Component of Statistics Canada Catalogue no. 82-003-X
Health Reports

Article

Quality control and data reduction procedures for accelerometry-derived measures of physical activity

by Rachel Colley, Sarah Connor Gorber
and Mark S. Tremblay

January, 2010

Statistics Canada Statistique Canada

Canada

Quality control and data reduction procedures for accelerometry-derived measures of physical activity

by Rachel Colley, Sarah Connor Gorber and Mark S. Tremblay

Abstract

Background
This article describes four key quality control and data reduction issues that researchers should consider when using accelerometry to measure physical activity: monitor reliability, spurious data, monitor wear time, and number of valid days required for analysis.

Data source and methods
Exploratory analyses were conducted on an unweighted subsample (n=987) of the accelerometry data from the Canadian Health Measures Survey. Participants were asked to wear an accelerometer for 7 consecutive days. Calibration, reliability, biological plausibility and compliance issues were explored using descriptive statistics.

Results
Ongoing calibration is an effective method for identifying malfunctioning accelerometers. The percentage of files deemed viable for analysis depends on participant compliance, the allowable interruption period chosen and the minimum wear-time-per-day criterion. A 60-minute allowable interruption period and 10-hours-per-day wear time criteria resulted in 95% of the subsample having at least 1 valid day, and 84% having at least 4 valid days.

Interpretation
Before the derivation of physical activity outcomes, accelerometry data should undergo standardized quality control and data reduction procedures to prevent mis-representation of the results. Incomplete accelerometry data should be handled carefully, and strategies to improve compliance in the field are warranted.

Keywords
ambulation, data quality, error, health measurement, quality control

Authors
Rachel Colley (1-613-737-7600x4118; Rachel.Colley@statcan.gc.ca) and Mark Tremblay are with the Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario; Rachel Colley and Sarah Connor Gorber are with the Health Analysis Division at Statistics Canada, Ottawa, Ontario.

Considerable evidence indicates that sedentary behaviour is a major risk factor for obesity and several other chronic conditions. Population-level surveillance of physical activity has historically relied on questionnaires,[1] a method of assessing lifestyle behaviours that can be affected by measurement bias.[2-4] Objective measurement devices, notably accelerometers, have the potential to overcome many problems associated with self-reports, and they provide robust and detailed information about physical activity.[5,6] However, because small inconsistencies can have a substantial impact on outcome variables,[6-8] stringent quality control and data reduction procedures are necessary.

This article considers four critical quality control and data reduction procedures that should be addressed before physical activity information is derived from accelerometry data. The results are based on early experience gained in cycle 1 of the Canadian Health Measures Survey, a comprehensive direct health measures survey of a national sample of Canadians, which used accelerometry to collect data on physical activity.

Intra- and inter-monitor variability[9,10] and the possibility that accelerometers will malfunction warrant the development and maintenance of rigorous calibration procedures. In 2006, Esliger and Tremblay[11] reported the results of a technical reliability assessment of three accelerometer models: Actical (Mini Mitter - Respironics, Oregon, USA), Actigraph model 7164 (Actigraph, Fort Walton Beach, FL), and RT3 (Stayhealthy, Inc., Monrovia, CA). The Actical had the best intra- and inter-monitor reliability, but discrepancies were observed, which confirms the need for ongoing calibration checks.

Accelerometers are designed to record counts within a defined range of movement that is plausible for humans.

Even so, monitor malfunctions and random spurious data points do occur during field measurements and must be managed. The threshold for defining spurious data must be low enough to exclude incorrect high values (for example, monitor aberrations), but high enough to include legitimate values that reflect vigorous activity. The maximum saturation value for the Actigraph is 32,767 counts per minute (cpm), and the recommended threshold for biological plausibility is less than 15,000 cpm.[6] According to the manufacturer, the maximum saturation value for the Actical is 28,404 cpm, and to the authors' knowledge, a threshold for biological plausibility has yet to be suggested.

