.5)	109 (33.3)	74 (36.3)	
Waist circumference percentile[3]				
<75th percentile	327 (63.5)	207 (65.1)	120 (60.9)	0.34
≥75th percentile	188 (36.5)	111 (34.9)	77 (39.1)	
Percent body fat				
≤32%	303 (60.8)	193 (62.7)	110 (57.9)	0.29
>32%	195 (39.2)	115 (37.3)	80 (42.1)	
Blood pressure percentile[3]				
<90th SBP and DBP percentile	434 (92.3)	267 (91.8)	167 (93.3)	0.54
≥90th SBP or DBP percentile	36 (7.7)	24 (8.2)	12 (6.7)	
Total cholesterol				
<170 mg/dl	325 (67.0)	209 (67.9)	116 (65.5)	0.60
≥170 mg/dl	160 (33.0)	99 (32.1)	61 (34.5)	
HDL-cholesterol				
>50 mg/dl	311 (64.1)	191 (62.0)	120 (67.8)	0.20
≤50 mg/dl	174 (35.9)	117 (38.0)	57 (32.2)	
Total in TAAG	**7910 (100.0)**	**4696 (100.0)**	**3214 (100.0)**	
Age (year)				
12	16 (0.2)	9 (0.2)	7 (0.3)	<0.0001
13	4234 (56.7)	2743 (58.5)	1491 (53.7)	
14	2890 (38.7)	1785 (38.1)	1105 (39.8)	
15	283 (3.8)	131 (2.8)	152 (5.5)	
16	43 (0.6)	19 (0.4)	24 (0.9)	
Race/Ethnicity				
White, non-Hispanic	3520 (46.3)	2190 (46.7)	1330 (45.6)	<0.0001
Hispanic	1637 (21.5)	1017 (21.7)	620 (21.3)	
Black, non-Hispanic	1540 (20.3)	866 (18.5)	674 (23.1)	
Asian	379 (5.0)	286 (6.1)	93 (3.2)	
American Indian and multiple race	525 (6.9)	327 (7.0)	198 (6.8)	
BMI percentile[3]				
<85th percentile	4896 (65.6)	3044 (64.9)	1852 (66.7)	0.13
≥85th percentile	2569 (34.4)	1643 (35.1)	926 (33.3)	

Characteristic	n (%) All girls	Included[1] Yes	Included[1] No	p-value[2]
Percent body fat				
≤32%	3989 (53.5)	2511 (53.6)	1478 (53.2)	<0.0001
>32%	3474 (46.6)	2175 (46.4)	1299 (46.8)	

BMI: body mass index; SBP: systolic blood pressure; DBP: diastolic blood pressure; HDL: high-density lipoprotein

[1] Girls were included in analyses if they had at least two valid (i.e., at least 10 hours of wear time) weekdays and one valid weekend day of accelerometer data.

[2] P-values are from chi-squared tests, except if the expected value for any one cell is less than 5. Then p-values are from Fisher's exact tests.

[3] BMI percentiles are age- and sex-specific (CDC, 2000). Waist circumference percentiles are age-, sex-, and ethnicity-specific (Fernandez et al., 2004). Blood pressure percentiles are age-, sex-, and height-specific (National High Blood Pressure Education Program, 2005).

Appendix D. **Spearman's rank correlations between disease risk factor and average daily minutes of physical activity determined using previously suggested and potential new accelerometer cut points by population subgroup and sample in NHANES**

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Body mass index (BMI) percentile						
All girls						
Development sample (N = 245)						
Treuth[1]	0.055	−0.040	−0.037	0.043	−0.049	−0.006
Puyau[1]	0.030	−0.048	−0.064	0.009	−0.058	−0.026
New (not determined)[2]	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 82)						
Treuth	-0.028	-0.169	-0.281	-0.055	-0.204	-0.147
Puyau	-0.164	-0.179	-0.200	-0.171	-0.193	-0.181
New (not evaluated)[2]	n/a	n/a	n/a	n/a	n/a	n/a
12−13 y						
Development sample (N = 126)						
Treuth	0.104	−0.132	−0.160	0.076	−0.155	−0.053
Puyau	−0.029	−0.146	−0.140	−0.064	−0.166	−0.109
New (1900, 4300, 10000)	−0.145	−0.188	−0.191	−0.159	−0.191	−0.175
Evaluation sample (N = 44)						
Treuth	−0.145	−0.260	−0.358	−0.162	−0.299	−0.245
Puyau	−0.107	−0.249	−0.129	−0.132	−0.259	−0.175
New (1900, 4300, 10000)	−0.254	−0.315	−0.156	−0.274	−0.324	−0.264
14−15 y						
Development sample (N = 119)						
Treuth	−0.039	0.037	0.073	−0.041	0.041	0.014
Puyau	0.059	0.037	−0.005	0.045	0.033	0.034
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 38)						
Treuth	0.121	−0.082	−0.210	0.100	−0.100	−0.034
Puyau	−0.258	−0.104	−0.290	−0.224	−0.129	−0.201
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	
White						
Development sample (N = 63)						
Treuth	−0.154	−0.158	−0.129	−0.162	−0.173	−0.155
Puyau	−0.163	−0.163	−0.114	−0.184	−0.169	−0.158
New (1300, 2300, 6200)	−0.185	−0.202	−0.166	−0.211	−0.220	−0.197
Evaluation sample (N = 21)						
Treuth	−0.095	−0.070	−0.202	−0.060	−0.052	−0.096
Puyau	−0.133	−0.103	−0.254	−0.078	−0.070	−0.127
New (1300, 2300, 6200)	−0.131	−0.068	−0.274	−0.081	−0.065	−0.124
Hispanic						
Development sample (N = 96)						
Treuth	−0.002	−0.071	−0.031	−0.024	−0.068	−0.039
Puyau	0.045	−0.053	−0.077	0.007	−0.063	−0.028
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 28)						
Treuth	0.016	−0.135	−0.121	0.030	−0.160	−0.074
Puyau	0.043	−0.165	0.006	0.048	−0.152	−0.044
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Black						
Development sample (N = 76)						
Treuth	0.244	0.095	0.008	0.229	0.073	0.130
Puyau	0.128	0.055	−0.077	0.118	0.038	0.052
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 31)						
Treuth	−0.047	−0.360	−0.449	−0.129	−0.379	−0.273
Puyau	−0.383	−0.347	−0.313	−0.436	−0.378	−0.371
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Waist circumference percentile						
All girls						
Development sample (N = 238)						
Treuth	0.123	0.000	−0.001	0.117	−0.016	0.045
Puyau	0.076	−0.010	−0.021	0.059	−0.024	0.016
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
Evaluation sample (N = 80)						
Treuth	0.049	-0.137	-0.168	0.036	-0.160	-0.076
Puyau	-0.064	-0.151	-0.072	-0.078	-0.148	-0.102
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
12−13 y						
Development sample (N = 123)						
Treuth	0.180	−0.115	−0.107	0.161	−0.134	−0.003
Puyau	−0.007	−0.124	−0.066	−0.036	−0.144	−0.075
New (1900, 4000, 5000)	−0.122	−0.166	−0.127	−0.140	−0.172	−0.145
Evaluation sample (N = 44)						
Treuth	−0.136	−0.217	−0.265	−0.142	−0.244	−0.201
Puyau	−0.060	−0.214	−0.044	−0.078	−0.202	−0.120
New (1900, 4000, 5000)	−0.202	−0.199	−0.280	−0.221	−0.238	−0.228
14−15 y						
Development sample (N = 115)						
Treuth	0.030	0.097	0.104	0.033	0.092	0.071
Puyau	0.139	0.095	0.026	0.127	0.086	0.095
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 36)						
Treuth	0.255	0.003	−0.052	0.251	−0.004	0.091
Puyau	−0.091	−0.027	−0.118	−0.091	−0.048	−0.075
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
White						
Development sample (N = 63)						
Treuth	−0.156	−0.086	−0.049	−0.154	−0.099	−0.109
Puyau	−0.130	−0.087	−0.064	−0.140	−0.096	−0.103
New (500, 1800, 2800)	−0.120	−0.178	−0.113	−0.152	−0.153	−0.143
Evaluation sample (N = 21)						
Treuth	0.096	−0.266	−0.263	0.101	−0.225	−0.111
Puyau	−0.026	−0.255	−0.206	−0.029	−0.229	−0.149
New (500, 1800, 2800)	0.090	−0.134	−0.205	0.055	−0.208	−0.081

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Hispanic						
Development sample (N = 90)						
Treuth	0.047	−0.035	−0.026	0.030	−0.054	−0.008
Puyau	0.106	−0.030	−0.053	0.063	−0.051	0.007
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 27)						
Treuth	0.069	−0.132	−0.107	0.084	−0.150	−0.047
Puyau	0.088	−0.157	−0.018	0.082	−0.147	−0.031
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Black						
Development sample (N = 75)						
Treuth	0.327	0.120	0.016	0.311	0.094	0.174
Puyau	0.204	0.076	−0.014	0.194	0.059	0.104
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 30)						
Treuth	−0.016	−0.151	−0.185	−0.079	−0.158	−0.118
Puyau	−0.241	−0.144	−0.092	−0.258	−0.140	−0.175
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Percent body fat						
All girls						
Development sample (N = 227)						
Treuth	0.016	−0.035	−0.069	0.007	−0.053	−0.027
Puyau	−0.026	−0.054	−0.094	−0.038	−0.064	−0.055
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 81)						
Treuth	-0.056	-0.207	-0.270	-0.089	-0.233	-0.171
Puyau	-0.205	-0.219	-0.149	-0.230	-0.221	-0.205
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
12−13 y						
Development sample (N = 117)						
Treuth	0.084	−0.070	−0.154	0.068	−0.109	−0.036
Puyau	−0.048	−0.103	−0.133	−0.067	−0.127	−0.096
New (1600, 4100, 9500)	−0.110	−0.165	−0.197	−0.126	−0.179	−0.155

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity: Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
Evaluation sample (N = 44)						
Treuth	−0.183	−0.295	−0.359	−0.208	−0.329	−0.275
Puyau	−0.136	−0.295	−0.135	−0.185	−0.289	−0.208
New (1600, 4100, 9500)	−0.245	−0.342	−0.091	−0.276	−0.331	−0.257
14−15 y						
Development sample (N = 110)						
Treuth	−0.010	0.037	0.074	−0.013	0.042	0.026
Puyau	0.061	0.033	0.022	0.045	0.037	0.040
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 37)						
Treuth	0.129	−0.110	−0.145	0.110	−0.113	−0.026
Puyau	−0.312	−0.127	−0.144	−0.276	−0.144	−0.200
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
White						
Development sample (N = 58)						
Treuth	−0.133	−0.168	−0.163	−0.163	−0.190	−0.164
Puyau	−0.138	−0.179	−0.139	−0.167	−0.182	−0.161
New (1300, 2400, 7500)	−0.164	−0.207	−0.198	−0.187	−0.216	−0.195
Evaluation sample (N = 21)						
Treuth	−0.143	−0.152	−0.134	−0.107	−0.078	−0.123
Puyau	−0.166	−0.175	−0.093	−0.142	−0.118	−0.139
New (1300, 2400, 7500)	−0.218	−0.092	−0.152	−0.148	−0.078	−0.138
Hispanic						
Development sample (N = 88)						
Treuth	−0.162	−0.082	−0.025	−0.180	−0.073	−0.104
Puyau	−0.067	−0.073	−0.048	−0.094	−0.075	−0.072
New (100, 4300, 9500)	−0.170	−0.116	−0.104	−0.177	−0.109	−0.135
Evaluation sample (N = 28)						
Treuth	0.014	−0.154	−0.151	0.018	−0.176	−0.090
Puyau	0.039	−0.182	−0.012	0.040	−0.168	−0.057
New (100, 4300, 9500)	0.010	−0.165	0.104	0.015	−0.150	−0.037

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	
Black						
Development sample (N = 72)						
Treuth	0.253	0.128	−0.029	0.235	0.086	0.135
Puyau	0.116	0.076	−0.125	0.110	0.053	0.046
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 30)						
Treuth	−0.077	−0.384	−0.346	−0.161	−0.374	−0.269
Puyau	−0.381	−0.344	−0.190	−0.423	−0.352	−0.338
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Systolic blood pressure (SBP) percentile						
All girls						
Development sample (N = 242)						
Treuth	0.124	−0.071	−0.109	0.100	−0.082	−0.008
Puyau	0.013	−0.086	−0.073	−0.016	−0.087	−0.050
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 81)						
Treuth	0.067	0.100	0.011	0.087	0.076	0.068
Puyau	0.092	0.090	0.006	0.096	0.084	0.074
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
12−13 y						
Development sample (N = 123)						
Treuth	0.095	−0.121	−0.159	0.060	−0.137	−0.052
Puyau	−0.063	−0.125	−0.181	−0.097	−0.140	−0.121
New (2100, 3200, 7700)	−0.119	−0.126	−0.213	−0.140	−0.140	−0.148
Evaluation sample (N = 44)						
Treuth	0.052	−0.049	−0.084	0.078	−0.061	−0.013
Puyau	0.039	−0.055	−0.096	−0.003	−0.062	−0.036
New (2100, 3200, 7700)	−0.109	−0.048	−0.113	−0.111	−0.062	−0.088
14−15 y						
Development sample (N = 119)						
Treuth	0.135	−0.039	−0.087	0.116	−0.040	0.017
Puyau	0.088	−0.063	0.006	0.058	−0.053	0.007
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	
Evaluation sample (N = 37)						
Treuth	0.052	0.322	0.144	0.103	0.272	0.179
Puyau	0.184	0.274	0.191	0.243	0.261	0.230
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
White						
Development sample (N = 63)						
Treuth	0.151	−0.350	−0.361	0.077	−0.360	−0.169
Puyau	−0.085	−0.364	−0.282	−0.167	−0.367	−0.253
New (2000, 3200, 4900)	−0.231	−0.374	−0.368	−0.302	−0.367	−0.328
Evaluation sample (N = 21)						
Treuth	0.058	0.177	0.099	0.100	0.168	0.120
Puyau	0.183	0.130	0.075	0.144	0.139	0.134
New (2000, 3200, 4900)	0.242	0.142	0.070	0.151	0.139	0.149
Hispanic						
Development sample (N = 95)						
Treuth	0.097	0.099	0.047	0.081	0.096	0.084
Puyau	0.089	0.080	0.052	0.098	0.085	0.081
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 28)						
Treuth	0.003	0.255	0.099	0.074	0.229	0.132
Puyau	0.207	0.245	0.102	0.258	0.258	0.214
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Black						
Development sample (N = 74)						
Treuth	0.094	0.064	0.031	0.107	0.034	0.066
Puyau	−0.005	0.060	−0.008	−0.005	0.043	0.017
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 30)						
Treuth	0.087	−0.038	−0.108	0.105	−0.077	−0.006
Puyau	−0.058	−0.059	−0.113	−0.092	−0.108	−0.086
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity: Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
< 85th BMI percentile						
Development sample (N = 163)						
Treuth	0.094	0.005	−0.072	0.082	−0.014	0.019
Puyau	0.017	−0.022	−0.020	−0.002	−0.022	−0.010
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 53)						
Treuth	−0.103	0.032	0.056	−0.077	0.023	−0.014
Puyau	−0.061	0.025	0.108	−0.034	0.039	0.016
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
≥ 85th BMI percentile						
Development sample (N = 79)						
Treuth	0.221	−0.166	−0.147	0.200	−0.167	−0.012
Puyau	0.076	−0.158	−0.111	0.050	−0.158	−0.060
New (3000, 4500, 7700)	−0.155	−0.186	−0.200	−0.167	−0.185	−0.179
Evaluation sample (N = 28)						
Treuth	0.368	0.299	0.030	0.386	0.257	0.268
Puyau	0.372	0.310	−0.114	0.374	0.245	0.237
New (3000, 4500, 7700)	0.305	0.182	−0.093	0.257	0.129	0.156
Diastolic blood pressure (DBP) percentile						
All girls						
Development sample (N = 218)						
Treuth	−0.089	−0.091	−0.046	−0.096	−0.073	−0.079
Puyau	−0.132	−0.070	−0.055	−0.127	−0.065	−0.090
New (900, 2200, 7200)	−0.129	−0.108	−0.076	−0.127	−0.099	−0.108
Evaluation sample (N = 73)						
Treuth	-0.103	0.076	0.020	-0.089	0.062	-0.007
Puyau	-0.011	0.039	0.023	0.002	0.048	0.020
New (900, 2200, 7200)	-0.058	0.067	0.026	0.019	0.074	0.026
12−13 y						
Development sample (N = 106)						
Treuth	−0.074	−0.075	−0.126	−0.077	−0.080	−0.086
Puyau	−0.114	−0.076	−0.103	−0.085	−0.077	−0.091
New (600, 4000, 5200)	−0.103	−0.091	−0.126	−0.099	−0.096	−0.103

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Evaluation sample (N = 38)						
Treuth	−0.141	0.070	0.103	−0.147	0.096	−0.004
Puyau	−0.161	0.077	−0.001	−0.107	0.093	−0.019
New (600, 4000, 5200)	−0.124	0.050	0.103	−0.097	0.081	0.003
14−15 y						
Development sample (N = 112)						
Treuth	−0.098	−0.092	0.041	−0.111	−0.062	−0.064
Puyau	−0.141	−0.055	0.007	−0.149	−0.048	−0.077
New (1100, 2200, 3300)	−0.173	−0.095	−0.053	−0.152	−0.090	−0.112
Evaluation sample (N = 35)						
Treuth	−0.053	−0.049	−0.072	−0.073	−0.074	−0.064
Puyau	0.177	−0.098	0.027	0.121	−0.075	0.030
New (1100, 2200, 3300)	0.148	0.165	−0.068	0.107	−0.019	0.067
White						
Development sample (N = 57)						
Treuth	−0.115	−0.186	−0.173	−0.115	−0.169	−0.152
Puyau	−0.192	−0.158	−0.188	−0.175	−0.143	−0.171
New (1600, 2600, 8000)	−0.291	−0.195	−0.217	−0.229	−0.182	−0.223
Evaluation sample (N = 19)						
Treuth	−0.440	0.019	0.216	−0.419	0.091	−0.107
Puyau	−0.384	0.028	0.063	−0.368	0.040	−0.124
New (1600, 2600, 8000)	−0.416	0.009	0.031	−0.298	0.030	−0.129
Hispanic						
Development sample (N = 88)						
Treuth	−0.256	−0.087	0.082	−0.267	−0.028	−0.111
Puyau	−0.205	−0.020	0.062	−0.203	−0.010	−0.075
New (100, 1100, 2100)	−0.220	−0.184	−0.085	−0.268	−0.169	−0.185
Evaluation sample (N = 26)						
Treuth	0.048	−0.184	−0.353	0.006	−0.219	−0.141
Puyau	−0.096	−0.227	−0.379	−0.108	−0.224	−0.207
New (100, 1100, 2100)	0.028	−0.059	−0.158	0.013	−0.095	−0.054

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Black						
Development sample (N = 64)						
Treuth	0.034	0.103	0.047	0.067	0.092	0.068
Puyau	0.008	0.079	−0.019	0.034	0.072	0.035
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 26)						
Treuth	0.119	0.513	0.363	0.236	0.480	0.342
Puyau	0.430	0.480	0.320	0.511	0.483	0.445
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
< 85th BMI percentile						
Development sample (N = 149)						
Treuth	−0.091	−0.099	−0.048	−0.095	−0.076	−0.082
Puyau	−0.132	−0.071	−0.057	−0.125	−0.064	−0.090
New (1100, 2100, 7300)	−0.144	−0.121	−0.081	−0.137	−0.114	−0.119
Evaluation sample (N = 48)						
Treuth	−0.155	−0.010	−0.068	−0.145	−0.025	−0.080
Puyau	−0.012	−0.019	−0.042	−0.009	−0.020	−0.020
New (1100, 2100, 7300)	−0.014	−0.006	−0.042	0.033	−0.006	−0.007
≥ 85th BMI percentile						
Development sample (N = 69)						
Treuth	−0.084	−0.040	−0.018	−0.097	−0.034	−0.055
Puyau	−0.133	−0.038	−0.028	−0.118	−0.035	−0.071
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 25)	25.000					
Treuth	−0.019	0.345	0.322	0.026	0.345	0.204
Puyau	−0.026	0.273	0.064	−0.026	0.309	0.119
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Total cholesterol (mg/dl)						
All girls						
Development sample (N = 233)						
Treuth	0.039	−0.053	−0.016	0.026	−0.055	−0.012
Puyau	0.031	−0.044	−0.036	0.013	−0.050	−0.017
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity: Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
Evaluation sample (N = 75)						
Treuth	0.046	0.019	0.011	0.056	0.021	0.031
Puyau	0.080	0.006	-0.019	0.083	0.003	0.031
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
12−13 y						
Development sample (N = 117)						
Treuth	0.015	−0.016	0.050	0.007	−0.012	0.009
Puyau	0.003	−0.001	0.008	0.010	−0.006	0.003
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 41)						
Treuth	−0.032	0.064	−0.011	−0.029	0.058	0.010
Puyau	0.041	0.039	−0.124	0.066	0.046	0.014
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
14−15 y						
Development sample (N = 116)						
Treuth	0.064	−0.091	−0.084	0.042	−0.105	−0.035
Puyau	0.063	−0.097	−0.080	0.017	−0.106	−0.041
New (2100, 3100, 10200)	−0.032	−0.102	−0.186	−0.082	−0.110	−0.102
Evaluation sample (N = 34)						
Treuth	0.200	−0.101	0.062	0.183	−0.043	0.060
Puyau	0.177	−0.092	0.134	0.148	−0.080	0.057
New (2100, 3100, 10200)	−0.017	−0.071	0.368	−0.007	−0.048	0.045
White						
Development sample (N = 60)						
Treuth	0.009	−0.076	0.041	−0.009	−0.065	−0.020
Puyau	−0.010	−0.076	0.048	−0.018	−0.066	−0.024
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 18)						
Treuth	−0.146	0.117	0.129	−0.152	0.119	0.013
Puyau	−0.179	0.098	−0.008	−0.135	0.121	−0.021
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Hispanic						
Development sample (N = 91)						
Treuth	0.097	−0.124	−0.042	0.063	−0.121	−0.025
Puyau	0.070	−0.114	−0.079	0.017	−0.116	−0.044
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 27)						
Treuth	0.266	−0.194	−0.219	0.219	−0.217	−0.029
Puyau	0.119	−0.261	−0.247	0.065	−0.251	−0.115
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Black						
Development sample (N = 72)						
Treuth	−0.100	−0.012	−0.044	−0.115	−0.034	−0.061
Puyau	−0.104	−0.007	−0.111	−0.094	−0.020	−0.067
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 28)						
Treuth	−0.062	0.101	0.069	0.051	0.081	0.048
Puyau	0.134	0.096	0.103	0.196	0.072	0.120
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
< 85th BMI percentile						
Development sample (N = 154)						
Treuth	0.133	−0.014	0.031	0.118	−0.011	0.051
Puyau	0.141	0.007	0.024	0.115	0.000	0.057
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 50)						
Treuth	−0.005	−0.017	0.051	−0.013	−0.009	0.001
Puyau	0.065	−0.026	0.058	0.071	−0.018	0.030
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
≥ 85th BMI percentile						
Development sample (N = 74)						
Treuth	−0.111	−0.102	−0.070	−0.118	−0.110	−0.102
Puyau	−0.100	−0.107	−0.130	−0.103	−0.112	−0.110
New (1400, 2400, 8700)	−0.123	−0.124	−0.167	−0.119	−0.128	−0.132

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	
Evaluation sample (N = 24)						
Treuth	0.197	0.125	-0.032	0.179	0.105	0.115
Puyau	0.050	0.144	-0.198	0.060	0.055	0.022
New (1400, 2400, 8700)	0.034	0.052	-0.152	0.010	0.034	-0.004
High density lipoprotein (HDL) cholesterol (mg/dl)						
All girls						
Development sample (N = 233)						
Treuth	0.046	0.113	0.095	0.064	0.115	0.087
Puyau	0.113	0.108	0.044	0.129	0.110	0.101
New (1300, 2300, 4300)	0.123	0.120	0.112	0.150	0.128	0.126
Evaluation sample (N = 75)						
Treuth	-0.041	-0.034	-0.031	-0.042	-0.038	-0.037
Puyau	-0.054	-0.035	-0.116	-0.021	-0.034	-0.052
New (1300, 2300, 4300)	-0.056	-0.086	-0.037	-0.041	-0.073	-0.058
12−13 y						
Development sample (N = 117)						
Treuth	0.018	0.215	0.174	0.045	0.216	0.134
Puyau	0.160	0.204	0.118	0.196	0.211	0.178
New (1300, 2300, 4600)	0.210	0.219	0.193	0.250	0.231	0.220
Evaluation sample (N = 41)						
Treuth	−0.215	−0.105	−0.088	−0.224	−0.101	−0.147
Puyau	−0.186	−0.118	−0.165	−0.130	−0.128	−0.146
New (1300, 2300, 4600)	−0.192	−0.149	−0.072	−0.137	−0.127	−0.135
14−15 y						
Development sample (N = 116)						
Treuth	0.112	0.036	0.031	0.118	0.041	0.068
Puyau	0.104	0.035	0.005	0.105	0.039	0.058
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 34)						
Treuth	0.338	0.027	0.062	0.323	0.038	0.158
Puyau	0.205	0.037	−0.045	0.201	0.039	0.087
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	
White						
Development sample (N = 60)						
Treuth	0.093	0.166	0.236	0.097	0.191	0.156
Puyau	0.229	0.168	0.203	0.250	0.170	0.204
New (1300, 2500, 5200)	0.233	0.185	0.236	0.253	0.209	0.223
Evaluation sample (N = 18)						
Treuth	−0.030	−0.029	−0.009	−0.034	−0.093	−0.039
Puyau	0.048	−0.021	−0.058	0.090	−0.034	0.005
New (1300, 2500, 5200)	0.048	−0.130	−0.009	−0.050	−0.145	−0.057
Hispanic						
Development sample (N = 91)						
Treuth	0.042	0.087	0.170	0.055	0.120	0.095
Puyau	0.065	0.095	0.072	0.082	0.120	0.087
New (1300, 2300, 5300)	0.106	0.104	0.173	0.142	0.127	0.130
Evaluation sample (N = 27)						
Treuth	−0.157	−0.157	−0.044	−0.163	−0.119	−0.128
Puyau	−0.113	−0.117	−0.134	−0.064	−0.100	−0.105
New (1300, 2300, 5300)	−0.135	−0.184	−0.010	−0.089	−0.100	−0.104
Black						
Development sample (N = 72)						
Treuth	0.050	0.082	−0.103	0.037	0.039	0.021
Puyau	0.038	0.057	−0.128	0.066	0.039	0.014
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 28)						
Treuth	0.012	0.098	−0.065	0.070	0.064	0.036
Puyau	−0.067	0.063	−0.187	−0.047	0.044	−0.039
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
< 85th BMI percentile						
Development sample (N = 154)						
Treuth	0.078	0.050	0.030	0.081	0.048	0.057
Puyau	0.104	0.050	−0.023	0.100	0.048	0.056
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	
Evaluation sample (N = 50)						
Treuth	−0.191	−0.153	−0.138	−0.195	−0.150	−0.165
Puyau	−0.170	−0.151	−0.237	−0.126	−0.143	−0.165
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
≥ 85th BMI percentile						
Development sample (N = 74)						
Treuth	0.009	0.142	0.230	0.034	0.158	0.115
Puyau	0.107	0.149	0.153	0.133	0.152	0.139
New (2200, 4300, 5900)	0.126	0.166	0.247	0.160	0.218	0.183
Evaluation sample (N = 24)						
Treuth	0.106	0.052	−0.040	0.098	0.047	0.053
Puyau	−0.006	0.037	0.016	0.029	0.038	0.023
New (2200, 4300, 5900)	−0.091	0.048	−0.026	−0.058	0.049	−0.016

[1] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 101-2999, 3000-5200, 5201, 101, and 3000 counts/minute, respectively. Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800-3199, 3200-8199, 8200, 800, and 3200 counts/min, respectively.

[2] New cut points were not determined if (a) the maximum magnitude of average correlations for valid cut point combinations (i.e., at least 1000 counts/min apart) was less than .100 in the expected direction and (b) less than 2 of intensity categories with a single cut point (i.e., vigorous, light-to-vigorous, moderate-to-vigorous) had a maximum magnitude of at least .11 in the first iteration. New cut points were not evaluated if they were not developed.

Appendix E. Spearman's rank correlations between disease risk factors and average daily minutes of physical activity determined using previously suggested and potential new accelerometer cut points by population subgroup and sample in TAAG

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Body mass index (BMI) percentile						
All girls						
Development sample (N = 3516)						
Treuth[1]	–0.047	–0.085	–0.119	–0.062	–0.099	–0.082
Puyau[1]	0.038	–0.104	–0.082	–0.001	–0.106	–0.051
New (not determined)[2]	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 1171)						
Treuth	–0.091	–0.111	–0.159	–0.111	–0.129	–0.120
Puyau	0.019	–0.136	–0.129	–0.023	–0.140	–0.082
New (not evaluated)[2]	n/a	n/a	n/a	n/a	n/a	n/a
White						
Development sample (N = 1649)						
Treuth	–0.114	–0.083	–0.130	–0.125	–0.098	–0.110
Puyau	–0.022	–0.102	–0.087	–0.050	–0.104	–0.073
New (50, 2100, 2600)	–0.118	–0.113	–0.130	–0.126	–0.124	–0.122
Evaluation sample (N = 541)						
Treuth	–0.131	–0.054	–0.147	–0.143	–0.087	–0.113
Puyau	–0.010	–0.090	–0.143	–0.035	–0.099	–0.075
New (50, 2100, 2600)	–0.131	–0.092	–0.147	–0.144	–0.131	–0.129
Hispanic						
Development sample (N = 770)						
Treuth	–0.006	–0.035	–0.065	–0.014	–0.048	–0.033
Puyau	0.079	–0.054	–0.031	0.043	–0.054	–0.003
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 247)						
Treuth	–0.081	0.030	–0.066	–0.074	0.018	–0.034
Puyau	0.071	0.004	–0.036	0.060	0.003	0.020
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity: Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
Black						
Development sample (N = 649)						
Treuth	0.033	–0.070	–0.085	0.024	–0.081	–0.036
Puyau	0.066	–0.090	–0.062	0.034	–0.091	–0.029
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 217)						
Treuth	–0.074	–0.248	–0.226	–0.106	–0.261	–0.183
Puyau	–0.041	–0.259	–0.159	–0.091	–0.263	–0.163
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Asian						
Development sample (N = 204)						
Treuth	0.071	0.058	0.059	0.082	0.057	0.065
Puyau	0.129	0.064	0.062	0.116	0.055	0.085
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 82)						
Treuth	0.088	–0.056	0.017	0.082	–0.045	0.017
Puyau	0.194	–0.067	0.080	0.138	–0.057	0.058
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Percent body fat						
All girls						
Development sample (N = 3516)						
Treuth	–0.059	–0.080	–0.125	–0.070	–0.096	–0.086
Puyau	0.017	–0.101	–0.091	–0.015	–0.103	–0.059
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 1170)						
Treuth	–0.118	–0.113	–0.178	–0.136	–0.134	–0.136
Puyau	–0.023	–0.142	–0.139	–0.057	–0.145	–0.101
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
White						
Development sample (N = 1649)						
Treuth	–0.137	–0.104	–0.147	–0.147	–0.118	–0.130
Puyau	–0.046	–0.122	–0.104	–0.074	–0.123	–0.094
New (50, 2100, 2600)	–0.140	–0.129	–0.147	–0.148	–0.140	–0.141

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity: Light	Moderate	Vigorous	Light-to-Vigorous	Moderate-to-Vigorous	Average Correlation
Evaluation sample (N = 541)						
Treuth	–0.186	–0.092	–0.181	–0.197	–0.124	–0.156
Puyau	–0.068	–0.128	–0.162	–0.091	–0.136	–0.117
New (50, 2100, 2600)	–0.185	–0.127	–0.181	–0.199	–0.167	–0.172
Hispanic						
Development sample (N = 770)						
Treuth	–0.021	–0.068	–0.095	–0.034	–0.081	–0.060
Puyau	0.065	–0.087	–0.055	0.023	–0.088	–0.028
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 247)						
Treuth	–0.109	0.004	–0.105	–0.105	–0.011	–0.065
Puyau	0.051	–0.029	–0.049	0.037	–0.028	–0.004
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a
Black						
Development sample (N = 649)						
Treuth	0.041	–0.090	–0.115	0.029	–0.105	–0.048
Puyau	0.069	–0.114	–0.095	0.032	–0.115	–0.044
New (1450, 1950, 2450)	–0.056	–0.123	–0.123	–0.099	–0.130	–0.106
Evaluation sample (N = 217)						
Treuth	–0.066	–0.269	–0.238	–0.103	–0.276	–0.191
Puyau	–0.060	–0.276	–0.169	–0.111	–0.278	–0.179
New (1450, 1950, 2450)	–0.244	–0.299	–0.257	–0.274	–0.288	–0.272
Asian						
Development sample (N = 204)						
Treuth	0.014	0.020	–0.003	0.020	0.009	0.012
Puyau	0.072	0.018	0.001	0.057	0.008	0.031
New (not determined)	n/a	n/a	n/a	n/a	n/a	n/a
Evaluation sample (N = 82)						
Treuth	0.088	–0.029	–0.044	0.084	–0.047	0.010
Puyau	0.176	–0.063	0.013	0.118	–0.064	0.036
New (not evaluated)	n/a	n/a	n/a	n/a	n/a	n/a

	Physical Activity Intensity					
Risk factor / Subgroup / Sample / Cut point	**Light**	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	**Average Correlation**
Physical Work Capacity (watts/kg) at 170 beats per minute (PWC–170)						
All girls						
Development sample (N = 503)						
Treuth	0.092	0.190	0.210	0.120	0.202	0.163
Puyau	0.101	0.204	0.156	0.139	0.206	0.161
New (1450, 2200, 2700)	0.184	0.208	0.212	0.204	0.217	0.205
Evaluation sample (N = 189)						
Treuth	0.174	0.290	0.345	0.225	0.334	0.274
Puyau	0.133	0.331	0.229	0.213	0.340	0.249
New (1450, 2200, 2700)	0.256	0.352	0.337	0.325	0.363	0.326
White						
Development sample (N = 240)						
Treuth	0.099	0.091	0.135	0.116	0.113	0.111
Puyau	0.079	0.107	0.098	0.094	0.117	0.099
New (50, 1000, 2700)	0.090	0.119	0.143	0.115	0.133	0.120
Evaluation sample (N = 85)						
Treuth	0.147	0.196	0.301	0.196	0.263	0.221
Puyau	0.207	0.256	0.213	0.227	0.271	0.235
New (50, 1000, 2700)	0.125	0.193	0.301	0.195	0.223	0.207
Hispanic						
Development sample (N = 106)						
Treuth	0.251	0.267	0.304	0.272	0.280	0.275
Puyau	0.242	0.270	0.244	0.261	0.283	0.260
New (550, 2150, 2650)	0.272	0.329	0.307	0.280	0.329	0.303
Evaluation sample (N = 34)						
Treuth	0.198	0.437	0.532	0.297	0.485	0.390
Puyau	0.255	0.480	0.285	0.379	0.489	0.378
New (550, 2150, 2650)	0.336	0.521	0.534	0.370	0.554	0.463
Black						
Development sample (N = 110)						
Treuth	0.006	0.315	0.227	0.040	0.291	0.176
Puyau	0.108	0.306	0.148	0.165	0.288	0.203
New (1050, 1550, 2050)	0.214	0.327	0.260	0.273	0.297	0.274

Risk factor / Subgroup / Sample / Cut point	Physical Activity Intensity					Average Correlation
	Light	**Moderate**	**Vigorous**	**Light-to-Vigorous**	**Moderate-to-Vigorous**	
Evaluation sample (N = 46)						
Treuth	0.217	0.327	0.230	0.231	0.331	0.267
Puyau	0.155	0.298	0.258	0.201	0.314	0.245
New (1050, 1550, 2050)	0.245	0.365	0.264	0.281	0.332	0.297
< 85th BMI percentile						
Development sample (N = 321)						
Treuth	0.079	0.147	0.152	0.094	0.155	0.125
Puyau	0.134	0.153	0.102	0.147	0.153	0.138
New (1000, 2150, 2700)	0.146	0.179	0.158	0.158	0.172	0.163
Evaluation sample (N = 121)						
Treuth	0.054	0.164	0.269	0.103	0.220	0.162
Puyau	0.130	0.213	0.172	0.154	0.227	0.179
New (1000, 2150, 2700)	0.127	0.245	0.264	0.186	0.271	0.218
≥ 85th BMI percentile						
Development sample (N = 182)						
Treuth	0.152	0.247	0.256	0.194	0.262	0.222
Puyau	0.180	0.267	0.234	0.217	0.274	0.234
New (1000, 1650, 3650)	0.221	0.264	0.265	0.270	0.277	0.259
Evaluation sample (N = 68)						
Treuth	0.288	0.348	0.227	0.329	0.354	0.309
Puyau	0.158	0.353	0.130	0.256	0.356	0.251
New (1000, 1650, 3650)	0.253	0.341	0.124	0.334	0.346	0.280

[1] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 51-1499, 1500-2600, 2601, 51, and 1500 counts/30 sec, respectively. Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800, 3200, and 8200 counts/min, respectively. The cut points used were 400-1599, 1600-4099, and 4100, 400, and 1600 counts/30 sec, respectively.

