Published in final edited form as:

J Phys Act Health. 2010 January ; 7(1): 45–53.

Defining Accelerometer Thresholds for Physical Activity in Girls Using ROC Analysis

Sofiya Alhassan and

Dept of Kinesiology, University of Massachusetts, Amherst, MA

Thomas N. Robinson

Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, CA

Abstract

Background—Receiver operating characteristic (ROC) analysis is a common method used in diagnostic and screening tests to define thresholds levels of a factor that discriminates between 2 levels of another factor. The purpose of this analysis was to use ROC analysis to determine the optimal accelerometer-measured physical activity (PA) thresholds for predicting selective cardiovascular disease (CVD) risk factors.

Methods—ROC was performed using data from Stanford Girls Health Enrichment Multisite Studies trial. PA was assessed for multiple days using accelerometers. CVD variables were overweight, elevated triglyceride, reduced HDL-C, hypertension, impaired fasting glucose, fasting insulin, and clustering of multiple CVD risk factors.

Results—A sample of 261 girls participated, of which 208 had complete CVD risk measures (mean $\pm$ SD age = 9.4 ± 0.9 yrs, BMI = 20.7 ± 4.8 kg/m²). An average of 11.1 minutes/day at 2,600 counts/min was the maximally sensitive and specific threshold for discriminating girls who were overweight, 16.6 minutes/day at 2,000 counts/min for hyperinsulinemia or with 2 CVD risk factors. The Area Under the Curve for overweight, hyperinsulinemia, and 2 CVD risk factors was of 0.66, 0.58, and 0.60, respectively. The sensitivity and specificity associated with overweight, hyperinsulinemia, and 2 CVD risk factors were 60.3% and 72.9%, 53.3% and 83.9%, 44.0% and 84.7%, respectively.

Conclusion—Empirically-derived thresholds of PA to optimally discriminate between girls with and without CVD risk were lower in this sample than generally recommended. This ROC approach should be repeated in other populations to determine optimal PA thresholds with clinical validity for research, surveillance and program evaluation.

Keywords

receiver operating characteristics; ActiGraph; African-American; girls

The health benefits of physical activity (PA) have been widely reported in children.¹ In general, increased PA has been associated with improvements in cardiovascular disease (CVD) risk factors such as obesity, hypertension, dyslipidemia, fitness, fasting insulin and glucose.^{2,3} To potentially reduce these risk factors, several organizations have published public health guidelines for PA. In 1988, the first PA guidelines for children were published by the American College of Sports Medicine; recommending that young individuals should accumulate approximately 20 to 30 minutes/day of vigorous PA.⁴ These guidelines were based on adult PA guidelines. Since then others have also published PA guidelines for children.^{5,6} Currently, it is recommended that school-age children should participate in at least 60 minutes or more of moderate to vigorous PA (MVPA) on most days of the week.^{7,8} These recommendations were based on systematic review of the available evidence on the

relationships between PA and several health outcomes in children. However, it is important to note that many of the studies linking PA and health outcomes have relied on subjective measures of PA. In addition, the thresholds defining both the duration and intensity of PA have been derived in a number of different ways, relating to energy expenditure in experimental contexts,^{9,10} according to population or sample distributions,⁸ and to different types of PA.⁸

Many studies of PA have used subjective measures such as self-reports due to their low cost and ease of use. The utilization of self-reports are limited, because self-reports rely on children's and adolescents' abilities to report and/or recall their past PA.¹¹ This inherent problem, which leads to social desirability bias as well as recall bias, therefore makes it difficult to accurately examine the relationship between PA and health outcomes. More studies are using accelerometers to provide a more objective means for quantifying PA.¹² Measuring accelerations associated with activity, accelerometers report a count value over a specified unit of time, which can be translated to different activity intensity thresholds.

In the scientific literature different counts/min thresholds are being used to define PA intensities (eg, MVPA). Differences in these threshold values can lead to very different estimates of time spent in MVPA and correlations with measures of other health behaviors.¹³ At present, neither the accelerometer counts/min thresholds nor the recommended dose of PA have been systematically tested to determine if they are able to optimally discriminate between individuals that have or do not have various CVD risk factors.

A common method to define thresholds in diagnostic and screening tests uses a signal detection technique known as receiver operating characteristic (ROC) analysis. ROC analysis is able to identify a potentially optimal threshold level of a factor (eg, PA) that most efficiently discriminates between 2 levels of another factor (eg, CVD risk factors) with the greatest sensitivity and specificity. Therefore, the primary objective of this analysis was to use ROC analysis to determine the optimal accelerometer measured PA thresholds with greatest sensitivity and specificity for predicting CVD risk factors and clustering of CVD risk factors.