Minimum daily wear time is another critical data reduction issue, because it affects the proportion of files that can be included in analyses. The minimum must be high enough to eliminate days when the monitor was clearly not worn long enough to accurately depict physical activity, but low enough to prevent too many days from being eliminated, which would bias the sample and reduce sample size and statistical power. Several studies have used 10 hours per day as a minimum requirement,[7] an approach that appears to be common in the broader research community.[6,8,12-14]

Calculation of wear time is complicated by the fact that inactivity is part of normal behaviour. Instead of simply deleting zero count values from the dataset, it is necessary to apply a decision rule that allows for a certain number and pattern of consecutive zeros throughout the day in order to capture and assess true inactivity. This is referred to as the "allowable interruption period" and ranges from 10 to 60 minutes.[7]

Population-level surveillance studies typically ask participants to wear an accelerometer for 7 full days, but because of non-compliance, the number of valid days varies among participants. To achieve some consistency, researchers have used various minimums for the number of valid days recommended for inclusion in analyses, ranging from fewer than 3[15-17] up to 7 full days.[6,18] No consensus has been reached on the minimum number of days required to gain an accurate picture of an individual's physical activity.

The information in this article will contribute to the development of quality control and data reduction procedures for researchers measuring physical activity with accelerometry, particularly those wishing to work with or compare their results to Canadian Health Measures Survey data. The hypothesis is that variations in quality control and data reduction procedures can have a substantial impact on which files are deemed acceptable for inclusion in analyses. This, in turn, affects physical activity outcomes.

Methods

Data source

From March 2007 through February 2009, cycle 1 of the Canadian Health Measures Survey collected data from a representative sample of Canadians aged 6 to 79 years. The survey was conducted by Statistics Canada in partnership with Health Canada and the Public Agency of Canada. Details on the background and rationale are available elsewhere.[19]

The survey involved an interview in the participant's household and a visit to a mobile examination centre where he or she underwent a series of physical measurements. Afterwards, ambulatory participants aged 6 years or older were asked to wear an Actical accelerometer over their right hip on an elasticized belt for 7 days. They were to wear the device during their waking hours and take it off only for bathing and swimming. The monitors were returned to Statistics Canada, where the data were downloaded and the monitor was checked to determine if it was still within the manufacturer's calibration specifications. An Actical calibrator was custom built for the Canadian Health Measures Survey.

The estimated sample for cycle 1 was about 5,000 individuals. An unweighted subsample (1,033) from the first four sites was used in the current study of accelerometer performance. As a result of monitor malfunctions, 46 files were unsuitable for analysis, reducing the sample to 987.

The measurement device

The Actical (dimensions: 2.8 x 2.7 x 1.0 centimetres; weight: 17 grams) is designed to measure and record acceleration in all directions, providing an indication of the intensity of physical activity. The digitized values are summed over a user-specified interval of 1 minute, resulting in a count value per minute (cpm). Accelerometer signals were also converted into steps accumulated per minute.[20]

Intra- and inter-monitor reliability

The Canadian Health Measures Survey calibration system simulates a dynamic motion that elicits a known value of 1,950 cpm from the Actical. Between each field measurement, the monitors were re-tested to ensure they recorded values within ±10% of 1,950 cpm. Those that failed the calibration procedures three times were returned to the manufacturer for repair or replacement.

Although 16 monitors can be placed on the calibration wheel at once, study flow meant that the calibration sometimes had to be performed with fewer than 16. Four different conditions were tested to compare the results of conducting the calibration with 16, 12, 8 and 4 monitors. The monitors were evenly distributed for each measurement, and each monitor underwent three testing cycles for each condition.

Identifying and managing spurious data

To determine a spurious data threshold, additional analyses were completed on previously published data that described the accuracy of the step-count function in the Actical.[20] The data were from a convenience sample of 38 participants aged 9 to 59 years with body mass indices from 19.9 to 36.6 kg/m^2. They were asked to walk or jog at three different speeds on a treadmill. The highest treadmill speed used in this study was 8 km/h, a speed representative of slow jogging and certainly not the fastest that could be obtained within a population sample. To account for this, the present study used linear regression to extrapolate the accelerometer counts to a high-level running speed (approximately 14 to 15 km/h) to determine a plausible upper threshold for identifying spurious data.