[2] New cut points were not determined if (a) the maximum magnitude of average correlations for valid cut point combinations (i.e., at least 500 counts/30 sec apart) was less than .100 in the expected direction and (b) less than 2 of intensity categories with a single cut point (i.e., vigorous, light-to-vigorous, moderate-to-vigorous) had a maximum magnitude of at least .11 in the first iteration. New cut points were not evaluated if they were not developed.

Appendix F. **Misclassification for potential new cut points and previously suggested cut points (counts/min) and area under the receiver operating characteristic (ROC) curve by disease risk factor and physical activity recommendation by population subgroup in NHANES development (N = 250) and evaluation (N = 83) samples**

Risk factor / Recommendation / Sample / Subgroup	New cut point		Misclassified[1] [n (%)] Treuth cut point[2] Light	Moderate	Vigorous	Puyau cut point[2] Light	Moderate	Vigorous	Area under ROC curve
At risk for overweight or Overweight									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	79 (32.2)	79 (32.2)	165 (67.4)	165 (67.4)	126 (51.4)	165 (67.4)	165 (67.4)	0.53
12-13 y	**200**	41 (32.5)	41 (32.5)	84 (66.7)	84 (66.7)	66 (52.4)	84 (66.7)	84 (66.7)	0.49
14-15 y	**100**	38 (31.9)	38 (31.9)	81 (68.1)	81 (68.1)	60 (50.4)	81 (68.1)	81 (68.1)	0.58
White	**600**	22 (34.9)	24 (38.1)	39 (61.9)	39 (61.9)	24 (38.1)	39 (61.9)	39 (61.9)	0.68
Hispanic	**200**	26 (27.1)	26 (27.1)	70 (72.9)	70 (72.9)	53 (55.2)	70 (72.9)	70 (72.9)	0.52
Black	**100**	27 (35.5)	27 (35.5)	48 (63.2)	48 (63.2)	44 (57.9)	48 (63.2)	48 (63.2)	0.45
Evaluation sample									
All girls	**100**	29 (35.4)	29 (35.4)	53 (64.6)	53 (64.6)	41 (50.0)	53 (64.6)	53 (64.6)	0.55
12-13 y	**200**	15 (34.1)	15 (34.1)	30 (68.2)	30 (68.2)	25 (56.8)	30 (68.2)	30 (68.2)	0.50
14-15 y	**100**	14 (36.8)	14 (36.8)	23 (60.5)	23 (60.5)	16 (42.1)	23 (60.5)	23 (60.5)	0.60
White	**600**	9 (42.9)	8 (38.1)	13 (61.9)	13 (61.9)	10 (47.6)	13 (61.9)	13 (61.9)	0.57
Hispanic	**100**	11 (39.3)	10 (35.7)	18 (64.3)	18 (64.3)	15 (53.6)	18 (64.3)	18 (64.3)	0.50
Black	**200**	11 (35.5)	11 (35.5)	20 (64.5)	20 (64.5)	14 (45.2)	20 (64.5)	20 (64.5)	0.60
2005 Dietary Guidelines									
Development sample									
All girls	**400**	78 (31.8)	79 (32.2)	165 (67.4)	165 (67.4)	92 (37.6)	165 (67.4)	165 (67.4)	0.56
12-13 y	**400**	41 (32.5)	41 (32.5)	84 (66.7)	84 (66.7)	46 (36.5)	84 (66.7)	84 (66.7)	0.60
14-15 y	**400**	37 (31.1)	38 (31.9)	81 (68.1)	81 (68.1)	46 (38.7)	81 (68.1)	81 (68.1)	0.53
White	**600**	23 (36.5)	24 (38.1)	39 (61.9)	39 (61.9)	25 (39.7)	39 (61.9)	39 (61.9)	0.61
Hispanic	**400**	26 (27.1)	26 (27.1)	70 (72.9)	70 (72.9)	30 (31.3)	70 (72.9)	70 (72.9)	0.57
Black	**400**	26 (34.2)	27 (35.5)	48 (63.2)	48 (63.2)	34 (44.7)	48 (63.2)	48 (63.2)	0.50

Evaluation sample									
All girls	**400**	29 (35.4)	29 (35.4)	53 (64.6)	53 (64.6)	32 (39.0)	53 (64.6)	53 (64.6)	0.58
12-13 y	**400**	14 (31.8)	14 (31.8)	30 (68.2)	30 (68.2)	16 (36.4)	30 (68.2)	30 (68.2)	0.62
14-15 y	**400**	15 (39.5)	15 (39.5)	23 (60.5)	23 (60.5)	16 (42.1)	23 (60.5)	23 (60.5)	0.57
White	**600**	6 (28.6)	8 (38.1)	13 (61.9)	13 (61.9)	7 (33.3)	13 (61.9)	13 (61.9)	0.63
Hispanic	**400**	10 (35.7)	10 (35.7)	18 (64.3)	18 (64.3)	13 (46.4)	18 (64.3)	18 (64.3)	0.41
Black	**400**	11 (35.5)	11 (35.5)	20 (64.5)	20 (64.5)	11 (35.5)	20 (64.5)	20 (64.5)	0.70
Healthy People 2010									
Development sample									
All girls	**600**	78 (31.8)	80 (32.7)	160 (65.3)	165 (67.4)	80 (32.7)	160 (65.3)	165 (67.4)	0.57
12-13 y	**700**	41 (32.5)	42 (33.3)	81 (64.3)	84 (66.7)	42 (33.3)	81 (64.3)	84 (66.7)	0.61
14-15 y	**600**	37 (31.1)	38 (31.9)	79 (66.4)	81 (68.1)	38 (31.9)	79 (66.4)	81 (68.1)	0.54
White	**1600**	23 (36.5)	24 (38.1)	38 (60.3)	39 (61.9)	25 (39.7)	38 (60.3)	39 (61.9)	0.62
Hispanic	**600**	26 (27.1)	26 (27.1)	70 (72.9)	70 (72.9)	27 (28.1)	70 (72.9)	70 (72.9)	0.58
Black	**800**	26 (34.2)	28 (36.8)	46 (60.5)	48 (63.2)	26 (34.2)	46 (60.5)	48 (63.2)	0.47
Evaluation sample									
All girls	**600**	29 (35.4)	29 (35.4)	51 (62.2)	53 (64.6)	29 (35.4)	53 (64.6)	53 (64.6)	0.56
12-13 y	**700**	14 (31.8)	14 (31.8)	29 (65.9)	30 (68.2)	14 (31.8)	30 (68.2)	30 (68.2)	0.58
14-15 y	**600**	15 (39.5)	15 (39.5)	22 (57.9)	23 (60.5)	15 (39.5)	23 (60.5)	23 (60.5)	0.54
White	**1600**	12 (57.1)	8 (38.1)	13 (61.9)	13 (61.9)	8 (38.1)	13 (61.9)	13 (61.9)	0.47
Hispanic	**800**	10 (35.7)	10 (35.7)	18 (64.3)	18 (64.3)	10 (35.7)	18 (64.3)	18 (64.3)	0.50
Black	**600**	11 (35.5)	11 (35.5)	18 (58.1)	20 (64.5)	11 (35.5)	20 (64.5)	20 (64.5)	0.68
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**400**	79 (32.2)	80 (32.7)	164 (66.9)	165 (67.4)	89 (36.3)	164 (66.9)	165 (67.4)	0.56
12-13 y	**500**	42 (33.3)	42 (33.3)	83 (65.9)	84 (66.7)	45 (35.7)	83 (65.9)	84 (66.7)	0.58
14-15 y	**400**	37 (31.1)	38 (31.9)	81 (68.1)	81 (68.1)	44 (37.0)	81 (68.1)	81 (68.1)	0.56
White	**600**	23 (36.5)	24 (38.1)	38 (60.3)	39 (61.9)	25 (39.7)	38 (60.3)	39 (61.9)	0.62
Hispanic	**500**	26 (27.1)	26 (27.1)	70 (72.9)	70 (72.9)	30 (31.3)	70 (72.9)	70 (72.9)	0.57
Black	**400**	27 (35.5)	28 (36.8)	48 (63.2)	48 (63.2)	31 (40.8)	48 (63.2)	48 (63.2)	0.53

Evaluation sample									
All girls	**400**	29 (35.4)	29 (35.4)	53 (64.6)	53 (64.6)	25 (30.5)	53 (64.6)	53 (64.6)	0.59
12-13 y	**500**	14 (31.8)	14 (31.8)	30 (68.2)	30 (68.2)	13 (29.6)	30 (68.2)	30 (68.2)	0.57
14-15 y	**400**	15 (39.5)	15 (39.5)	23 (60.5)	23 (60.5)	12 (31.6)	23 (60.5)	23 (60.5)	0.62
White	**600**	6 (28.6)	8 (38.1)	13 (61.9)	13 (61.9)	5 (23.8)	13 (61.9)	13 (61.9)	0.63
Hispanic	**400**	10 (35.7)	10 (35.7)	18 (64.3)	18 (64.3)	12 (42.9)	18 (64.3)	18 (64.3)	0.41
Black	**500**	11 (35.5)	11 (35.5)	20 (64.5)	20 (64.5)	7 (22.6)	20 (64.5)	20 (64.5)	0.73
2005 Dietary Guidelines									
Development sample									
All girls	**700**	80 (32.7)	80 (32.7)	158 (64.5)	164 (66.9)	81 (33.1)	161 (65.7)	165 (67.4)	0.57
12-13 y	**800**	42 (33.3)	42 (33.3)	80 (63.5)	83 (65.9)	42 (33.3)	82 (65.1)	84 (66.7)	0.60
14-15 y	**700**	38 (31.9)	38 (31.9)	78 (65.6)	81 (68.1)	39 (32.8)	79 (66.4)	81 (68.1)	0.55
White	**1600**	23 (36.5)	24 (38.1)	36 (57.1)	38 (60.3)	24 (38.1)	38 (60.3)	39 (61.9)	0.61
Hispanic	**700**	26 (27.1)	26 (27.1)	69 (71.9)	70 (72.9)	27 (28.1)	69 (71.9)	70 (72.9)	0.60
Black	**800**	28 (36.8)	28 (36.8)	46 (60.5)	48 (63.2)	28 (36.8)	47 (61.8)	48 (63.2)	0.50
Evaluation sample									
All girls	**700**	29 (35.4)	29 (35.4)	50 (61.0)	53 (64.6)	29 (35.4)	50 (61.0)	53 (64.6)	0.61
12-13 y	**800**	14 (31.8)	14 (31.8)	29 (65.9)	30 (68.2)	14 (31.8)	29 (65.9)	30 (68.2)	0.60
14-15 y	**700**	15 (39.5)	15 (39.5)	21 (55.3)	23 (60.5)	15 (39.5)	21 (55.3)	23 (60.5)	0.59
White	**1600**	9 (42.9)	8 (38.1)	13 (61.9)	13 (61.9)	8 (38.1)	13 (61.9)	13 (61.9)	0.63
Hispanic	**800**	10 (35.7)	10 (35.7)	18 (64.3)	18 (64.3)	10 (35.7)	18 (64.3)	18 (64.3)	0.47
Black	**700**	11 (35.5)	11 (35.5)	17 (54.8)	20 (64.5)	11 (35.5)	17 (54.8)	20 (64.5)	0.72
Healthy People 2010									
Development sample									
All girls	**900**	80 (32.7)	80 (32.7)	147 (60.0)	163 (66.5)	80 (32.7)	152 (62.0)	164 (66.9)	0.58
12-13 y	**1100**	42 (33.3)	42 (33.3)	73 (57.9)	82 (65.1)	42 (33.3)	77 (61.1)	83 (65.9)	0.60
14-15 y	**900**	38 (31.9)	38 (31.9)	74 (62.2)	81 (68.1)	38 (31.9)	75 (63.0)	81 (68.1)	0.56
White	**1900**	22 (34.9)	24 (38.1)	34 (54.0)	38 (60.3)	24 (38.1)	35 (55.6)	39 (61.9)	0.62
Hispanic	**900**	26 (27.1)	26 (27.1)	64 (66.7)	69 (71.9)	26 (27.1)	64 (66.7)	69 (71.9)	0.61
Black	**1100**	28 (36.8)	28 (36.8)	44 (57.9)	48 (63.2)	28 (36.8)	47 (61.8)	48 (63.2)	0.48
Evaluation sample									
All girls	**900**	29 (35.4)	29 (35.4)	44 (53.7)	53 (64.6)	29 (35.4)	47 (57.3)	53 (64.6)	0.59
12-13 y	**1100**	14 (31.8)	14 (31.8)	26 (59.1)	30 (68.2)	14 (31.8)	28 (63.6)	30 (68.2)	0.60
14-15 y	**900**	15 (39.5)	15 (39.5)	18 (47.4)	23 (60.5)	15 (39.5)	19 (50.0)	23 (60.5)	0.58
White	**1900**	8 (38.1)	8 (38.1)	13 (61.9)	13 (61.9)	8 (38.1)	13 (61.9)	13 (61.9)	0.62
Hispanic	**1100**	10 (35.7)	10 (35.7)	15 (53.6)	18 (64.3)	10 (35.7)	17 (60.7)	18 (64.3)	0.48
Black	**900**	11 (35.5)	11 (35.5)	15 (48.4)	20 (64.5)	11 (35.5)	15 (48.4)	20 (64.5)	0.70

Risk factor / Recommendation / Sample / Subgroup		Misclassified[1] [n (%)]							Area under ROC curve
			Treuth cut point[2]			Puyau cut point[2]			
		New cut point	Light	Moderate	Vigorous	Light	Moderate	Vigorous	
Moderate or High central adiposity									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	80 (33.6)	80 (33.6)	160 (67.2)	160 (67.2)	133 (55.9)	160 (67.2)	160 (67.2)	0.52
12-13 y	**200**	46 (37.4)	46 (37.4)	78 (63.4)	78 (63.4)	65 (52.9)	78 (63.4)	78 (63.4)	0.54
14-15 y	**100**	34 (29.6)	34 (29.6)	82 (71.3)	82 (71.3)	68 (59.1)	82 (71.3)	82 (71.3)	0.49
White	**800**	25 (39.7)	31 (49.2)	32 (50.8)	32 (50.8)	25 (39.7)	32 (50.8)	32 (50.8)	0.59
Hispanic	**200**	23 (25.6)	23 (25.6)	67 (74.4)	67 (74.4)	56 (62.2)	67 (74.4)	67 (74.4)	0.59
Black	**100**	23 (30.7)	23 (30.7)	54 (72.0)	54 (72.0)	46 (61.3)	54 (72.0)	54 (72.0)	0.43
Evaluation sample									
All girls	**100**	35 (43.8)	35 (43.8)	47 (58.8)	47 (58.8)	39 (48.8)	47 (58.8)	47 (58.8)	0.52
12-13 y	**200**	21 (47.7)	21 (47.7)	24 (54.6)	24 (54.6)	23 (52.3)	24 (54.6)	24 (54.6)	0.50
14-15 y	**100**	14 (38.9)	14 (38.9)	23 (63.9)	23 (63.9)	16 (44.4)	23 (63.9)	23 (63.9)	0.56
White	**800**	11 (52.4)	11 (52.4)	10 (47.6)	10 (47.6)	11 (52.4)	10 (47.6)	10 (47.6)	0.45
Hispanic	**100**	13 (48.2)	12 (44.4)	15 (55.6)	15 (55.6)	13 (48.2)	15 (55.6)	15 (55.6)	0.52
Black	**200**	11 (36.7)	11 (36.7)	21 (70.0)	21 (70.0)	14 (46.7)	21 (70.0)	21 (70.0)	0.56
2005 Dietary Guidelines									
Development sample									
All girls	**400**	79 (33.2)	79 (33.2)	160 (67.2)	160 (67.2)	97 (40.8)	160 (67.2)	160 (67.2)	0.54
12-13 y	**400**	46 (37.4)	46 (37.4)	78 (63.4)	78 (63.4)	50 (40.7)	78 (63.4)	78 (63.4)	0.59
14-15 y	**400**	33 (28.7)	33 (28.7)	82 (71.3)	82 (71.3)	47 (40.9)	82 (71.3)	82 (71.3)	0.48
White	**1200**	25 (39.7)	31 (49.2)	32 (50.8)	32 (50.8)	30 (47.6)	32 (50.8)	32 (50.8)	0.60
Hispanic	**400**	23 (25.6)	23 (25.6)	67 (74.4)	67 (74.4)	33 (36.7)	67 (74.4)	67 (74.4)	0.55
Black	**400**	22 (29.3)	22 (29.3)	54 (72.0)	54 (72.0)	30 (40.0)	54 (72.0)	54 (72.0)	0.48
Evaluation sample									
All girls	**400**	33 (41.3)	33 (41.3)	47 (58.8)	47 (58.8)	36 (45.0)	47 (58.8)	47 (58.8)	0.59
12-13 y	**400**	20 (45.5)	20 (45.5)	24 (54.6)	24 (54.6)	20 (45.5)	24 (54.6)	24 (54.6)	0.63
14-15 y	**400**	13 (36.1)	13 (36.1)	23 (63.9)	23 (63.9)	16 (44.4)	23 (63.9)	23 (63.9)	0.55
White	**1200**	10 (47.6)	11 (52.4)	10 (47.6)	10 (47.6)	10 (47.6)	10 (47.6)	10 (47.6)	0.59
Hispanic	**400**	12 (44.4)	12 (44.4)	15 (55.6)	15 (55.6)	13 (48.2)	15 (55.6)	15 (55.6)	0.54
Black	**400**	9 (30.0)	9 (30.0)	21 (70.0)	21 (70.0)	11 (36.7)	21 (70.0)	21 (70.0)	0.67

Healthy People 2010									
Development sample									
All girls	**200**	78 (32.8)	78 (32.8)	159 (66.8)	160 (67.2)	83 (34.9)	159 (66.8)	160 (67.2)	0.55
12-13 y	**200**	45 (36.6)	45 (36.6)	77 (62.6)	78 (63.4)	47 (38.2)	77 (62.6)	78 (63.4)	0.60
14-15 y	**600**	33 (28.7)	33 (28.7)	82 (71.3)	82 (71.3)	36 (31.3)	82 (71.3)	82 (71.3)	0.48
White	**1900**	24 (38.1)	31 (49.2)	31 (49.2)	32 (50.8)	32 (50.8)	31 (49.2)	32 (50.8)	0.63
Hispanic	**600**	23 (25.6)	23 (25.6)	69 (76.7)	67 (74.4)	26 (28.9)	69 (76.7)	67 (74.4)	0.55
Black	**200**	21 (28.0)	21 (28.0)	52 (69.3)	54 (72.0)	22 (29.3)	52 (69.3)	54 (72.0)	0.45
Evaluation sample									
All girls	**200**	33 (41.3)	33 (41.3)	45 (56.3)	47 (58.8)	33 (41.3)	47 (58.8)	47 (58.8)	0.58
12-13 y	**200**	20 (45.5)	20 (45.5)	23 (52.3)	24 (54.6)	20 (45.5)	24 (54.6)	24 (54.6)	0.64
14-15 y	**600**	13 (36.1)	13 (36.1)	22 (61.1)	23 (63.9)	13 (36.1)	23 (63.9)	23 (63.9)	0.51
White	**1900**	9 (42.9)	11 (52.4)	10 (47.6)	10 (47.6)	11 (52.4)	10 (47.6)	10 (47.6)	0.50
Hispanic	**200**	12 (44.4)	12 (44.4)	15 (55.6)	15 (55.6)	12 (44.4)	15 (55.6)	15 (55.6)	0.65
Black	**600**	9 (30.0)	9 (30.0)	19 (63.3)	21 (70.0)	9 (30.0)	21 (70.0)	21 (70.0)	0.63
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**400**	78 (32.8)	78 (32.8)	159 (66.8)	160 (67.2)	92 (38.7)	159 (66.8)	160 (67.2)	0.54
12-13 y	**500**	45 (36.6)	45 (36.6)	77 (62.6)	78 (63.4)	49 (39.8)	77 (62.6)	78 (63.4)	0.58
14-15 y	**400**	33 (28.7)	33 (28.7)	82 (71.3)	82 (71.3)	43 (37.4)	82 (71.3)	82 (71.3)	0.48
White	**1200**	22 (34.9)	31 (49.2)	31 (49.2)	32 (50.8)	30 (47.6)	31 (49.2)	32 (50.8)	0.61
Hispanic	**500**	23 (25.6)	23 (25.6)	67 (74.4)	67 (74.4)	31 (34.4)	67 (74.4)	67 (74.4)	0.55
Black	**400**	21 (28.0)	21 (28.0)	54 (72.0)	54 (72.0)	27 (36.0)	54 (72.0)	54 (72.0)	0.49
Evaluation sample									
All girls	**400**	33 (41.3)	33 (41.3)	47 (58.8)	47 (58.8)	29 (36.3)	47 (58.8)	47 (58.8)	0.59
12-13 y	**500**	20 (45.5)	20 (45.5)	24 (54.6)	24 (54.6)	17 (38.6)	24 (54.6)	24 (54.6)	0.59
14-15 y	**400**	13 (36.1)	13 (36.1)	23 (63.9)	23 (63.9)	12 (33.3)	23 (63.9)	23 (63.9)	0.60
White	**1200**	8 (38.1)	11 (52.4)	10 (47.6)	10 (47.6)	8 (38.1)	10 (47.6)	10 (47.6)	0.56
Hispanic	**400**	12 (44.4)	12 (44.4)	15 (55.6)	15 (55.6)	12 (44.4)	15 (55.6)	15 (55.6)	0.55
Black	**500**	9 (30.0)	9 (30.0)	21 (70.0)	21 (70.0)	7 (23.3)	21 (70.0)	21 (70.0)	0.68

2005 Dietary Guidelines									
Development sample									
All girls	**700**	78 (32.8)	78 (32.8)	157 (66.0)	159 (66.8)	80 (33.6)	160 (67.2)	160 (67.2)	0.54
12-13 y	**1000**	45 (36.6)	45 (36.6)	76 (61.8)	77 (62.6)	45 (36.6)	78 (63.4)	78 (63.4)	0.59
14-15 y	**700**	33 (28.7)	33 (28.7)	81 (70.4)	82 (71.3)	35 (30.4)	82 (71.3)	82 (71.3)	0.49
White	**1800**	22 (34.9)	31 (49.2)	29 (46.0)	31 (49.2)	31 (49.2)	31 (49.2)	32 (50.8)	0.61
Hispanic	**700**	23 (25.6)	23 (25.6)	68 (75.6)	67 (74.4)	24 (26.7)	68 (75.6)	67 (74.4)	0.58
Black	**700**	21 (28.0)	21 (28.0)	52 (69.3)	54 (72.0)	22 (29.3)	53 (70.7)	54 (72.0)	0.46
Evaluation sample									
All girls	**700**	33 (41.3)	33 (41.3)	44 (55.0)	47 (58.8)	33 (41.3)	44 (55.0)	47 (58.8)	0.61
12-13 y	**1000**	21 (47.7)	20 (45.5)	23 (52.3)	24 (54.6)	20 (45.5)	23 (52.3)	24 (54.6)	0.62
14-15 y	**700**	13 (36.1)	13 (36.1)	21 (58.3)	23 (63.9)	13 (36.1)	21 (58.3)	23 (63.9)	0.57
White	**1800**	8 (38.1)	11 (52.4)	10 (47.6)	10 (47.6)	11 (52.4)	10 (47.6)	10 (47.6)	0.60
Hispanic	**700**	12 (44.4)	12 (44.4)	15 (55.6)	15 (55.6)	12 (44.4)	15 (55.6)	15 (55.6)	0.63
Black	**700**	9 (30.0)	9 (30.0)	18 (60.0)	21 (70.0)	9 (30.0)	18 (60.0)	21 (70.0)	0.67
Healthy People 2010									
Development sample									
All girls	**900**	78 (32.8)	78 (32.8)	148 (62.2)	160 (67.2)	78 (32.8)	151 (63.5)	161 (67.7)	0.55
12-13 y	**1300**	44 (35.8)	45 (36.6)	71 (57.7)	78 (63.4)	45 (36.6)	73 (59.4)	79 (64.2)	0.58
14-15 y	**900**	33 (28.7)	33 (28.7)	77 (67.0)	82 (71.3)	33 (28.7)	78 (67.8)	82 (71.3)	0.51
White	**2300**	23 (36.5)	31 (49.2)	29 (46.0)	31 (49.2)	31 (49.2)	28 (44.4)	32 (50.8)	0.63
Hispanic	**900**	23 (25.6)	23 (25.6)	63 (70.0)	68 (75.6)	23 (25.6)	63 (70.0)	68 (75.6)	0.58
Black	**1100**	21 (28.0)	21 (28.0)	50 (66.7)	54 (72.0)	21 (28.0)	53 (70.7)	54 (72.0)	0.46
Evaluation sample									
All girls	**900**	33 (41.3)	33 (41.3)	40 (50.0)	47 (58.8)	33 (41.3)	41 (51.3)	47 (58.8)	0.60
12-13 y	**1300**	20 (45.5)	20 (45.5)	22 (50.0)	24 (54.6)	20 (45.5)	22 (50.0)	24 (54.6)	0.63
14-15 y	**900**	13 (36.1)	13 (36.1)	18 (50.0)	23 (63.9)	13 (36.1)	19 (52.8)	23 (63.9)	0.54
White	**2300**	8 (38.1)	11 (52.4)	10 (47.6)	10 (47.6)	11 (52.4)	10 (47.6)	10 (47.6)	0.60
Hispanic	**1100**	12 (44.4)	12 (44.4)	12 (44.4)	15 (55.6)	12 (44.4)	14 (51.9)	15 (55.6)	0.63
Black	**900**	9 (30.0)	9 (30.0)	16 (53.3)	21 (70.0)	9 (30.0)	16 (53.3)	21 (70.0)	0.63

Risk factor / Recommendation / Sample / Subgroup		Misclassified[1] [n (%)]							Area under ROC curve
			Treuth cut point[2]			Puyau cut point[2]			
	New cut point		Light	Moderate	Vigorous	Light	Moderate	Vigorous	
High body fatness									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	86 (37.9)	86 (37.9)	143 (63.0)	143 (63.0)	114 (50.2)	143 (63.0)	143 (63.0)	0.52
12-13 y	**200**	38 (32.5)	38 (32.5)	80 (68.4)	80 (68.4)	56 (47.9)	80 (68.4)	80 (68.4)	0.54
14-15 y	**500**	47 (42.7)	48 (43.6)	63 (57.3)	63 (57.3)	58 (52.7)	63 (57.3)	63 (57.3)	0.49
White	**600**	23 (39.7)	25 (43.1)	33 (56.9)	33 (56.9)	25 (43.1)	33 (56.9)	33 (56.9)	0.59
Hispanic	**600**	32 (36.4)	33 (37.5)	55 (62.5)	55 (62.5)	43 (48.9)	55 (62.5)	55 (62.5)	0.59
Black	**100**	24 (33.3)	24 (33.3)	50 (69.4)	50 (69.4)	42 (58.3)	50 (69.4)	50 (69.4)	0.43
Evaluation sample									
All girls	**100**	31 (38.3)	31 (38.3)	50 (61.7)	50 (61.7)	39 (48.2)	50 (61.7)	50 (61.7)	0.55
12-13 y	**200**	18 (40.9)	18 (40.9)	27 (61.4)	27 (61.4)	24 (54.6)	27 (61.4)	27 (61.4)	0.49
14-15 y	**500**	15 (40.5)	13 (35.1)	23 (62.2)	23 (62.2)	15 (40.5)	23 (62.2)	23 (62.2)	0.63
White	**600**	10 (47.6)	9 (42.9)	12 (57.1)	12 (57.1)	11 (52.4)	12 (57.1)	12 (57.1)	0.52
Hispanic	**100**	13 (46.4)	11 (39.3)	17 (60.7)	17 (60.7)	14 (50.0)	17 (60.7)	17 (60.7)	0.48
Black	**600**	11 (36.7)	11 (36.7)	19 (63.3)	19 (63.3)	12 (40.0)	19 (63.3)	19 (63.3)	0.64
2005 Dietary Guidelines									
Development sample									
All girls	**400**	85 (37.4)	85 (37.4)	143 (63.0)	143 (63.0)	94 (41.4)	143 (63.0)	143 (63.0)	0.54
12-13 y	**400**	38 (32.5)	38 (32.5)	80 (68.4)	80 (68.4)	44 (37.6)	80 (68.4)	80 (68.4)	0.59
14-15 y	**600**	47 (42.7)	47 (42.7)	63 (57.3)	63 (57.3)	50 (45.5)	63 (57.3)	63 (57.3)	0.48
White	**1100**	22 (37.9)	25 (43.1)	33 (56.9)	33 (56.9)	26 (44.8)	33 (56.9)	33 (56.9)	0.60
Hispanic	**800**	31 (35.2)	33 (37.5)	55 (62.5)	55 (62.5)	31 (35.2)	55 (62.5)	55 (62.5)	0.55
Black	**400**	23 (31.9)	23 (31.9)	50 (69.4)	50 (69.4)	32 (44.4)	50 (69.4)	50 (69.4)	0.48
Evaluation sample									
All girls	**400**	31 (38.3)	31 (38.3)	50 (61.7)	50 (61.7)	32 (39.5)	50 (61.7)	50 (61.7)	0.62
12-13 y	**400**	17 (38.6)	17 (38.6)	27 (61.4)	27 (61.4)	17 (38.6)	27 (61.4)	27 (61.4)	0.67
14-15 y	**600**	14 (37.8)	14 (37.8)	23 (62.2)	23 (62.2)	15 (40.5)	23 (62.2)	23 (62.2)	0.57
White	**1100**	8 (38.1)	9 (42.9)	12 (57.1)	12 (57.1)	8 (38.1)	12 (57.1)	12 (57.1)	0.64
Hispanic	**400**	14 (50.0)	11 (39.3)	17 (60.7)	17 (60.7)	14 (50.0)	17 (60.7)	17 (60.7)	0.46
Black	**800**	11 (36.7)	11 (36.7)	19 (63.3)	19 (63.3)	9 (30.0)	19 (63.3)	19 (63.3)	0.75

Healthy People 2010									
Development sample									
All girls	**200**	84 (37.0)	84 (37.0)	142 (62.6)	143 (63.0)	87 (38.3)	142 (62.6)	143 (63.0)	0.55
12-13 y	**200**	37 (31.6)	37 (31.6)	77 (65.8)	80 (68.4)	39 (33.3)	77 (65.8)	80 (68.4)	0.60
14-15 y	**1200**	47 (42.7)	47 (42.7)	65 (59.1)	63 (57.3)	48 (43.6)	65 (59.1)	63 (57.3)	0.48
White	**1600**	20 (34.5)	25 (43.1)	32 (55.2)	33 (56.9)	26 (44.8)	32 (55.2)	33 (56.9)	0.63
Hispanic	**1200**	32 (36.4)	33 (37.5)	55 (62.5)	55 (62.5)	34 (38.6)	55 (62.5)	55 (62.5)	0.55
Black	**200**	22 (30.6)	22 (30.6)	50 (69.4)	50 (69.4)	23 (31.9)	50 (69.4)	50 (69.4)	0.45
Evaluation sample									
All girls	**200**	31 (38.3)	31 (38.3)	48 (59.3)	50 (61.7)	31 (38.3)	50 (61.7)	50 (61.7)	0.61
12-13 y	**200**	17 (38.6)	17 (38.6)	26 (59.1)	27 (61.4)	17 (38.6)	27 (61.4)	27 (61.4)	0.64
14-15 y	**1200**	16 (43.2)	14 (37.8)	22 (59.5)	23 (62.2)	14 (37.8)	23 (62.2)	23 (62.2)	0.56
White	**1600**	11 (52.4)	9 (42.9)	12 (57.1)	12 (57.1)	9 (42.9)	12 (57.1)	12 (57.1)	0.48
Hispanic	**200**	12 (42.9)	11 (39.3)	17 (60.7)	17 (60.7)	11 (39.3)	17 (60.7)	17 (60.7)	0.56
Black	**1200**	11 (36.7)	11 (36.7)	17 (56.7)	19 (63.3)	11 (36.7)	19 (63.3)	19 (63.3)	0.73
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**400**	84 (37.0)	84 (37.0)	142 (62.6)	143 (63.0)	93 (41.0)	142 (62.6)	143 (63.0)	0.54
12-13 y	**500**	37 (31.6)	37 (31.6)	79 (67.5)	80 (68.4)	43 (36.8)	79 (67.5)	80 (68.4)	0.58
14-15 y	**400**	47 (42.7)	47 (42.7)	63 (57.3)	63 (57.3)	50 (45.5)	63 (57.3)	63 (57.3)	0.48
White	**1200**	22 (37.9)	25 (43.1)	32 (55.2)	33 (56.9)	26 (44.8)	32 (55.2)	33 (56.9)	0.61
Hispanic	**800**	33 (37.5)	33 (37.5)	55 (62.5)	55 (62.5)	33 (37.5)	55 (62.5)	55 (62.5)	0.55
Black	**400**	22 (30.6)	22 (30.6)	50 (69.4)	50 (69.4)	29 (40.3)	50 (69.4)	50 (69.4)	0.49
Evaluation sample									
All girls	**400**	31 (38.3)	31 (38.3)	50 (61.7)	50 (61.7)	27 (33.3)	50 (61.7)	50 (61.7)	0.63
12-13 y	**500**	17 (38.6)	17 (38.6)	27 (61.4)	27 (61.4)	16 (36.4)	27 (61.4)	27 (61.4)	0.62
14-15 y	**400**	14 (37.8)	14 (37.8)	23 (62.2)	23 (62.2)	11 (29.7)	23 (62.2)	23 (62.2)	0.64
White	**1200**	8 (38.1)	9 (42.9)	12 (57.1)	12 (57.1)	6 (28.6)	12 (57.1)	12 (57.1)	0.62
Hispanic	**400**	13 (46.4)	11 (39.3)	17 (60.7)	17 (60.7)	13 (46.4)	17 (60.7)	17 (60.7)	0.47
Black	**800**	11 (36.7)	11 (36.7)	19 (63.3)	19 (63.3)	7 (23.3)	19 (63.3)	19 (63.3)	0.77

