Implementation of Raw Accelerometry in Physical Activity Epidemiology

Thesis · February 2013

DOI: 10.13140/RG.2.1.3314.7766

1 author:

Vincent T Van Hees

Netherlands eScience Center

Implementation of Raw Accelerometry in Physical Activity Epidemiology

Vincent T. van Hees

Wolfson College

This dissertation is submitted for the degree of Doctor of Philosophy

UNIVERSITY OF CAMBRIDGE

Department: MRC Epidemiology Unit

Original version: May 2012

This is a condensed version created in May 2016: All published chapters have been replaced by a link to the online article.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the word limit of 60,000 words as set by the degree committee.

Summary

The assessment of individual human daily physical activity in population research requires accurate, precise and feasible measurement methods. Accelerometers may fulfil these criteria. Traditionally used accelerometers express their output in brand specific values called “counts” for which the exact relation to SI units is unknown. Accelerometers that collect and store the raw acceleration signal in manufacturer independent units have only recently become feasible for use in epidemiological research. The objective of this thesis is to facilitate the implementation of raw accelerometry in physical activity epidemiology by developing signal processing techniques for deriving meaningful outcome measures.

First, I developed an approximation method for converting raw acceleration data into traditional accelerometer output in order to facilitate historical comparisons (chapter 2 & 3). Next, I developed methods to predict daily physical activity-related energy expenditure (PAEE) from a simple summary measure derived from wrist-worn raw accelerometer data, which explained 26% of the variance in PAEE (chapter 4). A raw acceleration signal contains three components; gravity, activity-related acceleration and noise. The separation of the gravitational component from the acceleration signal is difficult in the presence of rotation in a vertical plane. I therefore evaluated five methods for removing the gravitational component from an acceleration signal under the controlled conditions using an industrial robot arm (chapter 5). Results showed that estimation error can be strongly affected by method of choice. I also evaluated these methods for their ability to predict daily PAEE resulting in an improvement in the explained variance to 38% by the best method (chapter 5). I thereafter explored more sophisticated signal processing techniques. I performed a literature review on activity type classification from wearable sensors output and uncovered gaps in knowledge about the validity of experimental designs (chapter 6). I therefore conducted a study to assess the validity of commonly used experimental designs for developing and evaluating activity type classifiers by using real life simulations from empirical data collected in a laboratory. Results indicated that method performance as estimated by common laboratory experiments may cause considerable overestimations of classifier performance in real life scenarios (chapter 7). The final

chapter explores how knowledge about activity type can be used to improve the estimation of PAEE (chapter 8).

The work as presented in this thesis facilitates the implementation of raw accelerometry in epidemiological research by providing signal processing tools and by identifying potential sources of measurement error. An important challenge for the future is to preserve methodological consistency between studies based on raw accelerometry without prohibiting methodological innovation. Methodological consistency is challenged by a lack of consensus on which signal processing techniques to use, a lack of consensus on sensor location, a lack of clarity on how a classifier should be evaluated and the urgent need for signal processing tools for activity type classification and for profiling activity-related energy expenditure.

Outline

List of figures

List of tables

Abbreviations

PAEE	Physical activity-related energy expenditure
TEE	Total energy expenditure
REE	Resting energy expenditure
MJ	MegaJoule
EN	Euclidean norm = Vector magnitude
ENMO	Euclidean norm minus one
BFEN	Band-pass filtered signals followed by euclidean norm
HFEN	High-pass filtered signals followed by euclidean norm
MAEN	Moving average followed by euclidean norm
PA	Physical activity
Acc	acceleration
g	Unit of acceleration where 1 g = gravitational acceleration (on average 9.81 m $\cdot$ s^{-2})
mg	milli-g
metric	short signal processing script or algorithm intended to provide one specific parameter
QDA	Quadratic discriminant analysis
HMM	Hidden Markov Models

Contributions

Chapter specific contribution:

	2	3	4	5	6	7	8
Idea and experimental design	6	6	1	6	5	6	1
Application for ethical approval	1	3	1	n/a	n/a	5	1
Data collecting in humans	1	7	1	1	n/a	7	1
Data collection on robot	3	1	n/a	4	n/a	n/a	n/a
Data management and cleaning	7	7	6	7	n/a	7	6
Design of analysis	6	6	5	6	n/a	5	5
Performing analysis	7	7	7	7	n/a	7	7
Writing manuscript	4	6	4	6	6	6	2

[7 = work fully done by student without involvement of others;

6 = work fully done by student, but in correspondence with others;

5 = work lead by student with minor support from others;

4 = work lead by student with important advice or support from others;

3 = work done by others based on specific instructions from student;

2 = work done by others in correspondence with student;

1 = work fully done by others]

1 Introduction

1.1 Why try to measure physical activity?

Accurate and precise assessment of human daily physical activity is required to understand population levels of physical activity and to monitor temporal trends in physical activity [1-3]. Furthermore, accuracy and precision are important to determine the dose-response relationship between physical activity and metabolic function or disease and to examine which aspect of physical activity is most important for a given health outcome [1-3]. Accuracy and precision is also important to assess the effect of physical activity interventions, and to more efficiently conduct studies of putative associations between physical activity and health, including the interaction with genetic factors [1-4]. Finally, clinical studies of specific biological mechanisms in humans would benefit from better characterisation of most recent (previous days or weeks) physical activity to ensure or at least check that the distribution of this important exposure is not a source of bias in small randomised controlled trials or large non-randomised controlled trials. With the advent of modern genotyping technology, the discovery of novel physiological mechanisms may also stem from genome-wide scans in large population studies, includinsg the interaction between physical activity and genes on physiological traits. This thesis focuses on the assessment of physical activity in large-scale population research (epidemiology). Other than accurate and precise measurement, methods for population research are required to have a high level of method feasibility in order to minimize participant burden, selection bias, and study costs.

1.2 Definition of physical activity

The development of methods for the assessment of physical activity requires a measureable definition of physical activity. The most commonly used definition of physical activity was introduced by an epidemiologist, Dr. Caspersen, and reads “any bodily movement produced by skeletal muscles that results in energy expenditure” [5]. Physical activity is often described in the dimensions: intensity, type, frequency and duration. Intensity refers to the energy turnover rate required to perform an activity and frequency refers to the number of times an activity is done in a period of time, e.g. day or week.

The term ‘physical activity’ could literarily be interpreted as ‘that what an individual does’ or maybe even aspects of ‘that what happens to the individual’, which raises the question whether Caspersen’s definition is sufficiently broad. For example, gait characteristics like stride length or the fashion with which certain activities are performed like handedness may as well be considered characteristics of what a person physically does and relevant for certain fields of science. Furthermore, the restriction to body movement in Caspersen’s definition may be undesirable as characteristics of non-movement are in some way also descriptions of what a person does, e.g. body posture. In order to ease communication about the actual science, clear terminology is needed that is on the one hand sufficiently broad to capture the common interests of various scientific fields and on the other hand specific about the actual quantity of measurement.

The definition of physical activity as proposed by Caspersen could be referred to as a possible definition of physical activity-related energy expenditure rather than physical activity in general. Physical activity is a construct similar to constructs like ‘fitness’ and ’health’. They all refer to a group of characteristics related to a certain phenomenon and facilitate the communication about this phenomenon, but are by no means directly measurable quantities or constrained by specific borders. A specification of the actual quantity of measurement is then needed to facilitate investigation. In the remaining part of this thesis the term ‘physical activity’ will be accompanied by a specification of the measurement quantity and the method used to measure the quantity. If the term ‘physical activity’ is used on its own then it refers to a broad spectrum of what individuals do, including all potentially relevant scientific areas, e.g.: energy expenditure, occupational activity, sedentary behaviour, sleep, exercise, patterns in activity types, body movement, inactivity, posture, mobility, gait, and global positioning.