Defining wear time

Previously reported data reduction procedures were used in the present study. Troiano et al.[14] defined a valid day as 10 or more hours of monitor wear. Wear time was defined by subtracting nonwear time from 24 hours. Nonwear was defined as a period of a least 60 consecutive minutes of zero counts, with allowance for 1 to 2 minutes of counts between 0 and 100. A range of interruption periods (10, 20, 30, 60 minutes) was compared to demonstrate the effect of altering this value on the wear time obtained and the percentage of files deemed viable for analysis. A range of wear time criteria (6, 8, 10, 12, 14 hours per day) was also compared to demonstrate the effect that altering this value would have on the percentage of files that would be accepted for analysis.

Defining minimum number of days required for analysis

The data reduction procedures used by Troiano et al.[14] were applied to the Canadian Health Measures Survey subsample to determine what proportion of participants would have data available for further physical activity analyses, specifically, the percentages of participants with 1 valid day and 4 or more valid days, as this is consistent with how data have been published from the 2003-2004 National Health and Nutritional Examination Survey.[14]

Results

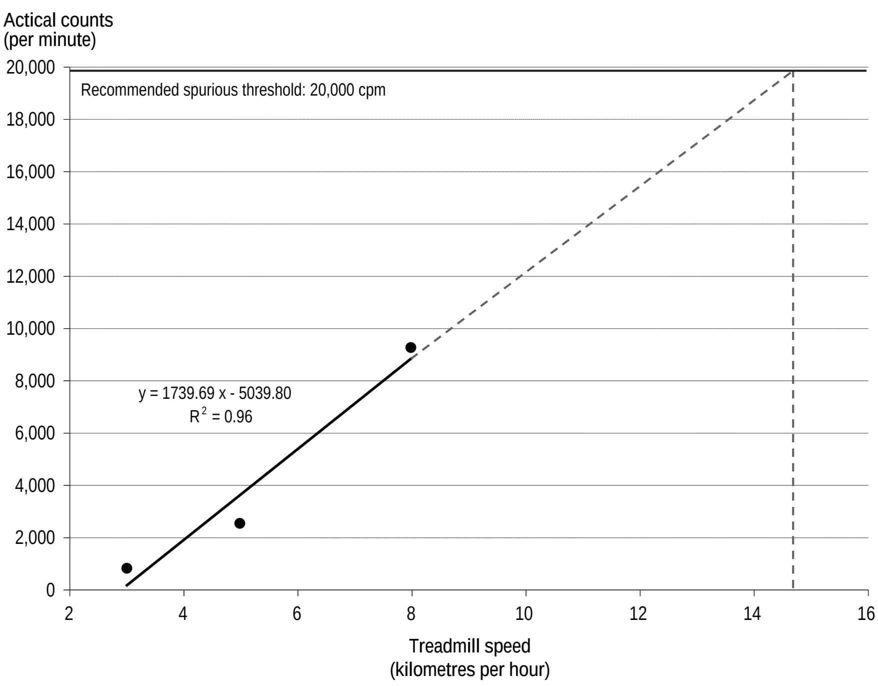
Intra- and inter-monitor reliability

As long as the monitors were evenly distributed, no differences in calibration results emerged between using 16, 12, 8 or 4 monitors on the calibration wheel. Cycle 1 of the calibration study found that 6 of the 16 monitors were outside the acceptable limits of ±10%. Two more cycles were completed (maximum number of cycles allowed was 3), which resulted in 4 of the failed monitors passing and 2 being identified as malfunctioning and requiring repair.

Spurious data threshold

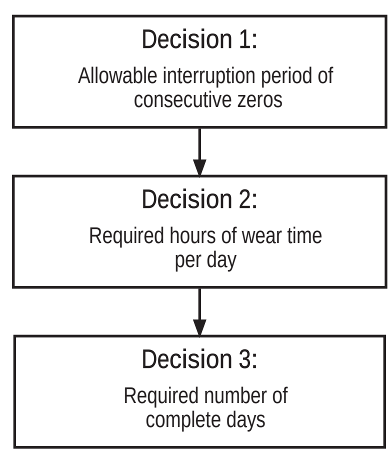
A review of the literature on the Actical revealed that no spurious data threshold has been suggested. Initial quality control procedures used 15,000 cpm, but this proved to be too low, given that it can be obtained by some individuals jogging at a moderate pace. According to the extrapolation procedure, a count value of 15,000 cpm corresponded to a running speed of approximately 12 km/h. A more appropriate spurious data threshold that would capture high-speed running, yet still exclude biologically implausible

Figure 1
Extrapolation of a linear regression relationship created between treadmill speed and Actical accelerometer counts



Source: 2007 to 2009 Canadian Health Measures Survey, first four sites.