Methods

Participants

This study used baseline data from Stanford GEMS (Girls Health Enrichment Multisite Studies), a randomized controlled excess weight gain prevention trial among lower socioeconomic status African-American preadolescent girls. Girls were recruited from low-income areas of Oakland, CA. Participants were required to be 8, 9 or 10 years of age, and have a BMI $\geq$ 25th percentile for their age and/or at least 1 parent/guardian with a BMI $\geq$ 25 kg/m². Girls were excluded from the study if they were taking medications and/or had a medical condition affecting their growth; had a condition limiting their participation in the interventions; were unable to understand or complete informed consent; or planned to move from the San Francisco Bay, CA area within the next 24 months. A parent/guardian provided written informed consent and girls gave assent to participate in the study. The study was approved by the Stanford University Panel on the Protection of Human Subjects in Medical Research, was implemented as a cooperative agreement with the National Heart, Lung, and Blood Institute (NHLBI), and received oversight from an independent Data and Safety Monitoring Board formed by NHLBI. A description of the study has been previously published.¹⁴

Measures

Physical Activity Monitoring—Participants' PA levels were monitored using the ActiGraph (Model 7164, Manufacturing Technologies Inc. Health Services, Ft Walton Beach, FL, previously known as the Computer Science Application (CSA) accelerometer). The ActiGraph accelerometer is a single axis accelerometer designed to measure and record vertical accelerations ranging in magnitude from 0.05 to 2.00 G with a frequency response from 0.25 to 2.5 Hz.¹⁵ Within these set parameters, the ActiGraph is capable of detecting normal human motion and rejecting motions associated with high frequency vibrations. The ActiGraph was attached to a belt and worn around the participants' waist over their right hip. The protocol called for monitoring of 4 days including 1 weekend day. To ensure that participants did not forget to put the monitor on each morning, they were instructed to wear the accelerometer at all times, including during sleep, and to remove them only when showering, bathing, swimming, or when monitors would get completely wet. Accelerations were measured in 1-minute epochs in this study.

Accelerometer Data Clean Up and Reduction—Data clean up and reduction protocols and validation have been previously described.¹⁶ Minute by minute accelerometer data were first reviewed visually to determine whether: 1) the number of days with accelerometer data appeared sufficient and matched study protocol, 2) sleep and awake times were logical, 3) if there were any error codes indicating that the monitor might have malfunctioned and 4) any count/min values $\geq 15,000$ were flagged and participant's parent/guardian was contacted to verify if the participant was involved in any activity that might have resulted in such a high count/min value. Other researchers have also flagged counts/min values $\geq 15,000$ as being biologically implausible.¹⁷ Data points with error codes and 1 minute on either side of those values were changed to missing data. Data were scanned to find periods of at least 20 consecutive minutes during which the ActiGraph measured only zeros, and these were considered periods when the monitor was not being worn and were changed to missing data.¹⁶ If there was a minute of nonzero data flanked by 2 periods of ≥ 20 consecutive zeros, the nonzero data point was also considered part of the nonworn period and set to missing.¹⁶ This latter step was intended to exclude occasional periods where the recorded movement was likely due to vibrations or incidental jostling. After the above procedures were completed, all remaining data were classified as acceptable data.

For all accelerometer data reduction procedures, weekdays and weekend days were treated separately. Next, within the weekday/weekend day data, each participant's accelerometer data were collapsed across days by averaging the counts/min for each minute of the day for each day with data for that minute. For each participant, this procedure created a composite estimate for a weekday and weekend day. The composite weekday and weekend day were defined as 6:00 AM to 10:00 PM (960 minutes) and 10:00 AM to 10:00 PM (720 minutes), respectively, based on the data from this sample.¹⁶ We have previously shown, in this same sample, that this data algorithm significantly improves the amount of usable data with greater correlational validity compared with other algorithms.¹⁶

Anthropometric Measurements—Body weight was measured twice in light clothing, to the nearest 0.1 kg, using a calibrated portable digital scale (Scaletonix 5602 Model scale; White Plains, NY). Standing height was measured twice to the nearest millimeter, using a portable direct reading stadiometer (Shorr Height Measuring Board; Olney, MD). For each assessment, the mean of the replicate measures was used in analysis. BMI (kg/m^2) was computed as the body weight (kg) divided by squared height (meters squared).