1.3 Commonly used measurement methods

A range of methods have been used to measure physical activity. This paragraph describes the most commonly used methods for the assessment of physical activity in the epidemiological field.

Self-report methods including questionnaires and diaries have been used for the assessment of physical activity-related energy expenditure (PAEE), time spent in activity types such as exercise and TV-viewing, and time spent in intensity levels [6, 7]. The subjective characteristic of self-report methods limits their accuracy at the individual level [6, 8], but various studies suggest that self-report methods do allow for ranking of individuals with respect to their daily PAEE [9, 10]. Questionnaires are cheap to apply, but potentially expensive to process depending on the length of the questionnaire, whether the questions are open or closed and the way the questionnaire is administered (on paper vs. electronically).

Pedometers are generally cheap in price and aim to record the total number of steps taken per measurement period or per day [11]. Pedometers have been suggested to be more accurate in ranking individuals on daily PAEE compared with questionnaires [12]. Some pedometers use entered body characteristics to improve the estimation of total energy expenditure which complicates interpretation [13]. The accuracy of pedometers to count the number of steps out of the laboratory under real life conditions has only been evaluated by concurrent validity assessment [14-16]. Concurrent validity assessment only assesses the degree to which two methods give the same answer but not if the answer is correct; since measurement error is likely to be correlated between pedometers, this can give the false impression of true validity, resulting in potential confusion with accuracy estimates. No gold standard exists for the detection of steps under real life conditions.

Accelerometers have received increasing popularity and are currently widely used. In this thesis a distinction is made between traditional accelerometers and raw accelerometers to be described in the next paragraph. Traditional accelerometers, e.g. Actigraph (The Actigraph, Inc., Pensacola, Florida, US), RT3 (Stayhealthy,Inc., Monrovia, California, US) and Actical (Philips Electronics N.V., Philips Healthcare, Andover, MA, US) express their output in brand-specific values called "counts" captured over a pre-specified period of time, called "epoch". An epoch is usually set at a few seconds, although older data are typically collected with a 60-sec epoch setting [17]. Traditional accelerometers have been used for the estimation of total daily PAEE and time spent in intensity levels [18-20]. Additionally, pilot studies have been done to explore the possibility of activity type detection from traditional accelerometer data [21, 22]. The calculation of a count is brand specific and its computation is usually kept confidential by the manufacturer of the accelerometer, which complicates the comparison and interpretation of data. Another limitation of some traditional accelerometers is that their dynamic range and frequency sensitivity are inappropriately set by their manufacturers. For example, the Actigraph has limited sensitivity to more vigorous movements [23, 24], while the RT3 has limited sensitivity to low frequency movements [25]. Various studies have been done to develop methods to predict PAEE from traditional accelerometer output and to classify time spent in intensity levels. Most of these studies suffer from potential selection bias, a lack of transparency on the contribution of individual method parameters and a lack of standardisation in the comparison between methods across studies. Nonetheless, there is a general consensus in the literature that accelerometers are useful for estimating PAEE and time spent in intensity levels [3, 13, 26]. However, the variance in measurement error between individuals is often more than 30% of the variance in reference PAEE between individuals [27, 28], which indicates that there is considerable space for improvement.

Heart rate monitoring on its own is nowadays less used for habitual PA assessment, but is being used in combination with accelerometers [29, 30]. A disadvantage of heart rate monitoring on its own when used for estimating PAEE is that it requires calibration to correct for between-individual differences in the relation between PAEE and heart rate. Another disadvantage of heart rate monitoring is that it is affected by emotions, in particular under low intensities, which limits the accuracy of heart rate monitoring for PAEE assessment. The combination of heart rate monitoring and acceleration sensing has been proposed to address the limited accuracy of heart rate monitoring in the lower intensity range and the limitations of some traditional accelerometers in the higher intensity range [31]. The additional value of heart rate monitoring on top of acceleration sensing for the estimation of PAEE, expressed in $kJ \cdot kg^{-1} \cdot min^{-1}$, has been shown in a 22hr room calorimeter protocol [32] and in walking and running experiments on a treadmill [33]. Assah et al. compared the accuracy and precision of the combined sensor method and an accelerometer-only method under free-living conditions using the doubly labelled water method as a reference [27]. In contrast to the previous studies this study showed that the accelerometer-only method provides a higher explained variance in PAEE ($kJ \cdot kg^{-1} \cdot min^{-1}$) compared with the combined sensor method, even though the latter had a lower mean bias [34]. The lower bias as reported for the combined sensor method may be explained by population specific calibration of the combined sensor method [34]. More research is needed to verify the source of the lower explained variance, whether it is population specific, and whether selecting different analytical techniques could address it. The required calibration procedures and the required continuous contact with the skin make the combined sensor method less feasible compared with pedometers and accelerometers.

