Technical report on accelerometry data processing in the Millennium Cohort Study

Authors

Marco Geraci, Carly Rich, Francesco Sera, Mario Cortina-Borja, Lucy J Griffiths, Carol Dezateux

Affiliation

Centre for Paediatric Epidemiology and Biostatistics/ MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, London WC1N 1EH, UK

Date of production

November 2012

Scope

This document applies to the preparation of a Standard Operating Procedure (SOP) for processing routines and basic statistical analysis of accelerometry data coming from the Physical Activity Study carried out as part of the UK Millennium Cohort Study (MCS).

Objective

To define a collection of routines that can be applied to accelerometry data.

Background

The Millennium Cohort Study (MCS) is a UK-wide prospective study of approximately 19,000 British children born at the start of the new century. Parents/ guardians of the children were interviewed when the children were aged nine months (first sweep or MCS1), three (MCS2), five (MCS3) and seven years (MCS4). The population was stratified by UK country and additional stratification by socioeconomic and demographic variables within country was carried out using data available at the electoral ward level. In particular, families were taken from a random sample of electoral wards, the primary sampling unit, disproportionately stratified to ensure an adequate representation of disadvantaged areas and ethnic minority groups. The original cohort comprised 18,819 children. Detailed information was obtained on demographic, social, and health factors relating to the children, and the children's siblings and parents through interviews of the main respondents and their partners in the home.

At aged seven years, accelerometers were introduced into the study to measure the levels and patterns of physical activity (PA) and sedentary behaviour (SB) in the MCS cohort children. All children participating in the MCS4 interviews (14,043) were invited to participate in wearing an accelerometer. Children were asked to wear an accelerometer for seven consecutive days (including two weekend days). Accelerometers were programmed to measure data using a short sampling period (15 second epochs), and were set to record both counts and steps. Accelerometry data were collected between May 2008 and August 2009 and comprised 13,219 consenting (94.1%) children.

The MCS4 accelerometer study also included a sub-study that investigated seasonal variation in levels and patterns of children's PA and SB. Children who wore their accelerometer during winter 2008/09 (November 2008 - January 2009) and returned their accelerometer for the

main physical activity data collection were invited to participate in a seasonal study. They were asked to wear their accelerometer on four further occasions using the same measurement protocol, so that a measurement was made in each season throughout a single calendar year. The MCS4 seasonal accelerometer study took place between November 2008 and October 2009. Accelerometer measurements were also obtained in a further winter season between December 2009 and January 2010 to investigate any variation in levels and patterns of children's PA and SB between winter 2008/09 and winter 2009/10.

Data Processing Procedure

Raw data coming from accelerometers must undergo a series of operations in order to be used in subsequent statistical analyses. Three stages of increasing level of data refining were identified:

1) **Stage**: pre-processing. **Input**: level-0 data (raw data). **Criteria**: standard. **Output**: level-1 data.
    a. Data cleaning and integrity. Detection and marking of aberrant values (all 5 digit values; all one value; values equal to 32,767 which indicates voltage signal saturation within the piezosensor – Esliger et al, 2005; file that does not return to baseline – Sherar et al, 2011); negative values; other errors in data recording.
    b. Data and metadata consistency. Correspondence between accelerometer metadata and study-defined parameters – e.g., start time and date, epoch, mode.
    c. Storage.
2) **Stage**: processing. **Input**: level-1 data. **Criteria**: recommended. **Output**: level-2 data.
    a. Time stamping. Date and time attribution to values.
    b. Labelling. Classification of wear and non-wear time status on epoch-by-epoch basis using number of consecutive zero-counts.
    c. Truncation. Exclusion of low and high end days..
    d. Storage.
3) **Stage**: post-processing. **Input**: level-2 data. **Criteria**: recommended. **Output**: level-3 data.
    a. Censoring. Detection and marking of valid day and minimum number of valid days per child.

b. Classification of valid data. Definition of sedentary behaviour and of light, moderate, and vigorous activity. Definition of behaviourally implausible values.
c. Collapsing data. Creation of ready-to-use datasets.
d. Storage.

MCS Accelerometry Data

Main survey

12,625 accelerometers were sent for the main survey (Winter 2008/09, hereafter S1) of which 10,034 were returned. Individual raw accelerometer files were stored as ".dat" files (ASCII).