Blood Pressure—Blood pressure was measured on the right arm supported at heart level using an automated blood pressure monitor (Dinamap Pro 100, GE Medical Systems;

Tampa, FL) with an appropriate-sized cuff, after 3 minutes of quiet rest in the seated position. Blood pressure measures were repeated 2 more times separated by 1 minute rest intervals. The average of the 3 measures was used in the analysis.

Blood Sampling and Biochemical Analysis—Blood samples were collected in participants' homes after an 8-hour fast. Blood samples were allowed to clot in a serum separator tube at room temperature for 30 minutes. Serum samples were then isolated by centrifugation at $1500 \times g$ for 15 minutes, aliquoted into microcentrifugation tubes and then placed on ice while in the field. Serum samples were then stored at -70°C for analysis within 24 hours at the Stanford University Hospital Biochemistry Laboratory, which participates in the Center for Disease Control—NHLBI Lipid Standardization Program. High-density lipoprotein cholesterol (HDL-C) was directly measured using a homogeneous assay that eliminates the need to first isolate HDL lipoprotein from serum sample. HDL-C and triglyceride (TG) were analyzed using Beckman Synchron LX20 reagents (Beckman Coulter, Inc., Fullerton, CA). Insulin was measured using an automated 2 site chemiluminescent immunometric (sandwich) assay (DPC 2000 Immulite, Diamond Diagnostics, Inc., Holliston, MA). Glucose was determined by the oxygen rate methods employing a Beckman Oxygen electrode using a Beckman Synchron LX20 (Beckman Coulter, Inc., Fullerton, CA).

Outcome and Predictor Variables

Selected CVD risk factors were used as the outcome variables of interest for this ROC analysis. Overweight status was defined as BMI $\geq$ the 95th percentile for their age.¹⁸ This cut-off was chosen because it is associated with obesity-related comorbidities.^{18,19} Elevated TG was defined as values $\geq$ 95th percentile for age according to National Cholesterol Education Program references.^{20,21} Abnormal HDL-C was defined as values <35 mg/dL.²⁰ The above cut-points for dyslipidemia have been associated with early atherosclerotic lesions in adolescents and young adults.²² Total cholesterol and low-density lipoprotein cholesterol were not included as outcome variables in the present analysis due to the weak association between them and PA.⁷ Elevated blood pressure was defined as systolic and/or diastolic blood pressure $\geq$ 90th percentile for age and height percentile for girls.²³ Elevated blood pressure in children is currently considered a risk factor for hypertension in early adulthood.²³ Impaired fasting plasma glucose was defined as values ≥ 100 mg/dL, which is considered a risk factor for Type 2 diabetes mellitus in children and adolescents.²⁴ Currently, there are no standardized cut-points for determining hyperinsulinemia in children. However, within the literature hyperinsulinemia is commonly defined as fasting serum insulin > 20 mU/L²⁵ and has been independently associated with other CVD risk factors in children.²⁶ Risk factor score clusters were defined as the presence of any 2 or more (Cluster 2), any 3 or more (Cluster 3), or any 4 or more (Cluster 4) CVD risk factors. To derive the PA (predictor) variables for the ROC analysis, we used weighted averages of weekday and weekend average counts $[(5 \times \text{weekday}) + (2 \times \text{weekend day})/7]$ for a total possible 891.4 minutes to define (1) the average daily counts/min for each participant and (2) the percent of time spent at or above every threshold of activity intensity from 0 to 18,309 counts/min, in multiples of 100 counts (ie, percent of time spent at or above 0, 100, 200, 300, ... 18,300 counts/min) for each participant.

Statistical Analysis

Descriptive data are expressed as means $\pm$ standard deviations (SD). ROC analyses are generally used to evaluate the accuracy of a screening or diagnostic test or to identify the characteristics of a subgroup at risk for a disease, due to the test's ability to discriminate between the presence and absence of a health condition over the complete spectrum of the operating condition.^{27,28} In this study, ROC analysis was used to determine the optimal cut-

point of PA, expressed as average counts/min and as percent of time at or above a threshold intensity (count/ min) level, needed to discriminate between the presence and absence of the selected CVD risk factors.

