Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Science and Medicine in Sport 15 (2012) 311–321

Journal of Science and Medicine in Sport

www.elsevier.com/locate/jsams

Review

Everything you wanted to know about selecting the "right" Actigraph accelerometer cut-points for youth, but. . .: A systematic review

Youngwon Kim [a,*], Michael W. Beets [a], Gregory J. Welk [b]

[a] *Department of Exercise Science, University of South Carolina, USA*
[b] *Department of Kinesiology, Iowa State University, USA*

Received 29 April 2011; received in revised form 22 November 2011; accepted 2 December 2011

Abstract

Objectives: The purpose of this study is to provide an overview of the evidence on the calibration of ActiGraph accelerometers to quantify moderate-to-vigorous physical activity (MVPA) for youth through the use of cut-points and describe the independent validation studies comparing the accuracy of the developed cut-points to a criterion measure.
Design: A systematic review.
Methods: Studies were identified that: (a) developed ActiGraph accelerometer cut-points for children and youth (calibration study); or (b) performed an independent validation of already established cut-points (validation study). Both calibration studies and independent validation studies were retrieved through a systematic search of online databases. According to proposed guidelines for designing accelerometer calibration studies, each calibration study was evaluated on the following criteria: quality of a criterion measure employed; epoch length; inclusion of a variety of activities; and sample size.
Results: A total of 11 calibration studies were identified. Two studies met all four criteria for a calibration study. A total of 4 independent validation studies were identified. Three of them reported that no cut-points accurately classified moderate-to-vigorous physical activity (MVPA) across all ranges of physical activity intensity levels in comparison to a criterion measure. The fourth study reported two sets of cut-points that under laboratory conditions, accurately classified moderate-to-vigorous physical activity (MVPA) compared to indirect calorimetry.
Conclusions: Limited evidence suggests that two sets of cut-points correctly classify ActiGraph counts into moderate-to-vigorous physical activity (MVPA). However, limitations with calibration and validation studies indicate greater efforts aimed at designing high quality studies are needed to confirm these findings.

Keywords: Calibration; Child; Moderate-to-vigorous physical activity; Validation studies; Epoch; Adolescent

1. Introduction

The lack of standardization in collecting and processing Actigraph accelerometer data continues to plague physical activity research with children and youth. There are multiple cut-points available and evidence indicates that the application of different cut-points results in vastly different estimates of moderate-to-vigorous physical activity (MVPA).[1–3] Without consensus it is not possible to compare estimates of children and youth physical activity across studies or to accurately estimate population prevalence rates of children and youth meeting physical activity guidelines, although efforts have been made toward converting MVPA across different cut-points.[4] Researchers and practitioners continually to be challenged with the question "Which cut-points should be applied?" This paper is intended to help provide information to inform these decisions.

Accelerometers are a widely used objective method to measure "free-living" physical activity of children and youth. Their widespread use is attributed to their ability to measure intensity and duration, thereby providing estimates of time spent in MVPA. Among various types of accelerometers, ActiGraph Models 7164, GT1M, GT3X (formerly known as Computer Science and Applications (CSA) and Manufacturing Technology, Inc. (MTI), Pensacola, FL) are one of

* Corresponding author.
E-mail address: myungshingo@gmail.com (Y. Kim).

1440-2440/$ – see front matter
doi:10.1016/j.jsams.2011.12.001

the most broadly used objective monitors. ActiGraph is a monitor that uses a piezoelectric acceleration sensor to filter and convert the signals produced from the sensor in samples collected at a preset frequency in hertz. The samples are summed over a user-specified time sampling interval, called an "epoch". Activity "counts", converted from the accelerations over a given epoch, are recorded to the internal memory of accelerometers.[5]

Numerous calibration studies have focused on translating accelerometer output (counts per a given epoch) into time spent in various physical activity intensity levels (i.e., sedentary, light, moderate, and vigorous). Establishing "cut-points" has been the most widely used method to create the link between accelerometer counts and physical activity intensity. Cut-points are developed in calibration studies in which participants perform various types of field (e.g., skipping, running) or laboratory (e.g., treadmill) activities while wearing an accelerometer along with the concurrent collection of energy expenditure from a criterion measure (e.g., indirect calorimetry). Accelerometer activity counts are then compared to the criterion measure (e.g., metabolic equivalents, METs). Counts corresponding to defined values for energy expenditure of MVPA (e.g., for moderate intensity, METs = 4 or VO_2 = 20 mL/kg/min) are then regarded as "cut-points" for moderate and above intensity physical activity. Based on these, accelerometer counts for a given epoch at or exceeding the defined cut-points are considered MVPA.