Figure 2
Sequence of decisions for determining viability of accelerometry data for further analysis



movement, was needed. The count value attained when a linear regression line was extrapolated to a high-level running speed (14 to 16 km/h) was approximately 20,000 cpm (Figure 1).

Wear time

Defining acceptable wear time involves a series of decisions about allowable interruption periods and the number of hours needed for a valid day (Figure 2). The goal is to preserve both true activity and inactivity.

Lengthening the allowable interruption period increases average wear time (Table 1). Similarly, shortening the allowable interruption period decreases the number of files attaining the 10-hours-per-day wear time criterion. For example, if 4 out of 7 days are required for analysis, the difference between setting the allowable interruption period at 10 minutes (38% of individuals have at least 4 viable days) rather than 60 minutes (84% of individuals have at least 4 viable days) is substantial.

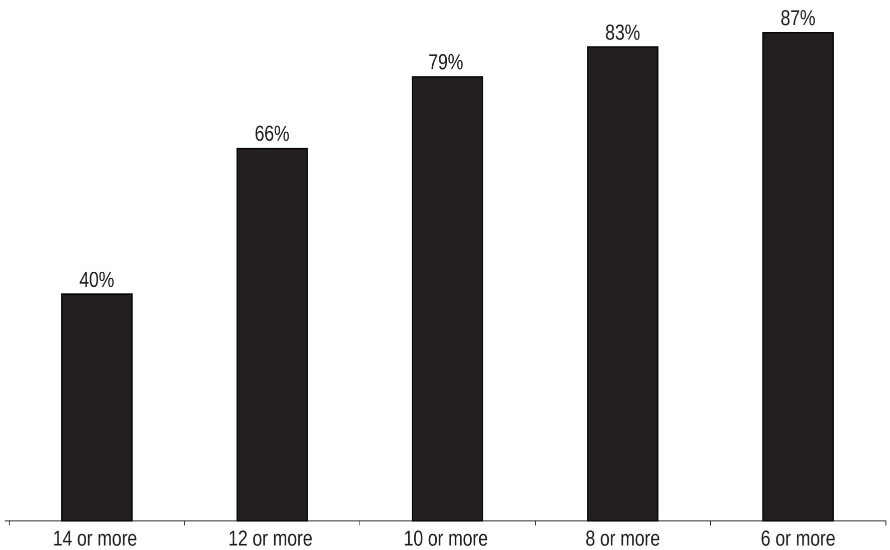
Figure 3 shows the percentages of individuals meeting the wear-time criteria when the minimum number of hours for a valid day is altered (with an allowable interruption period of 60 minutes). The gain in acceptable files between 14 and 10 hours of wear time per day is larger than between 10 and 6 hours per day. In other words, lowering the minimum number of hours needed for a valid day from 14 to 10 hours resulted in a marked increase of acceptable files; lowering the minimum from 10 to 6 hours yielded a smaller difference.

Table 1
Mean calculated accelerometer wear time, by allowable interruption period

Allowable interruption period (minutes)	Mean wear time (hours per day)
10	8.5 ± 4.4
20	9.7 ± 4.7
30	10.6 ± 4.9
60	12.0 ± 5.1

Source: 2007 to 2009 Canadian Health Measures Survey, first four sites.

Figure 3
Percentage of individuals with acceptable files when minimum criterion (hours) for "valid day" is altered



Note: Allowable interruption period is 60 minutes.
Source: 2007 to 2009 Canadian Health Measures Survey, first four sites.