ROC analysis was performed with ROC4 software, which can be found at the following public domain website: <http://mirecc.stanford.edu>. The first step in performing a ROC analysis is to define the clinically relevant binary outcome variables of interest.²⁷ Next, all the predictor variables and their possible cut-points are examined to find the combination with the optimal sensitivity (probability of correctly detecting true-positive results) and specificity (probability of correctly detecting true-negative results) for the outcome of interest. In this analysis equal weight was given to false positives and negatives in determining an optimal cut-point. The ROC analysis tests every predictor variable and every possible cut-point in every participant to identify the optimal predictor variable and its associated cut-points. The association between the outcome variable and predictor variables and their associated cut-points are tested until a “stopping rule” is reached: 1) there are less than 10 participants with or without the outcome variable of interest and 2) a 2×2 Chi-Square testing the association between the outcome variable and the predictor variable is not significant at less than the 0.05 level.²⁷ All CVD risk factor outcome variables were tested independently.

The area under the ROC curve (AUC) was calculated to assess the overall accuracy of the predictors. AUC reflects the probability that the test correctly discriminates the presences or absences of a defined CVD risk factor.^{27,29} AUC ranges from 0 to 1. An AUC of 0.5 indicates that the test is not better than chance and an AUC of 1 indicates a perfect test. The optimal cut-point of the test is generally identified as the point closest to the top left corner of the ROC curve (plotted as sensitivity versus 1-specificity), where the balance between sensitivity and specificity is maximized.²⁸ Odds ratios (assessing how close the association between the predictor and outcome variable) and risk difference (the difference in the percent of participants with the outcome in the high risk versus low risk group) were also calculated. All other statistical analyses were performed using SAS version 9.1 (SAS Institutes Inc., Cary, NC).

Results

A total of 261 participants were enrolled in the study. Complete ActiGraph data were obtained from 260 of whom fasting blood samples were successfully obtained from 208 (80%) and therefore included in this analysis. Participants' mean BMI was 20.7 kg/m². Their mean BMI percentile was 74.7 ± 26.2 and the prevalence of overweight status (BMI 95th percentile) was 32.7%. The sample's mean systolic and diastolic blood pressures percentiles were below the 50th percentile. Only 5.3% of the participants met the criteria for elevated blood pressure. Approximately 2% and 7.2% had elevated fasting glucose and insulin levels, respectively. Of the 208 participants, 6.3% met the criteria for elevated TG and 2.4% for reduced HDL-C. Participants' average daily accelerometer counts/min ranged from 525.8 to 738.3 with a mean of 642.6 ± 189.0 . The highest number of coexisting CVD risk factors (cluster) was 4. The percentage of participants classified as having any cluster of at least 2, 3, or 4 CVD risk factors was 12.0%, and 2.9%, 1.4%, respectively.

For this ROC analysis, equal weight was given to false positives and negatives in determining optimal cut-offs. As noted above, there were few children with abnormal HDL-C, TG, systolic and diastolic blood pressures, fasting glucose, or cluster of 3 or more CVD risk factors, and the ROC analysis for these outcomes did not identify optimal accelerometer cut-points for these variables before invoking the stopping rules. Optimal thresholds for percent of time spent at or above a count/ min level of intensity, where sensitivity best

approximates specificity for identifying overweight, elevated fasting insulin and clustering of 2 or more CVD risk factors are presented in Table 1. The table also includes the rate of positive cases corresponding to the threshold values (the percent of those above and below the threshold who are positive for the outcome), the chi-square, odds ratio, risk difference and AUC. The ROC curves (sensitivity versus 1-specificity) for BMI 95th percentile for age and fasting insulin 20 mU/L are shown in Figures 1 and 2, respectively. Only about 1 in 5 girls who spent at least 1.25% of their days (the equivalent of at least 11.1 minutes/day) at an activity level of 2600 counts/min or greater were overweight, compared with more than half of girls who did not meet that threshold of PA. A threshold of 1.86% of time (the equivalent of 16.6 minutes/day) at a level of at least 2,000 counts/min, optimally discriminated participants with and without elevated fasting insulin and also the clustering of 2 or more risk factors.

Table 2 includes descriptive characteristics of participants above and below each of the identified cut-offs. These results show that girls who were physically active at a level at or above the identified thresholds comprise a significantly younger group with, on average, lower levels of multiple CVD risk factors.

Discussion

The primary purpose of the present analysis was to use ROC analysis to determine the optimal accelerometer-measured PA thresholds, with the best combination of sensitivity and specificity, for predicting CVD risk factors and clustering of CVD risk factors in a sample of low-income African-American preadolescent girls. Our results indicate that the range of PA thresholds with the best tradeoff between true-positive and true-negative rates for identifying participants that were overweight, with hyperinsulinemia, or with clustering of 2 or more CVD risk factors, was the accumulation of about 11 to 17 minutes/day of activity at an intensity level of at least 2,000 to 2,600 counts/min.

