Gait & Posture 41 (2015) 516–521

Contents lists available at ScienceDirect

Gait & Posture

journal homepage: www.elsevier.com/locate/gaitpost

# A sensitivity analysis on the variability in accelerometer data processing for monitoring physical activity

Paul H. Lee *

*School of Nursing, Hong Kong Polytechnic University, Hong Kong*

ARTICLE INFO

*Article history:*
Received 25 June 2014
Received in revised form 2 December 2014
Accepted 3 December 2014

*Keywords:*
Accelerometry
Data analysis
Lifestyles
Motor activity
NHANES

ABSTRACT

*Background:* Accelerometers are gaining popularity for measuring physical activity, but there are many different ways to process accelerometer data. A sensitivity analysis was conducted to study the effect of varying accelerometer data processing protocols on estimating the association between PA level and socio-demographic characteristics using the National Health and Nutrition Examination Survey (NHANES) accelerometer data.

*Methods:* The NHANES waves 2003–2004 and 2005–2006 accelerometer data ($n$ = 14,072) were used to investigate the effect of changing the accelerometer non-wearing time and valid day definitions on the demographic composition of the filtered datasets and the association between physical activity (PA) and socio-demographic characteristics (sex, age, race, educational level, marital status).

*Results:* Under different filtering rules (minimum number of valid day and definition of non-wear time), the demographic characteristics of the final sample varied. The proportion of participants aged 20–29 decreased from 18.9% to 15.8% when the minimum number of valid days required increased from 1 to 4 ($p$ for trend $< 0.001$), whereas that for aged $\geq$70 years increased from 18.9% to 20.6% ($p$ for trend $< 0.001$). Furthermore, with different filters, the effect of these demographic variables and PA varied, with some variables being significant under certain filtering rules but becoming insignificant under some other rules.

*Conclusions:* The sensitivity analysis showed that the significance of the association between socio-demographic variables and PA could be varied with the definition of non-wearing time and minimum number of valid days.



## 1. Introduction

High levels of sedentary behaviors and physical inactivity have become a worldwide epidemic despite their association with increased chronic diseases [1–4]. In evaluating intervention programs on physical activity promotion, reliable and valid physical activity (PA) measures are essential. Self-report questionnaires were traditional instruments to measure PA, but their validity varies [5,6]. Accelerometers, electronic devices that measure the acceleration as a proxy of PA intensity and a measure of duration, have become more popular in large scale cohort studies [7,8] as they are valid instruments of PA measurement [9].

Unlike PA questionnaires that have standardized data-processing protocols, the processing of accelerometer data has not been standardized. The accelerometer is only a proxy to human body movement, nevertheless, it can represent the types of human activity by suitable transformation of data [10,11], yet the performance was not without errors.

The majority of the accelerometers used in assessment of human PA counts the number of movements that accelerate faster than a pre-specified threshold [12]. These accelerometer counts data were shown to be associated with various quantifiable PA outcomes, including time spent in PA [12], total energy expenditure [13], and time spent sitting [14]. Many calibration studies had been carried out to build transformation algorithms linking raw accelerometer count data to these PA outcomes. For example, there are well over 14 different algorithms to transform accelerometer counts data to total energy expenditure [15]. Given so many possibilities, researchers could have manipulated the data to generate favorable results, and this phenomenon was also observed in other research fields [16]. For example, in reporting the effectiveness of an intervention program to increase PA level,

* Correspondence to: School of Nursing, PQ433, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. Tel.:+852 3400 8275; fax: +852 2364 9663.
*E-mail address:* paul.h.lee@polyu.edu.hk

http://dx.doi.org/10.1016/j.gaitpost.2014.12.008
0966-6362/

it is tempting to choose the transformation algorithm that maximizes the absolute effect of the intervention.

A recent review showed great variability in processing accelerometer data on studies analyzing the open-to-public National Health and Nutrition Examination Survey (NHANES) waves 2003–2004 and 2005–2006 data [17]. This review showed that the greatest variability in data processing lies on the cutoff definition used for time spent in different PA level. Another sensitivity study showed that varying on the rules of two variables: (a) definition of accelerometer non-wearing time and (b) minimum number of valid days required for a participant could affect different PA outcomes including accelerometer counts, sedentary behaviors, and moderate-to-vigorous PA [18]. However, the effect of changing the rules of (a) and (b) on the socio-demographic characteristics of the sample and their associations with PA level remains unknown, and this study aims to assess this through a sensitivity analysis using the NHANES waves 2003–2004 and 2005–2006 data.

## 2. Materials and methods

### 2.1. Participants

The NHANES waves 2003–2004 and 2005–2006 data were used in this study. This survey, conducted by the National Center for Health Statistics, was designed to assess the health and nutrition status in the United States (http://www.cdc.gov/nchs/nhanes/about_nhanes.htm). The sample was selected with a multistage probability cluster design and was representative of the United States population. Participants were invited by the National Center for Health Statistics to complete a survey and a health examination; the details could be obtained in the NHANES Website (http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm). Consent was obtained from the participants by the National Center for Health Statistics, and the study was approved by the National Center for Health Statistics ethics review board.