Valid days suitable for analysis

Overall, 95% of the participants had at least 1 valid day, and 84% had 4 or more (Table 2). Thus, if researchers are satisfied that 1 valid day is sufficient to answer their question, they would only have to exclude approximately 5% of the sample.

Table 2
Percentage distribution of participants, by number of valid days of accelerometer wear, group and sex

Age group and sex	Valid days of accelerometer wear 0	1	2	3	4	5	6	7	1 or more	4 or more
6 to 11										
Male	3	3	5	3	11	13	27	35	97	86
Female	2	6	3	1	10	15	21	42	98	88
12 to 19										
Male	10	4	1	8	12	12	26	27	90	77
Female	11	1	6	1	11	13	23	34	89	81
20 to 39										
Male	7	6	0	8	4	17	11	47	93	79
Female	6	6	6	4	10	9	17	42	94	78
40 to 59										
Male	3	2	3	4	2	15	22	49	97	88
Female	4	4	2	5	4	9	19	53	96	85
60 to 79										
Male	0	2	3	4	9	19	19	44	100	91
Female	7	4	2	3	4	5	17	58	93	84
Overall group means	...	...	...	...	...	...	...	...	95	84

... not applicable
Note: A valid day is defined as 10 or more hours of accelerometer wear time with allowed interruption period of 60 minutes
Source: 2007 to 2009 Canadian Health Measures Survey, first four sites.

The percentage of participants achieving 7 valid days varied by age group from 27% to 58%. Complete nonwear (0 valid days) was most common among 12- to 19-year-olds; the highest compliance was among participants aged 40 to 79 years. No consistent sex difference was evident in any age group.

Discussion

The quality of accelerometry-derived data can be maximized through systematic, consistent quality control and data reduction procedures. This article presents four issues that researchers should address when establishing physical activity measurement and analysis protocols: intra- and inter-monitor reliability, spurious data thresholds, derivation of wear time, and number of valid days required for analysis.

Few studies have examined the technical reliability of accelerometers, largely because of the lack of commercially available calibration units. Unless researchers have the resources to design and build custom calibration equipment, they must trust that their accelerometers provide stable within- and between-monitor measurements of physical activity. However, apparent reliability issues suggest that calibration procedures are warranted. For example, in the present study, 2 out of the 16 monitors measured outside an acceptable reliability range. Without a way to identify these out-of-calibration monitors, they would have been sent back into the field and potentially contaminated physical activity outcomes. The 2006 study by Esliger and Tremblay[11] outlined calibration recommendations based on their work using a hydraulic shaker table to test the reliability of the Actical, Actigraph and RT3 accelerometers. They found that all three were susceptible to reliability problems, and of relevance to the present study, 7 out of 39 Acticals were too variable for use in the field.[11] The calibration protocol in place for the Canadian Health Measures Survey will decrease the likelihood that contaminated data are included in any analyses, thereby strengthening the quality of the physical activity measure.

The present study found that a spurious data threshold of 15,000 cpm (recommended for the Actigraph) is too low for the Actical and that 20,000 cpm is more appropriate. For instance, high-performance runners could legitimately accumulate accelerometer counts close to 20,000 cpm. Unless such data are captured, the level of activity of these participants would be underestimated.

A quality control step that detects spurious data can identify and manage both malfunctioning monitors and biologically implausible data. While most Canadian Health Measures Survey files contained no spurious data (more than 20,000 cpm), an occasional file had a small number (fewer than 5) of spurious observations. In rare circumstances, excessive spurious data appeared in a file, and upon further investigation, these files were found to be completely unusable because of a monitor malfunction. Thus, a simple quality control step that sums spurious observations by participant is a useful way of identifying monitors that may require technical attention. When occasional spurious observations occur in appropriately functioning monitors, they can be managed by techniques such as replacing the elevated observations with the mean of the two closest non-spurious data points on either side.[14]

Observations equal to zero pose a special problem. Excluding them would discard important information about inactivity. However, maintaining all these observations may lead to the inclusion of non-wear time in the final analysis and dilute physical activity outcomes. The assumption is that periods of consecutive zeros lasting longer than the allowable interruption period represent times when the accelerometer has been removed. Intervals of continuous zero counts that are shorter than the allowable interruption period are preserved as wear time, and are believed to indicate sedentary behaviour.