Risk differences indicated that the identified thresholds defined groups with dramatically different rates of CVD risk factors, about 52% versus 21% for overweight, 21% versus 4% for elevated fasting insulin, and 28% versus 8% for the presence of 2 or more risk factors. The groups defined by these PA thresholds also appeared to generally differ on average, across multiple CVD risk factors. Of note, the total average daily counts/min was never chosen for defining the optimal cut-point as compared with a threshold combining duration and intensity. This suggests that the amounts of time spent above the defined count/min thresholds are more highly associated with CVD risk factors than averages of accelerometer counts across the entire day.

The ROC analysis identified an optimal PA threshold that is substantially lower than the durations and intensities currently recommended. The current PA recommendation of 60 minutes or more of MVPA on most days of the week are based on systematic review of the available evidence on the influence of PA on several health outcomes in children.^{7,8} As a result, these recommended thresholds are also frequently used for purposes of research, surveillance and program evaluation. Within the literature, however, different intensity thresholds are used to define MVPA. For example, Trost et al³⁰ define MVPA as any activity that is above 3 metabolic equivalents (MET) based on Freedson et al's age-specific equation.⁹ Three MET is equivalent to 1952 counts/min based on ActiGraph accelerometer.³¹ However, Puyau et al¹⁰ defined MVPA as activity counts greater than 3200 counts/min using the ActiGraph accelerometer in 6 to 16 year old boys and girls. Whereas, Trueth et al³² defined MVPA as activity counts greater than 3000 counts/min using energy expenditure in 13 to 14 year old boys and girls. The discrepancies in the accelerometer cut-point for MVPA derived by these studies can lead to different estimates of the amount of

time children spend engaged in MVPA.¹³ Regardless of the definition of MVPA derived in different studies, they would all still result in a recommendation that is substantially greater in both time and intensity than the optimal threshold found for discriminating those girls with and without CVD risk factors in our sample.

Do our findings suggest that current PA recommendations need to be reduced? Clinical and public health guidelines and recommendations are intended to improve the health of individuals, groups or the entire population, and may not equally value sensitivity and specificity. For example, if the recommended action (behavior, treatment, etc.) is very safe, acceptable, accessible, available and affordable, as well as very effective in reducing risk, it may be desirable to base recommendations on a cut-off that values sensitivity for identifying adverse outcomes more than specificity, so that more of the population would receive the benefits of the action as well as more people who do not need or benefit from the recommended action. For PA, this reasoning would result in a recommendation for a greater level of PA than the average minimum level or optimally sensitive and specific level necessary to produce benefits. In contrast, if the recommended action has severe adverse side effects and other undesirable outcomes for some people, and is not very effective in producing benefits, it may be more desirable to base recommendations on a cut-off that values specificity more than sensitivity, so that fewer susceptible people will be harmed but more people who would benefit from the action also would not get it. For PA, this reasoning would result in a recommendation for a lower level of PA. In this study, we equally weighted sensitivity and specificity to find the optimal cut-off for maximally discriminating those with a CVD risk factor from those without a CVD risk factor. Depending on the balance of potential costs and benefits, and supported by the substantial risk differences identified, this may be an appropriate balance to use when making public health recommendations regarding PA (benefiting the greatest number possible while also minimizing the number who will not benefit and/or be harmed).

For research, surveillance, and program evaluation purposes, it may also be appropriate to use the maximally efficient thresholds that define clinically-relevant risk. The thresholds that we identified, about 11 minutes/day at an intensity of at least 2,600 accelerometer counts/min for overweight, about 17 minutes/day at an intensity of at least 2,000 accelerometer counts/min for fasting insulin and clustering of 2 or more risk factors, most efficiently separate those girls with and without these CVD risk factors. If one assesses the prevalence of “adequate” PA in a population or the success of an intervention using a threshold that is higher than the optimal cut-point, one will underestimate the proportion of participants who may be benefiting from physical activity’s association with CVD risk factors. In contrast, using a lower threshold will result in an overestimate of the proportion of participants who may be benefiting. Therefore, when the purpose is to define “adequate” PA duration and intensity for its relationship with CVD risk factors, we suggest using thresholds derived from ROC analysis with clinically-relevant CVD risk factors as outcomes, such as those from this study. Our results suggest that these thresholds may be substantially lower than those that are commonly used in research, surveillance and program evaluation.