Seasonal surveys

Overall, a total of 705 out of 1289 (55.0%) invited parents gave consent for their child to participate in the MCS4 seasonal accelerometer study (spring, summer, autumn 2009 and winter 2009/10, hereafter S2, S3, S4, S5 respectively). Individual raw accelerometer files were stored as ".dat" files (ASCII).

Results

In this section we report the results of the application of data processing procedure to the MCS accelerometry data.

Stage 1: Pre-processing

This stage is regarded as basic. The criteria should be considered as standard.

a) Data cleaning and integrity

There were 11 files with 237 negative counts in S1. In one file there were 211 counts equal to -32,768. Overall, negative counts were equal to -32,768 (234), -32,161 (1), -21582 (1), -9,049 (1). Files with negative measurements for counts had also negative measurements for steps. Aberrant values were in files with all zero counts. Files with zero counts had also zero steps.

Variables defined: file identifier; acceleration count values; step values; data error coding (codes: 0 – no error, 1 – aberrant value, 2 – negative value, 3 – other errors).

Table 1 (Stage 1.a). Data errors report: Number of files (number of count measurements) by survey sweep.

	S1	S2	S3	S4	S5
Total processed files	8,968*	587	554	507**	435
Aberrant values	31 (2,962,154)	2 (85,748)	2 (110,888)	2 (206,220)	1 (59,358)
Negative values	11 (237)	1 (1)	1 (1)	-	-
Other errors	-	-	-	-	-

*This includes 4 MCS singletons each with two different 'dat' files. If 1 of the 2 files (either with later starting date or with no data in it) is excluded, the number of remaining files is 8,964.
**One file was corrupted and not-readable.

b) Data and metadata consistency

Dates were expected to be in ranges 20/05/2008 to 18/05/2009 (S1), 31/03/2009 to 23/10/2009 (S2, S3, S4), 12/01/2010 to 25/01/2010 (S5) and in d/m/y format. Mode was expected to be set to 1 (counts and steps). All files had epoch set to 15 seconds.

Variables defined: file identifier; total number of measurements; epoch; start date and time; time zone (GMT, BST); coded survey sweep (S1 - S5); accelerometer mode; dummy variable for marking odd number of measurements when mode set to 1 (codes: 0 – false, 1 – true).

Table 2 (Stage 1.b). Metadata consistency report: Number of files by survey sweep.

Number of files	S1	S2	S3	S4	S5
Date as expected	8,985	590	556	508	428
Date in m/d/y format	22	-	-	-	-
Outside date range	2	-	-	1	9
Mode 0	23	-	-	18	-
Odd total number of values	-	-	-	-	-

c) Storage

Stage 1 output files were stored as ".Rdata" files (R format).

Stage 2: Processing

This stage is regarded as intermediate. The criteria are strongly recommended.

a) Time stamping

Variables defined: time and date for each measurement.

b) Labelling

Non-wear time defined as any time period of consecutive zero-counts for a minimum of 20 minutes.

Variables defined: dummy variable to identify wear time (codes: 1 – non-wear, 2 – wear).

Table 3 (Stage 2.a). Wear time report: Number of epochs and proportion of mean wear time per child by survey sweep.

	S1	S2	S3	S4	S5
Total number of epochs	854,078,139	48,053,824	57,772,538	63,126,298	39,913,381
Wear	148,923,106	12,084,905	11,344,388	10,343,553	9,106,509
Non-wear	705,155,033	35,968,919	46,428,150	52,782,745	30,806,872
Mean wear time (as % of total time) per child	22.2%	28.3%	21.9%	17.2%	24.6%

c) Truncation

The threshold was set to ≥ 150 minutes per day[1]. To determine the first and the last day of wear, total wear time (TWT) for each day was assessed. The first day with TWT exceeding the threshold was set as start date. The last day with TWT above the threshold was set as end date.

Variables defined: dummy variable to identify start date (codes: 0 – false, 1 – true); dummy variable to identify end date (codes: 0 – false, 1 – true); dummy variable to identify truncated days (codes: 0 – false, 1 – true).

d) Storage

Stage 2 output files were stored as ".Rdata" files (R format).

[1] A study (Sera et al, 2011. Unpublished Technical Report) was performed to discriminate between days when the accelerometer was worn and days when it was likely not worn (e.g., during postal delivery) on the basis of the minimum number of 'valid' hours per day. Initially, a 'manual' screening of a sample of files to identify start and end time points was conducted. Then, the output produced by automatic procedures for several cutoffs was compared to the results of manual detection. This produced an optimal cut off equal to 2.5 hours per day.