This study compared the impact of various allowable interruption periods on the likelihood that a file would

What is already known on this subject?

- Accelerometry-derived measures of physical activity continue to be published in the research literature. However, the implementation and reporting of data reduction and analytical methods is inconsistent.
- Given the potential impact that data reduction procedures can have on physical activity outcomes, consensus is needed among researchers using these devices.
- Publication of recommendations about processing accelerometry data from the National Health and Nutritional Examination Survey facilitated the establishment of consistent procedures for the Canadian Health Measures Survey.

What does this study add?

- One of the primary challenges in using accelerometers to derive information about physical activity is low compliance with wearing the devices. The resultant incomplete data create interpretation issues and require consistent quality control and data reduction procedures.
- Four important quality control and data reduction steps are presented that help address incomplete accelerometry data and should be considered before deriving physical activity information: intra- and inter-monitor reliability, spurious data thresholds, derivation of wear time, and number of valid days required for analysis.
- The information is particularly relevant for researchers who work with or compare their results to Canadian Health Measures Survey accelerometry data.

achieve the daily wear time criterion. As expected, longer interruption periods resulted in higher mean wear time values and a higher percentage of files being deemed acceptable for analysis. Earlier studies have used interruption periods ranging from 10[21,22] to 60[7,14] minutes. Masse et al.[7] found that a less restrictive interruption period (60 minutes) results in more minutes of inactivity being reported than a shorter interruption period (20 minutes). Changing the interruption period may affect power and sample size because this affects the number of days that reach the wear time criterion. However, consensus on a single allowable interruption period is unlikely. Rather, it appears that this decision should be made *a priori*, based on the research design (population surveillance versus intervention monitoring) and the age of the population being studied. Researchers measuring physical activity in children have used shorter interruption periods, possibly reflecting an assumption that children are more active than adults. For example, the European Youth Heart Study uses a 10-minute interruption period when reducing accelerometry data collected from children and youth.[13,22,23]

The finding that at least one valid day was available for 95% of the sample is encouraging and suggests that the data reduction procedures tested are not so stringent as to dramatically reduce sample size. Like the National Health and Nutritional Examination Survey 2003-2004 accelerometry analysis, the present study found compliance in wearing a monitor was highest among older adults, and lowest among adolescents. This demonstrates the need to encourage compliance among adolescents when accelerometers are used to measure physical activity.

This article presents practical issues for researchers to consider when using accelerometers. It also provides methodological information on the survey accelerometer procedures, which will be useful to researchers working with or comparing their own data with Canadian Health Measures Survey results. A limitation is the small sample size on which recommendations about an appropriate spurious data threshold are based. This value should be tested on a larger sample to confirm that it is appropriate for all ages and abilities. Future investigations might explore whether lower thresholds should be used for children, the elderly or people with known physical limitations. Conversely, higher thresholds might be appropriate for individuals who self-identify as high-performance athletes.

Conclusions

A number of issues must be considered to ensure that valid data are reported when using accelerometers to measure physical activity. A limitation of many studies that report accelerometry-based estimates of physical activity is that they lack a description of the procedures used to calibrate the accelerometers and the data reduction procedures used before derivation of physical activity information. Monitor malfunctions can and do occur. Therefore, it is essential that researchers continually check the reliability of the measurement tool itself. A series of steps can be followed during data collection to prevent defective monitors from being sent into the field for data collection. Similarly, before analysis of derived variables, it is important to have procedures in place to confirm that the data are biologically plausible and that compliance with wearing the device is acceptable. ■