It is important to note that the ROC analysis performed in this study was exploratory, and these thresholds also should be validated in other similar samples. Because these thresholds were derived from our specific sample of low-income African-American girls, they may not apply as well to other samples. To help determine the optimal recommendations for PA, it would be appropriate to test the thresholds derived in this study, and those from future studies using ROC analysis in other samples, to inform the intensities and durations of PA that should be subsequently tested in experimental dose-response intervention trials with clinically-relevant outcomes.

Our findings differ from those of other studies due to both conceptual and methodological differences. Conceptually, our purpose was to determine the most clinically-relevant thresholds, including a combination of both duration and intensity that is most strongly associated with presence or absence of CVD risk factors. Accelerometer count thresholds from some previous studies were based on laboratory or field calibration of accelerometers with activities of different energy expenditure.^{10,32} Although a variety of activities were used to derive these counts/ min thresholds, they were not based on their abilities to discriminate between those with and without a clinical outcome. When clinically-relevant outcomes have been used in prior research, such as BMI and hyperinsulinemia, they tend to be for examining correlations with predetermined PA thresholds, often based on the current recommendations.^{3,18} These studies are important for testing whether those recommended levels of PA are correlates or risk factors for clinical outcomes but such studies may miss important associations that exist at lower or higher thresholds that discriminate better (a form of misclassification bias). In contrast, ROC analysis allowed us to empirically derive the thresholds with maximal discriminating power, placing equal weight on sensitivity (true-positives) and specificity (true-negatives).

A limitation of the present analysis is the utilization of cross-sectional data to identify PA cut-points. Although these PA cut-points are valid for identifying girls with and without the CVD risk factors, the associations found may not be causal, and whether the CVD risk factors preceded or followed PA is unknown. In addition, in the absence of more data on subsequent clinical outcomes related to CVD risk factors in children and adolescents, the definitions of outcomes remains somewhat subjective. The use of different definitions and the balance between sensitivity and specificity may result in different ROC results, and different PA thresholds. In this study we defined the outcome variables based on current national standards, and in the absence of such standards for fasting hyperinsulinemia we used a definition that has been frequently used by other researchers in pediatric populations. Despite the potential uncertainty about defining CVD risk factors, finding optimal thresholds that were so similar for the 3 outcomes gives us some confidence in the robustness of our results. In addition, while accelerometers provide a more objective measure of PA than self-reports, accelerometers also have some limitations. For example, accelerometers do not capture upper body activities or activities that occur mostly in the horizontal plane with minimal vertical acceleration such as cycling. Accelerometers also have to be removed during water activities such as swimming. Participants' accelerometer counts were measured in 1-minute epochs; this could have potentially masked some short bursts of activities and reduced the estimated minutes spent in moderate and vigorous activity. The most appropriate epoch length for different populations and different contexts is in need of additional study. We chose to perform this analysis in a sample of low-income preadolescent African-American girls in a single city to enhance internal validity, and avoid the potential problem of obscuring true associations via racial and geographic heterogeneity. Although our sample was an advantage for purposes of internal validity, it is unknown whether our study results will generalize to other groups of African-American girls, to girls of other ages and races, or to boys.

Regardless of these limitations, this study has demonstrated the potential usefulness of ROC analysis to derive optimal PA thresholds for clinically-relevant outcomes based on objective measures of PA, and a novel method to conceptualize and use both accelerometer intensity and duration data simultaneously in ROC analysis. This approach can be repeated in other populations to determine optimal PA thresholds with clinical validity, for research, surveillance and program evaluation, and to develop clinical and public health guidelines and recommendations.

Acknowledgments

This study was funded by a grant from National Heart, Lung, and Blood Institute (UO1 HL062663). We thank the Stanford GEMS girls and families for their participation in this study, the Stanford GEMS field staff for recruitment and data collection, K. Farish Haydel and Ann Varady, MS, for database management and assistance with statistical analysis, and Helena C. Kraemer, PhD, for statistical advice. Authors do not have any conflicts or financial interest in this research to disclose.