Table 4 (Stage 2.b). Truncation report: Number of days by survey sweep.

	S1	S2	S3	S4	S5
Total number of days	154,452	8,707	10,406	11,317	7,212
Truncated days	45,105	2,304	4,403	4,305	2,474
Percentage total	29.2%	26.5%	42.3%	38.0%	34.3%

Stage 3: Post-processing

This stage is regarded as advanced. The criteria are recommended.

a) Censoring

Valid days were defined as days with TWT ≥ 360 minutes of wear time, excluding time for measurements classified as errors in Stage 1. Valid files were defined as accelerometer files with at least 2 valid days (Rich et al, 2012a).

Variables defined: dummy variable to identify valid days (codes: 0 – false, 1 – true); dummy variable to identify valid files (codes: 0 – false, 1 – true);.

b) Classification of valid data

Accelerometer count thresholds as detailed in Pulsford et al (2011) were applied to define sedentary behaviour (< 100/4 counts/15 seconds) and light (≤ 2240/4), moderate (≤ 3840/4) (PA type 2), and vigorous activity (> 3840/4). The threshold for extremely high counts was set to ≥ 11,715/4 counts per 15 seconds (Rich et al, 2012b).

c) Collapsing data

Data were collapsed into smaller and ready-to-use datasets to be used in statistical analyses. Data were collapsed (i.e., aggregated) by identifier, date and hour. The following logical expressions were used: exclusion of errors, exclusion of extreme values, exclusion of non-wear time, exclusion of truncated days (i.e., before start date and after end date). Files with non-valid data were reported with NAs.

Variables defined: file identifier; MCS identifier and child number; dummy variable to identify files to be excluded (see note Table 1); dummy variable to identify empty files (codes: 0 – false, 1 – true); time stamp; time zone, day of the week; weekday; dummy variable to identify weekend days (codes: 0 – false, 1 – true); month, calendar season based on actual equinoxes and solstices; season as defined in Mattocks et al, 2007; survey sweep;

total wear time (minutes) per day for counts classified at different PA intensities; total wear time (minutes) per day including extreme counts; dummy variable to identify valid days (codes: 0 – false, 1 – true); time (minutes) spent in sedentary behaviour; time (minutes) spent in light activity; time (minutes) spent in moderate activity; time (minutes) spent in vigorous activity; time (minutes) spent at PA intensity classified as extremely high; non-wear time (minutes); total counts per day; total steps per day; total counts per day divided by number of days; total counts per day divided by total wear time per day; total steps per day divided by total wear time per day; multiple birth status; MCS survey weights adjusted for missing data – country-specific; MCS survey weights adjusted for missing data – all of UK.

Further details on MCS sampling weights and adjustment for non-response/ non-compliance are given in Plewis (2007) and Geraci et al (2012).

d) Storage

Stage 3 output files were stored as ".Rdata" files (R format).

Data and software

The present document describes the data processing procedures used in the MCS physical activity study and is integrating part of the technical documentation. The raw and processed data, together with additional accompanying documentation, will be available in the UK Data Archive starting late 2012. The processing software is publicly available (Geraci, 2012).

HOW TO ACKNOWLEDGE THIS DOCUMENT

Geraci M, Rich C, Sera F, Cortina-Borja M, Griffiths LJ, and Dezateux C (2012). Technical report on accelerometry data processing in the Millennium Cohort Study. London: University College London. Available at http://discovery.ucl.ac.uk/1361699.

APPENDIX A – ACTIGRAPH GT1M TECHNICAL CHARACTERISTICS

Name: Actigraph GT1M (Actigraph, Pensacola, Florida)

Axes: bi-axial

Dimensions: L 3.8 X W 3.7 X H 1.8 cm

Weight: 24 g

Solid state sensor: micro-electro-mechanical systems (MEMS)

Dynamic range: 0.05-2.5 g

Frequency range: 0.25-2.5 Hz

Current firmware: v7.5.0

MEMS undergoes a precise batch manufacturing process to ensure high repeatability. A filter is implemented within the software GT1M to remove unit-to-unit variability. Accelerometer vendor manufactures the device to ensure the initial tolerance specification on sensitivity varies by ± 10% (Silva et al, 2010).