2005 Dietary Guidelines									
Development sample									
All girls	**700**	84 (37.0)	84 (37.0)	140 (61.7)	142 (62.6)	86 (37.9)	141 (62.1)	143 (63.0)	0.54
12-13 y	**800**	37 (31.6)	37 (31.6)	76 (65.0)	79 (67.5)	37 (31.6)	78 (66.7)	80 (68.4)	0.59
14-15 y	**700**	47 (42.7)	47 (42.7)	64 (58.2)	63 (57.3)	49 (44.6)	63 (57.3)	63 (57.3)	0.49
White	**1600**	20 (34.5)	25 (43.1)	30 (51.7)	32 (55.2)	25 (43.1)	32 (55.2)	33 (56.9)	0.61
Hispanic	**1100**	33 (37.5)	33 (37.5)	54 (61.4)	55 (62.5)	34 (38.6)	54 (61.4)	55 (62.5)	0.58
Black	**700**	22 (30.6)	22 (30.6)	50 (69.4)	50 (69.4)	23 (31.9)	49 (68.1)	50 (69.4)	0.46
Evaluation sample									
All girls	**700**	31 (38.3)	31 (38.3)	47 (58.0)	50 (61.7)	31 (38.3)	47 (58.0)	50 (61.7)	0.64
12-13 y	**800**	17 (38.6)	17 (38.6)	26 (59.1)	27 (61.4)	17 (38.6)	26 (59.1)	27 (61.4)	0.66
14-15 y	**700**	14 (37.8)	14 (37.8)	21 (56.8)	23 (62.2)	14 (37.8)	21 (56.8)	23 (62.2)	0.60
White	**1600**	10 (47.6)	9 (42.9)	12 (57.1)	12 (57.1)	9 (42.9)	12 (57.1)	12 (57.1)	0.63
Hispanic	**700**	12 (42.9)	11 (39.3)	17 (60.7)	17 (60.7)	11 (39.3)	17 (60.7)	17 (60.7)	0.53
Black	**1100**	11 (36.7)	11 (36.7)	16 (53.3)	19 (63.3)	11 (36.7)	16 (53.3)	19 (63.3)	0.77
Healthy People 2010									
Development sample									
All girls	**900**	84 (37.0)	84 (37.0)	131 (57.7)	141 (62.1)	84 (37.0)	134 (59.0)	142 (62.6)	0.55
12-13 y	**1300**	37 (31.6)	37 (31.6)	71 (60.7)	78 (66.7)	37 (31.6)	73 (62.4)	79 (67.5)	0.58
14-15 y	**900**	47 (42.7)	47 (42.7)	60 (54.6)	63 (57.3)	47 (42.7)	61 (55.5)	63 (57.3)	0.51
White	**2000**	20 (34.5)	25 (43.1)	28 (48.3)	32 (55.2)	25 (43.1)	29 (50.0)	33 (56.9)	0.63
Hispanic	**900**	33 (37.5)	33 (37.5)	51 (58.0)	54 (61.4)	33 (37.5)	51 (58.0)	54 (61.4)	0.58
Black	**1100**	22 (30.6)	22 (30.6)	46 (63.9)	50 (69.4)	22 (30.6)	49 (68.1)	50 (69.4)	0.46
Evaluation sample									
All girls	**900**	31 (38.3)	31 (38.3)	41 (50.6)	50 (61.7)	31 (38.3)	44 (54.3)	50 (61.7)	0.64
12-13 y	**1300**	17 (38.6)	17 (38.6)	23 (52.3)	27 (61.4)	17 (38.6)	25 (56.8)	27 (61.4)	0.66
14-15 y	**900**	14 (37.8)	14 (37.8)	18 (48.7)	23 (62.2)	14 (37.8)	19 (51.4)	23 (62.2)	0.59
White	**2000**	9 (42.9)	9 (42.9)	12 (57.1)	12 (57.1)	9 (42.9)	12 (57.1)	12 (57.1)	0.61
Hispanic	**1100**	11 (39.3)	11 (39.3)	14 (50.0)	17 (60.7)	11 (39.3)	16 (57.1)	17 (60.7)	0.54
Black	**900**	11 (36.7)	11 (36.7)	14 (46.7)	19 (63.3)	11 (36.7)	14 (46.7)	19 (63.3)	0.75

Risk factor / Recommendation / Sample / Subgroup	New cut point		Misclassified[1] [n (%)] Treuth cut point[2] Light	Moderate	Vigorous	Puyau cut point[2] Light	Moderate	Vigorous	Area under ROC curve
Borderline or High total cholesterol									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	73 (31.3)	73 (31.3)	156 (67.0)	156 (67.0)	122 (52.4)	156 (67.0)	156 (67.0)	0.52
12-13 y	**400**	32 (27.4)	33 (28.2)	83 (70.9)	83 (70.9)	56 (47.9)	83 (70.9)	83 (70.9)	0.56
14-15 y	**100**	40 (34.5)	40 (34.5)	73 (62.9)	73 (62.9)	66 (56.9)	73 (62.9)	73 (62.9)	0.48
White	**400**	21 (35.0)	25 (41.7)	35 (58.3)	35 (58.3)	30 (50.0)	35 (58.3)	35 (58.3)	0.57
Hispanic	**200**	16 (17.6)	16 (17.6)	75 (82.4)	75 (82.4)	58 (63.7)	75 (82.4)	75 (82.4)	0.34
Black	**500**	26 (36.1)	27 (37.5)	41 (56.9)	41 (56.9)	28 (38.9)	41 (56.9)	41 (56.9)	0.65
UW/NW	**200**	44 (28.6)	44 (28.6)	108 (70.1)	108 (70.1)	80 (52.0)	108 (70.1)	108 (70.1)	0.53
AR/OW	**100**	28 (37.8)	28 (37.8)	44 (59.5)	44 (59.5)	40 (54.1)	44 (59.5)	44 (59.5)	0.51
Evaluation sample									
All girls	**100**	20 (26.7)	20 (26.7)	53 (70.7)	53 (70.7)	47 (62.7)	53 (70.7)	53 (70.7)	0.42
12-13 y	**400**	9 (22.0)	9 (22.0)	31 (75.6)	31 (75.6)	24 (58.5)	31 (75.6)	31 (75.6)	0.51
14-15 y	**100**	11 (32.4)	11 (32.4)	22 (64.7)	22 (64.7)	23 (67.7)	22 (64.7)	22 (64.7)	0.32
White	**400**	6 (33.3)	6 (33.3)	12 (66.7)	12 (66.7)	10 (55.6)	12 (66.7)	12 (66.7)	0.56
Hispanic	**500**	10 (37.0)	9 (33.3)	18 (66.7)	18 (66.7)	18 (66.7)	18 (66.7)	18 (66.7)	0.34
Black	**200**	9 (32.1)	5 (17.9)	21 (75.0)	21 (75.0)	17 (60.7)	21 (75.0)	21 (75.0)	0.39
UW/NW	**200**	16 (32.0)	15 (30.0)	34 (68.0)	34 (68.0)	32 (64.0)	34 (68.0)	34 (68.0)	0.42
AR/OW	**100**	5 (20.8)	5 (20.8)	18 (75.0)	18 (75.0)	14 (58.3)	18 (75.0)	18 (75.0)	0.43
2005 Dietary Guidelines									
Development sample									
All girls	**500**	76 (32.6)	76 (32.6)	156 (67.0)	156 (67.0)	84 (36.1)	156 (67.0)	156 (67.0)	0.56
12-13 y	**400**	33 (28.2)	33 (28.2)	83 (70.9)	83 (70.9)	38 (32.5)	83 (70.9)	83 (70.9)	0.57
14-15 y	**500**	41 (35.3)	43 (37.1)	73 (62.9)	73 (62.9)	46 (39.7)	73 (62.9)	73 (62.9)	0.53
White	**700**	24 (40.0)	25 (41.7)	35 (58.3)	35 (58.3)	26 (43.3)	35 (58.3)	35 (58.3)	0.55
Hispanic	**500**	16 (17.6)	16 (17.6)	75 (82.4)	75 (82.4)	25 (27.5)	75 (82.4)	75 (82.4)	0.48
Black	**1400**	27 (37.5)	30 (41.7)	41 (56.9)	41 (56.9)	29 (40.3)	41 (56.9)	41 (56.9)	0.67
UW/NW	**400**	46 (29.9)	46 (29.9)	108 (70.1)	108 (70.1)	52 (33.8)	108 (70.1)	108 (70.1)	0.54
AR/OW	**600**	28 (37.8)	29 (39.2)	44 (59.5)	44 (59.5)	31 (41.9)	44 (59.5)	44 (59.5)	0.57

Evaluation sample									
All girls	**500**	22 (29.3)	22 (29.3)	53 (70.7)	53 (70.7)	31 (41.3)	53 (70.7)	53 (70.7)	0.37
12-13 y	**400**	10 (24.4)	10 (24.4)	31 (75.6)	31 (75.6)	15 (36.6)	31 (75.6)	31 (75.6)	0.30
14-15 y	**500**	12 (35.3)	12 (35.3)	22 (64.7)	22 (64.7)	16 (47.1)	22 (64.7)	22 (64.7)	0.41
White	**700**	8 (44.4)	6 (33.3)	12 (66.7)	12 (66.7)	8 (44.4)	12 (66.7)	12 (66.7)	0.44
Hispanic	**1400**	9 (33.3)	9 (33.3)	18 (66.7)	18 (66.7)	11 (40.7)	18 (66.7)	18 (66.7)	0.43
Black	**500**	19 (67.9)	7 (25.0)	21 (75.0)	21 (75.0)	11 (39.3)	21 (75.0)	21 (75.0)	0.26
UW/NW	**400**	16 (32.0)	16 (32.0)	34 (68.0)	34 (68.0)	21 (42.0)	34 (68.0)	34 (68.0)	0.35
AR/OW	**600**	8 (33.3)	6 (25.0)	18 (75.0)	18 (75.0)	9 (37.5)	18 (75.0)	18 (75.0)	0.49
Healthy People 2010									
Development sample									
All girls	**700**	75 (32.2)	77 (33.1)	151 (64.8)	156 (67.0)	78 (33.5)	151 (64.8)	156 (67.0)	0.57
12-13 y	**700**	33 (28.2)	34 (29.1)	82 (70.1)	83 (70.9)	34 (29.1)	82 (70.1)	83 (70.9)	0.57
14-15 y	**1000**	40 (34.5)	43 (37.1)	69 (59.5)	73 (62.9)	44 (37.9)	69 (59.5)	73 (62.9)	0.57
White	**1200**	24 (40.0)	25 (41.7)	34 (56.7)	35 (58.3)	26 (43.3)	34 (56.7)	35 (58.3)	0.55
Hispanic	**700**	15 (16.5)	16 (17.6)	75 (82.4)	75 (82.4)	17 (18.7)	75 (82.4)	75 (82.4)	0.54
Black	**1700**	26 (36.1)	31 (43.1)	39 (54.2)	41 (56.9)	30 (41.7)	39 (54.2)	41 (56.9)	0.64
UW/NW	**700**	45 (29.2)	46 (29.9)	104 (67.5)	108 (70.1)	47 (30.5)	104 (67.5)	108 (70.1)	0.56
AR/OW	**1200**	27 (36.5)	30 (40.5)	43 (58.1)	44 (59.5)	30 (40.5)	43 (58.1)	44 (59.5)	0.58
Evaluation sample									
All girls	**700**	23 (30.7)	22 (29.3)	55 (73.3)	53 (70.7)	23 (30.7)	53 (70.7)	53 (70.7)	0.40
12-13 y	**700**	10 (24.4)	10 (24.4)	32 (78.1)	31 (75.6)	10 (24.4)	31 (75.6)	31 (75.6)	0.34
14-15 y	**1000**	14 (41.2)	12 (35.3)	23 (67.7)	22 (64.7)	13 (38.2)	22 (64.7)	22 (64.7)	0.45
White	**1200**	7 (38.9)	6 (33.3)	12 (66.7)	12 (66.7)	6 (33.3)	12 (66.7)	12 (66.7)	0.46
Hispanic	**1700**	9 (33.3)	9 (33.3)	18 (66.7)	18 (66.7)	9 (33.3)	18 (66.7)	18 (66.7)	0.45
Black	**700**	17 (60.7)	7 (25.0)	23 (82.1)	21 (75.0)	8 (28.6)	21 (75.0)	21 (75.0)	0.27
UW/NW	**700**	16 (32.0)	16 (32.0)	36 (72.0)	34 (68.0)	16 (32.0)	34 (68.0)	34 (68.0)	0.39
AR/OW	**1200**	9 (37.5)	6 (25.0)	18 (75.0)	18 (75.0)	7 (29.2)	18 (75.0)	18 (75.0)	0.48

Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**600**	75 (32.2)	77 (33.1)	156 (67.0)	156 (67.0)	85 (36.5)	156 (67.0)	156 (67.0)	0.55
12-13 y	**500**	34 (29.1)	34 (29.1)	83 (70.9)	83 (70.9)	37 (31.6)	83 (70.9)	83 (70.9)	0.57
14-15 y	**600**	40 (34.5)	43 (37.1)	73 (62.9)	73 (62.9)	48 (41.4)	73 (62.9)	73 (62.9)	0.51
White	**600**	24 (40.0)	25 (41.7)	35 (58.3)	35 (58.3)	26 (43.3)	35 (58.3)	35 (58.3)	0.55
Hispanic	**600**	16 (17.6)	16 (17.6)	75 (82.4)	75 (82.4)	25 (27.5)	75 (82.4)	75 (82.4)	0.47
Black	**700**	27 (37.5)	31 (43.1)	41 (56.9)	41 (56.9)	30 (41.7)	41 (56.9)	41 (56.9)	0.63
UW/NW	**600**	46 (29.9)	46 (29.9)	108 (70.1)	108 (70.1)	54 (35.1)	108 (70.1)	108 (70.1)	0.52
AR/OW	**600**	28 (37.8)	30 (40.5)	44 (59.5)	44 (59.5)	30 (40.5)	44 (59.5)	44 (59.5)	0.56
Evaluation sample									
All girls	**600**	25 (33.3)	22 (29.3)	53 (70.7)	53 (70.7)	28 (37.3)	53 (70.7)	53 (70.7)	0.40
12-13 y	**500**	10 (24.4)	10 (24.4)	31 (75.6)	31 (75.6)	14 (34.2)	31 (75.6)	31 (75.6)	0.31
14-15 y	**600**	13 (38.2)	12 (35.3)	22 (64.7)	22 (64.7)	14 (41.2)	22 (64.7)	22 (64.7)	0.47
White	**600**	8 (44.4)	6 (33.3)	12 (66.7)	12 (66.7)	8 (44.4)	12 (66.7)	12 (66.7)	0.54
Hispanic	**700**	9 (33.3)	9 (33.3)	18 (66.7)	18 (66.7)	10 (37.0)	18 (66.7)	18 (66.7)	0.43
Black	**600**	9 (32.1)	7 (25.0)	21 (75.0)	21 (75.0)	9 (32.1)	21 (75.0)	21 (75.0)	0.27
UW/NW	**600**	16 (32.0)	16 (32.0)	34 (68.0)	34 (68.0)	17 (34.0)	34 (68.0)	34 (68.0)	0.38
AR/OW	**600**	9 (37.5)	6 (25.0)	18 (75.0)	18 (75.0)	10 (41.7)	18 (75.0)	18 (75.0)	0.45
2005 Dietary Guidelines									
Development sample									
All girls	**800**	76 (32.6)	77 (33.1)	150 (64.4)	156 (67.0)	76 (32.6)	153 (65.7)	156 (67.0)	0.55
12-13 y	**800**	34 (29.1)	34 (29.1)	80 (68.4)	83 (70.9)	34 (29.1)	82 (70.1)	83 (70.9)	0.55
14-15 y	**1000**	42 (36.2)	43 (37.1)	70 (60.3)	73 (62.9)	42 (36.2)	71 (61.2)	73 (62.9)	0.53
White	**800**	25 (41.7)	25 (41.7)	33 (55.0)	35 (58.3)	25 (41.7)	35 (58.3)	35 (58.3)	0.54
Hispanic	**800**	15 (16.5)	16 (17.6)	76 (83.5)	75 (82.4)	15 (16.5)	76 (83.5)	75 (82.4)	0.49
Black	**1600**	29 (40.3)	31 (43.1)	39 (54.2)	41 (56.9)	31 (43.1)	40 (55.6)	41 (56.9)	0.62
UW/NW	**800**	45 (29.2)	46 (29.9)	103 (66.9)	108 (70.1)	45 (29.2)	106 (68.8)	108 (70.1)	0.54
AR/OW	**1100**	29 (39.2)	30 (40.5)	43 (58.1)	44 (59.5)	30 (40.5)	43 (58.1)	44 (59.5)	0.55

Evaluation sample									
All girls	**800**	22 (29.3)	22 (29.3)	55 (73.3)	53 (70.7)	22 (29.3)	55 (73.3)	53 (70.7)	0.40
12-13 y	**800**	10 (24.4)	10 (24.4)	32 (78.1)	31 (75.6)	10 (24.4)	32 (78.1)	31 (75.6)	0.31
14-15 y	**1000**	13 (38.2)	12 (35.3)	23 (67.7)	22 (64.7)	12 (35.3)	23 (67.7)	22 (64.7)	0.46
White	**800**	6 (33.3)	6 (33.3)	12 (66.7)	12 (66.7)	6 (33.3)	12 (66.7)	12 (66.7)	0.44
Hispanic	**1600**	9 (33.3)	9 (33.3)	18 (66.7)	18 (66.7)	9 (33.3)	18 (66.7)	18 (66.7)	0.48
Black	**800**	16 (57.1)	7 (25.0)	23 (82.1)	21 (75.0)	7 (25.0)	23 (82.1)	21 (75.0)	0.29
UW/NW	**800**	16 (32.0)	16 (32.0)	36 (72.0)	34 (68.0)	16 (32.0)	36 (72.0)	34 (68.0)	0.39
AR/OW	**1100**	8 (33.3)	6 (25.0)	18 (75.0)	18 (75.0)	6 (25.0)	18 (75.0)	18 (75.0)	0.46
Healthy People 2010									
Development sample									
All girls	**1200**	76 (32.6)	77 (33.1)	145 (62.2)	157 (67.4)	77 (33.1)	144 (61.8)	157 (67.4)	0.55
12-13 y	**1200**	34 (29.1)	34 (29.1)	82 (70.1)	84 (71.8)	34 (29.1)	79 (67.5)	84 (71.8)	0.53
14-15 y	**1400**	42 (36.2)	43 (37.1)	63 (54.3)	73 (62.9)	43 (37.1)	65 (56.0)	73 (62.9)	0.56
White	**1900**	25 (41.7)	25 (41.7)	34 (56.7)	35 (58.3)	25 (41.7)	33 (55.0)	35 (58.3)	0.55
Hispanic	**1100**	15 (16.5)	16 (17.6)	69 (75.8)	76 (83.5)	16 (17.6)	70 (76.9)	76 (83.5)	0.50
Black	**2000**	29 (40.3)	31 (43.1)	38 (52.8)	41 (56.9)	31 (43.1)	38 (52.8)	41 (56.9)	0.60
UW/NW	**1200**	45 (29.2)	46 (29.9)	99 (64.3)	109 (70.8)	46 (29.9)	98 (63.6)	109 (70.8)	0.53
AR/OW	**1400**	29 (39.2)	30 (40.5)	43 (58.1)	44 (59.5)	30 (40.5)	42 (56.8)	44 (59.5)	0.57
Evaluation sample									
All girls	**1200**	24 (32.0)	22 (29.3)	49 (65.3)	53 (70.7)	22 (29.3)	52 (69.3)	53 (70.7)	0.42
12-13 y	**1200**	11 (26.8)	10 (24.4)	31 (75.6)	31 (75.6)	10 (24.4)	33 (80.5)	31 (75.6)	0.31
14-15 y	**1400**	14 (41.2)	12 (35.3)	18 (52.9)	22 (64.7)	12 (35.3)	19 (55.9)	22 (64.7)	0.50
White	**1900**	9 (50.0)	6 (33.3)	12 (66.7)	12 (66.7)	6 (33.3)	12 (66.7)	12 (66.7)	0.45
Hispanic	**2000**	9 (33.3)	9 (33.3)	13 (48.2)	18 (66.7)	9 (33.3)	15 (55.6)	18 (66.7)	0.53
Black	**1100**	15 (53.6)	7 (25.0)	23 (82.1)	21 (75.0)	7 (25.0)	23 (82.1)	21 (75.0)	0.30
UW/NW	**1200**	16 (32.0)	16 (32.0)	31 (62.0)	34 (68.0)	16 (32.0)	34 (68.0)	34 (68.0)	0.41
AR/OW	**1400**	9 (37.5)	6 (25.0)	17 (70.8)	18 (75.0)	6 (25.0)	17 (70.8)	18 (75.0)	0.48

Risk factor / Recommendation / Sample / Subgroup		Misclassified[1] [n (%)]							Area under ROC curve
			Treuth cut point[2]			Puyau cut point[2]			
		New cut point	Light	Moderate	Vigorous	Light	Moderate	Vigorous	
Borderline or Low HDL-cholesterol									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	95 (40.8)	95 (40.8)	140 (60.1)	140 (60.1)	112 (48.1)	140 (60.1)	140 (60.1)	0.54
12-13 y	**300**	47 (40.2)	48 (41.0)	68 (58.1)	68 (58.1)	59 (50.4)	68 (58.1)	68 (58.1)	0.51
14-15 y	**100**	47 (40.5)	47 (40.5)	72 (62.1)	72 (62.1)	53 (45.7)	72 (62.1)	72 (62.1)	0.57
White	**600**	21 (35.0)	21 (35.0)	39 (65.0)	39 (65.0)	26 (43.3)	39 (65.0)	39 (65.0)	0.63
Hispanic	**300**	43 (47.3)	44 (48.4)	47 (51.7)	47 (51.7)	46 (50.6)	47 (51.7)	47 (51.7)	0.49
Black	**500**	27 (37.5)	28 (38.9)	46 (63.9)	46 (63.9)	33 (45.8)	46 (63.9)	46 (63.9)	0.55
UW/NW	**100**	51 (33.1)	51 (33.1)	105 (68.2)	105 (68.2)	81 (52.6)	105 (68.2)	105 (68.2)	0.52
AR/OW	**800**	29 (39.2)	43 (58.1)	31 (41.9)	31 (41.9)	29 (39.2)	31 (41.9)	31 (41.9)	0.56
Evaluation sample									
All girls	**100**	26 (34.7)	26 (34.7)	51 (68.0)	51 (68.0)	45 (60.0)	51 (68.0)	51 (68.0)	0.43
12-13 y	**300**	17 (41.5)	17 (41.5)	25 (61.0)	25 (61.0)	28 (68.3)	25 (61.0)	25 (61.0)	0.29
14-15 y	**100**	9 (26.5)	9 (26.5)	26 (76.5)	26 (76.5)	17 (50.0)	26 (76.5)	26 (76.5)	0.64
White	**600**	10 (55.6)	9 (50.0)	9 (50.0)	9 (50.0)	11 (61.1)	9 (50.0)	9 (50.0)	0.47
Hispanic	**500**	10 (37.0)	9 (33.3)	18 (66.7)	18 (66.7)	16 (59.3)	18 (66.7)	18 (66.7)	0.43
Black	**300**	10 (35.7)	8 (28.6)	22 (78.6)	22 (78.6)	16 (57.1)	22 (78.6)	22 (78.6)	0.34
UW/NW	**100**	15 (30.0)	15 (30.0)	36 (72.0)	36 (72.0)	32 (64.0)	36 (72.0)	36 (72.0)	0.32
AR/OW	**800**	13 (54.2)	10 (41.7)	15 (62.5)	15 (62.5)	13 (54.2)	15 (62.5)	15 (62.5)	0.56
2005 Dietary Guidelines									
Development sample									
All girls	**700**	85 (36.5)	92 (39.5)	140 (60.1)	140 (60.1)	90 (38.6)	140 (60.1)	140 (60.1)	0.59
12-13 y	**800**	45 (38.5)	48 (41.0)	68 (58.1)	68 (58.1)	45 (38.5)	68 (58.1)	68 (58.1)	0.61
14-15 y	**700**	40 (34.5)	44 (37.9)	72 (62.1)	72 (62.1)	45 (38.8)	72 (62.1)	72 (62.1)	0.57
White	**700**	18 (30.0)	21 (35.0)	39 (65.0)	39 (65.0)	22 (36.7)	39 (65.0)	39 (65.0)	0.60
Hispanic	**1400**	37 (40.7)	44 (48.4)	47 (51.7)	47 (51.7)	39 (42.9)	47 (51.7)	47 (51.7)	0.60
Black	**700**	24 (33.3)	25 (34.7)	46 (63.9)	46 (63.9)	26 (36.1)	46 (63.9)	46 (63.9)	0.57
UW/NW	**400**	49 (31.8)	49 (31.8)	105 (68.2)	105 (68.2)	57 (37.0)	105 (68.2)	105 (68.2)	0.57
AR/OW	**2200**	30 (40.5)	42 (56.8)	31 (41.9)	31 (41.9)	32 (43.2)	31 (41.9)	31 (41.9)	0.59

Evaluation sample									
All girls	**700**	25 (33.3)	24 (32.0)	51 (68.0)	51 (68.0)	31 (41.3)	51 (68.0)	51 (68.0)	0.42
12-13 y	**800**	19 (46.3)	16 (39.0)	25 (61.0)	25 (61.0)	19 (46.3)	25 (61.0)	25 (61.0)	0.39
14-15 y	**700**	9 (26.5)	8 (23.5)	26 (76.5)	26 (76.5)	12 (35.3)	26 (76.5)	26 (76.5)	0.52
White	**700**	7 (38.9)	9 (50.0)	9 (50.0)	9 (50.0)	7 (38.9)	9 (50.0)	9 (50.0)	0.51
Hispanic	**700**	20 (74.1)	9 (33.3)	18 (66.7)	18 (66.7)	11 (40.7)	18 (66.7)	18 (66.7)	0.37
Black	**1400**	8 (28.6)	6 (21.4)	22 (78.6)	22 (78.6)	12 (42.9)	22 (78.6)	22 (78.6)	0.31
UW/NW	**400**	14 (28.0)	14 (28.0)	36 (72.0)	36 (72.0)	21 (42.0)	36 (72.0)	36 (72.0)	0.33
AR/OW	**2200**	15 (62.5)	9 (37.5)	15 (62.5)	15 (62.5)	10 (41.7)	15 (62.5)	15 (62.5)	0.48
Healthy People 2010									
Development sample									
All girls	**1300**	83 (35.6)	93 (39.9)	137 (58.8)	140 (60.1)	94 (40.3)	137 (58.8)	140 (60.1)	0.58
12-13 y	**1300**	42 (35.9)	49 (41.9)	69 (59.0)	68 (58.1)	49 (41.9)	69 (59.0)	68 (58.1)	0.61
14-15 y	**1400**	41 (35.3)	44 (37.9)	68 (58.6)	72 (62.1)	45 (38.8)	68 (58.6)	72 (62.1)	0.56
White	**1000**	17 (28.3)	21 (35.0)	40 (66.7)	39 (65.0)	22 (36.7)	40 (66.7)	39 (65.0)	0.61
Hispanic	**2200**	35 (38.5)	44 (48.4)	47 (51.7)	47 (51.7)	45 (49.5)	47 (51.7)	47 (51.7)	0.60
Black	**1200**	22 (30.6)	26 (36.1)	44 (61.1)	46 (63.9)	25 (34.7)	44 (61.1)	46 (63.9)	0.58
UW/NW	**900**	49 (31.8)	49 (31.8)	103 (66.9)	105 (68.2)	50 (32.5)	103 (66.9)	105 (68.2)	0.55
AR/OW	**2200**	29 (39.2)	43 (58.1)	30 (40.5)	31 (41.9)	43 (58.1)	30 (40.5)	31 (41.9)	0.60
Evaluation sample									
All girls	**1300**	33 (44.0)	24 (32.0)	51 (68.0)	51 (68.0)	25 (33.3)	51 (68.0)	51 (68.0)	0.41
12-13 y	**1300**	21 (51.2)	16 (39.0)	24 (58.5)	25 (61.0)	16 (39.0)	25 (61.0)	25 (61.0)	0.38
14-15 y	**1400**	13 (38.2)	8 (23.5)	27 (79.4)	26 (76.5)	9 (26.5)	26 (76.5)	26 (76.5)	0.49
White	**1000**	8 (44.4)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	0.46
Hispanic	**1200**	19 (70.4)	9 (33.3)	18 (66.7)	18 (66.7)	9 (33.3)	18 (66.7)	18 (66.7)	0.37
Black	**2200**	12 (42.9)	6 (21.4)	22 (78.6)	22 (78.6)	7 (25.0)	22 (78.6)	22 (78.6)	0.34
UW/NW	**900**	14 (28.0)	14 (28.0)	36 (72.0)	36 (72.0)	14 (28.0)	36 (72.0)	36 (72.0)	0.37
AR/OW	**2200**	16 (66.7)	9 (37.5)	15 (62.5)	15 (62.5)	10 (41.7)	15 (62.5)	15 (62.5)	0.42

Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**900**	85 (36.5)	93 (39.9)	140 (60.1)	140 (60.1)	87 (37.3)	140 (60.1)	140 (60.1)	0.60
12-13 y	**900**	40 (34.2)	49 (41.9)	68 (58.1)	68 (58.1)	44 (37.6)	68 (58.1)	68 (58.1)	0.63
14-15 y	**800**	43 (37.1)	44 (37.9)	72 (62.1)	72 (62.1)	43 (37.1)	72 (62.1)	72 (62.1)	0.58
White	**700**	19 (31.7)	21 (35.0)	39 (65.0)	39 (65.0)	20 (33.3)	39 (65.0)	39 (65.0)	0.61
Hispanic	**1300**	39 (42.9)	44 (48.4)	47 (51.7)	47 (51.7)	41 (45.1)	47 (51.7)	47 (51.7)	0.59
Black	**900**	22 (30.6)	26 (36.1)	46 (63.9)	46 (63.9)	23 (31.9)	46 (63.9)	46 (63.9)	0.61
UW/NW	**500**	49 (31.8)	49 (31.8)	105 (68.2)	105 (68.2)	53 (34.4)	105 (68.2)	105 (68.2)	0.58
AR/OW	**2000**	28 (37.8)	43 (58.1)	31 (41.9)	31 (41.9)	33 (44.6)	31 (41.9)	31 (41.9)	0.60
Evaluation sample									
All girls	**900**	30 (40.0)	24 (32.0)	51 (68.0)	51 (68.0)	28 (37.3)	51 (68.0)	51 (68.0)	0.43
12-13 y	**900**	18 (43.9)	16 (39.0)	25 (61.0)	25 (61.0)	18 (43.9)	25 (61.0)	25 (61.0)	0.39
14-15 y	**800**	10 (29.4)	8 (23.5)	26 (76.5)	26 (76.5)	10 (29.4)	26 (76.5)	26 (76.5)	0.52
White	**700**	7 (38.9)	9 (50.0)	9 (50.0)	9 (50.0)	7 (38.9)	9 (50.0)	9 (50.0)	0.49
Hispanic	**900**	18 (66.7)	9 (33.3)	18 (66.7)	18 (66.7)	10 (37.0)	18 (66.7)	18 (66.7)	0.39
Black	**1300**	13 (46.4)	6 (21.4)	22 (78.6)	22 (78.6)	10 (35.7)	22 (78.6)	22 (78.6)	0.32
UW/NW	**500**	14 (28.0)	14 (28.0)	36 (72.0)	36 (72.0)	17 (34.0)	36 (72.0)	36 (72.0)	0.39
AR/OW	**2000**	15 (62.5)	9 (37.5)	15 (62.5)	15 (62.5)	11 (45.8)	15 (62.5)	15 (62.5)	0.44
2005 Dietary Guidelines									
Development sample									
All girls	**1300**	81 (34.8)	93 (39.9)	136 (58.4)	140 (60.1)	94 (40.3)	137 (58.8)	140 (60.1)	0.60
12-13 y	**1300**	40 (34.2)	49 (41.9)	67 (57.3)	68 (58.1)	49 (41.9)	67 (57.3)	68 (58.1)	0.63
14-15 y	**1300**	41 (35.3)	44 (37.9)	69 (59.5)	72 (62.1)	45 (38.8)	70 (60.3)	72 (62.1)	0.58
White	**1000**	19 (31.7)	21 (35.0)	39 (65.0)	39 (65.0)	21 (35.0)	39 (65.0)	39 (65.0)	0.61
Hispanic	**1300**	36 (39.6)	44 (48.4)	48 (52.8)	47 (51.7)	45 (49.5)	48 (52.8)	47 (51.7)	0.61
Black	**1300**	23 (31.9)	26 (36.1)	44 (61.1)	46 (63.9)	26 (36.1)	45 (62.5)	46 (63.9)	0.61
UW/NW	**1300**	49 (31.8)	49 (31.8)	102 (66.2)	105 (68.2)	50 (32.5)	103 (66.9)	105 (68.2)	0.58
AR/OW	**2400**	28 (37.8)	43 (58.1)	30 (40.5)	31 (41.9)	43 (58.1)	30 (40.5)	31 (41.9)	0.59