Participants aged 6 years and older were asked by the National Center for Health Statistics to wear an ActiGraph model 7164 (ActiGraph LLC) for 7 days over the right hip for all waking hours, removed only when bathing or sleeping. Out of the 20,470 participants who completed the 2003–2004 and 2005–2006 NHANES, 14,072 (68.7%) wore the accelerometer. These data were recorded in 1-min epochs. All participants with missing socio-demographic characteristics and outliers (>20,000 cpm) were removed from the analysis ($n = 6249$).

### 2.2. Sensitivity analysis

All analyses were performed using IBM-SPSS version 20. The primary outcome variable used throughout this study was the accelerometer counts per valid minute (cpm), defined as the total accelerometer counts in a 24-h period divided by the total valid accelerometer minutes (defined below) in the corresponding time period. In addition, secondary outcome variables including time spent per day in sedentary behaviors, light activity, moderate-intensity activity, and vigorous intensity activity were also used. A minute of accelerometer counts of <100, 100–1951, 1952–5724, and ≥5725 will be classified as sedentary, light activity, moderate-intensity activity, and vigorous-intensity activity respectively [12].

To demonstrate the effect on variability in accelerometer data processing, analyses were repeated under eight different data-filtering rules. A day was considered valid if the number of wearing hours was at least 10, in which 90.7% of the papers utilizing the NHANES 2003–2006 accelerometer data used this criterion [17]. The variability in adopting the two most commonly used non-wearing time criterion for this data [17], that is,

(i) ≥60 consecutive zeros and
(ii) ≥60 consecutive zeros allowing 1–2 counts ranging between 1 and 100, were tested. Besides non-wearing time, the minimum number of valid days (1–4) was also varied.

Under all these filtering rules, the demographic characteristics and the PA level with different outcomes (accelerometer cpm/time spent per day in different activities) were computed and the demographic composition across these filtering rules was compared using Pearson $\chi^2$ test. For demographic characteristics demonstrating a significant association with minimum number of valid days, Cochran–Armitage test for trend was used to test its linear trend across minimum number of valid days. Furthermore, a multi-variable regression examining the relationship between socio-demographic characteristics (sex, age, race, educational level, and marital status) and accelerometer cpm/time spent per day in different activities was performed under the aforementioned filtering rules.

## 3. Results

Table 1 shows the demographic characteristics of the sample under different filtering rules. There was a clear linear trend across minimum number of valid days required for the proportions of participants belonged to age groups 20–29 years and ≥70 years. When non-wear time was defined as 60 or more consecutive zeros, the proportion of participants aged 20–29 years decreased from 18.9% to 15.8% when the minimum number of valid days required increased from 1 to 4 ($p$ for trend < 0.001; Table 1), whereas that for aged ≥70 years increased from 18.9% to 20.6% ($p$ for trend < 0.001). Similar trends were found for non-wear time definition as 60 or more consecutive zeros allowing 1–2 min with counts between 1 and 100, with both $p$ for trend < 0.001 (Table 1).

Under different filtering rules, the estimation of population-level PA was stable. The accelerometer cpm, time spent in sedentary behaviors, light activity, moderate-intensity activity, and vigorous-intensity activity ranged from 333.39 to 375.97 cpm, from 537.05 to 553.31 min/day, from 316.14 to 330.89 min/day, from 30.44 to 31.14 min/day, and from 2.52 to 3.75 min/day, respectively.

Table 2 shows the regression results of sex, age, race, educational level, and marital status on accelerometer cpm. The difference between age 30–39 years and ≥70 years in accelerometer cpm varied from 167.31 cpm ($p < 0.001$, ≥1 valid days, ≥60 consecutive zeros) to 195.56 cpm ($p < 0.001$, ≥2 valid days, ≥60 consecutive zeros allowing 1–2 min with count between 1 and 100), representing a 16.9% increment. Furthermore, some demographic characteristics were significant under certain filtering rules, but became insignificant under some other rules, including education (less than 9 years of high school education, 9–11 years of high school education, and completed high school, relative to college graduate or above) and marital status (separated and never married, relative to living with partner). The details are shown in Table 2.

Tables S1–S4 (available in the online supplementary materials) show the regression results of sex, age, race, educational level, and marital status on minutes spent per day in sedentary behaviors, light activity, moderate-intensity activity, and vigorous intensity activity, respectively. Similar to the results shown by the regression using accelerometer cpm as outcome, some variables were significant under certain filtering rules, but became insignificant under some other rules, including age (50–59 years old and 60–69 years old, relative to ≥70 years old), sex, race (Mexican American relative to other races), education (less than 9 years of high school education, 9–11 years of high school education, and completed high school, relative to college graduate