1.4 Raw accelerometry

Raw accelerometry distinguishes itself from traditional accelerometry as described in the previous paragraph in that it measures body acceleration in non-brand specific "g-units" at a sample frequency at least twice above the maximum frequency of interest. Here, "g-units" refers to the sensor's calibration against gravitational acceleration. One "g" equals about 9.81 m $\cdot s^{-2}$ subject to global latitude and the accuracy of the calibration procedure. Raw accelerometry is usually based on seismic or inertial sensors, which means that it is sensitive to gravitational acceleration under both static and dynamic conditions [35]. Seismic accelerometers use a seismic or a proof mass suspended by a spring structure in a case. When the case is accelerated, the proof mass is also accelerated by the force transmitted through the spring structure. The displacement of the spring, the displacement of the mass within the case, or the force transmitted by the spring is transduced into an electrical signal proportional to the acceleration of the case. The orientation of the case also affects the position of the mass. If the case is held still and out of the horizontal plane, the mass moves downwards and stays downwards resulting in a constant offset in the signal, which reflects gravity. The magnitude of the offset and direction of the offset relate to the alignment of gravity with the axis/axes of the accelerometer [36]. Therefore, the measured signal reflects both sensor orientation and movement. Although, it should be noted that under dynamic conditions involving rotation in the vertical plane these components are hard to distinguish. A capacitor or a resistor can be used as transduction element [37].

Sensitivity to sensor orientation relative to gravity under static conditions is useful for detecting the orientation of body segments and subsequent posture detection [35]. Technological developments over the past two decades have allowed raw accelerometry to evolve from bulky expensive devices to low cost and small devices [37-40]. The ability to use raw accelerometry for continuous periods of seven days on one battery sealed within a small compact device has only been possible for a couple of years now and has made raw accelerometry applicable in the epidemiological field.

However, raw accelerometry is not a method on its own; it only becomes useful when combined with analytical tools that transform the raw data into meaningful outcome measures. The availability of raw data allows for the extraction of a wide range of signal features which can be used to train classifiers for automatic detection of events or patterns in the signals [41-45]. Various studies have been published on data collected with the older, more bulky generations of raw accelerometry and showed that raw accelerometry in combination with the right algorithms can be used to extract gait characteristics [46, 47], to classify activity types [45, 48] and to estimate energy expenditure [49].

Both traditional accelerometry and raw accelerometry use the same basic information; the main difference is that traditional accelerometry provides a processed form of this information derived with proprietary algorithms while raw accelerometry provides a minimally processed form of the data, e.g. just anti-aliasing filter and A/D conversion. Therefore, all studies that have shown the value of traditional accelerometry implicitly have proven the minimum potential value of raw accelerometry. Finally, raw accelerometry has the advantage that innovations in signal processing can be applied to raw data previously collected. Therefore, raw accelerometry has the potential of enhancing the assessment of physical activity in large-scale population studies, and by that make a continual contribution to the inference on the role of physical activity with respect to disease endpoints that typically take very long time to develop. For example, to study the role of physical activity on incident type 2 diabetes in the UK, one would need to recruit >1,000 adults who were free of diabetes, measure their activity, and then wait >5 years to accrue >300 incident cases [50]. A schematic model was created to show the relation between raw- and traditional accelerometry and their usage for the assessment of physical activity outcomes, see Figure 1-1.