Evaluation sample									
All girls	**1300**	28 (37.3)	24 (32.0)	51 (68.0)	51 (68.0)	24 (32.0)	51 (68.0)	51 (68.0)	0.42
12-13 y	**1300**	19 (46.3)	16 (39.0)	24 (58.5)	25 (61.0)	16 (39.0)	24 (58.5)	25 (61.0)	0.38
14-15 y	**1300**	9 (26.5)	8 (23.5)	27 (79.4)	26 (76.5)	8 (23.5)	27 (79.4)	26 (76.5)	0.51
White	**1000**	8 (44.4)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	0.43
Hispanic	**1300**	8 (29.6)	9 (33.3)	18 (66.7)	18 (66.7)	9 (33.3)	18 (66.7)	18 (66.7)	0.42
Black	**1300**	12 (42.9)	6 (21.4)	22 (78.6)	22 (78.6)	6 (21.4)	22 (78.6)	22 (78.6)	0.34
UW/NW	**1300**	17 (34.0)	14 (28.0)	36 (72.0)	36 (72.0)	14 (28.0)	36 (72.0)	36 (72.0)	0.40
AR/OW	**2400**	16 (66.7)	9 (37.5)	15 (62.5)	15 (62.5)	9 (37.5)	15 (62.5)	15 (62.5)	0.43
Healthy People 2010									
Development sample									
All girls	**1700**	85 (36.5)	93 (39.9)	129 (55.4)	141 (60.5)	93 (39.9)	134 (57.5)	141 (60.5)	0.59
12-13 y	**1800**	41 (35.0)	49 (41.9)	61 (52.1)	69 (59.0)	49 (41.9)	66 (56.4)	69 (59.0)	0.62
14-15 y	**1700**	44 (37.9)	44 (37.9)	68 (58.6)	72 (62.1)	44 (37.9)	68 (58.6)	72 (62.1)	0.56
White	**1300**	19 (31.7)	21 (35.0)	36 (60.0)	39 (65.0)	21 (35.0)	37 (61.7)	39 (65.0)	0.61
Hispanic	**1800**	36 (39.6)	44 (48.4)	45 (49.5)	48 (52.8)	44 (48.4)	46 (50.6)	48 (52.8)	0.60
Black	**1700**	24 (33.3)	26 (36.1)	43 (59.7)	46 (63.9)	26 (36.1)	45 (62.5)	46 (63.9)	0.59
UW/NW	**1100**	49 (31.8)	49 (31.8)	96 (62.3)	106 (68.8)	49 (31.8)	101 (65.6)	106 (68.8)	0.57
AR/OW	**3300**	29 (39.2)	43 (58.1)	30 (40.5)	31 (41.9)	43 (58.1)	29 (39.2)	31 (41.9)	0.58
Evaluation sample									
All girls	**1700**	30 (40.0)	24 (32.0)	47 (62.7)	51 (68.0)	24 (32.0)	48 (64.0)	51 (68.0)	0.43
12-13 y	**1800**	19 (46.3)	16 (39.0)	23 (56.1)	25 (61.0)	16 (39.0)	23 (56.1)	25 (61.0)	0.44
14-15 y	**1700**	11 (32.4)	8 (23.5)	24 (70.6)	26 (76.5)	8 (23.5)	25 (73.5)	26 (76.5)	0.43
White	**1300**	8 (44.4)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	9 (50.0)	0.44
Hispanic	**1700**	10 (37.0)	9 (33.3)	17 (63.0)	18 (66.7)	9 (33.3)	17 (63.0)	18 (66.7)	0.41
Black	**1800**	12 (42.9)	6 (21.4)	20 (71.4)	22 (78.6)	6 (21.4)	20 (71.4)	22 (78.6)	0.37
UW/NW	**1100**	14 (28.0)	14 (28.0)	31 (62.0)	36 (72.0)	14 (28.0)	32 (64.0)	36 (72.0)	0.39
AR/OW	**3300**	15 (62.5)	9 (37.5)	16 (66.7)	15 (62.5)	9 (37.5)	16 (66.7)	15 (62.5)	0.43
Overweight – All girls									
Frequency and duration									
Strong et al., 2005									
Development sample	**100**	42 (17.1)	42 (17.1)	206 (84.1)	206 (84.1)	141 (57.6)	206 (84.1)	206 (84.1)	0.49
Evaluation sample	**100**	12 (14.6)	12 (14.6)	72 (87.8)	72 (87.8)	50 (61.0)	72 (87.8)	72 (87.8)	0.37
2005 Dietary Guidelines									
Development sample	**400**	39 (15.9)	40 (16.3)	206 (84.1)	206 (84.1)	71 (29.0)	206 (84.1)	206 (84.1)	0.57
Evaluation sample	**400**	10 (12.2)	10 (12.2)	72 (87.8)	72 (87.8)	21 (25.6)	72 (87.8)	72 (87.8)	0.48

Healthy People 2010									
Development sample	**600**	39 (15.9)	39 (15.9)	201 (82.0)	206 (84.1)	43 (17.6)	201 (82.0)	206 (84.1)	0.59
Evaluation sample	**600**	10 (12.2)	10 (12.2)	70 (85.4)	72 (87.8)	12 (14.6)	72 (87.8)	72 (87.8)	0.46
Accumulated time									
Strong et al., 2005									
Development sample	**400**	38 (15.5)	39 (15.9)	205 (83.7)	206 (84.1)	64 (26.1)	205 (83.7)	206 (84.1)	0.55
Evaluation sample	**400**	10 (12.2)	10 (12.2)	72 (87.8)	72 (87.8)	16 (19.5)	72 (87.8)	72 (87.8)	0.51
2005 Dietary Guidelines									
Development sample	**700**	39 (15.9)	39 (15.9)	199 (81.2)	205 (83.7)	40 (16.3)	202 (82.5)	206 (84.1)	0.57
Evaluation sample	**700**	10 (12.2)	10 (12.2)	69 (84.2)	72 (87.8)	10 (12.2)	69 (84.2)	72 (87.8)	0.49
Healthy People 2010									
Development sample	**900**	39 (15.9)	39 (15.9)	182 (74.3)	204 (83.3)	39 (15.9)	189 (77.1)	205 (83.7)	0.56
Evaluation sample	**900**	10 (12.2)	10 (12.2)	61 (74.4)	72 (87.8)	10 (12.2)	64 (78.1)	72 (87.8)	0.47

[1] Misclassification includes false positives (i.e., low risk and inadequate physical activity) and false negatives (i.e., high risk and adequate physical activity).

[2] Treuth et al. (2004) cut points for light, moderate, and vigorous intensity physical activity were 101, 3000, and 5201 counts/min, respectively. Puyau et al. (2002) cut points for light, moderate, and vigorous intensity physical activity were 800, 3200, and 8200 counts/min, respectively.

UW/NW is underweight or normal weight (< 85^{th} BMI percentile).

AR/OW is at risk for overweight or overweight (≥ 85^{th} BMI percentile).

HDL is high-density lipoprotein.

Appendix G. Misclassification for potential new cut points and previously suggested cut points (counts/min) and area under the receiver operating characteristic (ROC) curve by disease risk factor and physical activity recommendation by population subgroup in TAAG development (N = 3522) and evaluation (N = 1174) samples

Risk factor / Recommendation / Sample / Subgroup	New cut point		Misclassified[1] [n (%)] Treuth cut point[2] Light	Treuth cut point[2] Moderate	Treuth cut point[2] Vigorous	Puyau cut point[2] Light	Puyau cut point[2] Moderate	Puyau cut point[2] Vigorous	Area under ROC curve
At risk for overweight or Overweight									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	1253 (35.6)	1253 (35.6)	2279 (64.8)	2279 (64.8)	1783 (50.7)	2279 (64.8)	2279 (64.8)	0.50
White	**100**	476 (28.9)	476 (28.9)	1182 (71.7)	1182 (71.7)	795 (48.2)	1182 (71.7)	1182 (71.7)	0.53
Hispanic	**100**	341 (44.3)	341 (44.3)	434 (56.4)	434 (56.4)	399 (51.8)	434 (56.4)	434 (56.4)	0.47
Black	**100**	284 (43.8)	284 (43.8)	369 (56.9)	369 (56.9)	349 (53.8)	369 (56.9)	369 (56.9)	0.46
Asian	**100**	58 (28.4)	58 (28.4)	148 (72.6)	148 (72.6)	128 (62.8)	148 (72.6)	148 (72.6)	0.42
Evaluation sample									
All girls	**100**	413 (35.3)	413 (35.3)	765 (65.3)	765 (65.3)	581 (49.6)	765 (65.3)	765 (65.3)	0.49
White	**100**	156 (28.8)	156 (28.8)	387 (71.5)	387 (71.5)	263 (48.6)	387 (71.5)	387 (71.5)	0.48
Hispanic	**100**	103 (41.7)	103 (41.7)	146 (59.1)	146 (59.1)	129 (52.2)	146 (59.1)	146 (59.1)	0.46
Black	**100**	102 (47.0)	102 (47.0)	116 (53.5)	116 (53.5)	99 (45.6)	116 (53.5)	116 (53.5)	0.53
Asian	**100**	17 (20.7)	17 (20.7)	65 (79.3)	65 (79.3)	49 (59.8)	65 (79.3)	65 (79.3)	0.37
2005 Dietary Guidelines									
Development sample									
All girls	**200**	1241 (35.3)	1241 (35.3)	2271 (64.6)	2278 (64.8)	1400 (39.8)	2275 (64.7)	2279 (64.8)	0.51
White	**200**	466 (28.3)	467 (28.3)	1176 (71.3)	1182 (71.7)	554 (33.6)	1180 (71.6)	1182 (71.7)	0.53
Hispanic	**500**	340 (44.2)	340 (44.2)	434 (56.4)	434 (56.4)	376 (48.8)	434 (56.4)	434 (56.4)	0.49
Black	**400**	281 (43.3)	282 (43.5)	367 (56.6)	369 (56.9)	289 (44.5)	368 (56.7)	369 (56.9)	0.50
Asian	**100**	56 (27.5)	56 (27.5)	148 (72.6)	148 (72.6)	84 (41.2)	148 (72.6)	148 (72.6)	0.44
Evaluation sample									
All girls	**200**	410 (35.0)	410 (35.0)	762 (65.1)	765 (65.3)	495 (42.3)	764 (65.2)	765 (65.3)	0.48
White	**200**	156 (28.8)	156 (28.8)	385 (71.2)	387 (71.5)	204 (37.7)	387 (71.5)	387 (71.5)	0.47
Hispanic	**500**	102 (41.3)	101 (40.9)	146 (59.1)	146 (59.1)	111 (44.9)	146 (59.1)	146 (59.1)	0.41
Black	**400**	103 (47.5)	102 (47.0)	115 (53.0)	116 (53.5)	107 (49.3)	115 (53.0)	116 (53.5)	0.57
Asian	**100**	17 (20.7)	17 (20.7)	65 (79.3)	65 (79.3)	33 (40.2)	65 (79.3)	65 (79.3)	0.39

Healthy People 2010									
Development sample									
All girls	**200**	1241 (35.3)	1241 (35.3)	2198 (62.5)	2277 (64.8)	1261 (35.9)	2230 (63.4)	2279 (64.8)	0.53
White	**100**	466 (28.3)	466 (28.3)	1119 (67.9)	1181 (71.6)	478 (29.0)	1141 (69.2)	1182 (71.7)	0.55
Hispanic	**300**	340 (44.2)	341 (44.3)	426 (55.3)	434 (56.4)	344 (44.7)	431 (56.0)	434 (56.4)	0.49
Black	**700**	280 (43.1)	282 (43.5)	363 (55.9)	369 (56.9)	281 (43.3)	368 (56.7)	369 (56.9)	0.52
Asian	**100**	56 (27.5)	56 (27.5)	147 (72.1)	147 (72.1)	60 (29.4)	147 (72.1)	148 (72.6)	0.50
Evaluation sample									
All girls	**200**	410 (35.0)	410 (35.0)	745 (63.6)	764 (65.2)	425 (36.3)	755 (64.5)	765 (65.3)	0.49
White	**100**	156 (28.8)	156 (28.8)	374 (69.1)	387 (71.5)	163 (30.1)	381 (70.4)	387 (71.5)	0.46
Hispanic	**300**	101 (40.9)	101 (40.9)	145 (58.7)	145 (58.7)	103 (41.7)	145 (58.7)	146 (59.1)	0.42
Black	**700**	102 (47.0)	101 (46.5)	112 (51.6)	116 (53.5)	104 (47.9)	114 (52.5)	116 (53.5)	0.58
Asian	**100**	17 (20.7)	17 (20.7)	64 (78.1)	65 (79.3)	18 (22.0)	65 (79.3)	65 (79.3)	0.43
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**100**	1240 (35.3)	1240 (35.3)	2267 (64.5)	2278 (64.8)	1355 (38.5)	2271 (64.6)	2278 (64.8)	0.52
White	**400**	466 (28.3)	466 (28.3)	1174 (71.2)	1182 (71.7)	525 (31.8)	1177 (71.4)	1182 (71.7)	0.54
Hispanic	**100**	340 (44.2)	340 (44.2)	432 (56.1)	434 (56.4)	361 (46.9)	432 (56.1)	434 (56.4)	0.48
Black	**400**	282 (43.5)	282 (43.5)	368 (56.7)	369 (56.9)	287 (44.2)	369 (56.9)	369 (56.9)	0.50
Asian	**100**	56 (27.5)	56 (27.5)	148 (72.6)	148 (72.6)	80 (39.2)	148 (72.6)	148 (72.6)	0.44
Evaluation sample									
All girls	**100**	409 (34.9)	409 (34.9)	761 (65.0)	764 (65.2)	466 (39.8)	761 (65.0)	765 (65.3)	0.50
White	**400**	154 (28.5)	156 (28.8)	384 (71.0)	386 (71.4)	189 (34.9)	385 (71.2)	387 (71.5)	0.49
Hispanic	**100**	101 (40.9)	101 (40.9)	146 (59.1)	146 (59.1)	103 (41.7)	145 (58.7)	146 (59.1)	0.43
Black	**400**	102 (47.0)	101 (46.5)	115 (53.0)	116 (53.5)	105 (48.4)	115 (53.0)	116 (53.5)	0.58
Asian	**100**	17 (20.7)	17 (20.7)	65 (79.3)	65 (79.3)	29 (35.4)	65 (79.3)	65 (79.3)	0.39
2005 Dietary Guidelines									
Development sample									
All girls	**200**	1237 (35.2)	1237 (35.2)	2151 (61.2)	2275 (64.7)	1245 (35.4)	2189 (62.3)	2278 (64.8)	0.53
White	**400**	465 (28.2)	465 (28.2)	1072 (65)	1180 (71.6)	468 (28.4)	1101 (66.8)	1182 (71.7)	0.55
Hispanic	**200**	338 (43.9)	338 (43.9)	421 (54.7)	434 (56.4)	341 (44.3)	427 (55.5)	434 (56.4)	0.50
Black	**800**	282 (43.5)	282 (43.5)	366 (56.4)	369 (56.9)	282 (43.5)	368 (56.7)	369 (56.9)	0.53
Asian	**200**	56 (27.5)	56 (27.5)	149 (73.0)	147 (72.1)	59 (28.9)	149 (73.0)	148 (72.6)	0.47

Evaluation sample									
All girls	**200**	409 (34.9)	409 (34.9)	722 (61.7)	763 (65.2)	413 (35.3)	740 (63.2)	764 (65.2)	0.52
White	**400**	156 (28.8)	156 (28.8)	362 (66.9)	386 (71.4)	156 (28.8)	372 (68.8)	386 (71.4)	0.50
Hispanic	**200**	101 (40.9)	101 (40.9)	144 (58.3)	145 (58.7)	102 (41.3)	143 (57.9)	146 (59.1)	0.43
Black	**800**	102 (47.0)	101 (46.5)	108 (49.8)	116 (53.5)	102 (47.0)	113 (52.1)	116 (53.5)	0.59
Asian	**200**	17 (20.7)	17 (20.7)	61 (74.4)	65 (79.3)	17 (20.7)	63 (76.8)	65 (79.3)	0.42
Healthy People 2010									
Development sample									
All girls	**100**	1236 (35.2)	1236 (35.2)	1874 (53.3)	2264 (64.4)	1241 (35.3)	1947 (55.4)	2275 (64.7)	0.55
White	**700**	464 (28.1)	465 (28.2)	885 (53.7)	1169 (70.9)	466 (28.3)	929 (56.3)	1180 (71.6)	0.56
Hispanic	**200**	336 (43.6)	336 (43.6)	386 (50.1)	433 (56.2)	340 (44.2)	404 (52.5)	434 (56.4)	0.51
Black	**1100**	282 (43.5)	282 (43.5)	351 (54.1)	368 (56.7)	282 (43.5)	351 (54.1)	369 (56.9)	0.55
Asian	**300**	56 (27.5)	56 (27.5)	130 (63.7)	147 (72.1)	57 (27.9)	139 (68.1)	147 (72.1)	0.49
Evaluation sample									
All girls	**100**	407 (34.8)	407 (34.8)	635 (54.2)	759 (64.8)	409 (34.9)	667 (57.0)	764 (65.2)	0.53
White	**700**	156 (28.8)	156 (28.8)	301 (55.6)	382 (70.6)	156 (28.8)	324 (59.9)	386 (71.4)	0.52
Hispanic	**200**	101 (40.9)	101 (40.9)	136 (55.1)	146 (59.1)	101 (40.9)	142 (57.5)	146 (59.1)	0.44
Black	**1100**	101 (46.5)	100 (46.1)	105 (48.4)	115 (53.0)	101 (46.5)	104 (47.9)	116 (53.5)	0.61
Asian	**300**	17 (20.7)	17 (20.7)	57 (69.5)	65 (79.3)	17 (20.7)	57 (69.5)	65 (79.3)	0.45
High body fatness									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	1617 (46.0)	1617 (46.0)	1897 (54.0)	1897 (54.0)	1761 (50.1)	1897 (54.0)	1897 (54.0)	0.51
White	**100**	691 (41.9)	691 (41.9)	959 (58.2)	959 (58.2)	786 (47.7)	959 (58.2)	959 (58.2)	0.54
Hispanic	**2200**	333 (43.3)	434 (56.4)	333 (43.3)	333 (43.3)	404 (52.5)	333 (43.3)	333 (43.3)	0.48
Black	**100**	285 (43.9)	285 (43.9)	366 (56.4)	366 (56.4)	348 (53.6)	366 (56.4)	366 (56.4)	0.47
Asian	**300**	87 (42.7)	87 (42.7)	119 (58.3)	119 (58.3)	115 (56.4)	119 (58.3)	119 (58.3)	0.45
Evaluation sample									
All girls	**100**	535 (45.7)	535 (45.7)	634 (54.2)	634 (54.2)	560 (47.9)	634 (54.2)	634 (54.2)	0.51
White	**100**	232 (42.9)	232 (42.9)	305 (56.4)	305 (56.4)	249 (46.0)	305 (56.4)	305 (56.4)	0.52
Hispanic	**2200**	118 (47.8)	131 (53.0)	118 (47.8)	118 (47.8)	121 (49.0)	118 (47.8)	118 (47.8)	0.51
Black	**100**	101 (46.5)	101 (46.5)	115 (53.0)	115 (53.0)	104 (47.9)	115 (53.0)	115 (53.0)	0.51
Asian	**300**	31 (37.8)	28 (34.2)	54 (65.9)	54 (65.9)	48 (58.5)	54 (65.9)	54 (65.9)	0.39

2005 Dietary Guidelines									
Development sample									
All girls	**200**	1617 (46.0)	1617 (46.0)	1895 (53.9)	1896 (53.9)	1648 (46.9)	1899 (54.0)	1897 (54.0)	0.52
White	**600**	683 (41.4)	690 (41.8)	957 (58.0)	959 (58.2)	699 (42.4)	961 (58.3)	959 (58.2)	0.55
Hispanic	**3500**	332 (43.1)	435 (56.5)	333 (43.3)	333 (43.3)	441 (57.3)	333 (43.3)	333 (43.3)	0.49
Black	**400**	284 (43.8)	285 (43.9)	366 (56.4)	366 (56.4)	290 (44.7)	367 (56.6)	366 (56.4)	0.50
Asian	**100**	85 (41.7)	85 (41.7)	119 (58.3)	119 (58.3)	103 (50.5)	119 (58.3)	119 (58.3)	0.44
Evaluation sample									
All girls	**200**	536 (45.8)	536 (45.8)	633 (54.1)	634 (54.2)	569 (48.6)	633 (54.1)	634 (54.2)	0.50
White	**600**	248 (45.8)	236 (43.6)	305 (56.4)	305 (56.4)	250 (46.2)	305 (56.4)	305 (56.4)	0.50
Hispanic	**3500**	117 (47.4)	129 (52.2)	118 (47.8)	118 (47.8)	127 (51.4)	118 (47.8)	118 (47.8)	0.46
Black	**400**	102 (47.0)	101 (46.5)	114 (52.5)	115 (53.0)	108 (49.8)	114 (52.5)	115 (53.0)	0.56
Asian	**100**	28 (34.2)	28 (34.2)	54 (65.9)	54 (65.9)	40 (48.8)	54 (65.9)	54 (65.9)	0.43
Healthy People 2010									
Development sample									
All girls	**1100**	1612 (45.9)	1617 (46.0)	1846 (52.5)	1897 (54.0)	1621 (46.1)	1868 (53.1)	1897 (54.0)	0.53
White	**1100**	684 (41.5)	689 (41.8)	914 (55.4)	960 (58.2)	695 (42.2)	932 (56.5)	959 (58.2)	0.56
Hispanic	**3300**	332 (43.1)	436 (56.6)	335 (43.5)	333 (43.3)	433 (56.2)	334 (43.4)	333 (43.3)	0.50
Black	**700**	283 (43.6)	285 (43.9)	362 (55.8)	366 (56.4)	284 (43.8)	367 (56.6)	366 (56.4)	0.52
Asian	**100**	85 (41.7)	85 (41.7)	118 (57.8)	118 (57.8)	87 (42.7)	118 (57.8)	119 (58.3)	0.45
Evaluation sample									
All girls	**1100**	553 (47.3)	536 (45.8)	622 (53.2)	633 (54.1)	549 (46.9)	630 (53.9)	634 (54.2)	0.50
White	**1100**	249 (46.0)	236 (43.6)	300 (55.5)	305 (56.4)	241 (44.6)	305 (56.4)	305 (56.4)	0.49
Hispanic	**3300**	119 (48.2)	129 (52.2)	117 (47.4)	117 (47.4)	131 (53.0)	117 (47.4)	118 (47.8)	0.45
Black	**700**	101 (46.5)	100 (46.1)	111 (51.2)	115 (53.0)	103 (47.5)	113 (52.1)	115 (53.0)	0.56
Asian	**100**	28 (34.2)	28 (34.2)	53 (64.6)	54 (65.9)	29 (35.4)	54 (65.9)	54 (65.9)	0.46
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**100**	1618 (46.0)	1618 (46.0)	1893 (53.8)	1896 (53.9)	1627 (46.3)	1895 (53.9)	1896 (53.9)	0.52
White	**700**	683 (41.4)	689 (41.8)	955 (57.9)	959 (58.2)	688 (41.7)	958 (58.1)	959 (58.2)	0.55
Hispanic	**2700**	331 (43.0)	437 (56.8)	333 (43.3)	333 (43.3)	432 (56.1)	333 (43.3)	333 (43.3)	0.50
Black	**500**	285 (43.9)	285 (43.9)	367 (56.6)	366 (56.4)	288 (44.4)	366 (56.4)	366 (56.4)	0.50
Asian	**100**	85 (41.7)	85 (41.7)	119 (58.3)	119 (58.3)	101 (49.5)	119 (58.3)	119 (58.3)	0.45

Evaluation sample									
All girls	**100**	535 (45.7)	535 (45.7)	632 (54.0)	633 (54.1)	560 (47.9)	632 (54.0)	634 (54.2)	0.52
White	**700**	246 (45.5)	236 (43.6)	304 (56.2)	304 (56.2)	247 (45.7)	305 (56.4)	305 (56.4)	0.53
Hispanic	**2700**	119 (48.2)	129 (52.2)	118 (47.8)	118 (47.8)	127 (51.4)	117 (47.4)	118 (47.8)	0.48
Black	**500**	102 (47.0)	100 (46.1)	114 (52.5)	115 (53.0)	106 (48.9)	114 (52.5)	115 (53.0)	0.57
Asian	**100**	28 (34.2)	28 (34.2)	54 (65.9)	54 (65.9)	34 (41.5)	54 (65.9)	54 (65.9)	0.42
2005 Dietary Guidelines									
Development sample									
All girls	**200**	1617 (46.0)	1617 (46.0)	1833 (52.1)	1895 (53.9)	1619 (46.1)	1851 (52.7)	1896 (53.9)	0.53
White	**1200**	683 (41.4)	688 (41.7)	889 (53.9)	959 (58.2)	689 (41.8)	906 (54.9)	959 (58.2)	0.56
Hispanic	**3900**	330 (42.9)	437 (56.8)	342 (44.4)	333 (43.3)	436 (56.6)	340 (44.2)	333 (43.3)	0.51
Black	**800**	285 (43.9)	285 (43.9)	365 (56.2)	366 (56.4)	285 (43.9)	367 (56.6)	366 (56.4)	0.52
Asian	**200**	85 (41.7)	85 (41.7)	120 (58.8)	118 (57.8)	88 (43.1)	120 (58.8)	119 (58.3)	0.45
Evaluation sample									
All girls	**200**	535 (45.7)	535 (45.7)	607 (51.9)	632 (54.0)	539 (46.1)	619 (52.9)	633 (54.1)	0.53
White	**1200**	245 (45.3)	236 (43.6)	296 (54.7)	304 (56.2)	236 (43.6)	300 (55.5)	304 (56.2)	0.52
Hispanic	**3900**	119 (48.2)	129 (52.2)	116 (47.0)	117 (47.4)	130 (52.6)	115 (46.6)	118 (47.8)	0.49
Black	**800**	101 (46.5)	100 (46.1)	107 (49.3)	115 (53.0)	101 (46.5)	112 (51.6)	115 (53.0)	0.59
Asian	**200**	28 (34.2)	28 (34.2)	50 (61.0)	54 (65.9)	28 (34.2)	52 (63.4)	54 (65.9)	0.46
Healthy People 2010									
Development sample									
All girls	**2200**	1603 (45.6)	1618 (46.0)	1722 (49.0)	1890 (53.8)	1619 (46.1)	1739 (49.5)	1897 (54.0)	0.54
White	**1700**	686 (41.6)	688 (41.7)	802 (48.6)	954 (57.9)	689 (41.8)	820 (49.7)	961 (58.3)	0.56
Hispanic	**4900**	331 (43.0)	437 (56.8)	347 (45.1)	332 (43.1)	437 (56.8)	345 (44.8)	333 (43.3)	0.52
Black	**2100**	283 (43.6)	285 (43.9)	350 (53.9)	365 (56.2)	285 (43.9)	350 (53.9)	366 (56.4)	0.54
Asian	**300**	85 (41.7)	85 (41.7)	117 (57.4)	118 (57.8)	86 (42.2)	120 (58.8)	118 (57.8)	0.45
Evaluation sample									
All girls	**2200**	546 (46.7)	535 (45.7)	568 (48.6)	628 (53.7)	535 (45.7)	580 (49.6)	633 (54.1)	0.54
White	**1700**	248 (45.8)	236 (43.6)	263 (48.6)	300 (55.5)	236 (43.6)	276 (51.0)	304 (56.2)	0.53
Hispanic	**4900**	117 (47.4)	129 (52.2)	118 (47.8)	118 (47.8)	129 (52.2)	118 (47.8)	118 (47.8)	0.49
Black	**2100**	90 (41.5)	101 (46.5)	102 (47.0)	114 (52.5)	100 (46.1)	103 (47.5)	115 (53.0)	0.60
Asian	**300**	28 (34.2)	28 (34.2)	52 (63.4)	54 (65.9)	28 (34.2)	50 (61.0)	54 (65.9)	0.49

Risk factor / Recommendation / Sample / Subgroup		Misclassified[1] [n (%)]							Area under ROC curve
			Treuth cut point[2]			Puyau cut point[2]			
	New cut point		Light	Moderate	Vigorous	Light	Moderate	Vigorous	
Overweight									
Frequency and duration									
Strong et al., 2005									
Development sample									
All girls	**100**	672 (19.1)	672 (19.1)	2892 (82.3)	2892 (82.3)	1770 (50.3)	2892 (82.3)	2892 (82.3)	0.50
White	**100**	216 (13.1)	216 (13.1)	1456 (88.3)	1456 (88.3)	801 (48.6)	1456 (88.3)	1456 (88.3)	0.53
Hispanic	**100**	202 (26.2)	202 (26.2)	579 (75.2)	579 (75.2)	380 (49.4)	579 (75.2)	579 (75.2)	0.47
Black	**100**	181 (27.9)	181 (27.9)	476 (73.3)	476 (73.3)	344 (53.0)	476 (73.3)	476 (73.3)	0.49
Asian	**100**	21 (10.3)	21 (10.3)	189 (92.7)	189 (92.7)	127 (62.3)	189 (92.7)	189 (92.7)	0.44
Evaluation sample									
All girls	**100**	213 (18.2)	213 (18.2)	975 (83.3)	975 (83.3)	591 (50.5)	975 (83.3)	975 (83.3)	0.48
White	**100**	73 (13.5)	73 (13.5)	476 (88.0)	476 (88.0)	254 (47.0)	476 (88.0)	476 (88.0)	0.50
Hispanic	**100**	51 (20.7)	51 (20.7)	198 (80.2)	198 (80.2)	135 (54.7)	198 (80.2)	198 (80.2)	0.41
Black	**100**	61 (28.1)	61 (28.1)	161 (74.2)	161 (74.2)	116 (53.5)	161 (74.2)	161 (74.2)	0.48
Asian	**100**	9 (11.0)	9 (11.0)	73 (89.0)	73 (89.0)	45 (54.9)	73 (89.0)	73 (89.0)	0.43
2005 Dietary Guidelines									
Development sample									
All girls	**100**	648 (18.4)	648 (18.4)	2884 (82.0)	2891 (82.2)	977 (27.8)	2888 (82.1)	2892 (82.3)	0.51
White	**100**	205 (12.4)	205 (12.4)	1450 (87.9)	1456 (88.3)	366 (22.2)	1454 (88.2)	1456 (88.3)	0.53
Hispanic	**200**	199 (25.8)	199 (25.8)	579 (75.2)	579 (75.2)	263 (34.2)	579 (75.2)	579 (75.2)	0.50
Black	**200**	177 (27.3)	177 (27.3)	474 (73.0)	476 (73.3)	220 (33.9)	475 (73.2)	476 (73.3)	0.51
Asian	**100**	17 (8.3)	17 (8.3)	189 (92.7)	189 (92.7)	63 (30.9)	189 (92.7)	189 (92.7)	0.46
Evaluation sample									
All girls	**100**	206 (17.6)	206 (17.6)	972 (83.0)	975 (83.3)	349 (29.8)	974 (83.2)	975 (83.3)	0.48
White	**100**	69 (12.8)	69 (12.8)	474 (87.6)	476 (88.0)	137 (25.3)	476 (88.0)	476 (88.0)	0.50
Hispanic	**200**	49 (19.8)	49 (19.8)	198 (80.2)	198 (80.2)	77 (31.2)	198 (80.2)	198 (80.2)	0.37
Black	**200**	61 (28.1)	61 (28.1)	160 (73.7)	161 (74.2)	86 (39.6)	160 (73.7)	161 (74.2)	0.53
Asian	**100**	9 (11.0)	9 (11.0)	73 (89.0)	73 (89.0)	25 (30.5)	73 (89.0)	73 (89.0)	0.44

Healthy People 2010									
Development sample									
All girls	**200**	648 (18.4)	648 (18.4)	2785 (79.2)	2890 (82.2)	692 (19.7)	2823 (80.3)	2892 (82.3)	0.53
White	**100**	204 (12.4)	204 (12.4)	1383 (83.9)	1455 (88.2)	226 (13.7)	1405 (85.2)	1456 (88.3)	0.55
Hispanic	**300**	199 (25.8)	200 (26.0)	561 (72.9)	579 (75.2)	209 (27.1)	572 (74.3)	579 (75.2)	0.48
Black	**700**	175 (27.0)	177 (27.3)	466 (71.8)	476 (73.3)	178 (27.4)	471 (72.6)	476 (73.3)	0.54
Asian	**100**	17 (8.3)	17 (8.3)	188 (92.2)	188 (92.2)	25 (12.3)	188 (92.2)	189 (92.7)	0.53
Evaluation sample									
All girls	**200**	206 (17.6)	206 (17.6)	945 (80.7)	974 (83.2)	223 (19.0)	959 (81.9)	975 (83.3)	0.49
White	**100**	69 (12.8)	69 (12.8)	455 (84.1)	476 (88.0)	78 (14.4)	464 (85.8)	476 (88.0)	0.51
Hispanic	**300**	49 (19.8)	49 (19.8)	195 (79.0)	197 (79.8)	51 (20.7)	197 (79.8)	198 (80.2)	0.36
Black	**700**	61 (28.1)	60 (27.7)	157 (72.4)	161 (74.2)	63 (29.0)	159 (73.3)	161 (74.2)	0.54
Asian	**100**	9 (11.0)	9 (11.0)	72 (87.8)	73 (89.0)	10 (12.2)	73 (89.0)	73 (89.0)	0.47
Accumulated time									
Strong et al., 2005									
Development sample									
All girls	**100**	647 (18.4)	647 (18.4)	2880 (81.9)	2891 (82.2)	872 (24.8)	2884 (82.0)	2891 (82.2)	0.52
White	**100**	204 (12.4)	204 (12.4)	1448 (87.8)	1456 (88.3)	317 (19.2)	1451 (88.0)	1456 (88.3)	0.54
Hispanic	**100**	199 (25.8)	199 (25.8)	577 (74.9)	579 (75.2)	236 (30.7)	577 (74.9)	579 (75.2)	0.50
Black	**400**	177 (27.3)	177 (27.3)	475 (73.2)	476 (73.3)	202 (31.1)	476 (73.3)	476 (73.3)	0.51
Asian	**100**	17 (8.3)	17 (8.3)	189 (92.7)	189 (92.7)	51 (25.0)	189 (92.7)	189 (92.7)	0.47
Evaluation sample									
All girls	**100**	205 (17.5)	205 (17.5)	971 (82.9)	974 (83.2)	302 (25.8)	971 (82.9)	975 (83.3)	0.51
White	**100**	69 (12.8)	69 (12.8)	473 (87.4)	475 (87.8)	114 (21.1)	474 (87.6)	476 (88.0)	0.52
Hispanic	**100**	49 (19.8)	49 (19.8)	198 (80.2)	198 (80.2)	63 (25.5)	197 (79.8)	198 (80.2)	0.40
Black	**400**	61 (28.1)	60 (27.7)	160 (73.7)	161 (74.2)	80 (36.9)	160 (73.7)	161 (74.2)	0.54
Asian	**100**	9 (11.0)	9 (11.0)	73 (89.0)	73 (89.0)	21 (25.6)	73 (89.0)	73 (89.0)	0.45
2005 Dietary Guidelines									
Development sample									
All girls	**200**	642 (18.3)	642 (18.3)	2698 (76.7)	2888 (82.1)	656 (18.7)	2758 (78.4)	2891 (82.2)	0.55
White	**200**	201 (12.2)	201 (12.2)	1320 (80.1)	1454 (88.2)	208 (12.6)	1357 (82.3)	1456 (88.3)	0.56
Hispanic	**200**	197 (25.6)	197 (25.6)	542 (70.4)	579 (75.2)	200 (26.0)	558 (72.5)	579 (75.2)	0.53
Black	**800**	177 (27.3)	177 (27.3)	469 (72.3)	476 (73.3)	177 (27.3)	471 (72.6)	476 (73.3)	0.53
Asian	**200**	17 (8.3)	17 (8.3)	182 (89.2)	188 (92.2)	22 (10.8)	184 (90.2)	189 (92.7)	0.54

Evaluation sample									
All girls	**200**	205 (17.5)	203 (17.3)	908 (77.5)	973 (83.1)	213 (18.2)	936 (79.9)	974 (83.2)	0.53
White	**200**	69 (12.8)	69 (12.8)	433 (80.0)	475 (87.8)	73 (13.5)	449 (83.0)	475 (87.8)	0.55
Hispanic	**200**	49 (19.8)	49 (19.8)	192 (77.7)	197 (79.8)	50 (20.2)	195 (79.0)	198 (80.2)	0.42
Black	**800**	61 (28.1)	58 (26.7)	151 (69.6)	161 (74.2)	61 (28.1)	156 (71.9)	161 (74.2)	0.56
Asian	**200**	9 (11.0)	9 (11.0)	69 (84.2)	73 (89.0)	9 (11.0)	71 (86.6)	73 (89.0)	0.47
Healthy People 2010									
Development sample									
All girls	**100**	639 (18.2)	639 (18.2)	2191 (62.3)	2873 (81.7)	648 (18.4)	2350 (66.8)	2888 (82.1)	0.57
White	**100**	199 (12.1)	199 (12.1)	1027 (62.3)	1441 (87.4)	204 (12.4)	1113 (67.5)	1454 (88.2)	0.58
Hispanic	**200**	195 (25.3)	195 (25.3)	437 (56.8)	576 (74.8)	199 (25.8)	477 (62.0)	579 (75.2)	0.54
Black	**1100**	177 (27.3)	177 (27.3)	428 (66.0)	475 (73.2)	177 (27.3)	438 (67.5)	476 (73.3)	0.55
Asian	**300**	17 (8.3)	17 (8.3)	149 (73.0)	188 (92.2)	18 (8.8)	162 (79.4)	188 (92.2)	0.56
Evaluation sample									
All girls	**100**	201 (17.2)	201 (17.2)	729 (62.3)	969 (82.8)	205 (17.5)	803 (68.6)	974 (83.2)	0.56
White	**100**	69 (12.8)	69 (12.8)	328 (60.6)	471 (87.1)	69 (12.8)	369 (68.2)	475 (87.8)	0.58
Hispanic	**200**	49 (19.8)	49 (19.8)	156 (63.2)	198 (80.2)	49 (19.8)	172 (69.6)	198 (80.2)	0.44
Black	**1100**	60 (27.7)	57 (26.3)	136 (62.7)	160 (73.7)	60 (27.7)	141 (65.0)	161 (74.2)	0.58
Asian	**300**	9 (11.0)	9 (11.0)	63 (76.8)	73 (89.0)	9 (11.0)	65 (79.3)	73 (89.0)	0.48

[1] Misclassification includes false positives (i.e., low risk and inadequate physical activity) and false negatives (i.e., high risk and adequate physical activity).