References

1. Katzmarzyk P, Tremblay M. Limitations of Canada's physical activity data: implications for monitoring trends. *Applied Physiology, Nutrition and Metabolism / Canadian Journal of Public Health* 2007; 32 / 98: S185-S194.
2. Adamo K, Prince S, Tricco A, et al. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: A systematic review. *International Journal of Pediatric Obesity* 2009; 4(1): 2-27.
3. Prince S, Adamo K, Hamel M, et al. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity* 2008; 5: 56.
4. Walsh M, Hunter GR, Sirikul B, Gower B. Comparison of self-reported with objectively assessed energy expenditure in black and white women before and after weight loss. *American Journal of Clinical Nutrition* 2004; 79: 1013-9.
5. Montoye HJ, Kemper HCG, Saris WHM, Washburn RA. *Measuring Physical Activity and Energy Expenditure.* Champaign, Illinois: Human Kinetics, 1996.
6. Esliger D, Copeland J, Barnes J, Tremblay M. Standardizing and optimizing the use of accelerometer data for free-living physical activity monitoring. *Journal of Physical Activity and Health* 2005; 3: 366-83.
7. Masse L, Fuemmeler B, Anderson C, et al. Accelerometer data reduction: A comparison of four reduction algorithms on select outcome variables. *Medicine and Science in Sports and Exercise* 2005; 37: S544-S554.
8. Esliger D, Tremblay M. Physical activity and inactivity profiling: the next generation. *Applied Physiology, Nutrition and Metabolism/Canadian Journal of Public Health* 2007; 32/98: S195-S207.
9. Powell SM, Jones DI, Rowlands AV. Technical variability of the RT3 accelerometer. *Medicine and Science in Sports and Exercise* 2003; 35: 1773-8.
10. Powell SM, Rowlands AV. Intermonitor variability of the RT3 accelerometer during typical physical activities. *Medicine and Science in Sports and Exercise* 2004; 36: 324-30.
11. Esliger D, Tremblay M. Technical reliability assessment of three accelerometer models in a mechanical setup. *Medicine and Science in Sports and Execrise* 2006; 38: 2173-81.
12. Mattocks C, Ness A, Leary S, et al. Use of accelerometers in a large field-based study of children: protocols, design issues, and effects on precision. *Journal of Physical Activity and Health* 2008; 5: S98-S111.
13. Moller N, Kristensen P, Wedderkopp N, et al. Objectively measured habitual physical activity in 1997/1998 vs 2003/2004 in Danish children: The European Youth Heart Study. *Scandinavian Journal of Medicine and Science in Sports* 2009; 19(1): 19-29.
14. Troiano R, Berrigan D, Dodd K, et al. Physical activity in the United States measured by accelerometer. *Medicine and Science in Sports and Exercise* 2008; 40: 181-8.
15. Ness A, Leary S, Mattocks C, et al. Objectively measured physical activity and fat mass in a large cohort of children. *PLoS Medicine* 2007; 4: e97.
16. Moore L, Gao D, Bradlee M, et al. Does early physical activity predict body fat change throughout childhood? *Preventive Medicine* 2003; 37: 10-7.
17. Stevens J, Suchindran C, Ring K, et al. Physical activity as a predictor of body composition in American Indian children. *Obesity Research* 2004; 12: 1974-80.
18. Trost S, Pate R, Freedson P, et al. Using objective physical activity measures with youth: how many days of monitoring are needed. *Medicine and Science in Sports and Exercise* 2000; 32: 426-31.
19. Tremblay M, Wolfson M, Connor Gorber S. Canadian Health Measures Survey: Background, rationale and overview. *Health Reports* (Statistics Canada, Catalogue 82-003) 2007;18 (Suppl.): 7-20.
20. Esliger D, Probert A, Connor Gorber S, et al. Validity of Actical accelerometer step count function. *Medicine and Science in Sports and Exercise* 2007; 39: 1200-4.
21. Ekelund U, Yngve A, Brage S, et al. Body movement and physical activity energy expenditure in children and adolescents: how to adjust for differences in body size and age. *American Journal of Clinical Nutrition* 2004; 79: 851-6.
22. Riddoch C, Andersen LB, Wedderkopp N, et al. Physical activity levels and patterns of 9- and 15-yr-old European children. *Medicine and Science in Sports and Exercise* 2004; 36: 86-92.
23. Nilsson A, Anderssen S, Andersen LB, et al. Between- and within-day variability in physical activity and inactivity in 9- and 15-year-old European children. *Scandinavian Journal of Medicine and Science in Sports* 2009; 19(1): 10-8.