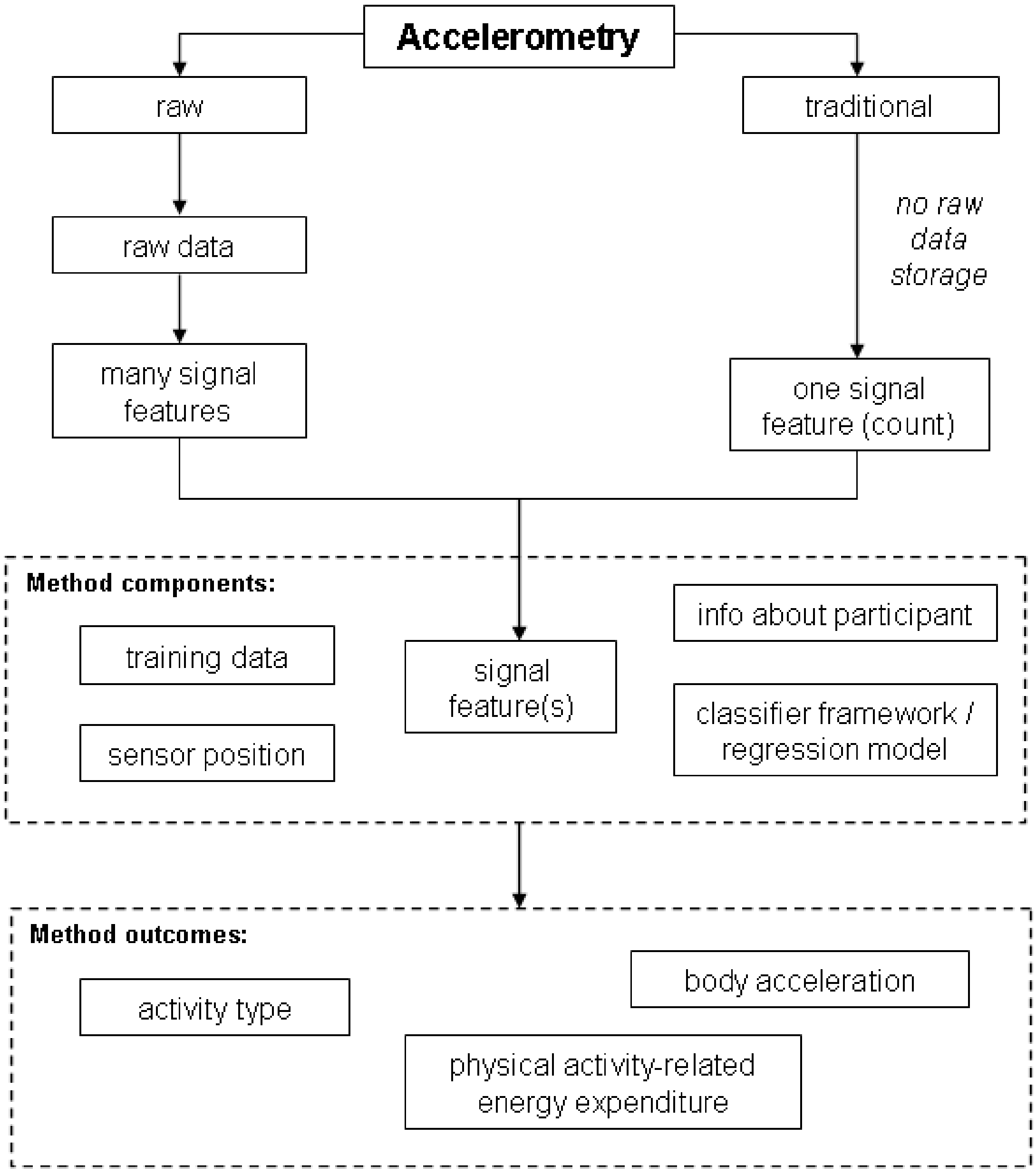

Figure 1-1: Diagram showing the relation between accerometry and their usage for the assessment of physical activity outcomes

1.5 Gaps in knowledge

1.5.1 Availability of algorithms

Most of the literature on the development of signal processing methods for raw accelerometry does not contain the actual algorithms and algorithm parameters, e.g. [44, 45, 48]. The lack of access to algorithms prohibits study replication, evaluation and continued development. Future research needs to develop and publish methods for the analysis of raw accelerometry data in particular for the classification of activity types and the estimation of PAEE.

