Science of the Total Environment 454–455 (2013) 184–188

Contents lists available at SciVerse ScienceDirect

Science of the Total Environment

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

Simple to complex modeling of breathing volume using a motion sensor

Dinesh John [a,*], John Staudenmayer [b], Patty Freedson [c]

[a] Department of Health Sciences, Northeastern University, 360 Huntington Avenue, 316 Robinson Bldg., Boston, MA 02115, United States
[b] Department of Mathematics and Statistics, University of Massachusetts, Amherst, United States
[c] Department of Kinesiology, University of Massachusetts, Amherst, United States

HIGHLIGHTS

- We developed simple and complex methods to predict breathing from accelerometers.
- Simple multiple regression and complex random forest techniques performed comparably.
- Accelerometry can be used to predict breathing volume, duration and frequency.
- The methods may improve the understanding of how toxin exposure impacts disease.

ARTICLE INFO

Article history:
Received 30 July 2012
Received in revised form 25 November 2012
Accepted 27 February 2013
Available online 27 March 2013

Keywords:
Breathing volume
Accelerometer
Machine learning

ABSTRACT

Purpose: To compare simple and complex modeling techniques to estimate categories of low, medium, and high ventilation (VE) from ActiGraph™ activity counts.
Methods: Vertical axis ActiGraph™ GT1M activity counts, oxygen consumption and VE were measured during treadmill walking and running, sports, household chores and labor-intensive employment activities. Categories of low (<19.3 l/min), medium (19.3 to 35.4 l/min) and high (>35.4 l/min) VEs were derived from activity intensity classifications (light <2.9 METs, moderate 3.0 to 5.9 METs and vigorous >6.0 METs). We examined the accuracy of two simple techniques (multiple regression and activity count cut-point analyses) and one complex (random forest technique) modeling technique in predicting VE from activity counts.
Results: Prediction accuracy of the complex random forest technique was marginally better than the simple multiple regression method. Both techniques accurately predicted VE categories almost 80% of the time. The multiple regression and random forest techniques were more accurate (85 to 88%) in predicting medium VE. Both techniques predicted the high VE (70 to 73%) with greater accuracy than low VE (57 to 60%). Actigraph™ cut-points for light, medium and high VEs were <1381, 1381 to 3660 and >3660 cpm.
Conclusions: There were minor differences in prediction accuracy between the multiple regression and the random forest technique. This study provides methods to objectively estimate VE categories using activity monitors that can easily be deployed in the field. Objective estimates of VE should provide a better understanding of the dose–response relationship between internal exposure to pollutants and disease.

1. Introduction

The etiology of several diseases is attributed to interactions among the physical, chemical, and biological characteristics of the environment with the human genome (Hunter, 2005). Inhalation is a common pathway for various chemical and biological toxins to enter the human body. Environmental researchers are interested in quantifying inhalation exposure to understand the relationship between toxin exposure and disease development (Boffetta et al., 1997; Dons et al., 2012; Mills et al., 2007). Various devices are now available to measure the concentrations and type of toxins in the environment. In addition to toxin concentration, the accurate assessment of inhalation exposure requires information on breathing rate and volume (VE). Breathing rate is measured as the total number of breaths per minute and VE is the total volume of air inspired per minute (breathing rate × tidal volume). Combining data on the type and concentration of inhaled toxins with VE is necessary to comprehensively understand the dose–response relationship between toxic exposure and disease development. Recent evidence suggests that motion sensors such as accelerometers may be useful to estimate VE (Kawahara et al., 2011; Rodes et al., 2012).

Accelerometers have been extensively used to estimate metabolic equivalents (METs) or energy expenditure (kcals) (Crouter et al., 2006; Freedson et al., 1998; Staudenmayer et al., 2009). Accelerometers vary in the number of axes that detect acceleration, data storing capacity, onboard data processing capabilities and battery life. Output from these monitors increase linearly with activity intensity during most light and moderate intensity activities (Freedson et al., 1998, 2011; Staudenmayer

* Corresponding author. Tel.: +1 617 373 5695; fax: +1 617 373 2968.

0048-9697/$ – see front matter
http://dx.doi.org/10.1016/j.scitotenv.2013.02.092

et al., 2009). A commonly used accelerometer is the uniaxial ActiGraph™ monitor (ActiGraph™, LLC, Pensacola, FL). Activity counts from ActiGraph ™ accelerometers have been used to develop both simple and complex modeling methods to quantify physical activity (Freedson et al., 1998, 2011; John et al., 2011; Staudenmayer et al., 2009). These techniques are typically developed and validated in the lab with measured VO_2 during various ambulatory and simulated free-living activities as the criterion measure. The models are then applied in the free-living environment to estimate physical activity variables. A similar approach may be useful to estimate breathing volume intensity categories using the ActiGraph™ accelerometer. Recently, an accelerometer was used to estimate VE in children (Kawahara et al., 2011) and Rodes et al. (2012) used simple linear regression to estimate VE from various accelerometers in adults.

Machine learning techniques using advanced statistical prediction models are trained on input data to predict an outcome variable. These techniques are novel because they detect underlying patterns in the input data and are adaptable to improve prediction accuracy. The potential of using machine learning techniques to estimate VE from accelerometer data has not been examined. The objective of this study was to use simple and complex modeling techniques to estimate categories of low, medium, and high VEs from ActiGraph™ activity counts during a variety of ambulatory and simulated free-living activities.