[2] Treuth et al. (2004) cut points for light, moderate, and vigorous intensity physical activity were 101, 3000, and 5201 counts/min, respectively. Puyau et al. (2002) cut points for light, moderate, and vigorous intensity physical activity were 800, 3200, and 8200 counts/min, respectively.

Appendix H. Concordance correlation coefficient (r_c), location shift (u), scale shift (v), bias correction factor (C_b), and Pearson's correlation coefficient (r) for minutes of physical activity determined using new accelerometer cut points and comparison measures by intensity and population subgroup in NHANES (N = 333)

Comparison / Intensity / Subgroup	Cut point (counts/min)	r_c (95% CI)	u	v	C_b	r
Questionnaire[1]						
Moderate						
12-13 y	4000-4999	0.00 (-0.027, 0.024)	-1.48	0.10	0.16	-0.01
	4300-9999	0.00 (-0.058, 0.058)	-0.87	0.23	0.37	0.00
Vigorous						
12-13 y	5000	0.01 (-0.020, 0.031)	-1.64	0.10	0.16	0.03
	10000	0.00 (-0.008, 0.007)	-3.19	0.03	0.05	-0.01
Moderate-to-vigorous						
All girls	100	0.01 (-0.006, 0.020)	3.92	0.98	0.12	0.06
	400	0.02 (-0.013, 0.050)	2.22	0.72	0.28	0.06
12-13 y	400	0.01 (-0.039, 0.051)	2.19	0.70	0.29	0.02
	4000	0.00 (-0.022, 0.031)	-1.77	0.12	0.17	0.03
	4300	0.00 (-0.019, 0.027)	-1.93	0.10	0.15	0.03
14-15 y	100	0.01 (-0.007, 0.028)	4.06	1.01	0.11	0.10
White	400	0.05 (-0.026, 0.118)	1.99	0.64	0.32	0.14
	700	0.10 (-0.022, 0.214)	1.10	0.50	0.54	0.18
	1200	0.13 (-0.011, 0.263)	0.20	0.37	0.64	0.20
	1800	0.09 (-0.011, 0.192)	-0.46	0.26	0.47	0.19
	2300	0.05 (-0.027, 0.132)	-0.84	0.22	0.36	0.14
Treuth cut point[2]						
Light						
12-13 y	1900-3999	0.01 (0.007, 0.014)	-9.33	0.23	0.02	0.48
	1900-4299	0.01 (0.007, 0.015)	-9.04	0.25	0.02	0.47
Moderate						
12-13 y	4000-4999	0.34 (0.294, 0.396)	-1.55	0.40	0.38	0.91
	4300-9999	0.56 (0.481, 0.631)	-0.86	0.84	0.72	0.77
Vigorous						
12-13 y	≥5000	0.99 (0.994, 0.996)	0.06	1.06	1.00	1.00
	≥10000	0.31 (0.241, 0.370)	-0.78	0.33	0.51	0.60
Light-to-vigorous						
All girls	100	1.00 (1.000, 1.000)	0.01	1.00	1.00	1.00
	400	0.24 (0.209, 0.268)	-2.37	0.74	0.26	0.92
12-13 y	400	0.24 (0.201, 0.284)	-2.37	0.79	0.26	0.93
	1900	0.02 (0.011, 0.022)	-7.94	0.30	0.03	0.55
14-15 y	100	1.00 (1.000, 1.000)	0.01	1.00	1.00	1.00
White	400	0.23 (0.175, 0.291)	-2.42	0.73	0.25	0.93
	700	0.09 (0.062, 0.120)	-3.93	0.58	0.11	0.81
	1200	0.03 (0.021, 0.049)	-5.75	0.42	0.06	0.63
	1800	0.02 (0.008, 0.024)	-7.56	0.30	0.03	0.50
	2300	0.01 (0.005, 0.017)	-8.70	0.25	0.03	0.43

Comparison / Intensity / Subgroup	Cut point (counts/min)	r_c (95% CI)	u	v	C_b	r
Moderate-to-vigorous						
All girls	100	0.00 (0.003, 0.006)	10.98	5.99	0.02	0.30
	400	0.02 (0.013, 0.022)	6.91	4.45	0.04	0.47
12-13 y	400	0.02 (0.013, 0.027)	6.78	4.23	0.04	0.50
	4000	0.71 (0.656, 0.755)	-0.74	0.70	0.75	0.94
	4300	0.60 (0.543, 0.657)	-0.92	0.62	0.65	0.92
14-15 y	100	0.00 (0.000, 0.005)	11.61	6.60	0.01	0.16
White	400	0.02 (0.012, 0.034)	6.34	3.81	0.05	0.51
	700	0.06 (0.034, 0.077)	4.42	3.00	0.09	0.63
	1200	0.17 (0.119, 0.222)	2.56	2.18	0.22	0.78
	1800	0.45 (0.370, 0.540)	1.33	1.56	0.50	0.90
	2300	0.77 (0.713, 0.833)	0.66	1.28	0.80	0.96
Puyau cut point[3]						
Light						
12-13 y	1900-3999	0.13 (0.104, 0.161)	-3.07	0.43	0.16	0.81
	1900-4299	0.14 (0.111, 0.172)	-2.93	0.46	0.18	0.80
Moderate						
12-13 y	4000-4999	0.35 (0.308, 0.400)	-1.36	0.31	0.37	0.95
	4300-9999	0.69 (0.636, 0.736)	-0.75	0.64	0.72	0.95
Vigorous						
12-13 y	≥5000	0.36 (0.290, 0.430)	0.69	2.62	0.57	0.63
	≥10000	0.96 (0.948, 0.966)	-0.16	0.82	0.97	0.99
Light to vigorous						
All girls	100	0.08 (0.063, 0.088)	4.25	1.83	0.10	0.77
	400	0.38 (0.337, 0.413)	1.71	1.36	0.40	0.94
12-13 y	400	0.39 (0.334, 0.440)	1.69	1.33	0.40	0.96
	1900	0.18 (0.147, 0.218)	-2.62	0.51	0.21	0.85
14-15 y	100	0.06 (0.044, 0.075)	4.54	2.01	0.09	0.69
White	400	0.36 (0.283, 0.433)	1.76	1.36	0.38	0.93
	700	0.93 (0.914, 0.954)	0.36	1.07	0.94	1.00
	1200	0.57 (0.495, 0.653)	-1.14	0.78	0.59	0.97
	1800	0.22 (0.161, 0.277)	-2.37	0.55	0.25	0.87
	2300	0.13 (0.089, 0.169)	-3.10	0.46	0.16	0.79
Moderate to vigorous						
All girls	100	0.00 (0.002, 0.006)	11.51	6.50	0.01	0.28
	400	0.02 (0.011, 0.019)	7.28	4.83	0.03	0.44
12-13 y	400	0.02 (0.011, 0.023)	7.11	4.54	0.04	0.47
	4000	0.81 (0.775, 0.847)	-0.55	0.75	0.84	0.97
	4300	0.71 (0.658, 0.755)	-0.72	0.66	0.74	0.95
14-15 y	100	0.00 (0.000, 0.004)	12.28	7.29	0.01	0.15
White	400	0.02 (0.010, 0.030)	6.58	4.02	0.04	0.49
	700	0.05 (0.029, 0.069)	4.62	3.16	0.08	0.61
	1200	0.15 (0.101, 0.195)	2.72	2.30	0.20	0.75
	1800	0.39 (0.311, 0.478)	1.48	1.64	0.45	0.88
	2300	0.69 (0.619, 0.766)	0.80	1.35	0.73	0.94

[1] Moderate, vigorous, and moderate-to-vigorous intensity in the NHANES questionnaire were defined as 3 to <6 METS, ≥6 METS, and ≥3 METS, respectively.

[2] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 101-2999, 3000-5200, 5201, 101, and 3000 counts/minute, respectively.

[3] Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800-3199, 3200-8199, 8200, 800, and 3200 counts/min, respectively.

Appendix I. **Concordance correlation coefficient (r_c), location shift (u), scale shift (v), bias correction factor (C_b), and Pearson's correlation coefficient (r) of minutes of physical activity determined using new accelerometer cut points and comparison measures by intensity and population subgroup in TAAG (N = 4696)**

Comparison / Intensity / Subgroup	Cut point (counts/30 sec)	r_c (95% CI)	u	v	C_b	r
3DPAR[1]						
Light						
Black	1050-1549	0.00 (-0.021, 0.019)	-1.92	0.37	0.30	0.00
	1450-1949	0.00 (-0.009, 0.009)	-3.19	0.24	0.14	0.00
Moderate						
Black	1550-2049	0.02 (0.001, 0.031)	-1.78	0.18	0.22	0.07
	1950-2449	0.01 (-0.002, 0.015)	-2.64	0.11	0.13	0.05
Vigorous						
Black	2050	0.01 (-0.006, 0.030)	-1.65	0.22	0.27	0.05
	2450	0.01 (-0.004, 0.019)	-2.29	0.15	0.17	0.05
Light-to-vigorous						
Overall	50	0.01 (0.010, 0.016)	4.16	1.85	0.10	0.13
White	50	0.01 (0.011, 0.018)	4.38	1.99	0.09	0.16
Black	100	0.02 (0.010, 0.035)	2.95	1.56	0.18	0.12
	1050	0.05 (0.027, 0.069)	-1.91	0.42	0.31	0.15
	1450	0.02 (0.013, 0.034)	-3.08	0.28	0.15	0.16
Moderate-to-vigorous						
Overall	50	0.01 (0.007, 0.011)	5.38	2.15	0.06	0.14
White	50	0.01 (0.007, 0.012)	5.60	2.33	0.06	0.17
Black	100	0.01 (0.006, 0.019)	4.14	1.83	0.10	0.12
	1550	0.07 (0.044, 0.093)	-1.32	0.29	0.37	0.19
	1950	0.04 (0.025, 0.053)	-2.08	0.19	0.21	0.19
Treuth cut point[2]						
Light						
Black	1050-1549	0.01 (0.008, 0.011)	-11.01	0.12	0.02	0.61
	1450-1949	0.00 (0.004, 0.005)	-13.99	0.08	0.01	0.46
Moderate						
Black	1550-2049	0.69 (0.668, 0.708)	-0.77	0.60	0.70	0.98
	1950-2449	0.32 (0.297, 0.338)	-1.74	0.38	0.33	0.96
Vigorous						
Black	2050	0.63 (0.598, 0.655)	-0.80	0.74	0.73	0.85
	2450	0.31 (0.283, 0.336)	-1.49	0.49	0.42	0.73
Light-to-vigorous						
Overall	50	1.00 (1.000, 1.000)	0.03	1.00	1.00	1.00
White	50	1.00 (1.000, 1.000)	0.03	1.00	1.00	1.00
Black	100	1.00 (1.000, 1.000)	0.02	1.00	1.00	1.00
	1050	0.02 (0.019, 0.024)	-7.34	0.24	0.03	0.63
	1450	0.01 (0.009, 0.012)	-9.58	0.16	0.02	0.53
Moderate-to-vigorous						
Overall	50	0.01 (0.010, 0.011)	9.43	5.63	0.02	0.49
White	50	0.01 (0.009, 0.012)	9.20	5.40	0.02	0.47
Black	100	0.01 (0.013, 0.017)	8.35	5.90	0.03	0.56
	1550	0.99 (0.990, 0.992)	-0.12	0.95	0.99	1.00
	1950	0.62 (0.592, 0.639)	-0.94	0.62	0.64	0.96
Puyau cut point[3]						
Light						
Black	1050-1549	0.06 (0.058, 0.070)	-4.77	0.27	0.07	0.86
	1450-1949	0.03 (0.025, 0.032)	-6.56	0.18	0.04	0.69

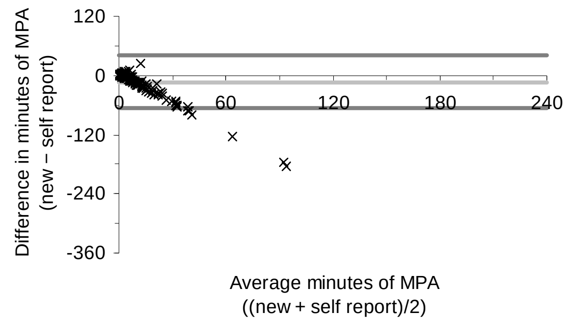
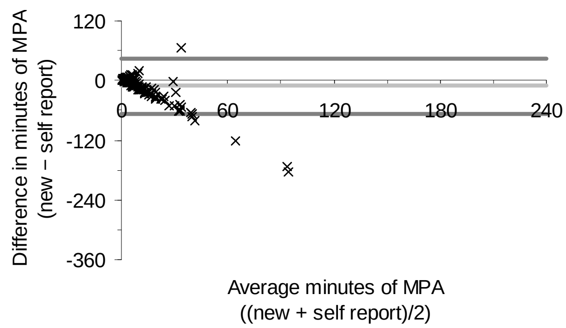
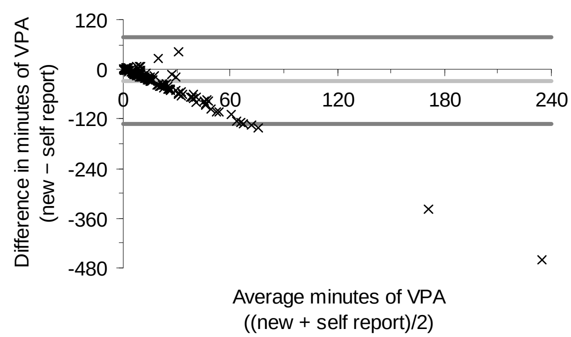
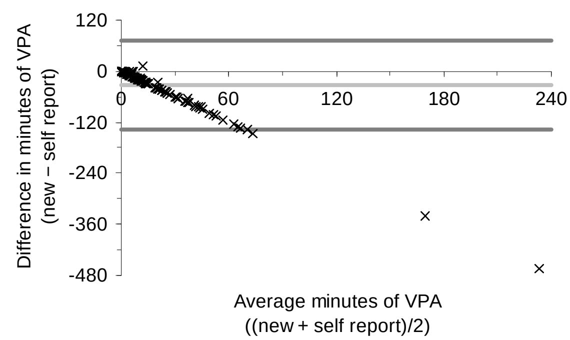
Comparison / Intensity / Subgroup	Cut point (counts/30 sec)	r_c (95% CI)	u	v	C_b	r
Moderate						
Black	1550-2049	0.63 (0.606, 0.652)	-0.78	0.54	0.67	0.94
	1950-2449	0.31 (0.294, 0.334)	-1.71	0.34	0.32	0.97
Vigorous						
Black	2050	0.67 (0.647, 0.693)	-0.80	0.67	0.71	0.94
	2450	0.35 (0.326, 0.375)	-1.46	0.45	0.41	0.85
Light-to-vigorous						
Overall	50	0.11 (0.105, 0.114)	3.57	1.74	0.13	0.82
White	50	0.11 (0.107, 0.121)	3.45	1.64	0.14	0.81
Black	100	0.19 (0.178, 0.210)	2.65	1.74	0.21	0.91
	1050	0.16 (0.142, 0.169)	-2.93	0.47	0.18	0.87
	1450	0.07 (0.061, 0.075)	-4.39	0.31	0.09	0.77
Moderate-to-vigorous						
Overall	50	0.01 (0.008, 0.010)	9.94	6.13	0.02	0.47
White	50	0.01 (0.008, 0.010)	9.67	5.85	0.02	0.46
Black	100	0.01 (0.011, 0.014)	8.92	6.59	0.02	0.54
	1550	0.99 (0.990, 0.992)	0.11	1.06	0.99	1.00
	1950	0.74 (0.723, 0.761)	-0.70	0.69	0.76	0.98

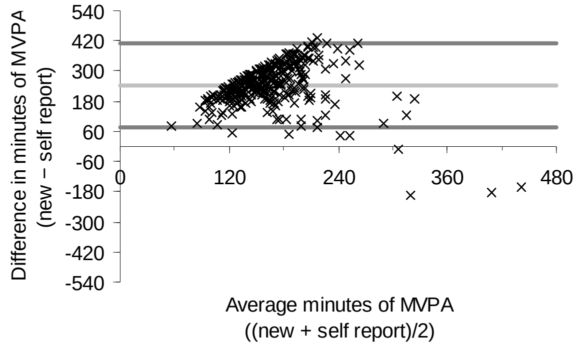
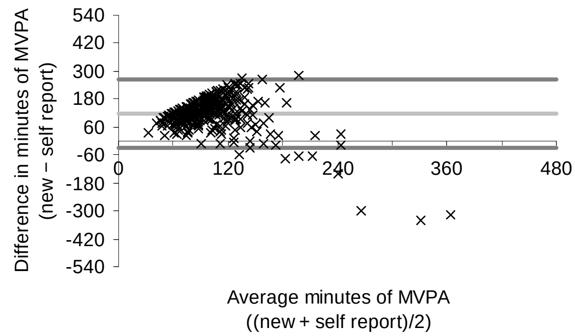
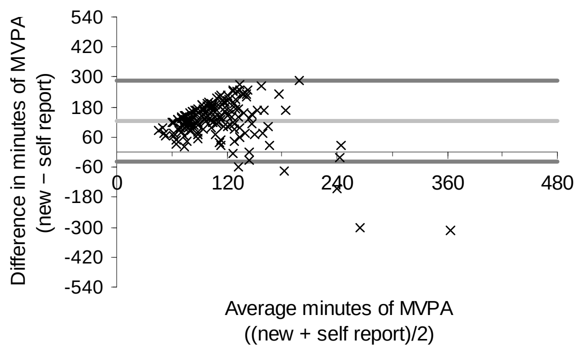
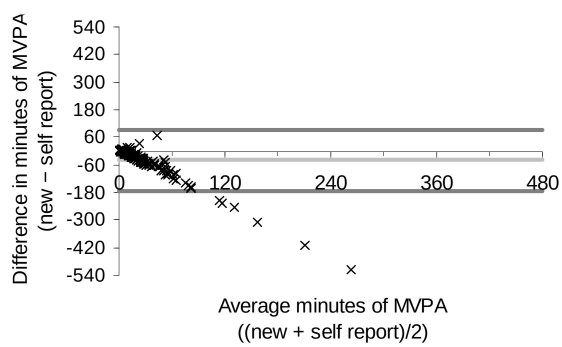
[1] Light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous intensity in the 3-day physical activity recall (3DPAR) were defined as 2 to <3 METS, 3 to <6 METS, ≥6 METS, ≥2 METS, and ≥3 METS, respectively.

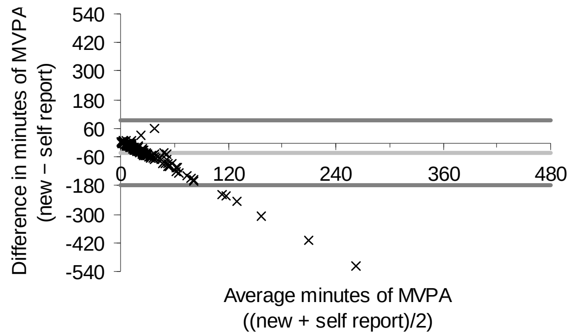
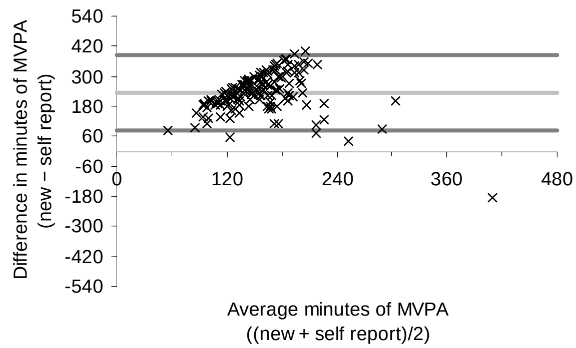
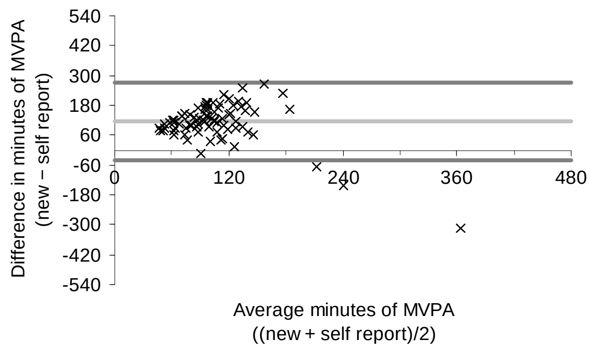
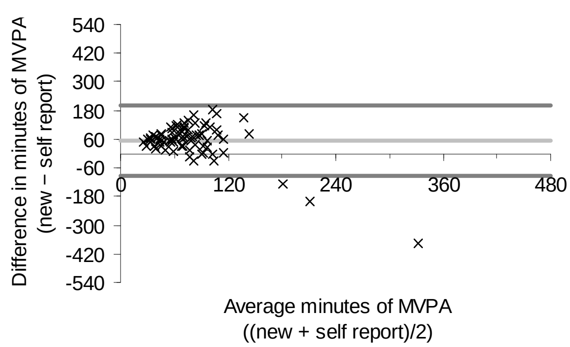
[2] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 51-1499, 1500-2600, 2601, 51, and 1500 counts/30 sec, respectively.

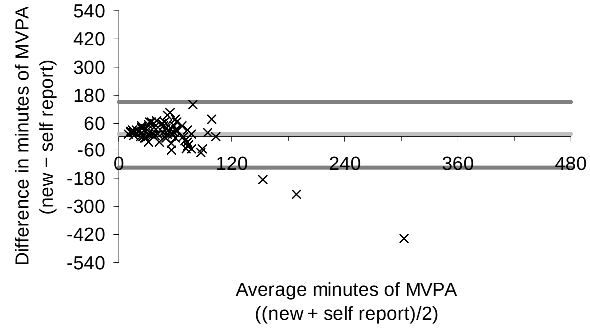
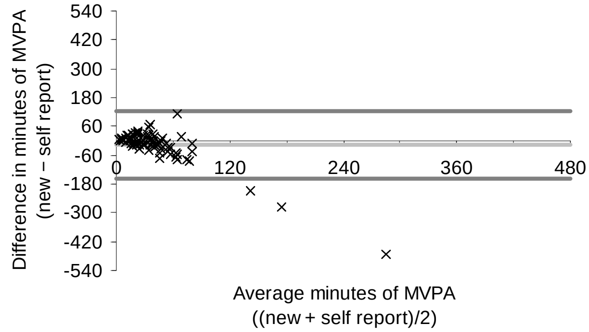
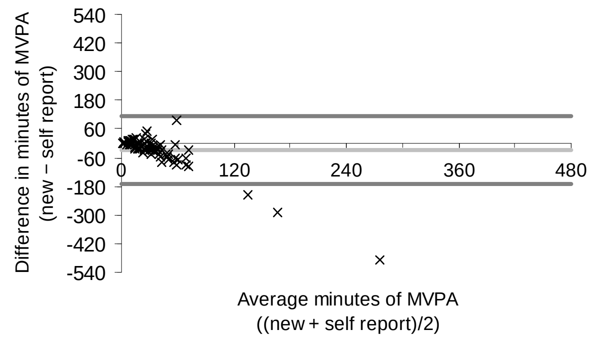
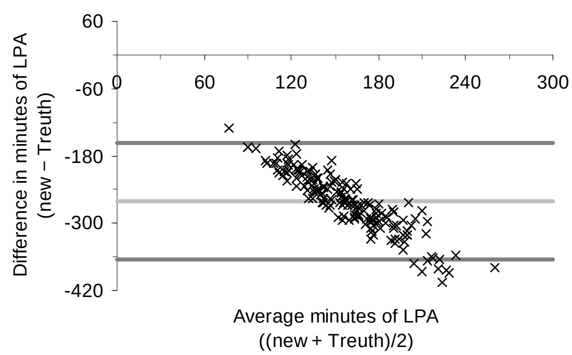
[3] Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800, 3200, and 8200 counts/min, respectively. The cut points used were 400-1599, 1600-4099, and 4100, 400, and 1600 counts/30 sec, respectively.

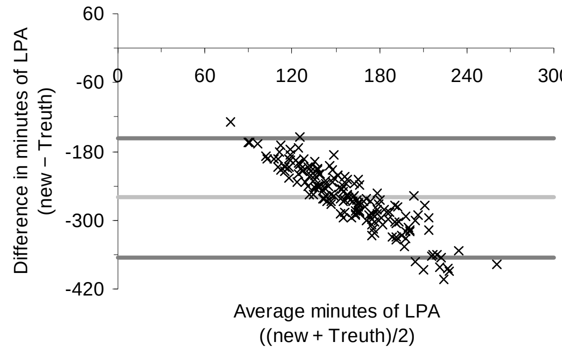
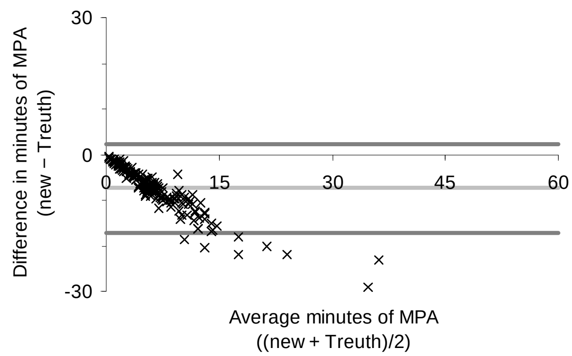
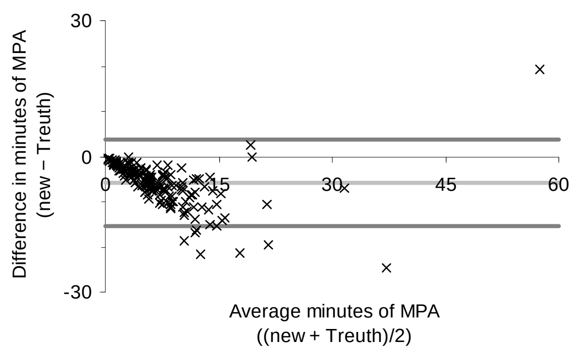
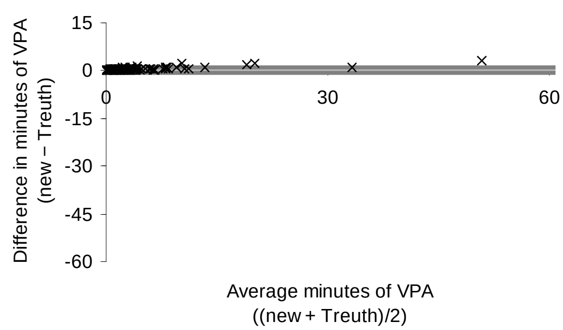
Appendix J. **Bland-Altman plots graphically presenting differences in minutes of physical activity between new cut points and comparison measures by average minutes of activity and mean difference (light line) ± 1.96 × SD (dark lines) by intensity and population subgroup in NHANES**

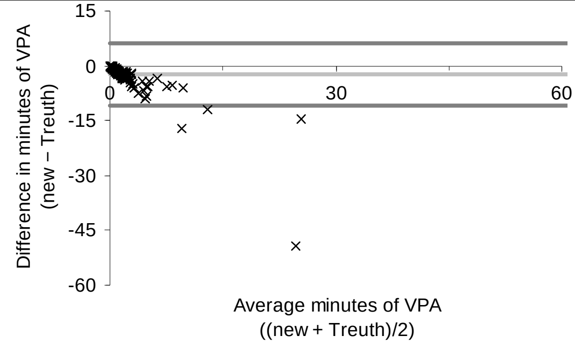
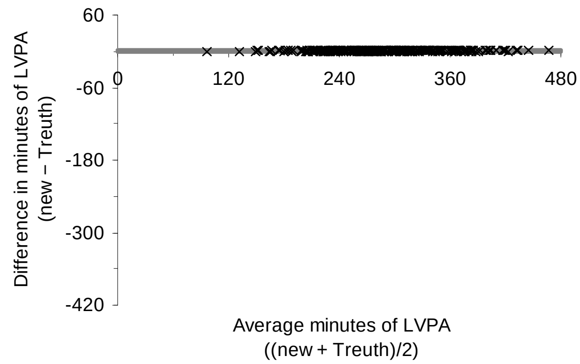
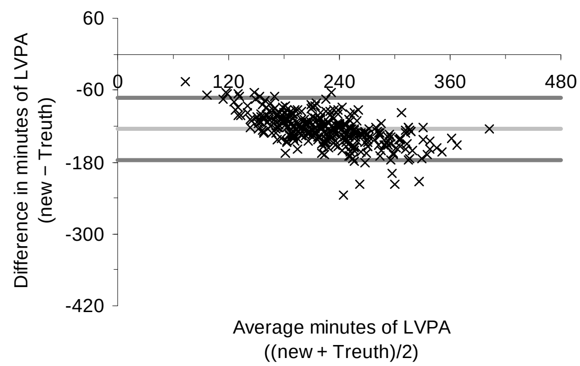
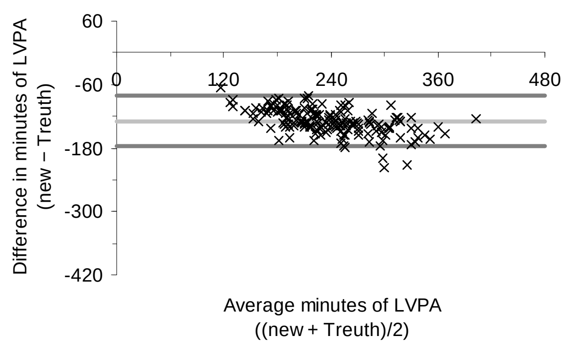
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
NHANES questionnaire		
Moderate (3 to less than 6 METS)		
12-13 y	**4000-4999**	
	4300-9999	
Vigorous (6 or more METS)		
12-13 y	**5000**	
	10000	

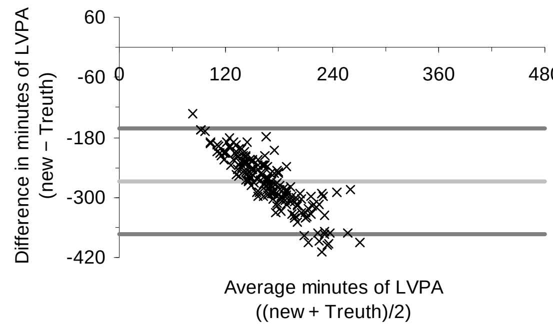
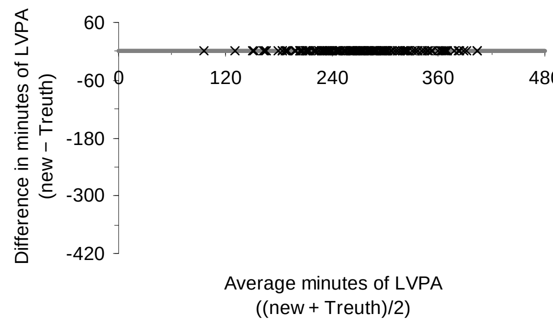
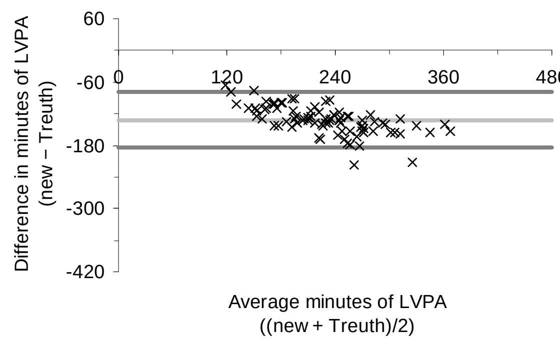
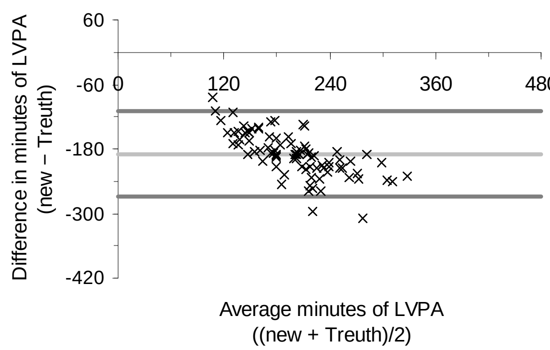
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
Moderate-to-vigorous (6 or more METS)		
All girls	**100**	
	400	
12-13 y	**400**	
	4000	