References

1. United States Department of Health and Human Services. Physical Activity and Health: A report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 1996.
2. Robinson T, Killen J, Kraemer H, et al. Dance and reducing television viewing to prevent weight gain in African-American girls: the Stanford GEMS pilot study. *Ethn Dis*. 2003; 13(Suppl 1):S65–S77. [PubMed: 12713212]
3. Brage S, Wedderkopp N, Ekelund U, et al. Objectively measured physical activity correlates with indices of insulin resistance in Danish children. The European Youth Heart Study (EYHS). *Int J Obes Relat Metab Disord*. 2004; 28:1503–1508. [PubMed: 15326467]
4. American College of Sports Medicine. Physical fitness in children and youth. *Med Sci Sports Exerc*. 1988; 20:422–423.
5. Sallis J, Patrick K. Physical activity guidelines for adolescents. *Pediatr Exerc Sci*. 1994; 6:302–314.
6. United States Department of Health and Human Services. With Understanding and Improving Health and Objectives for Improving Health. 2. Vol. 2. Washington, DC: U.S. Government Printing Office; 2000. Healthy People 2010.
7. Strong WB, Malina RM, Blimkie CJ, et al. Evidence based physical activity for school-age youth. *J Pediatr*. 2005; 146:732–737. [PubMed: 15973308]
8. United States Department of Health and Human Services U.S. Department of Agriculture. . Physical Activity. Dietary Guidelines for Americans. 6. Washington, DC: U.S. Government Printing Office; 2005. p. 30-33.
9. Freedson P, Sirard J, Debold E, et al. Calibration of the Computer Science and Application Inc. (CSA). *Med Sci Sports Exerc*. 1997; 29(Suppl):S45.
10. Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and Calibration of Physical Activity Monitors in Children. *Obes Res*. 2002; 10:150–157. [PubMed: 11886937]
11. Treuth MS, Sherwood NE, Baranowski T, et al. Physical activity self-report and accelerometry measures from the Girls Health Enrichment Multi-site Studies. *Prev Med*. 2004; 38(Supplement 1):43–49.
12. Stevens J, Murray DM, Catellier DJ, et al. Design of the Trial of Activity in Adolescent Girls (TAAG). *Contemp Clin Trials*. 2005; 26:223–233. [PubMed: 15837442]
13. Guinhouya CB, Hubert H, Soubrier S, Vilhelm C, Lemdani M, Durocher A. Moderate-to-vigorous physical activity among children: discrepancies in accelerometry-based cut-off points. *Obesity (Silver Spring)*. 2006; 14:774–777. [PubMed: 16855185]
14. Robinson TN, Kraemer HC, Matheson DM, et al. Stanford GEMS phase 2 obesity prevention trial for low-income African-American girls: design and sample baseline characteristics. *Contemp Clin Trials*. 2008; 29:56–69. [PubMed: 17600772]
15. Tryon W, Williams R. Fully proportional actigraph: a new instrument. *Behav Res Methods Instrum Comput*. 1996; 28:392–403.
16. Alhassan S, Sirard J, Spencer T, Varady A, Robinson TN. Estimating physical activity from incomplete accelerometer data in field studies. *J Phys Act Health*. 2008; 5(Suppl 1):s112–s125. [PubMed: 18364516]
17. Eslinger DW, Copeland JL, Barnes JD, Tremblay MS. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. *J Phys Act Health*. 2005; 3:366–383.
18. Himes JH, Dietz WH. Guidelines for Overweight in adolescent preventive services: recommendations from an expert committee. The Expert Committee on Clinical Guidelines for

- Overweight in Adolescent Preventive Services. *Am J Clin Nutr.* 1994; 59:307–316. [PubMed: 8310979]
19. Freedman DS, Dietz WH, Srinivasan SR, Berenson GS. The relation of overweight to cardiovascular risk factors among children and adolescents: The Bogalusa Heart Study. *Pediatrics.* 1999; 103:1175–1182. [PubMed: 10353925]
 20. National Cholesterol Education Program Expert Panel on Blood Cholesterol Levels in Children and Adolescents III. The individualized approach: detection/diagnosis/ evaluation. *Pediatrics.* 1992; 89:545–554.
 21. National Cholesterol Education Program Expert Panel on Blood Cholesterol Levels in Children and Adolescents I. Rationale for attention to cholesterol levels in children and adolescents. *Pediatrics.* 1992; 89:528–536.
 22. National Cholesterol Education Program Expert Panel on Blood Cholesterol Levels in Children and Adolescents. National Cholesterol Education Program (NCEP): Highlights of the report of the expert panel on blood cholesterol levels in children and adolescents. *Pediatrics.* 1992; 89:495–501. [PubMed: 1741227]
 23. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents. The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics.* 2004; 114:555–576. [PubMed: 15286277]
 24. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2005; 28(Suppl 1):S37–S42. [PubMed: 15618111]
 25. Decsi T, Molnar D. Insulin resistance syndrome in children: pathophysiology and potential management strategies. *Paediatr Drugs.* 2003; 55:291–299. [PubMed: 12716216]
 26. Ronnema T, Knip M, Lautala P, et al. Serum insulin and other cardiovascular risk indicators in children, adolescents and young adults. *Ann Med.* 1991; 23:67–72. [PubMed: 2036208]
 27. Kraemer, HC. *Evaluating Medical Tests: Objective and Quantitative Guidelines.* Newbury Park, CA: Sage Publications; 1992.
 28. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem.* 1993; 39:561–577. [PubMed: 8472349]
 29. Hanley JA. The robustness of the “binormal” assumptions used in fitting ROC curves. *Med Decis Making.* 1988; 8:197–203. [PubMed: 3398748]
 30. Trost SG, Pate RR, Sallis JF, et al. Age and gender differences in objectively measured physical activity in youth. *Med Sci Sports Exerc.* 2002; 34:350–355. [PubMed: 11828247]
 31. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998; 30:777–781. [PubMed: 9588623]
 32. Treuth MS, Schmitz K, Catellier DJ, et al. Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc.* 2004; 36:1259–1266. [PubMed: 15235335]