The accuracy of cut-points developed from calibration studies is subsequently tested in independent validation studies. Independent validation studies typically utilize different settings (e.g., criteria measures, participants, epoch lengths, types of activities included) from those used in the calibration studies, with the goal of determining how closely existing cut-points estimate MVPA in comparison to MVPA measured by a criterion measure. Cut-points that estimate MVPA most closely to that of the criterion measure are considered the most "accurate" and are recommended for widespread application.

Although considerable work on both these issues has been conducted, there exists a lack of agreement on which cut-points can be considered the "best". This is seen by the broad use of different cut-points in the literature. Furthermore, no formal review of existing calibration and corresponding validation studies has been performed. Therefore, this paper is intended to provide an overview of the evidence regarding the calibration of accelerometers to classify MVPA through the use of cut-points and to describe the independent validation studies comparing the accuracy of the developed cut-points.

2. Methods

Calibration studies and independent validation studies were retrieved by the author (Y.K.) from June 2011 through November 2011 through a systematic search of online databases (e.g., PubMed, Google Scholar) using the following keywords: 'accelerometer/accelerometry', 'calibration', 'cut-points/off', 'threshold', 'counts', 'children', 'adolescents', 'youth', 'physical activity', 'validation' and combinations thereof. EBSCOHost was used for cross referencing for any additional articles. Reference lists of identified studies and studies that utilized accelerometers to measure children and youth physical activity were also searched for relevant articles by another author (M.B.). Once potentially appropriate articles were collected, inclusion/exclusion criteria were applied to create a final list of studies included (Y.W. and M.B.). The inclusion criteria were Actigraph accelerometer calibration and/or validation studies published through June 2011. The sample age range was children and youth 18 yrs old and younger. Studies that developed cut-points with other brands of accelerometers (e.g., Actical, Actiwatch, RT3 Triaxial, armbands) were excluded. Discrepancies between reviewers were resolved through discussion. For the evaluation of calibration studies, activity counts for cut-points, types of a criterion measure, epoch length, protocol, demographic information on participants, and analytical procedures were extracted from the calibration studies retrieved (Y.W. and M.B.). For the evaluation of validation studies, demographic information on participants, protocols, analytical procedures and different types of cut-points validated were extracted from the validation studies retrieved (Y.W. and M.B.).

Each calibration study was assessed and scored based on the degree of compliance with published recommendations by Freedson et al.[6] and Welk et al.[7] The four criteria used for the evaluation were: (1) use of an appropriate biological standard, (2) inclusion of a wide variety of activities, (3) use of an epoch length less than 60 s and (4) sample size with at least 10 participants per age group.[6] Although direct observation has been used as a criterion measure the use of a biological standard, such as indirect calorimetry, provides a way to convert to more usable quantifiable metrics to then compare with accelerometer counts. Therefore, studies were weighted more favorably if they used indirect calorimetry rather than direct observation. The authors were unable to locate established guidelines to classify studies based on the degree to which they included "a wide variety of activities". Therefore, arbitrary standards were established that included: (1) at least a total of 6 activities that include at least 3 moderate or vigorous activities should be included in the calibration study, and (2) at least 50% of total activities included in the calibration study should be either free-living or over-ground activities. A short epoch length is an important, if not essential, aspect for children and youth calibration studies due to the intermittent pattern of physical activity behaviors (i.e., alternating short bouts of activity). Studies were rated more favorably if they used epochs less than 60 s.[8,9] The requirement of at least 10 children or youth per age group is somewhat arbitrary but reasonably large samples are needed to produce generalizable results.[6] Each study was scored with either a "+" or "−" based on these criteria.