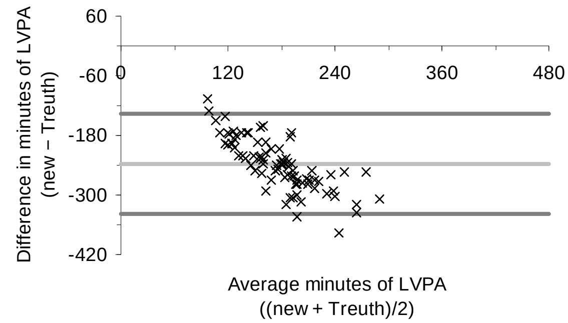
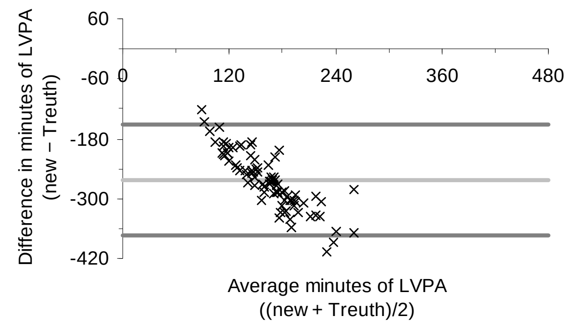
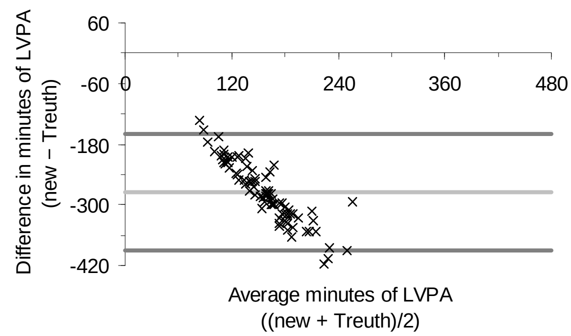
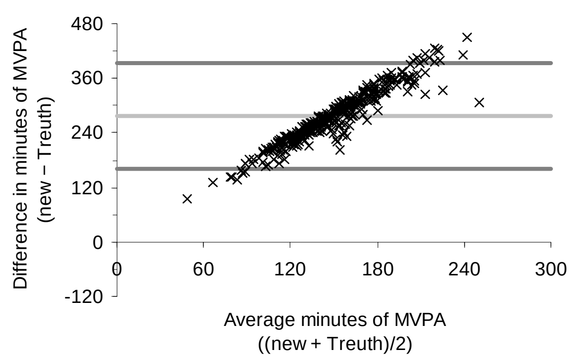
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	4300	
14-15 y	**100**	
White	**400**	
	700	

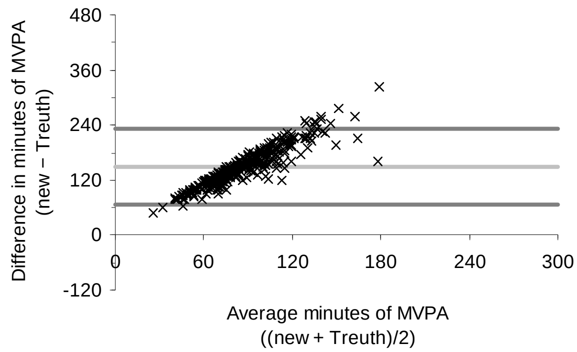
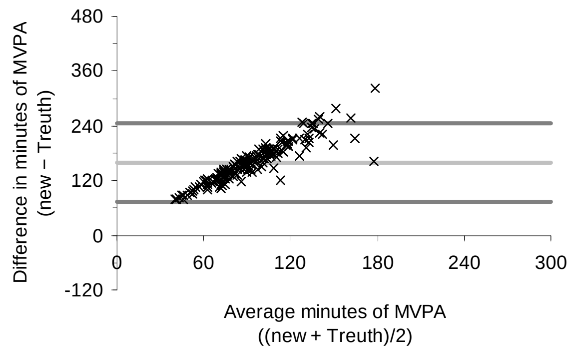
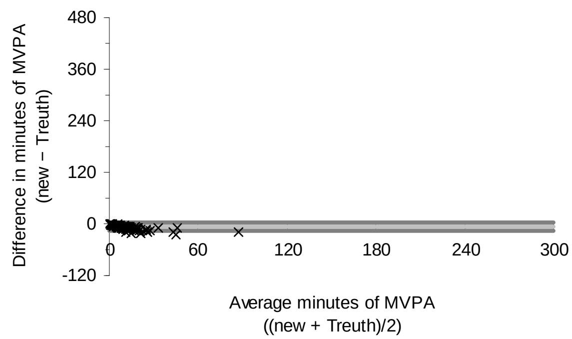
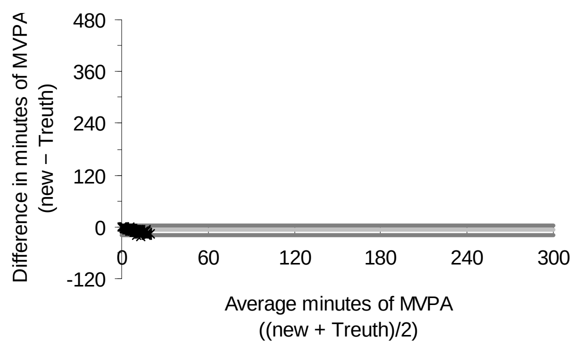
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	1200	
	1800	
	2300	
Treuth cut point ***Light (101-2999 counts/min)***		
12-13 y	**1900-3999**	

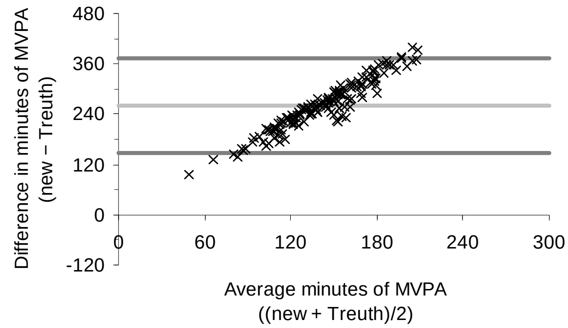
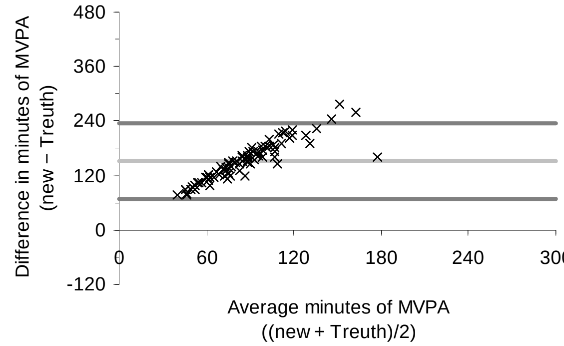
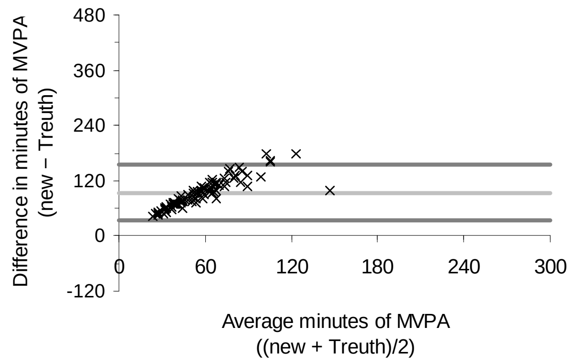
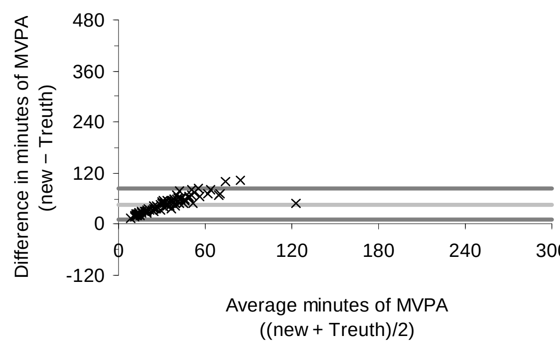
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	1900-4299	
Moderate (3000-5200 counts/min)		
12-13 y	4000-4999	
	4300-9999	
Vigorous (5201 counts/min)		
12-13 y	5000	

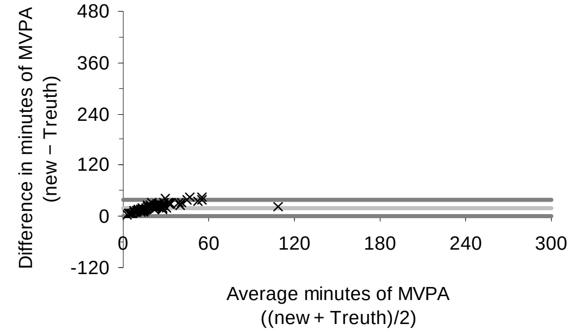
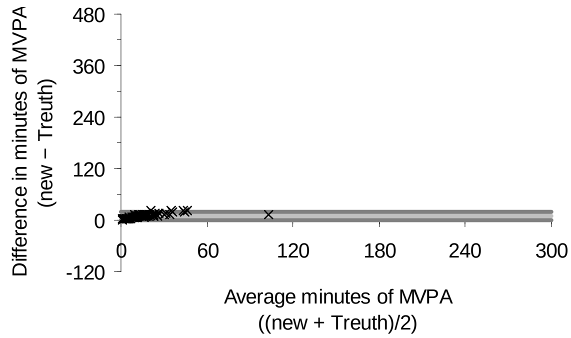
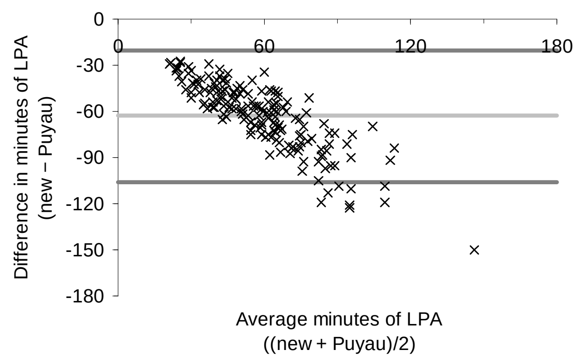
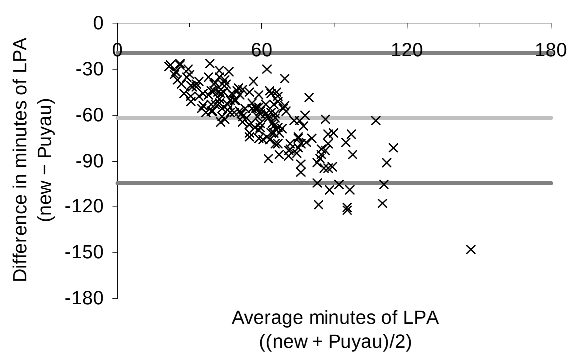
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	10000	
Light -to-vigorous (101 counts/min)		
All girls	**100**	
	400	
12-13 y	**400**	

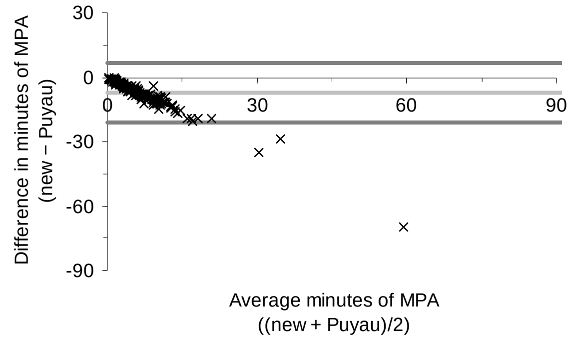
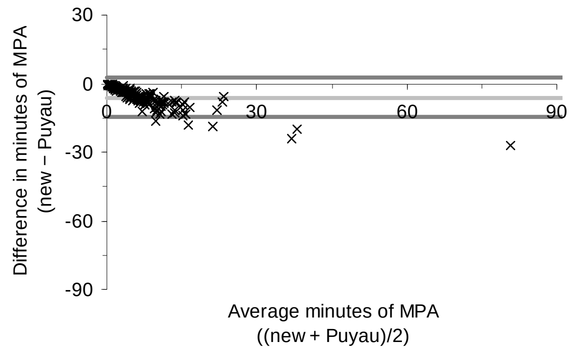
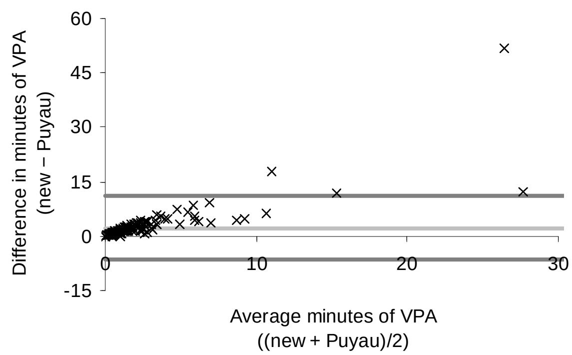
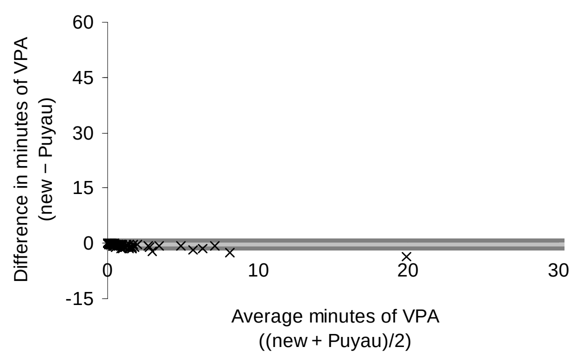
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	1900	
14-15 y	**100**	
White	**400**	
	700	

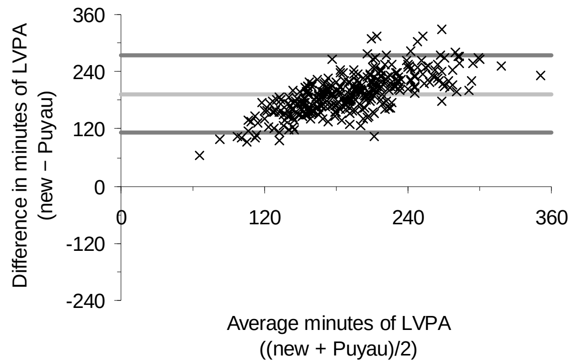
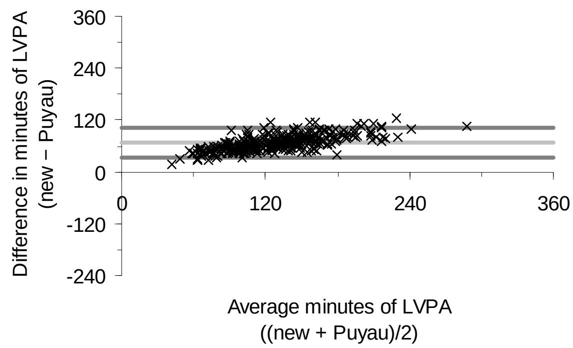
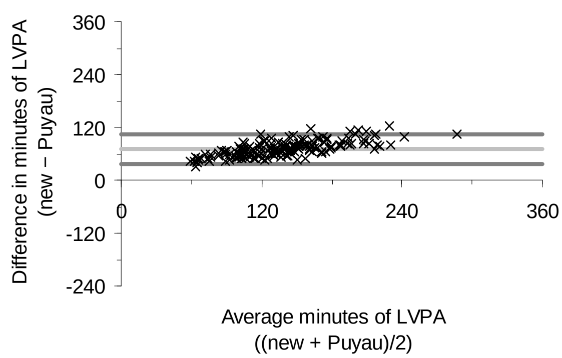
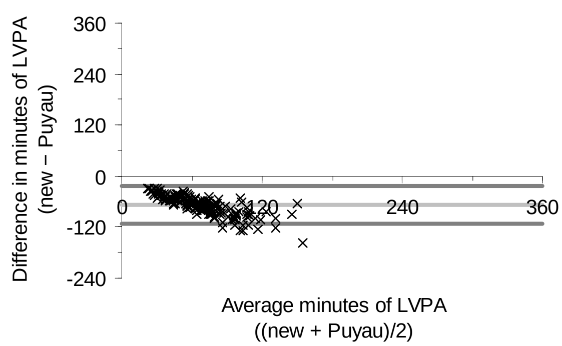
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	1200	
	1800	
	2300	
Moderate-to-vigorous (3000 counts/min)		
All girls	**100**	

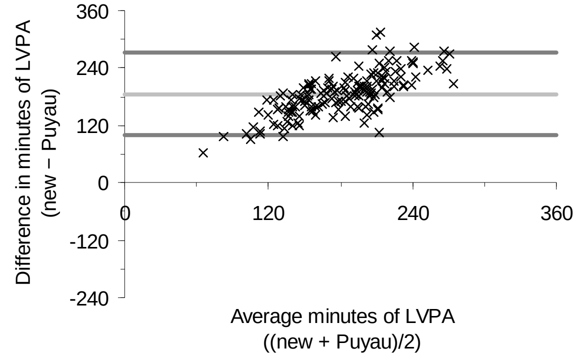
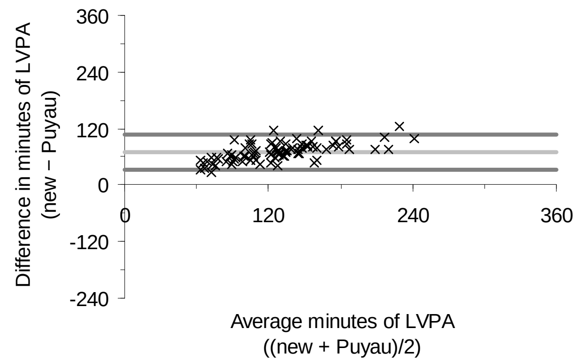
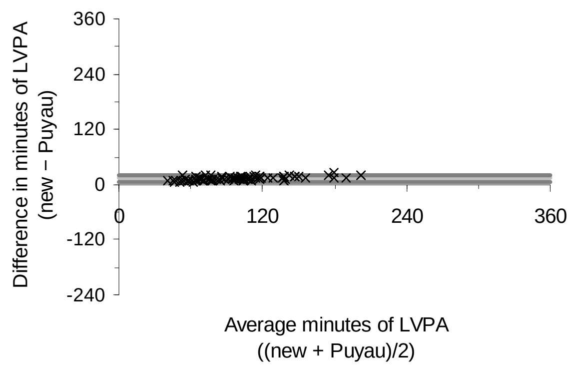
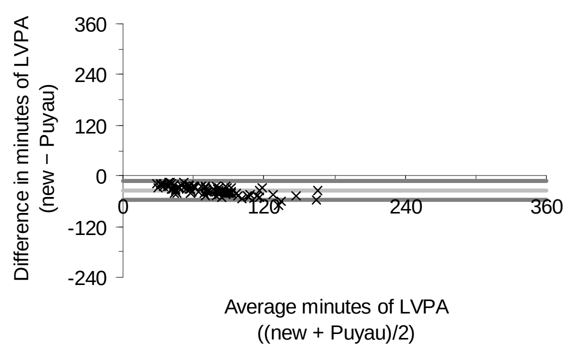
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	400	
12-13 y	**400**	
	4000	
	4300	

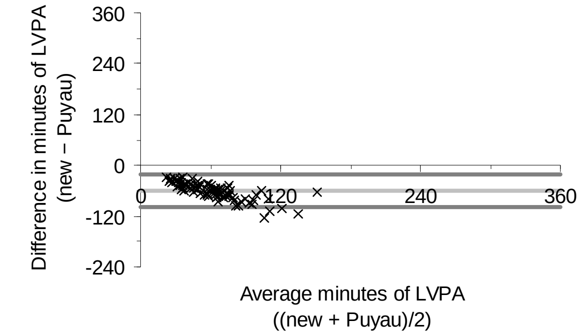
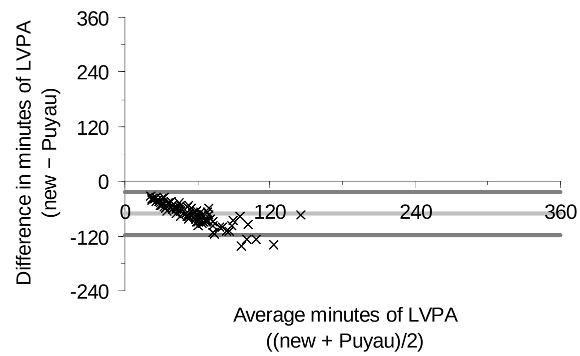
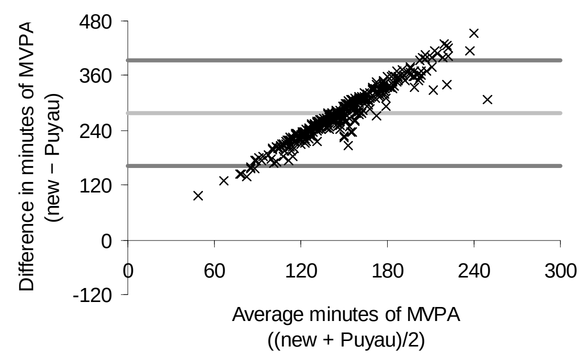
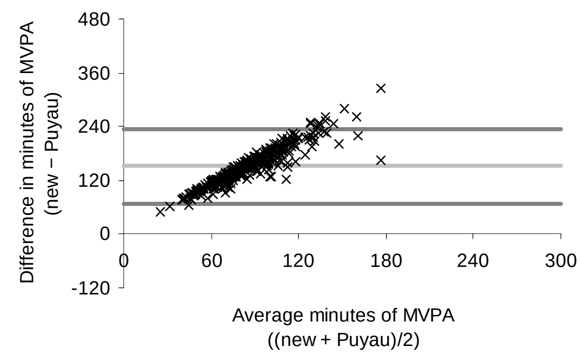
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
14-15 y	**100**	
White	**400**	
	700	
	1200	

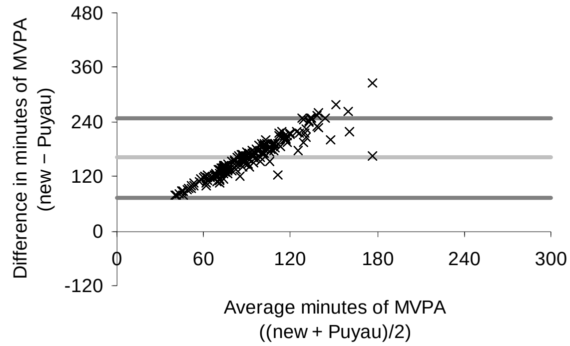
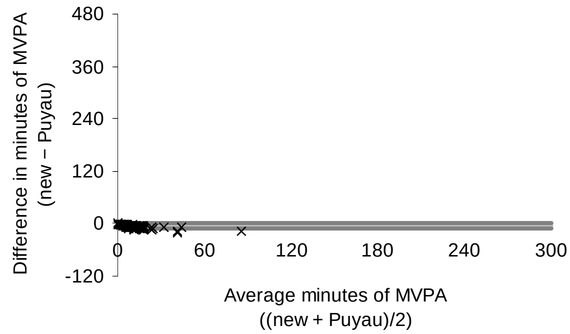
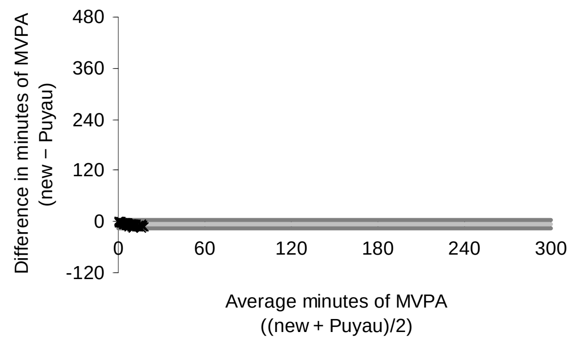
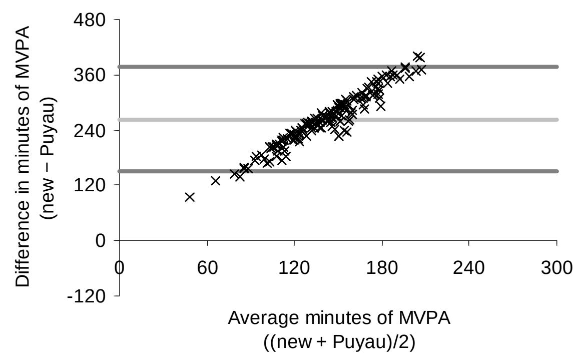
<table>
<tr><th>Comparison / Intensity / Subgroup</th><th>New cut point(s) (counts/min)</th><th>Plot</th></tr>
<tr><td></td><td>1800</td><td>
</td></tr>
<tr><td></td><td>2300</td><td>
</td></tr>
<tr><td>Puyau cut point
Light (400-1599 counts/30 sec)</td><td></td><td></td></tr>
<tr><td>12-13 y</td><td>1900-3999</td><td>
</td></tr>
<tr><td></td><td>1900-4299</td><td>
</td></tr>
</table>

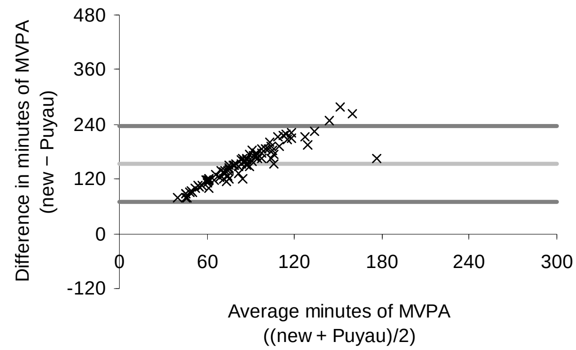
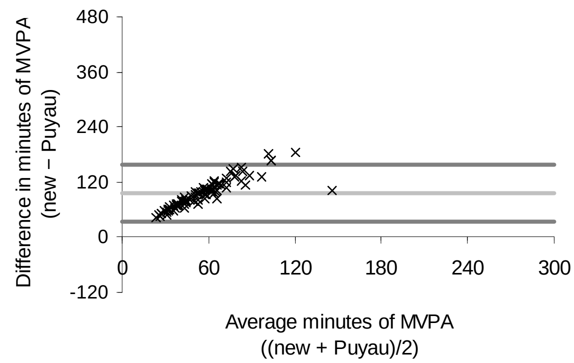
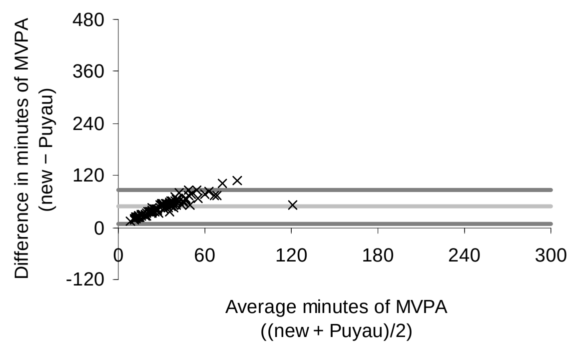
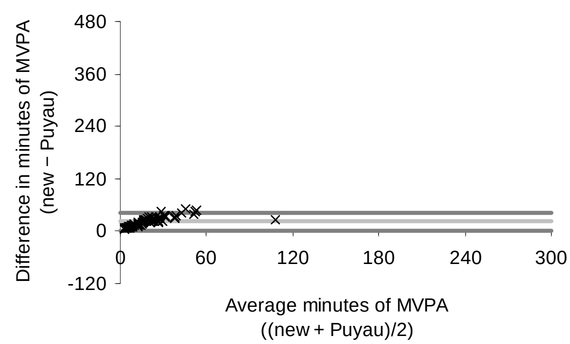
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
Moderate (1600-4099 counts/30 sec)		
12-13 y	**4000-4999**	
	4300-9999	
Vigorous (4100 counts/30 sec)		
12-13 y	**5000**	
	10000	

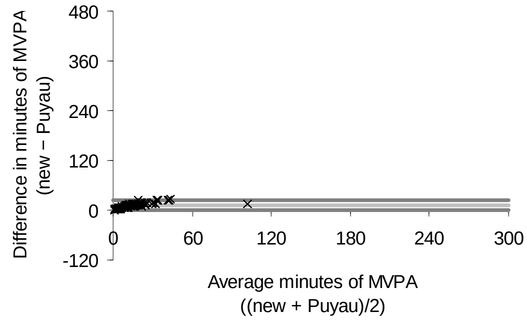
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
Light-to-vigorous (400 counts/30 sec)		
All girls	**100**	
	400	
12-13 y	**400**	
	1900	

Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
14-15 y	**100**	
White	**400**	
	700	
	1200	

<table>
<tr><th>Comparison / Intensity / Subgroup</th><th>New cut point(s) (counts/min)</th><th>Plot</th></tr>
<tr><td></td><td>**1800**</td><td>
</td></tr>
<tr><td></td><td>**2300**</td><td>
</td></tr>
<tr><td colspan="3">***Moderate-to-vigorous (1600 counts/30 sec)***</td></tr>
<tr><td>**All girls**</td><td>**100**</td><td>
</td></tr>
<tr><td></td><td>**400**</td><td>
</td></tr>
</table>

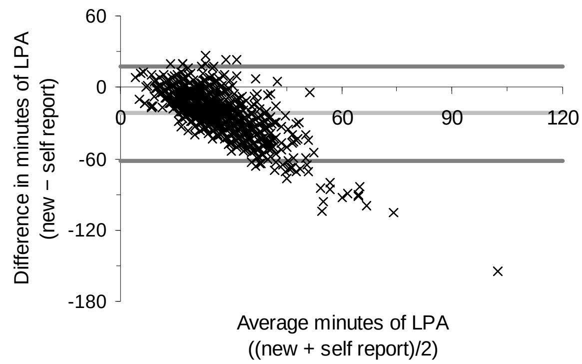
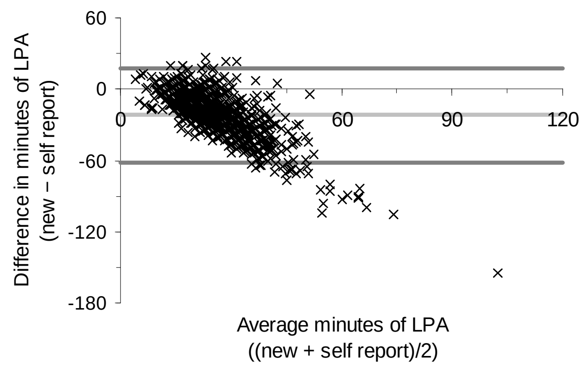
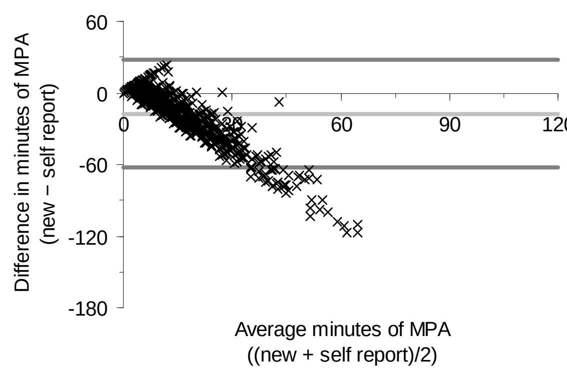
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
12-13 y	**400**	
	4000	
	4300	
14-15 y	**100**	

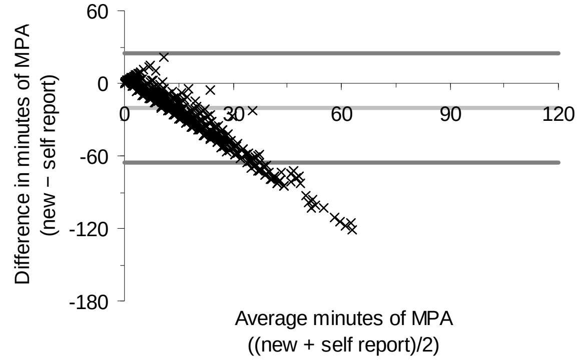
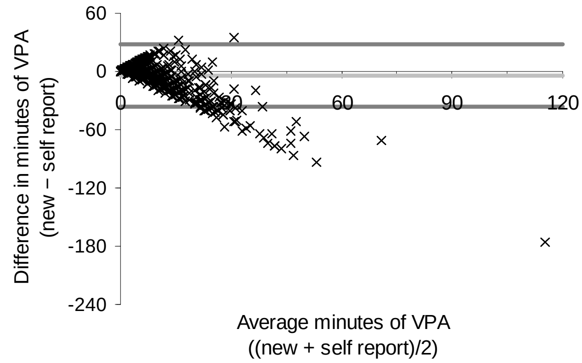
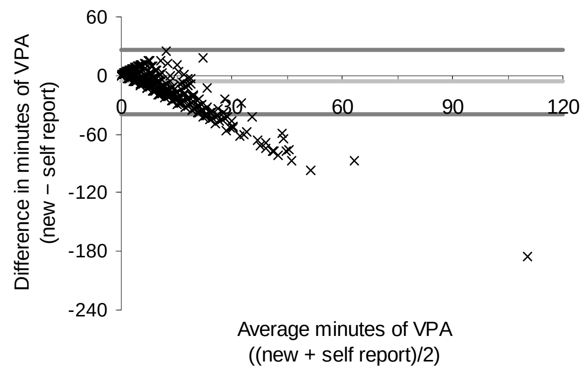
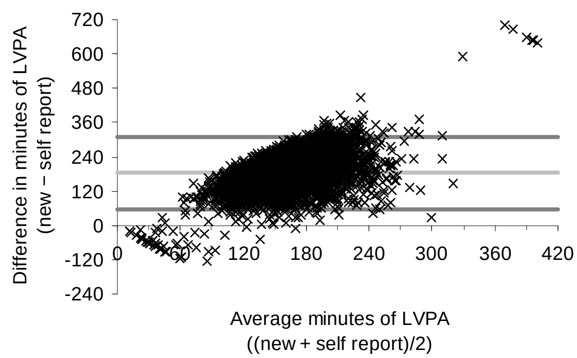
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
White	**400**	
	700	
	1200	
	1800	