Figure 1.
ROC curve for BMI Percentile.

Figure 2.
ROC curve for Insulin.

Table 1
Optimal Counts/Minute and Percent of Time Thresholds for Selected CVD Risk Factors and Risk Score

Outcome variables	Counts per minute	Percent of time spent	Positive for outcome (%)	χ^2	OR	Risk difference	AUC	Sensitivity (%)	Specificity (%)
BMI 95th percentile for age	2,600	<1.25 1.25	51.9 20.9	21.3	4.08	31	0.66	60.3	72.9
Insulin >20 mU/L	2,000	<1.86 1.86	20.5 4.1	12.7	6.03	16.4	0.58	53.3	83.9
Cluster of 2 or more risk factors	2,000	<1.86 1.86	28.2 8.3	11.4	4.34	19.9	0.60	44.0	84.7

Note. All χ^2 values were significant at $P < .01$. To calculate average minutes/day of PA multiply percent of time by 891.4 minutes.

Table 2

Descriptive Characteristics (Mean $\pm$ SD) for Participants Above and Below Each of the Identified Physical Activity Cut-offs

Counts/min	Identified physical activity cut-offs			
	2,600		2,000	
% Time spent	<1.25 (n = 79)	1.25 (n = 129)	<1.86 (n = 40)	1.86 (n = 168)
Age (yrs)	9.1 $\pm$ 0.8	8.8 $\pm$ 0.8*	9.3 $\pm$ 0.7	8.8 $\pm$ 0.8 [†]
Weight (kg)	45.8 $\pm$ 14.7	37.2 $\pm$ 9.9 [‡]	47.3 $\pm$ 14.5	38.8 $\pm$ 11.6 [‡]
BMI (kg/m ²)	22.3 $\pm$ 5.4	19.6 $\pm$ 4.2 [‡]	22.9 $\pm$ 5.3	20.1 $\pm$ 4.5 [†]
HDL-C (mg/dL)	54.0 $\pm$ 13.9	56.5 $\pm$ 11.9	52.0 $\pm$ 12.2	56.4 $\pm$ 12.7*
TG (mg/dL)	68.4 $\pm$ 35.9	62.0 $\pm$ 29.2	71.0 $\pm$ 38.1	62.9 $\pm$ 30.2
SBP (mmHg)	98.5 $\pm$ 11.1	97.4 $\pm$ 9.5	100.4 $\pm$ 11.1	97.2 $\pm$ 9.8
DBP (mmHg)	56.2 $\pm$ 6.8	56.0 $\pm$ 6.5	57.1 $\pm$ 7.3	55.8 $\pm$ 6.4
Insulin (mU/L)	13.9 $\pm$ 12.8	8.3 $\pm$ 5.4 [‡]	15.3 $\pm$ 15.2	9.2 $\pm$ 6.9 [†]
Glucose (mg/dL)	86.0 $\pm$ 7.3	84.3 $\pm$ 6.7	86.6 $\pm$ 7.0	84.5 $\pm$ 7.0

Abbreviations: BMI, body mass index; HDL-C, high-density lipoprotein cholesterol; TG, triglyceride; SBP, systolic blood pressure; DBP, diastolic blood pressure.

Note. Statistical testing of differences between participants above and below the percent of time spent within each of the identified PA cut-offs (2,600 and 2,000 counts/min) was conducted using *t* tests.

[‡]Significant difference at $P < .0001$.

[†]Significant difference at $P < .01$.

*Significant difference at $P < .05$.