1.5.2 Methodological consistency

The introduction of a new measurement technology complicates historical comparisons between newly collected data against previously published findings based on traditional technology. Algorithms that can emulate traditional accelerometry output from raw accelerometry output would ease the transition in measurement technology in the field, but are currently not available in the public domain.

1.5.3 Understanding the signal

The field of physical activity assessment has moved towards more complex signal processing techniques using machine learning and non-linear regression analysis [21, 51]. A complementary analytical approach could be to focus on the interpretation of acceleration signals and seek biomechanical understanding. The latter approach has been applied in gait analysis [47] and activity type classification [35]. Acceleration signals contain three main components: acceleration related to gravity, acceleration related to movement, and noise. A starting point for interpreting acceleration signals collected under real life situations would be to separate these three components [52, 53]. In recent years, two different methods have been used to remove or at least minimize the contribution of the gravitational component to the acceleration signals: the 'frequency filter'-method [37] and the 'minus one g'-method [54], for a more detailed description see chapter 5. Gaining insight in the difference between the two methods is important to understand potential differences in study findings.

1.5.4 Sensor location

In the past, accelerometers have predominantly been attached to the hip and to the lower back for the following reasons: (i) the notion that this position was closest to the theoretical centre of body mass, (ii) accelerometer dimensions were unfeasible for wrist or ankle attachment; (iii) the human trunk is a relatively safe location for non-waterproof devices, and; (iv) the success rate in deriving meaningful outcome measures from hip and lower back positioning. The smaller size of modern raw accelerometers and their increased protection against water allows for attachment to various other body locations of which the wrist and the ankle are potentially feasible. However, differences between sensor locations have hardly been investigated in the literature [48, 55]. The challenge when comparing method accuracy between sensor locations is that the optimal type of signal processing could be body location specific. More research is needed to systematically evaluate various analytical techniques for various sensor locations in order to facilitate informed decisions on sensor positioning.

1.5.5 Study design for evaluating an activity type classifier

All studies as published on the development of methods for the classification of activity types rely on experimental protocols under confined conditions [56-58]. It is unclear whether these study designs provide valid estimates of classifier performance under real life conditions. For example: the detection of the activity type 'standing' out of periods of walking and running is likely to be much easier than the detection of standing out of a dataset containing sitting on a chair, sitting in a bus, lying and sitting on a barstool. Various published articles have concluded that the development of classification algorithms should take place under conditions highly comparable to free-living conditions [21, 45, 48]. However, no ideal study design has so far been proposed.

1.5.6 Estimating physical activity-related energy expenditure

The estimation of PAEE based on raw accelerometry has so far only been investigated with a room calorimeter experiment [49]. Future studies need to investigate whether raw accelerometry allows for the estimation of PAEE under real life conditions. Further, information on the type of physical activity (e.g., sitting, standing, walking, driving) may improve the validity of physical activity energy expenditure estimates [49], in addition to providing an interesting new dimension to the objective characterisation of physical activity. However, research in this area has been limited to a few studies [49, 59]. Additional research is needed to explore the potential of using activity type classification for improving the estimation of energy expenditure [60].

1.6 Objectives

The general objective of this thesis is to contribute to the implementation of raw accelerometry in physical activity epidemiology by developing signal processing techniques for extracting meaningful outcome measures from the data. Specific objectives are listed below:

A. To investigate how raw accelerometry data can be converted to the output of the most commonly used traditional accelerometer, the Actigraph in order to facilitate historical data comparisons (chapter 2 and 3).