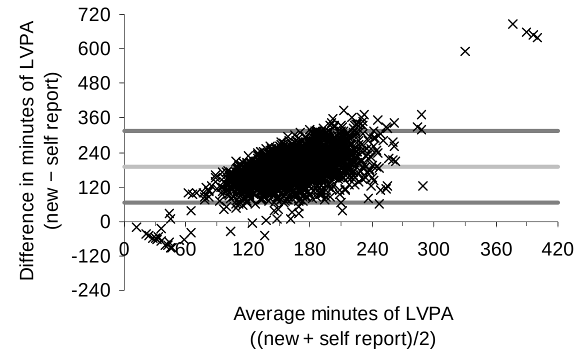
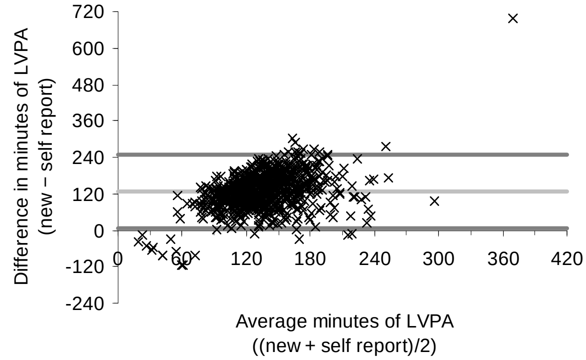
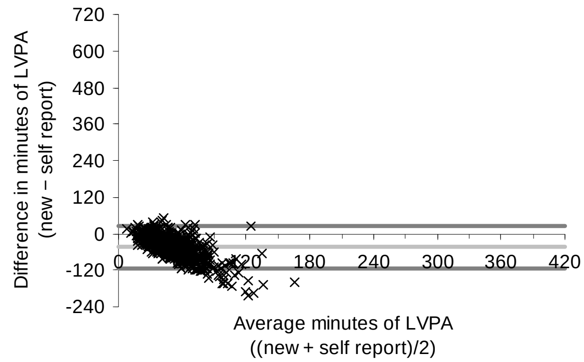
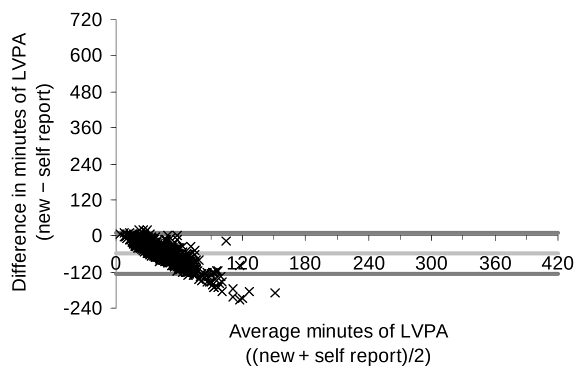
Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Plot
	2300	

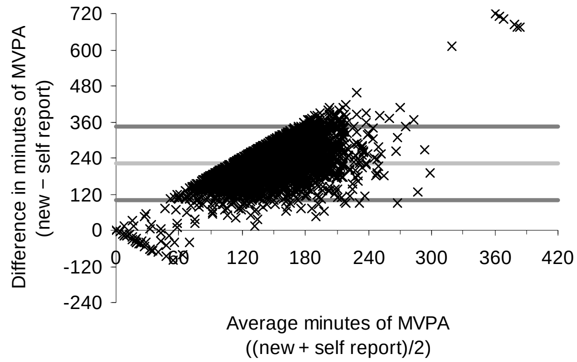
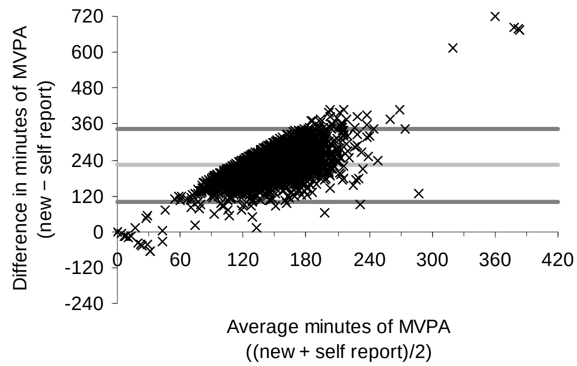
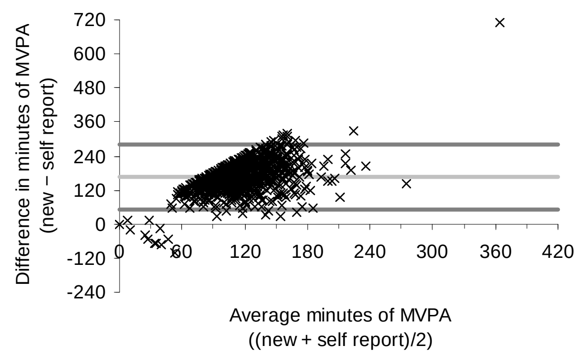
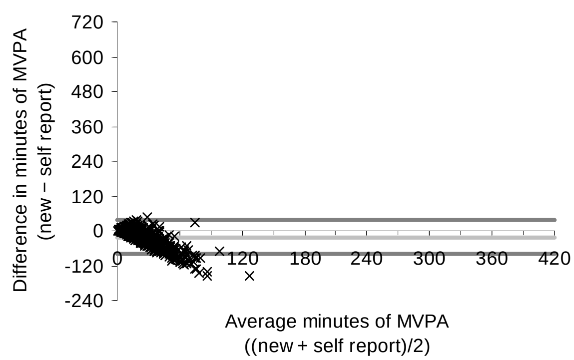
NHANES: National Health and Nutrition Examination Survey; LPA: light physical activity, MPA: moderate physical activity; VPA: vigorous physical activity; LVPA: light-to-vigorous physical activity; MVPA: moderate-to-vigorous physical activity

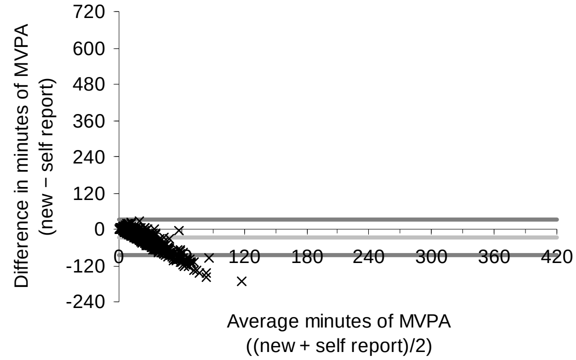
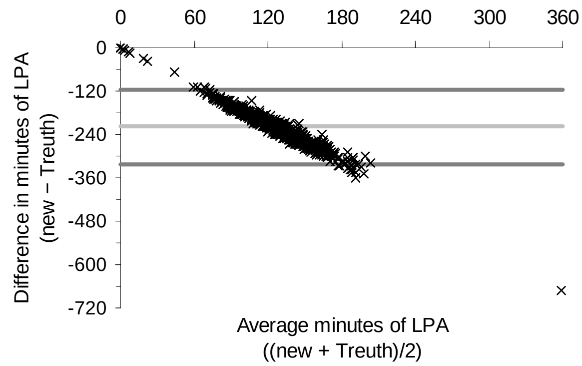
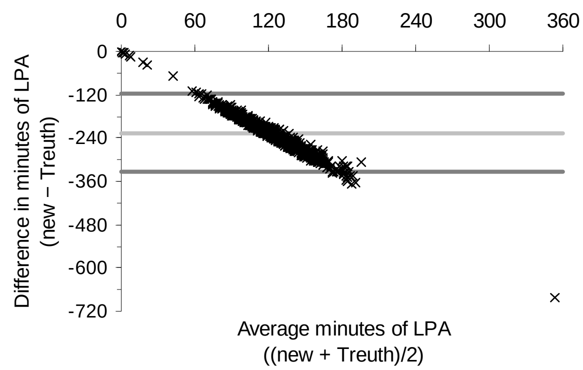
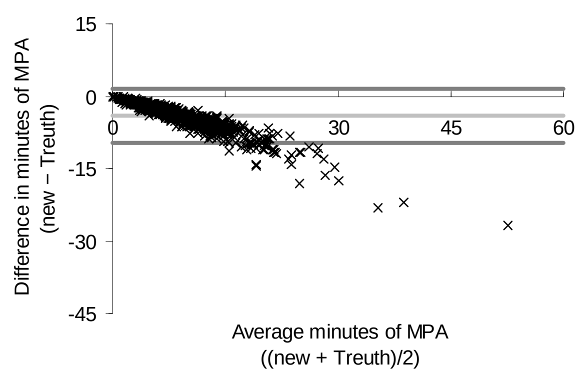
Appendix K. Bland-Altman plots graphically presenting differences in minutes of physical activity between new cut points and comparison measures by average minutes of activity and mean difference (light line) ± 1.96 × SD (dark lines) by intensity and population subgroup in TAAG

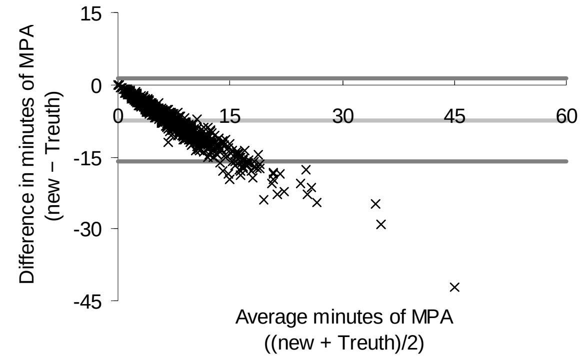
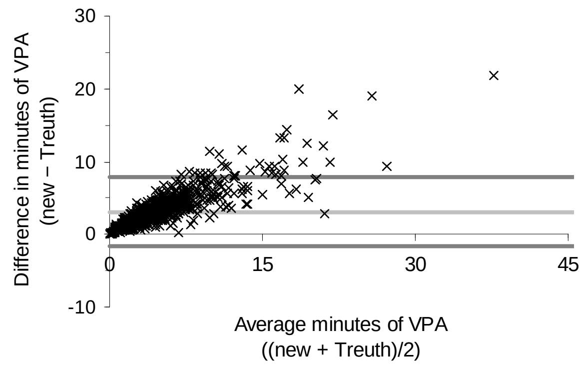
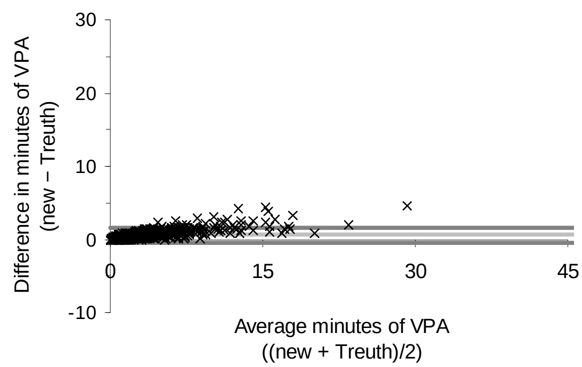
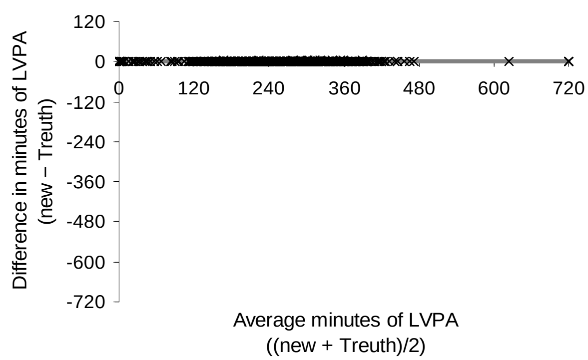
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
3DPAR		
Light (2 to less than 3 METS)		
Black	1050-1549	
	1450-1949	
Moderate (3 to less than 6 METS)		
Black	1550-2049	

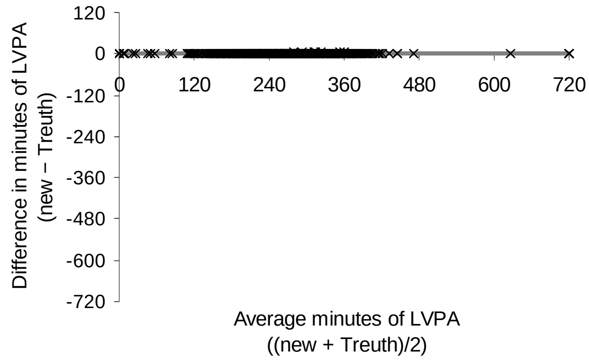
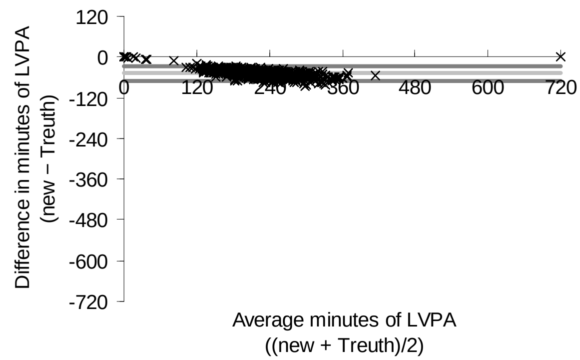
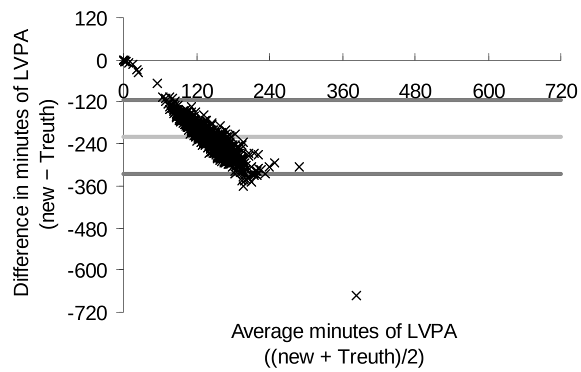
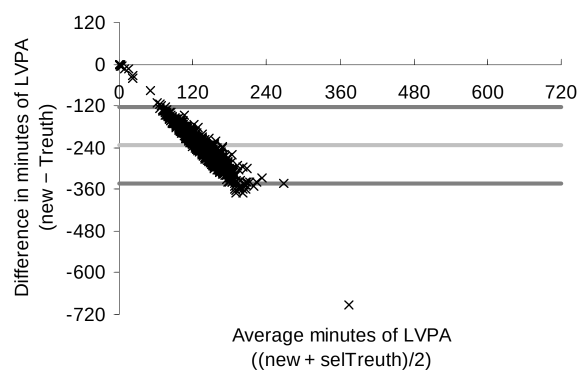
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950-2449	
Vigorous (6 or more METS)		
Black	2050	
	2450	
Light-to-vigorous (2 or more METS)		
All girls	50	

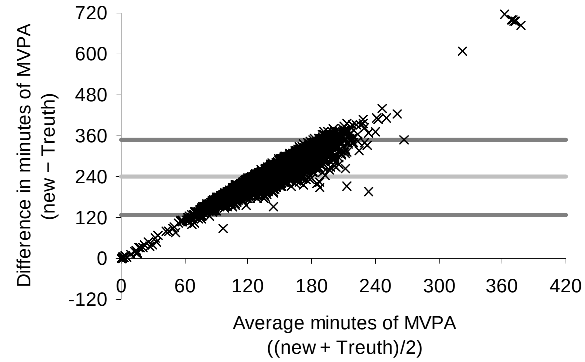
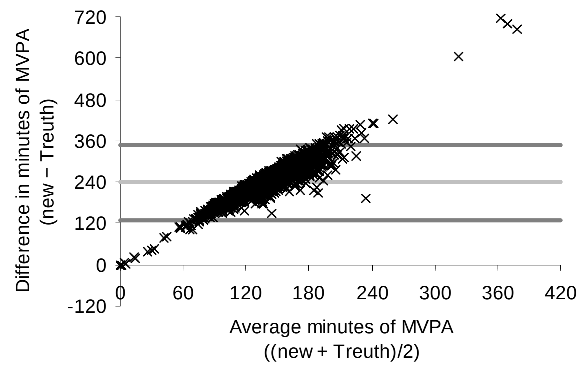
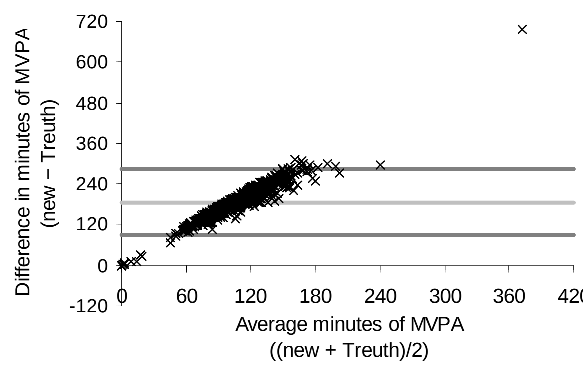
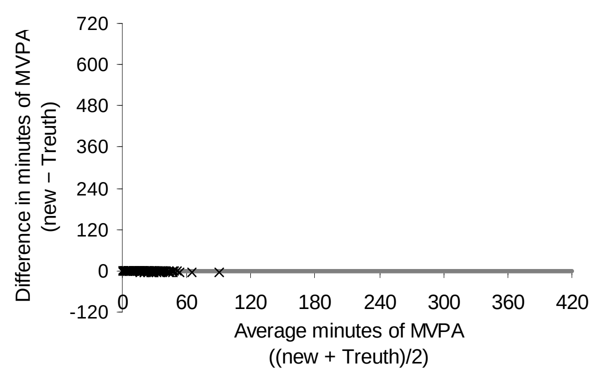
<table>
<tr><th>Comparison / Intensity / Subgroup</th><th>New cut point(s) (counts/30 sec)</th><th>Plot</th></tr>
<tr><td>White</td><td>50</td><td>
</td></tr>
<tr><td>Black</td><td>100</td><td>
</td></tr>
<tr><td></td><td>1050</td><td>
</td></tr>
<tr><td></td><td>1450</td><td>
</td></tr>
</table>

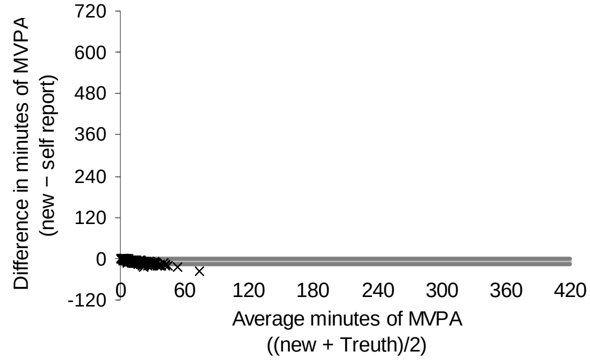
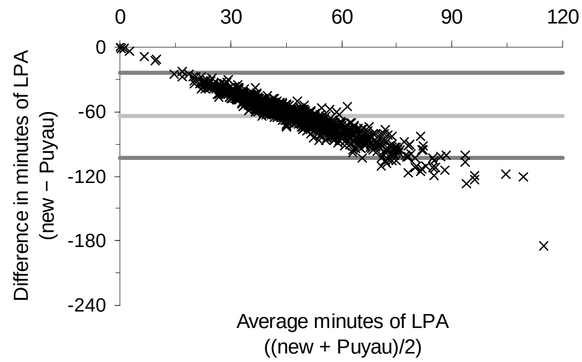
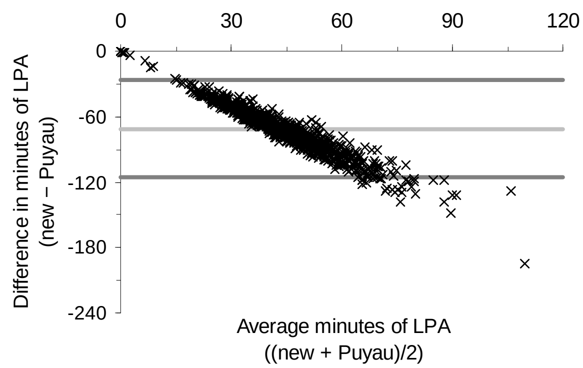
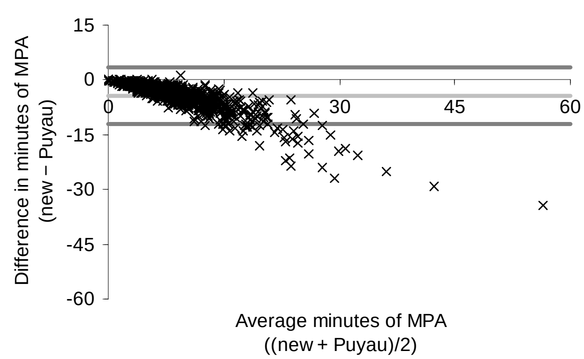
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
Moderate-to-vigorous (3 or more METS)		
All girls	50	
White	50	
Black	100	
	1550	

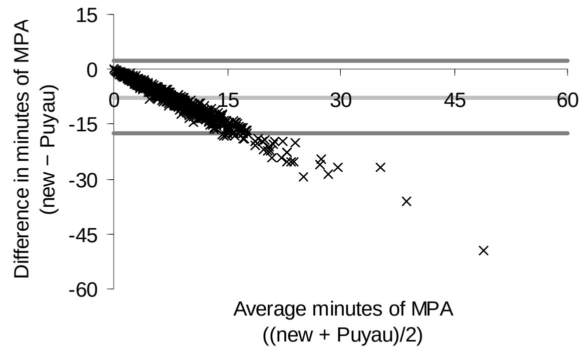
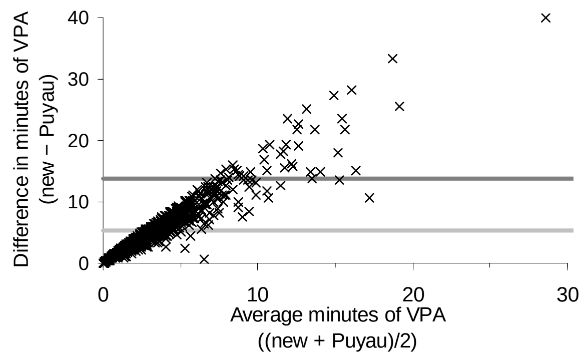
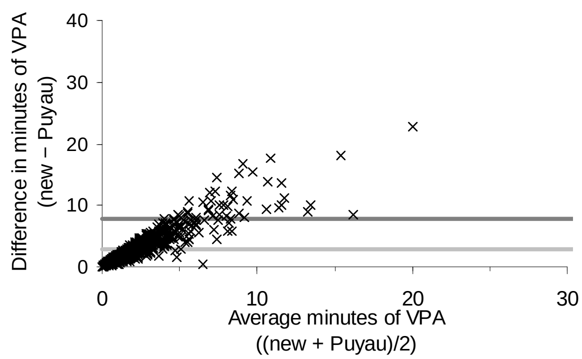
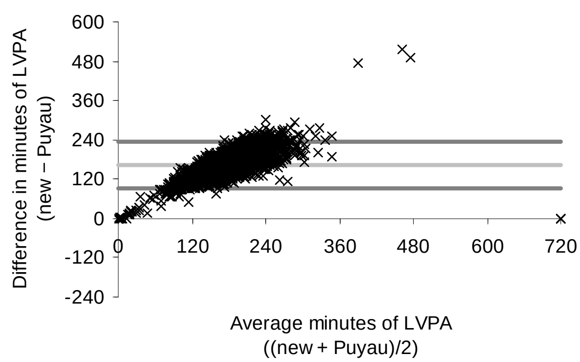
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950	
Treuth cut point		
Light (51-1499counts/30 sec)		
Black	1050-1549	
	1450-1949	
Moderate (1500-2600 counts/30 sec)		
Black	1550-2049	

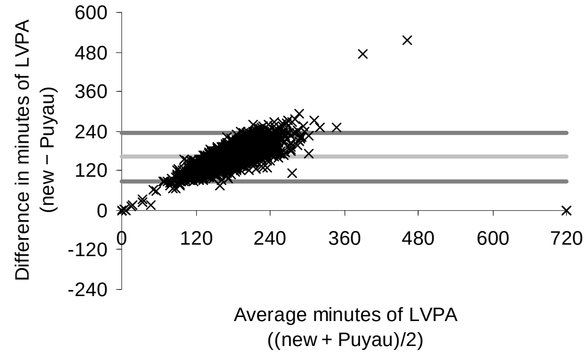
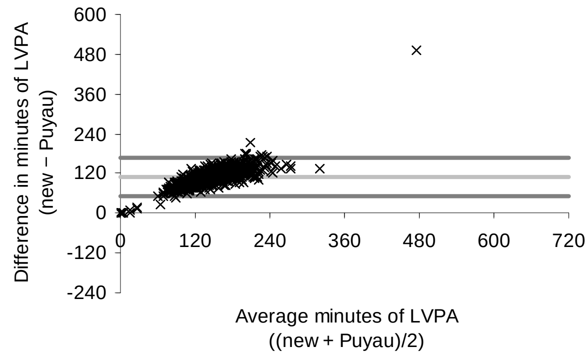
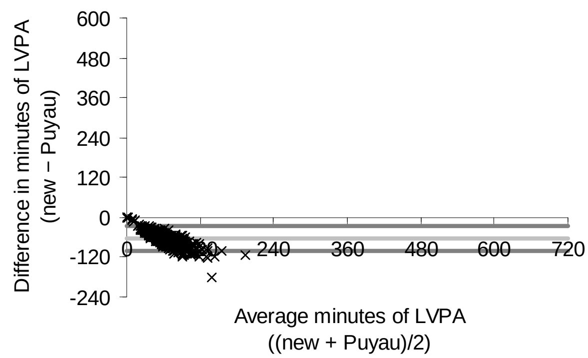
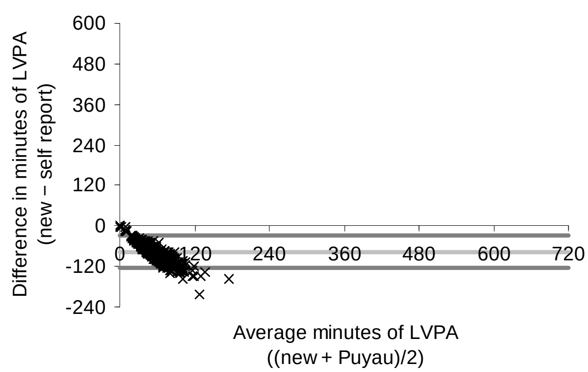
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950-2449	
Vigorous (2601 counts/30 sec)		
Black	2050	
	2450	
Light-to-vigorous (51 counts/30 sec)		
All girls	50	

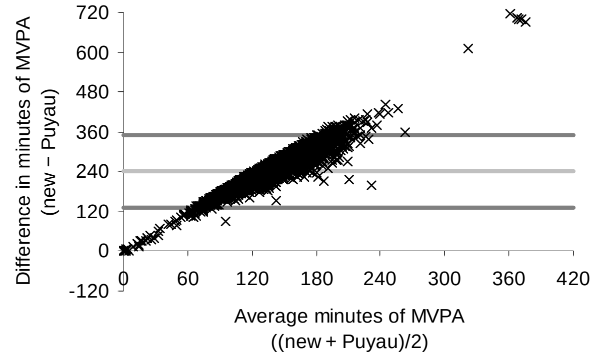
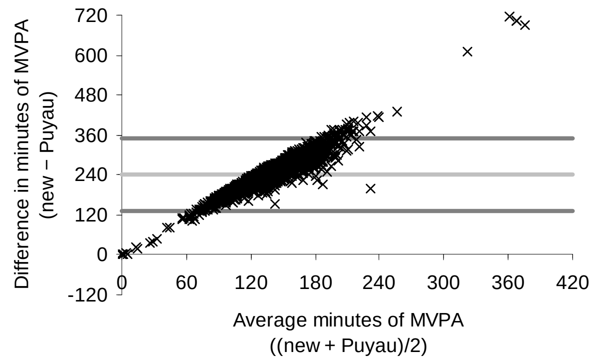
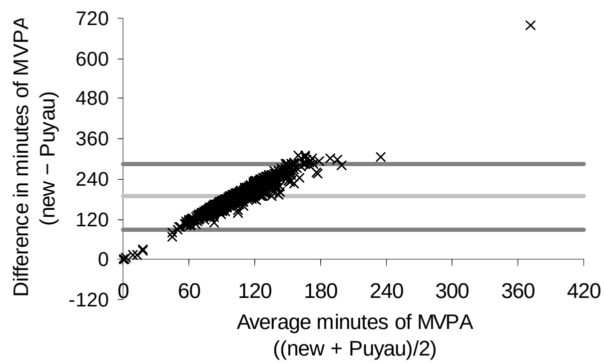
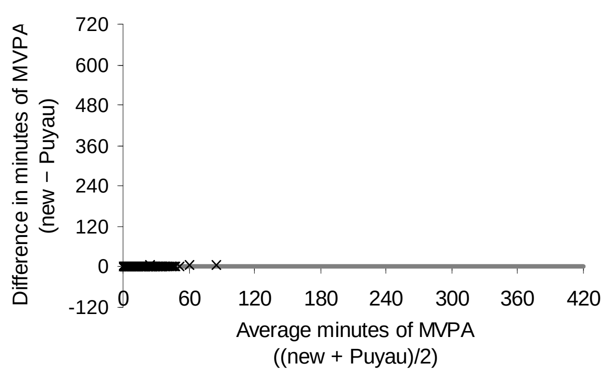
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
White	50	
Black	100	
	1050	
	1450	

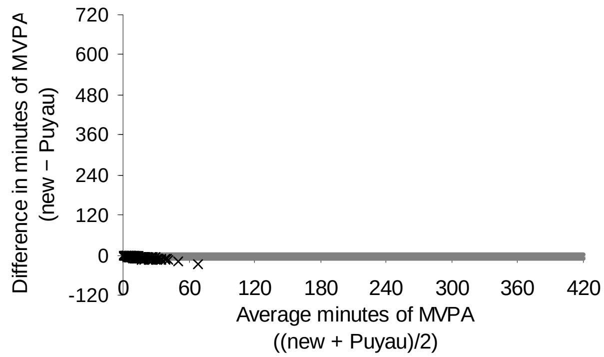
Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
Moderate-to-vigorous (1500 counts/30 sec)		
All girls	50	
White	50	
Black	100	
	1550	

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950	
Puyau cut point		
Light (400-1599 counts/30 sec)		
Black	1050-1549	
	1450-1949	
Moderate (1600-4099 counts/30 sec)		
Black	1550-2049	

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950-2449	
Vigorous (4100 counts/30 sec)		
Black	2050	
	2450	
Light-to-vigorous (400 counts/30 sec)		
All girls	50	

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
White	50	
Black	100	
	1050	
	1450	

<table>
<tr><th>Comparison / Intensity / Subgroup</th><th>New cut point(s) (counts/30 sec)</th><th>Plot</th></tr>
<tr><td colspan="3">Moderate-to-vigorous (1600 counts/30 sec)</td></tr>
<tr><td>All girls</td><td>50</td><td>
</td></tr>
<tr><td>White</td><td>50</td><td>
</td></tr>
<tr><td>Black</td><td>100</td><td>
</td></tr>
<tr><td></td><td>1550</td><td>
</td></tr>
</table>

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Plot
	1950	

TAAG: Trial of Activity for Adolescent Girls; 3DPAR: 3-day physical activity recall; LPA: light physical activity, MPA: moderate physical activity; VPA: vigorous physical activity; LVPA: light-to-vigorous physical activity; MVPA: moderate-to-vigorous physical activity

Appendix L. **Mean difference and 95% confidence interval (CI), limits of agreement (mean ± 1.96×SD), and average minutes with differences outside limits of agreement from Bland-Altman plots comparing minutes of physical activity using new cut points and comparison measures by intensity and population subgroup in NHANES (N = 333)**

Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Mean (95% CI)		Limits of agreement	Avg min below LL[2]	Avg min above UL[2]
Questionnaire[1]						
Moderate						
12-13 y	4000-4999	−13.0	(−17.4, −8.7)	−67.7 - 41.6	38-94	None
	4300-9999	−11.5	(−15.9, −7.1)	−67.3 - 44.2	40-94	34
Vigorous						
12-13 y	5000	−28.2	(−36.6, −19.9)	−133.1 - 76.7	72-234	None
	10000	−30.9	(−39.2, −22.6)	−135.7 - 73.9	70-233	None
Moderate-to-vigorous						
All girls	100	244.9	(235.4, 254.4)	76.8 - 413.0	124-441	211-217
	400	118.7	(110.3, 127.1)	−29.4 - 266.7	132-363	135-198
12-13 y	400	125.2	(112.3, 138.2)	−37.4 - 287.9	132-363	None
	4000	−41.3	(−52.0, −30.6)	−176.0 - 93.5	114-264	None
	4300	−42.4	(−53.1, −31.7)	−176.9 - 92.1	113-263	None
14-15 y	100	234.1	(221.7, 246.6)	81.8 - 386.5	0-409	194-206
White	400	114.6	(96.8, 132.3)	−42.5 - 271.6	212-363	None
	700	56.3	(39.7, 72.8)	−90.5 - 203.1	181-332	None
	1200	8.7	(−7.2, 24.6)	−132.2 - 149.6	153-303	None
	1800	−17.2	(−32.9, −1.4)	−156.6 - 122.3	141-285	None
	2300	−28.2	(−44.1, −12.4)	−168.9 - 112.4	136-276	None
Treuth cut point[2]						
Light						
12-13 y	1900-3999	−261.2	(−269.2, −253.2)	−365.3 - −157.2	204-260	77
	1900-4299	−260.1	(−268.1, −252.1)	−364.0 - −156.2	205-261	78-125
Moderate						
12-13 y	4000-4999	−7.4	(−8.2, −6.7)	−17.2 - 2.3	10-36	None
	4300-9999	−5.9	(−6.7, −5.2)	−15.5 - 3.6	10-37	57
Vigorous						
12-13 y	5000	0.3	(0.3, 0.4)	−0.5 - 1.2	None	4-51
	10000	−2.3	(−3.0, −1.7)	−10.8 - 6.2	9-25	None
Light -to-vigorous						
All girls	100	0.8	(0.8, 0.8)	0.0 - 1.6	291-388	261-404
	400	−125.5	(−128.4, −122.7)	−177.6 - −73.5	244-326	73-232
12-13 y	400	−129.1	(−132.8, −125.5)	−176.6 - −81.7	257-326	117
	1900	−268.3	(−276.4, −260.2)	−373.5 - −163.2	209-271	83
14-15 y	100	0.8	(0.7, 0.8)	0.0 - 1.6	None	261-404
White	400	−132.0	(−137.9, −126.2)	−185.0 - −79.0	261-326	118-150
	700	−189.9	(−198.7, −181.2)	−269.4 - −110.5	221-278	108-110
	1200	−237.4	(−248.5, −226.3)	−338.6 - −136.2	198-244	98-99
	1800	−263.4	(−275.7, −251.1)	−375.1 - −151.8	230-238	89-91
	2300	−274.7	(−287.4, −262.0)	−390.2 - −159.1	224-228	83-88

Comparison / Intensity / Subgroup	New cut point(s) (counts/min)	Mean (95% CI)		Limits of agreement	Avg min below LL[2]	Avg min above UL[2]
Moderate-to-vigorous						
All girls	100	275.9	(269.5, 282.3)	159.5 - 392.3	49-89	205-242
	400	149.6	(145.1, 154.1)	67.3 - 231.9	26-46	128-179
12-13 y	400	159.5	(152.9, 166.1)	73.5 - 245.5	None	128-179
	4000	−7.1	(−7.8, −6.4)	−16.3 - 2.1	11-87	None
	4300	−8.3	(−9.1, −7.4)	−19.3 - 2.8	13-82	None
14-15 y	100	261.7	(252.8, 270.6)	149.3 - 374.0	49-83	197-209
White	400	151.8	(142.7, 160.9)	69.2 - 234.3	None	146-162
	700	93.9	(87.2, 100.5)	33.7 - 154.1	None	102-123
	1200	46.4	(42.3, 50.5)	9.3 - 83.5	None	55-85
	1800	20.4	(18.3, 22.5)	1.4 - 39.3	None	30-55
	2300	9.1	(8.0, 10.2)	−0.8 - 19.1	None	20-46
Puyau cut point[3]						
Light						
12-13 y	1900-3999	−63.0	(−66.3, −59.7)	−105.9 - −20.2	83-146	None
	1900-4299	−61.9	(−65.2, −58.6)	−104.5 - −19.3	83-147	None
Moderate						
12-13 y	4000-4999	−7.5	(−8.5, −6.4)	−21.2 - 6.3	30-59	None
	4300-9999	−5.9	(−6.6, −5.3)	−14.5 - 2.6	10-81	None
Vigorous						
12-13 y	5000	2.3	(1.7, 3.0)	−6.4 - 11.1	None	11-28
	10000	−0.3	(−0.4, −0.2)	−1.2 - 0.6	1-20	None
Light-to-vigorous						
All girls	100	193.5	(189.0, 198.0)	112.0 - 275.0	65-212	206-280
	400	67.2	(65.3, 69.1)	32.1 - 102.3	42-72	119-287
12-13 y	400	71.1	(68.4, 73.7)	37.2 - 104.9	63	162-287
	1900	−68.1	(−71.6, −64.7)	−113.0 - −23.3	86-156	None
14-15 y	100	185.6	(178.8, 192.4)	100.0 - 271.2	65-132	206-267
White	400	70.7	(66.5, 74.9)	32.6 - 108.8	63-72	124-229
	700	12.8	(11.9, 13.6)	5.0 -- 20.6	None	175-179
	1200	−34.7	(−37.1, −32.3)	−56.3 - −13.1	132-135	None
	1800	−60.7	(−65.1, −56.4)	−100.3 - −21.1	106-135	None
	2300	−71.9	(−77.2, −66.7)	−119.4 - −24.5	96-123	None
Moderate-to-vigorous						
All girls	100	277.9	(271.4, 284.3)	160.8 - 394.9	49-88	204-240
	400	151.5	(146.9, 156.1)	68.3 - 234.8	25-45	128-177
12-13 y	400	161.5	(154.8, 168.2)	74.3 - 248.7	None	139-177
	4000	−5.1	(−5.6, −4.6)	−12.0 - 1.8	11-86	None
	4300	−6.2	(−6.9, −5.6)	−15.1 - 2.6	10-81	None
14-15 y	100	263.5	(254.6, 272.4)	150.9 - 376.1	49-82	196-206
White	400	153.4	(144.3, 162.6)	70.0 - 236.9	None	144-160
	700	95.5	(88.8, 102.3)	34.1 - 157.0	None	102-121
	1200	48.0	(43.8, 52.3)	9.4 - 86.7	None	72-83
	1800	22.0	(19.7, 24.3)	1.2 - 42.8	None	28-53
	2300	10.8	(9.5, 12.1)	−1.1 - 22.7	None	19-43

[1] Range of average minutes of activity for which there are differences above the upper limit (UL) of agreement or below the lower limit (LL) of agreement, if any.