APPENDIX B – GLOSSARY

Count (acceleration count). "Count" is an enduring term. Prior to solid-state analog-to-digital converter commercial availability, activity monitors utilized either the 'threshold crossing' or 'cycle count' activity measurement. The threshold crossing technique involved incrementing a "count" each time the magnitude of acceleration (activity) exceeded a given threshold. The cycle count approach produced a "count" when enough force was applied to move a mechanical lever through a full cycle (up and down). The latter of these two approaches is very similar in nature to the modern day pedometer measurement technique.

ActiGraph's original activity monitor, the 7164 model, utilized a mechanical lever capable of measuring the change in acceleration with respect to time (g/sec, where g is gravity or 9.806

m/s2). To suppress unwanted motion and enhance human activity, the acceleration signal was passed through an analog band-pass filter, the output of which yields a dynamic range of 4.26g/sec (+/-2.13g/sec) at 0.75Hz (center frequency of the filter). Using a sample rate of 10 samples-per-second, this filtered signal was then digitized into 256 distinct levels by an 8-bit solid-state analog-to-digital converter, producing 4.26g/sec per 256 levels or 0.01664 g/sec/count (each level is considered 1 count). When each filtered sample is multiplied by the sample window of 0.1sec, a resolution of 0.001664g/count is achieved (Source: http://support.theactigraph.com/faq/counts).

Epoch. On ActiGraph devices, "epochs" are whole-second numbers (e.g., 1s, 5s, 60s, 120s, etc.). For each epoch interval, data samples taken from the accelerometer inside the device at a rate of 30Hz are first filtered then accumulated before being stored in memory. For example, a device set to collect 60 second epoch data will store 30 × 60 = 1800 accumulated samples for each enabled axis on the device every 60 seconds (Source: http://support.theactigraph.com/faq/data-stored-actigraph-device).

Mode. Devices can be set to store data in raw mode by selecting a sample frequency (12Hz or 30Hz) rather than a whole-second number in the "Epoch" dropdown during the initialization process. When raw mode is set, devices store raw (non-filtered/accumulated) accelerometry data (in units of gravity – G's) directly into memory at the selected rate (12Hz or 30Hz). When raw mode is enabled, non-accelerometer channels (steps, inclinometer, and lux) are disabled (Source: http://support.theactigraph.com/faq/data-stored-actigraph-device).

REFERENCES

Esliger DW, Copeland JL, Barnes JD, and Tremblay MS (2005). Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. Journal of Physical Activity and Health 2(3): 366-383.

Everitt BS and Skrondal A (2010). *The Cambridge Dictionary of Statistics*. Cambridge: University Press Cambridge.

Geraci, M. (2012), Pawacc: Physical Activity with Accelerometers. Available at http://r-forge.r-project.org/projects/pawacc.

Geraci M, Sera F, Griffiths LJ, Cortina-Borja M, Rich C, Pouliou T, and Dezateux C (2012). Missing data in accelerometry-based surveys on physical activity: Adjusting for non-response and non-compliance in the Millennium Cohort Study (in preparation).

Mattocks C, Leary S, Ness A, Deere K, Saunders J, Kirkby J, et al (2007). Intraindividual variation of objectively measured physical activity in children. Medicine and science in sports and exercise 39(4): 622-9.

Plewis I (2007). Non-response in a birth cohort study: the case of the Millennium Cohort Study. International Journal of Social Research Methodology 10(5): 325-334.

Pulsford RM, Cortina-Borja M, Rich C, Kinnafick F-E, Dezateux C, and Griffiths LJ (2011). Actigraph accelerometer-defined boundaries for sedentary behaviour and physical activity intensities in 7 year old children. PLoS ONE 6(8): e21822. doi:10.1371/ journal.pone.0021822

Rich C, Geraci M, Griffiths LJ, Sera F, Dezateux C, and Cortina-Borja M (2012a). Quality control methods in accelerometer data processing: Defining minimum wear time (under review).

Rich C, Geraci M, Griffiths LJ, Sera F, Dezateux C, and Cortina-Borja M (2012b). Quality control methods in accelerometer data processing: identifying extreme counts (under review).

Sherar L, Griew P, Esliger DW, Cooper A, Ekelund U, Judge K, and Riddoch C (2011). International children's accelerometry database (ICAD): Design and methods. BMC Public Health 11(1): 485-497

Silva P, Esliger D, and Welk G (2010). Technical reliability assessment of the Actigraph GT1M Accelerometer. Measurement in Physical Education and Exercise Science 14: 79-91.