2. Material and methods

2.1. Participants

Two hundred and seventy-seven healthy men and women (mean ± SD: age = 38.0 ± 12.4 years, BMI = 24.6 ± 4.0 kg/m^2) were recruited from the University of Massachusetts, Amherst and surrounding areas. The study was approved by the University of Massachusetts, Amherst Institutional Review Board and all participants provided written informed consent. Participants were screened for chronic diseases including cardiovascular and pulmonary disorders and exercise readiness using a health history questionnaire and the Physical Activity Readiness Questionnaire (PAR-Q). Male and female participants above the age of 40 and 50 years, respectively, were screened for cardiovascular disease risk with a physician-supervised 12-lead ECG stress test in accordance with the American College of Sports Medicine Guidelines for Exercise Testing (2009). Participants reporting any contraindications to exercise on the health history and PAR-Q questionnaires, displaying symptoms of cardiovascular disease during the stress test, or were on any medication that alter metabolic rate were excluded from the study.

2.2. Resting metabolic rate

Resting metabolic rate (RMR) was measured using the MedGem Analyzer (HealtheTech, Inc., Golden, CO). The MedGem is a valid device for measuring RMR (Nieman et al., 2003). Following a 4-h restriction of food, caffeine and exercise, participants rested quietly for 15 min in the supine position. Measured RMR was used to calculate METs to indicate the intensity for each activity from the lab-based activity protocols (explained below).

2.3. ActiGraph™ accelerometer

The uniaxial ActiGraph™ GT1M (5.1 × 3.8 × 1.5 cm, 42.6 g) monitor was used in this study. The GT1M detects accelerations in the vertical plane ranging between 0.05 and 2.0 G that lie within a frequency range of 0.25 to 2.5 Hz. Accelerations are sampled at a rate of 30 Hz and then converted to activity counts for a user specified time interval (epoch). The monitors were initialized to collect data in 1-s epochs and the results were downloaded using software (ActiLife v. 3.1.0.) provided by the manufacturer. One-second activity counts were summed to obtain 1-min values (counts per minute or cpm) that were used in the data analyses. Participants wore the GT1M monitor snugly at the waist in line with the anterior axillary line using an elastic belt. ActiGraph™ monitors are commonly worn at the hip because of its proximity to the center of mass of the body and hip movement is representative of whole body movement.

2.4. Activity protocols

The activity protocol consisted of two routines of nine activities and the activities were performed in random order. Each routine consisted of treadmill activities and activities of daily living. Participants performed six treadmill activities at three speeds (1.34, 1.56, and 2.23 m.s^{-1}), each at 0 and 3% grade. Three activities of daily living were randomly selected from the following set of 15 activities: cleaning the room, dusting, gardening, laundry, mopping, moving a box, mowing, painting, raking, sweeping, trimming, vacuuming, washing dishes, basketball and tennis. These activities represent common household, leisure time and sporting activities and were performed at a self-selected pace. Each activity was performed for 7 min with 4 min of rest between activities. Participants were allowed to stop performing an activity if they were unable to maintain activity intensity (e.g. high treadmill speed).

2.5. Indirect calorimetry

Criterion VE and oxygen consumption (VO_2) were measured on a breath-by-breath basis using a portable metabolic measurement system (Oxycon Mobile™; CareFusion, Yorba Linda, California) during each activity. The Oxycon Mobile™ system consists of a facemask with a small flow-meter and sampling line (for expired air) connected to two small units mounted in a harness secured to the upper back. This system was calibrated using its automatic flow calibrator and a known gas mixture of oxygen (16%) and carbon dioxide (4%) before each use.

2.6. Data reduction and analyses

Activity data were not included in the analyses if a participant was unable to complete the activity or if either the Oxycon Mobile™ or ActiGraph™ activity monitor malfunctioned. Steady state VE, VO_2 and activity counts for each activity were obtained after discarding data for the first two minutes and averaging values for minutes 3 to 7. MET cut-offs (derived from measured VO_2 and RMR) for light (less than 3 METs), moderate (3 to 6 METs), and vigorous (greater than 6 METs) activities were used to determine low, medium and high VEs. Due to the near linear relationship between activity intensity and VE, simple linear regression between steady state METs (independent variable) and VE was used to determine VE values corresponding to 3 and 6 METs. Medium and high VE cut-offs corresponded to 19.3 ± 1.6 and 35.4 ± 0.14 L/min, respectively. Similar MET-based methodologies have been used by the United States Environment Protection Agency (EPA) to account for differences arising from individual variability (McCurdy, 2000).

Two simple techniques and one complex modeling technique were evaluated to predict VE categories from activity counts. The simple modeling techniques were multiple regression analysis and an activity count 'cut-point' method. Activity count cut-points for medium and high VE categories were determined using receiving operator characteristic (ROC) curves. We calculated two variables over a wide range of cpm cut-points: true positive percentage (y-axis) and false positive percentage (x-axis). Fig. 2A and B depict the ROC curves used to determine cut-points for medium and high VEs. We considered cut-points from 60 cpm to 12240 cpm. Each data-point in 2A and B represents different cut-points equally spaced by 60 cpm. In Fig. 2A, the true positive percentage (sensitivity) on the y-axis is the fraction of minutes of medium VE correctly detected by the cut-point. The false positive percentage (1-specificity) is the fraction of minutes that are not at least medium VE but were incorrectly determined to be at least medium VE by the cut-point. Fig. 2B is similar, but detects high VE. On each graph, we