[2] Moderate, vigorous, and moderate-to-vigorous intensity in the NHANES questionnaire were defined as 3 to <6 METS, ≥6 METS, and ≥3 METS, respectively.

[3] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 101-2999, 3000-5200, 5201, 101, and 3000 counts/min, respectively.

[4] Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800-3199, 3200-8199, 8200, 800, and 3200 counts/min, respectively.

Appendix M. **Mean difference and 95% confidence interval (CI), limits of agreement (mean ± 1.96×SD), and average minutes with differences outside limits of agreement from Bland-Altman plots comparing minutes of physical activity using new cut points and comparison measures by intensity and population subgroup in TAAG (N = 4696)**

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Mean	(95% CI)	Limits of agreement	Avg min below LL[1]	Avg min above UL[1]
3DPAR[2]						
Light						
Black	1050-1549	−21.9	(−23.2, −20.5)	−61.2 - 17.5	36-102	13-32
	1450-1949	−29.6	(−30.9, −28.3)	−67.5 - 8.4	36-96	15-15
Moderate						
Black	1550-2049	−17.2	(−18.8, −15.7)	−62.2 - 27.8	35-65	None
	1950-2449	−20.5	(−22.0, −18.9)	−65.4 - 24.4	34-63	None
Vigorous						
Black	2050	−4.0	(−5.1, −2.8)	−36.5 - 28.5	22-115	16-31
	2450	−6.4	(−7.6, −5.3)	−39.0 - 26.1	22-110	None
Light-to-vigorous						
All girls	50	184.1	(182.3, 186.0)	57.5 - 310.8	12-300	180-400
White	50	187.6	(185.0, 190.3)	63.6 - 311.7	12-246	180-400
Black	100	128.0	(123.9, 132.1)	8.9 - 247.1	0-219	155-370
	1050	−43.1	(−45.5, −40.7)	−113.0 - 26.8	59-166	18-71
	1450	−56.5	(−58.8, −54.2)	−124.5 - 11.5	69-151	23-30
Moderate-to-vigorous						
All girls	50	220.2	(218.4, 222.0)	97.4 - 343.1	0-268	173-383
White	50	222.0	(219.4, 224.6)	101.3 - 342.8	0-231	179-383
Black	100	165.4	(161.5, 169.4)	50.8 - 280.1	0-170	140-365
	1550	−21.2	(−23.2, −19.2)	−78.4 - 36.0	44-126	18-29
	1950	−26.9	(−28.9, −25.0)	−83.8 - 30.0	45-117	None
Treuth cut point[3]						
Light						
Black	1050-1549	−218.3	(−221.9, −214.8)	−322.2 - −114.5	177-358	0-69
	1450-1949	−226.1	(−229.7, −222.4)	−333.7 - −118.4	172-353	0-66
Moderate						
Black	1550-2049	−4.1	(−4.3, −3.9)	−9.8 - 1.6	16-53	None
	1950-2449	−7.4	(−7.7, −7.1)	−16.0 - 1.3	14-45	None
Vigorous						
Black	2050	3.1	(2.9, 3.2)	−1.7 - 7.8	None	7-38
	2450	0.6	(0.5, 0.6)	−0.5 - 1.6	None	5-29
Light-to-vigorous						
All girls	50	1.5	(1.5, 1.5)	0.5 - 2.5	0-720	168-445
White	50	1.6	(1.5, 1.6)	0.5 - 2.6	0-720	193-445
Black	100	−48.3	(−49.0, −47.5)	−70.8 - −25.7	187-336	0-720
	1050	−219.4	(−222.9, −215.8)	−323.7 - −115.1	185-383	0-81
	1450	−232.8	(−236.6, −229.1)	−343.8 - −121.9	189-374	0-69
Moderate-to-vigorous						
All girls	50	238.9	(237.3, 240.5)	129.9 - 348.0	0-97	182-378
White	50	239.1	(236.7, 241.4)	129.3 - 348.8	0-75	187-378
Black	100	185.7	(182.4, 189.0)	88.7 - 282.8	0-51	150-373
	1550	−1.0	(−1.1, −1.0)	−2.3 - 0.2	15-90	None
	1950	−6.8	(−7.1, −6.5)	−14.7 - 1.1	17-74	None
Puyau cut point[4]						
Light						
Black	1050-1549	−63.5	(−64.8, −62.1)	−102.7 - −24.2	71-115	0-17
	1450-1949	−71.2	(−72.7, −69.7)	−115.7 - −26.7	64-110	0-15

Comparison / Intensity / Subgroup	New cut point(s) (counts/30 sec)	Mean (95% CI)		Limits of agreement	Avg min below LL[1]	Avg min above UL[1]
Moderate						
Black	1550-2049	−4.3	(−4.6, −4.1)	−12.1 - 3.4	13-56	None
	1950-2449	−7.6	(−8.0, −7.3)	−17.6 - 2.4	14-49	None
Vigorous						
Black	2050	5.3	(5.0, 5.6)	−3.2 - 13.8	None	8-29
	2450	2.8	(2.7, 3.0)	−2.2 - 7.9	None	5-20
Light-to-vigorous						
All girls	50	162.8	(161.7, 163.8)	89.7 - 235.8	0-720	172-475
White	50	163.0	(161.4, 164.5)	89.3 - 236.6	0-720	193-462
Black	100	108.6	(106.7, 110.6)	51.8 - 165.4	0-88	200-475
	1050	−62.5	(−63.8, −61.2)	−101.1 - −23.9	76-195	0-22
	1450	−76.0	(−77.6, −74.3)	−123.8 - −28.1	70-175	0-19
Moderate-to-vigorous						
All girls	50	241.2	(239.6, 242.8)	131.2 - 351.3	0-96	181-375
White	50	241.4	(239.1, 243.8)	130.8 - 352.1	0-75	186-375
Black	100	187.7	(184.4, 191.1)	89.5 - 285.9	0-50	149-371
	1550	1.0	(0.9, 1.0)	−0.3 - 2.2	None	15-85
	1950	−4.8	(−5.0, −4.6)	−10.5 - 1.0	14-69	None

[1] Range of average minutes of activity for which there are differences above the upper limit (UL) of agreement or below the lower limit (LL) of agreement, if any.

[2] Light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous intensity for the 3-day physical activity recall (3DPAR) were defined as 2 to <3 METS, 3 to <6 METS, ≥6 METS, ≥2 METS, and ≥3 METS, respectively.

[3] Treuth et al. (2004) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 51-1499, 1500-2600, 2601, 51, and 1500 counts/30 sec, respectively.

[4] Puyau et al. (2002) light, moderate, vigorous, light-to-vigorous, and moderate-to-vigorous cut points were 800, 3200, and 8200 counts/min, respectively. The cut points used were 400-1599, 1600-4099, and 4100, 400, and 1600 counts/30 sec, respectively.

Appendix N. Percent classified as meeting recommendations using self-report physical activity recall questionnaire and Treuth, Puyau, and new accelerometer cut points for moderate-to-vigorous physical activity by population subgroup and dataset

Subgroup / Physical activity measure[1]	Frequency and duration			Accumulated time		
	Strong	Dietary Guidelines	Healthy People	Strong	Dietary Guidelines	Healthy People
All girls in NHANES						
NHANES questionnaire (N = 314)				20.1	39.2	56.7
Accelerometer cut point (N = 333)						
Treuth				0.3	3.6	12.6
Puyau				0.3	2.7	8.7
100 counts/min				100.0	100.0	100.0
400 counts/min				99.7	100.0	100.0
12-13 year old girls in NHANES						
NHANES questionnaire (N = 161)				23.6	42.2	58.4
Accelerometer cut point (N = 171)						
Treuth				0.6	4.1	12.3
Puyau				0.6	2.9	7.6
400 counts/min				100.0	100.0	100.0
4000 counts/min				0.6	1.2	2.9
4300 counts/min				0.6	1.2	2.9
14-15 year old girls in NHANES						
NHANES questionnaire (N = 153)				16.3	35.9	54.9
Accelerometer cut point (N = 162)						
Treuth				0.0	3.1	13.0
Puyau				0.0	2.5	9.9
100 counts/min				100.0	100.0	100.0
White girls in NHANES						
NHANES questionnaire (N = 81)				23.5	45.7	63.0
Accelerometer cut point (N = 85)						
Treuth				1.2	3.5	10.6
Puyau				1.2	1.2	7.1
400 counts/min				100.0	100.0	100.0
700 counts/min				90.6	100.0	100.0
1200 counts/min				43.5	76.5	96.5
1800 counts/min				7.1	41.2	65.9
2300 counts/min				1.2	11.8	40.0

Subgroup / Physical activity measure[1]	Frequency and duration			Accumulated time		
	Strong	Dietary Guidelines	Healthy People	Strong	Dietary Guidelines	Healthy People
All girls in TAAG						
TAAG 3DPAR (N = 4649)	7.2	19.4	27.9	18.0	43.1	64.3
Accelerometer cut point (N = 4696)						
Treuth	0.0	0.4	3.4	0.4	6.8	27.0
Puyau	0.0	0.2	2.2	0.3	4.4	19.2
50 counts/30 sec	97.6	98.9	98.9	99.0	99.2	99.3
White girls in TAAG						
TAAG 3DPAR (N = 2176)	6.1	18.4	27.2	16.4	44.0	65.4
Accelerometer cut point (N = 2190)						
Treuth	0.0	0.4	4.5	0.5	8.5	30.8
Puyau	0.0	0.1	3.1	0.3	5.9	23.2
50 counts/30 sec	98.0	99.0	99.0	99.1	99.3	99.4
Black girls in TAAG						
TAAG 3DPAR (N = 850)	8.0	18.7	26.1	17.4	39.1	59.5
Accelerometer cut point (N = 866)						
Treuth	0.0	0.3	1.8	0.2	3.3	15.8
Puyau	0.0	0.2	0.8	0.1	1.8	10.6
100 counts/30 sec	94.5	98.5	98.3	98.6	98.6	98.8
1550 counts/30 sec	0.0	0.2	1.4	0.2	2.9	13.2
1950 counts/30 sec	0.0	0.0	0.5	0.0	0.3	3.3

NHANES: National Health and Nutrition Examination Survey; TAAG: Trial of Activity for Adolescent Girls; 3DPAR: 3-day physical activity recall
Values not determined for frequency and duration recommendation for NHANES because they could not be determined for NHANES questionnaire.

[1] Moderate -to-vigorous intensity was defined as 3 or more METS for the NHANES questionnaire and for the TAAG 3DPAR, ≥ 3000 counts/min (≥ 1500 counts/30 sec) for Treuth et al. (2004), and 3200 counts/min (≥ 1600 counts/30 sec) for Puyau et al. (2002).

Appendix O. Proportions of overall, positive (met recommendation), and negative (did not meet recommendation) agreement and p-values from McNemar's tests for meeting physical activity recommendations using new accelerometer cut points (counts/minute) compared with using NHANES questionnaire and Treuth and Puyau moderate-to-vigorous cut points by population subgroup in NHANES (N = 333)

	Strong, accumulated time				Dietary Guidelines, accumulated time				Healthy People, accumulated time			
Cut point	**Overall[1]**	**Met[1]**	**Not met[1]**	**p-value**	**Overall**	**Met**	**Not met**	**p-value**	**Overall**	**Met**	**Not met**	**p-value**
All girls												
Questionnaire[2]												
100	0.201	0.334	0.000	+++	0.392	0.563	0.000	+++	0.567	0.724	0.000	+++
400	0.204	0.335	0.008	<.0001	0.392	0.563	0.000	+++	0.567	0.724	0.000	+++
Treuth[2]												
100	0.003	0.006	0.000	+++	0.036	0.070	0.000	+++	0.126	0.224	0.000	+++
400	0.006	0.006	0.006	<.0001	0.036	0.070	0.000	+++	0.126	0.224	0.000	+++
Puyau[2]												
100	0.003	0.006	0.000	+++	0.027	0.053	0.000	+++	0.087	0.160	0.000	+++
400	0.006	0.006	0.006	<.0001	0.027	0.053	0.000	+++	0.087	0.160	0.000	+++
12-13 year old girls												
Questionnaire												
400	0.236	0.382	0.000	+++	0.422	0.594	0.000	+++	0.584	0.737	0.000	+++
4000	0.758	0.000	0.862	<.0001	0.565	0.000	0.722	<.0001	0.416	0.041	0.580	<.0001
4300	0.758	0.000	0.862	<.0001	0.565	0.000	0.722	<.0001	0.416	0.041	0.580	<.0001
Treuth												
400	0.006	0.012	0.000	+++	0.041	0.079	0.000	+++	0.123	0.219	0.000	+++
4000	1.000	1.000	1.000	NDP	0.971	0.444	0.985	0.03	0.906	0.385	0.949	<.0001
4300	1.000	1.000	1.000	NDP	0.971	0.444	0.985	0.03	0.906	0.385	0.949	<.0001
Puyau												
400	0.006	0.012	0.000	+++	0.029	0.057	0.000	+++	0.076	0.141	0.000	+++
4000	1.000	1.000	1.000	NDP	0.982	0.571	0.991	0.0833	0.953	0.556	0.975	0.0047
4300	1.000	1.000	1.000	NDP	0.982	0.571	0.991	0.0833	0.953	0.556	0.975	0.0047
14-15 year old girls												
Questionnaire												
100	0.163	0.281	0.000	+++	0.359	0.529	0.000	+++	0.549	0.709	0.000	+++
Treuth												
100	0.000	0.000	0.000	+++	0.031	0.060	0.000	+++	0.130	0.230	0.000	+++
Puyau												
100	0.000	0.000	0.000	+++	0.025	0.048	0.000	+++	0.099	0.180	0.000	+++

Cut point	Strong, accumulated time				Dietary Guidelines, accumulated time				Healthy People, accumulated time			
	Overall[1]	**Met**[1]	**Not met**[1]	**p-value**	**Overall**	**Met**	**Not met**	**p-value**	**Overall**	**Met**	**Not met**	**p-value**
White girls												
Questionnaire												
400	0.235	0.380	0.000	+++	0.457	0.627	0.000	+++	0.630	0.773	0.000	+++
700	0.321	0.409	0.203	<.0001	0.457	0.627	0.000	+++	0.630	0.773	0.000	+++
1200	0.593	0.400	0.692	0.003	0.617	0.687	0.508	<.0001	0.642	0.775	0.121	<.0001
1800	0.753	0.167	0.855	0.002	0.617	0.563	0.659	0.59	0.630	0.712	0.483	0.72
2300	0.753	0.000	0.859	<.0001	0.556	0.217	0.690	<.0001	0.580	0.595	0.564	0.002
Treuth												
400	0.012	0.023	0.000	+++	0.035	0.068	0.000	+++	0.106	0.191	0.000	+++
700	0.106	0.026	0.174	<.0001	0.035	0.068	0.000	+++	0.106	0.191	0.000	+++
1200	0.576	0.053	0.727	<.0001	0.271	0.088	0.392	<.0001	0.141	0.198	0.076	<.0001
1800	0.941	0.286	0.969	0.0253	0.624	0.158	0.758	<.0001	0.447	0.277	0.552	<.0001
2300	1.000	1.000	1.000	NDP	0.918	0.462	0.955	0.0082	0.706	0.419	0.803	<.0001
Puyau												
400	0.012	0.023	0.000	+++	0.012	0.023	0.000	+++	0.071	0.132	0.000	+++
700	0.106	0.026	0.174	<.0001	0.012	0.023	0.000	+++	0.071	0.132	0.000	+++
1200	0.576	0.053	0.727	<.0001	0.247	0.030	0.385	<.0001	0.106	0.136	0.073	<.0001
1800	0.941	0.286	0.969	0.0253	0.600	0.056	0.746	<.0001	0.412	0.194	0.537	<.0001
2300	1.000	1.000	1.000	NDP	0.894	0.182	0.943	0.0027	0.671	0.300	0.785	<.0001

NHANES: National Health and Nutrition Examination Survey

[1] Overall agreement was the proportion of both measures producing the same classification overall. Positive/negative (met/not met) agreement was the proportion of both measures producing positive/negative classification among those classified as positive/negative by either measure.

[2] Moderate -to-vigorous intensity was defined as 3 or more METS for the NHANES questionnaire, ≥ 3000 counts/min for Treuth et al. (2004), and ≥ 3200 counts/min for Puyau et al. (2002).

+++ Value not determined because all participants classified as meeting that recommendation using that cut point.

NDP Value not determined because no discordant pairs.

Appendix P. Proportions of overall, positive (met recommendation), and negative (did not meet recommendation) agreement and p-values from McNemar's tests for meeting physical activity recommendations using new accelerometer cut points (counts/30 sec) compared with using 3-day physical activity recall (3DPAR) and Treuth and Puyau moderate-to-vigorous cut points by population subgroup in TAAG (N = 4696)

	Strong				Dietary Guidelines				Healthy People			
Cut point	**Overall[1]**	**Met[1]**	**Not met[1]**	**p-value**	**Overall**	**Met**	**Not met**	**p-value**	**Overall**	**Met**	**Not met**	**p-value**
All girls												
Frequency and duration												
3DPAR[2]												
50	0.091	0.133	0.045	<.0001	0.199	0.323	0.021	<.0001	0.285	0.436	0.024	<.0001
Treuth[2]												
50	0.024	0.000	0.046	---	0.015	0.007	0.022	<.0001	0.045	0.066	0.024	<.0001
Puyau[2]												
50	0.024	0.000	0.046	---	0.013	0.004	0.022	<.0001	0.033	0.043	0.023	<.0001
Accumulated time												
3DPAR												
50	0.184	0.302	0.017	<.0001	0.432	0.601	0.015	<.0001	0.642	0.781	0.017	<.0001
Treuth												
50	0.014	0.009	0.019	<.0001	0.076	0.127	0.018	<.0001	0.277	0.427	0.018	<.0001
Puyau												
50	0.012	0.005	0.019	<.0001	0.052	0.084	0.017	<.0001	0.199	0.324	0.017	<.0001
White girls												
Frequency and duration												
3DPAR												
50	0.077	0.114	0.037	<.0001	0.193	0.312	0.022	<.0001	0.280	0.430	0.024	<.0001
Treuth												
50	0.020	0.000	0.039	---	0.014	0.007	0.020	<.0001	0.054	0.086	0.020	<.0001
Puyau												
50	0.020	0.000	0.039	---	0.011	0.002	0.020	<.0001	0.040	0.060	0.020	<.0001
Accumulated time												
3DPAR												
50	0.170	0.281	0.017	<.0001	0.441	0.610	0.015	<.0001	0.655	0.791	0.021	<.0001
Treuth												
50	0.014	0.010	0.017	<.0001	0.093	0.158	0.016	<.0001	0.315	0.474	0.018	<.0001
Puyau												
50	0.012	0.006	0.017	<.0001	0.067	0.113	0.015	<.0001	0.238	0.379	0.017	<.0001

Cut point	Strong				Dietary Guidelines				Healthy People			
	Overall[1]	Met[1]	Not met[1]	p-value	Overall	Met	Not met	p-value	Overall	Met	Not met	p-value
Black girls												
Frequency and duration												
3DPAR												
100	0.131	0.152	0.109	<.0001	0.188	0.307	0.020	<.0001	0.265	0.409	0.028	<.0001
1550	0.920	0.000	0.958	---	0.813	0.012	0.897	<.0001	0.740	0.052	0.849	<.0001
1950	0.920	0.000	0.958	---	0.813	0.000	0.897	---	0.741	0.027	0.851	<.0001
Treuth												
100	0.055	0.000	0.105	---	0.018	0.007	0.030	<.0001	0.036	0.037	0.035	<.0001
1550	1.000	NMR	1.000	NMR	0.999	0.800	0.999	0.32	0.995	0.857	0.998	<0.05
1950	1.000	NMR	1.000	NMR	0.997	0.000	0.998	---	0.986	0.400	0.993	<0.05
Puyau												
100	0.055	0.000	0.105	---	0.017	0.005	0.030	<.0001	0.025	0.016	0.034	<.0001
1550	1.000	NMR	1.000	NMR	1.000	1.000	1.000	NDP	0.994	0.737	0.997	0.03
1950	1.000	NMR	1.000	NMR	0.998	0.000	0.999	---	0.997	0.727	0.998	0.08
Accumulated time												
3DPAR												
100	0.176	0.290	0.020	<.0001	0.386	0.554	0.015	<.0001	0.591	0.741	0.017	<.0001
1550	0.828	0.027	0.906	<.0001	0.612	0.073	0.754	<.0001	0.460	0.258	0.575	<.0001
1950	0.826	0.000	0.905	---	0.611	0.012	0.758	<.0001	0.429	0.093	0.584	<.0001
Treuth												
100	0.016	0.005	0.027	<.0001	0.047	0.066	0.028	<.0001	0.170	0.276	0.027	<.0001
1550	1.000	1.000	1.000	NDP	0.995	0.926	0.998	<0.05	0.973	0.908	0.984	<.0001
1950	0.998	0.000	0.999	---	0.970	0.188	0.985	<.0001	0.875	0.349	0.931	<.0001
Puyau												
100	0.015	0.002	0.027	<.0001	0.032	0.037	0.028	<.0001	0.118	0.194	0.026	<.0001
1550	0.999	0.667	0.999	0.32	0.990	0.780	0.995	0.003	0.975	0.893	0.986	<.0001
1950	0.999	0.000	0.999	---	0.985	0.316	0.992	0.0003	0.927	0.479	0.961	<.0001

TAAG: Trial of Activity for Adolescent Girls; 3DPAR: 3-day physical activity recall

[1] Overall agreement was the proportion of both measures producing the same classification overall. Positive/negative (met/not met) agreement was the proportion of both measures producing positive/negative classification among those classified as positive/negative by either measure.

[2] Moderate-to-vigorous intensity was defined as 3 or more METS for the TAAG 3DPAR, ≥ 1500 counts/30 sec for Treuth et al. (2004), and ≥ 1600 counts/30 sec (i.e., 3200 counts/min) for Puyau et al. (2002).

--- Value not determined because all participants classified as not meeting that recommendation using that cut point.

NDP Value not determined because no discordant pairs.

NMR Value not determined because no participants classified as meeting that recommendation using that cut point nor the comparison measure.

Reference List

ActiGraph (2009). What exactly is a "count"? http://www.theactigraph.com/index.php?option=com_content&view=article&id=146:whatisacount&catid=62:gt1m&Itemid=41 [On-line].

Allison, K. R., Dwyer, J. J., & Makin, S. (1999). Perceived barriers to physical activity among high school students. *Prev.Med., 28,* 608-615.

American Academy of Pediatrics Committee on Nutrition (1998). Cholesterol in childhood. *Pediatrics, 101,* 141-147.

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

Baranowski, T. (1988). Validity and reliability of self report measures of physical activity: an information-processing perspective. *Res.Q.Exerc.Sport, 59,* 314-327.

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

Bland, J. M. & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1,* 307-310.

Boreham, C. & Riddoch, C. (2001). The physical activity, fitness and health of children. *J.Sports Sci., 19,* 915-929.

Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med.Sci.Sports Exerc., 32,* 1601-1609.

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep., 100,* 126-131.

Centers for Disease Control and Prevention, National Center for Health Statistics (2003). National Health and Nutrition Examination Survey overview. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/overviewbrochure_0304.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2004a). National Health and Nutrition Examination Survey anthropometry procedures manual. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/BM.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2004b). National Health and Nutrition Examination Survey laboratory procedures manual. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/lab_pm.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2006). National Health and Nutrition Examination Survey 2003-2004 documentation, codebook, and frequencies: exam component physical activity monitor examination data. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/paxraw_c.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2009). National Health and Nutrition Examination Survey 2003-2004 documentation, codebook, and frequencies: demographic variables and sample weights. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/demo_c.pdf [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics (2008). 2003-2004 National Health and Nutrition Examination Survey: survey operations manuals, brochures, consent documents. http://www.cdc.gov/nchs/about/major/nhanes/nhanes2003-2004/current_nhanes_03_04.htm [On-line].

Centers for Disease Control and Prevention, National Center for Health Statistics(2000). CDC growth charts: United States. http://www.cdc.gov/growthcharts/ [On-line].

Chen, K. Y. & Bassett, D. R., Jr. (2005). The technology of accelerometry-based activity monitors: current and future. *Med.Sci.Sports Exerc., 37,* S490-S500.

de Ferranti, S. D., Gauvreau, K., Ludwig, D. S., Neufeld, E. J., Newburger, J. W., & Rifai, N. (2004). Prevalence of the metabolic syndrome in American adolescents: findings from the third National Health and Nutrition Examination Survey. *Circulation, 110,* 2494-2497.

Dunton, G. F. & Schneider, M. (2006). Perceived barriers to walking for physical activity. *Prev.Chronic Dis., 3,* A116.

Eisenmann, J. C., Welk, G. J., Wickel, E. E., & Blair, S. N. (2007). Combined influence of cardiorespiratory fitness and body mass index on cardiovascular disease risk factors among 8-18 year old youth: the Aerobics Center Longitudinal Study. *Int.J.Pediatr.Obes., 2,* 66-72.

Eisenmann, J. C., Wickel, E. E., Welk, G. J., & Blair, S. N. (2005). Relationship between adolescent fitness and fatness and cardiovascular disease risk factors in adulthood: the Aerobics Center Longitudinal Study (ACLS). *Am.Heart J., 149,* 46-53.

Ekelund, U., Anderssen, S. A., Froberg, K., Sardinha, L. B., Andersen, L. B., & European Youth Heart Study Group (2007). Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European Youth Heart Study. *Diabetologia, 50,* 1832-1840.

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *J.Sports Sci.,* 1-9.

Fernandez, J. R., Redden, D. T., Pietrobelli, A., & Allison, D. B. (2004). Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J.Pediatr., 145,* 439-444.

Freedson, P. S., Sirard, J., Debold, E., Pate, R., Dowda, M., Trost, S. et al. (1997). Calibration of the Computer Science and Applications, Inc. (CSA) accelerometer. *Med.Sci.Sports Exerc., 29,* S45.

Grieser, M., Vu, M. B., Bedimo-Rung, A. L., Neumark-Sztainer, D., Moody, J., Young, D. R. et al. (2006). Physical activity attitudes, preferences, and practices in African American, Hispanic, and Caucasian girls. *Health Educ.Behav., 33,* 40-51.

Healy, G. N., Dunstan, D. W., Salmon, J., Cerin, E., Shaw, J. E., Zimmet, P. Z. et al. (2007). Objectively measured light-intensity physical activity is independently associated with 2-h plasma glucose. *Diabetes Care, 30,* 1384-1389.

Healy, G. N., Wijndaele, K., Dunstan, D. W., Shaw, J. E., Salmon, J., Zimmet, P. Z. et al. (2008). Objectively measured sedentary time, physical activity, and metabolic risk. *Diabetes Care, 31,* 369-371.

Kimm, S. Y., Glynn, N. W., Kriska, A. M., Barton, B. A., Kronsberg, S. S., Daniels, S. R. et al. (2002). Decline in physical activity in black girls and white girls during adolescence. *N.Engl.J.Med., 347,* 709-715.

Kimm, S. Y. S., Barton, B. A., Obarzanek, E., McMahon, R. P., Sabry, Z. I., Waclawiw, M. A. et al. (2001). Racial divergence in adiposity during adolescence: the NHLBI Growth and Health Study. *Pediatrics, 107,* e34.

Korn, E. L. & Graubard, B. I. (1991). Epidemiologic studies using surveys: accounting for the sampling design. *Am.J.Public Health, 81,* 1166-1173.

Kraemer, H. C. (1988). Assessment of 2 x 2 associations: generalization of signal-detection methodology. *Am.Statistician, 42,* 37-49.

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45,* 255-268.

Lohman, T. G. & Falls, H. B. (2001). Body composition assessment. In G.J.Welk, J. R. Morrow, & H. B. Falls (Eds.), *FITNESSGRAM.Reference Guide* (pp. 87-92). Dallas: The Cooper Institute for Aerobics Research.

Lohman, T. G., Ring, K., Schmitz, K. H., Treuth, M. S., Loftin, M., Yang, S. et al. (2006). Associations of body size and composition with physical activity in adolescent girls. *Med.Sci.Sports Exerc., 38,* 1175-1181.

Malina, R. M. (2007). Physical fitness of children and adolescents in the United States: status and secular change. *Med.Sport Sci., 50,* 67-90.

Masse, L. C., Fuemmeler, B. F., Anderson, C. B., Matthews, C. E., Trost, S. G., Catellier, D. J. et al. (2005). Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. *Med.Sci.Sports Exerc., 37,* S544-S554.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. I., & Hu, J. (2003). Tracking of physical activity and aerobic power from childhood through adolescence. *Med.Sci.Sports Exerc., 35,* 1914-1922.

Moliner-Urdiales, D., Ruiz, J. R., Ortega, F. B., Rey-Lopez, J. P., Vicente-Rodriguez, G., España-Romero, V. et al. (2009). Association of objectively assessed physical activity with total and central body fat in Spanish adolescents; The HELENA Study. *Int J.Obes., advance online publication,* doi:10.1038.

Montoye, H. J., Kemper, H. C. G., Saris, W. H. M., & Washburn, R. A. (1996). Questionnaires and interviews. In *Measuring Physical Activity and Energy Expenditure* (pp. 42-71). Champaign, IL: Human Kinetics.

Morrison, J. A., Barton, B. A., Obarzanek, E., Crawford, P. B., Guo, S. S., Schreiber, G. B. et al. (2001). Racial differences in the sums of skinfolds and percentage of body fat estimated from impedance in black and white girls, 9 to 19 years of age: the National Heart, Lung, and Blood Institute Growth and Health Study. *Obes.Res., 9,* 297-305.

Nader, P. R., Bradley, R. H., Houts, R. M., McRitchie, S. L., & O'Brien, M. (2008). Moderate-to-vigorous physical activity from ages 9 to 15 years. *JAMA, 300,* 295-305.

Ogden, C. L., Carroll, M. D., & Flegal, K. M. (2008). High body mass index for age among US children and adolescents, 2003-2006. *JAMA, 299,* 2401-2405.

Ogden, C. L., Flegal, K. M., Carroll, M. D., & Johnson, C. L. (2002). Prevalence and trends in overweight among US children and adolescents, 1999-2000. *JAMA, 288,* 1728-1732.

Ondrak, K. S., McMurray, R. G., Bangdiwala, S. I., & Harrell, J. S. (2007). Influence of aerobic power and percent body fat on cardiovascular disease risk in youth. *J.Adolesc.Health, 41*, 146-152.

Pate, R. R., Ross, R., Dowda, M., Trost, S. G., & Sirard, J. R. (2003). Validation of a 3-day physical activity recall instrument in female youth. *Pediatr.Exerc.Sci., 15,* 257-265.

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

Puyau, M. R., Adolph, A. L., Vohra, F. A., & Butte, N. F. (2002). Validation and calibration of physical activity monitors in children. *Obes.Res., 10,* 150-157.

Reilly, J. J., Penpraze, V., Hislop, J., Davies, G., Grant, S., & Paton, J. Y. (2008). Objective measurement of physical activity and sedentary behaviour: review with new data. *Arch.Dis.Child., 93,* 614-619.

Strong, W. B., Malina, R. M., Blimkie, C. J., Daniels, S. R., Dishman, R. K., Gutin, B. et al. (2005). Evidence based physical activity for school-age youth. *J.Pediatr. 146,* 732-737.

Treuth, M. S., Schmitz, K., Catellier, D. J., McMurray, R. G., Murray, D. M., Almeida, M. J. et al. (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med.Sci.Sports Exerc., 36,* 1259-1266.

Troped, P. J., Wiecha, J. L., Fragala, M. S., Matthews, C. E., Finkelstein, D. M., Kim, J. et al. (2007). Reliability and validity of YRBS physical activity items among middle school students. *Med.Sci.Sports Exerc., 39,* 416-425.

Trost, S. G., Pate, R. R., Freedson, P. S., Sallis, J. F., & Taylor, W. C. (2000). Using objective physical activity measures with youth: how many days of monitoring are needed? *Med.Sci.Sports Exerc., 32,* 426-431.

Trost, S. G., Ward, D. S., Moorehead, S. M., Watson, P. D., Riner, W., & Burke, J. R. (1998). Validity of the Computer Science and Applications (CSA) activity monitor in children. *Med.Sci.Sports Exerc., 30,* 629-633.

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

U.S. Department of Health and Human Services (2000). *Healthy People 2010: understanding and improving health and objectives for improving health* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services & U.S. Department of Agriculture (2005). *Dietary guidelines for Americans*.

Ward, D. S., Brown Rodgers, A., & Vaughn, A. (2005). Objective monitoring of physical activity: closing the gaps in the science of accelerometry. *Med.Sci.Sports Exerc., 37*.

Welk, G. J., Wickel, E., Peterson, M., Heitzler, C. D., Fulton, J. E., & Potter, L. D. (2007). Reliability and validity of questions on the Youth Media Campaign Longitudinal Survey. *Med.Sci.Sports Exerc., 39,* 612-621.