B. To investigate whether a simple summary measure derived from raw accelerometry attached on the wrist allows for accurate estimation of daily PAEE (chapter 4)

C. To investigate how the gravitational component can be removed from an acceleration signal (chapter 5)

D. To investigate whether the procedure used to remove the gravitational component from an acceleration signal affects the estimation of daily PAEE (chapter 5)

E. To review the literature on the development of activity type classifiers based on wearable sensors (chapter 6)

F. To develop an empirical dataset for systematic training and evaluation of signal processing techniques and sensor locations for the classification of activity types (chapter 7)

G. To develop a simple activity type classifier from the dataset as developed under “F” and evaluate how classifier accuracy varies between sensor positions on the body (chapter 7).

H. To evaluate the validity of experimental protocols under confined conditions for estimating activity type classifier performance in real life (chapter 7)

I. To evaluate the additional value of combined activity type and raw accelerometry attached to the lower back for the estimation of PAEE in cycling, stair walking and level ground walking (chapter 8)

2 A method to compare new and traditional accelerometry data in physical activity monitoring: Pilot study

This chapter was published in June 2010 and accessible here: https://www.researchgate.net/publication/221567366_A_method_to_compare_new_and_traditional_accelerometry_data_in_physical_activity_monitoring

3 A method to compare new and traditional accelerometer data in physical activity monitoring: Follow-up study

3.1 Abstract

The previous chapter presented an initial attempt to develop a conversion algorithm to facilitate the comparison between raw accelerometer data and Actigraph data, the most widely used device brand in epidemiological studies. In this chapter this development is continued by addressing the main limitations of the previous chapter. The algorithm and corresponding parameters were optimized under controlled conditions of a different mechanical shaker device, after which an evaluation was carried out through physical activity monitoring in a newly collected dataset in 18 free-living adult participants (age: 32 ± 10 yrs, BMI: 23.4 ± 2.6 kg·m-2) wearing both devices on each hip for two days but this time using triaxial Actigraphs. The resulting algorithm for estimating Actigraph counts explained on average 94.9% of the minute-by-minute variance in Actigraph counts and 96.2% of the variance in 2-day averages. Sensitivity to classify data in the commonly used count-based intensity levels was > 56.6%. In conclusion, perfect emulation of Actigraph counts has not been achieved, although insight has been gained in how a conversion algorithm could be developed. Future studies should take into account possible differences between Actigraph counts as collected under different recording modes and potential differences between Actigraph versions and batches.

3.2 Introduction

In Chapter 2 the first attempt was made to develop an algorithm for the conversion of raw accelerometer data to Actigraph counts. The approach was mainly explorative and used an easy accessible shaker device and a ready available free-living dataset to get a first impression. Reflection on this approach allows for the design of a more systematic follow-up study that aims to address the following limitations of the initial pilot study.

In the pilot study a mechanical shaker device with three radii settings was used to investigate the relationship between both accelerometer outputs. The use of three radii settings is insufficient for detecting non-linear relationships between the frequency of acceleration, the magnitude of acceleration and the resulting count. Replication of the experiments on a different mechanical shaker device that allows for more radii conditions is needed.

The pilot study relied on the filter cut-off frequencies as specified by the manufacturer (0.25 Hz and 2.5 Hz) and only used one type of frequency filter (Butterworth). Evaluation of wider range of plausible cut-off frequencies, filter orders and three types of digital frequency filters may improve algorithm accuracy.

The order of the components in the signal processing scheme was not evaluated in the pilot study but better agreement may be achieved if it starts with limiting the dynamic range instead of with signal filtering.

The attachment of the accelerometers to opposite hips as done in the pilot study introduces uncertainty about the origin of observed differences in measurement; attachment to the same hip is needed. Further, the accelerometers as used in the free-living experiment were uni-axial; replacement by tri-axial accelerometers would take away the risk of misalignment.

The aim of the study as described in this chapter is to continue the development of an algorithm for conversion of raw accelerometer data into Actigraph counts by taking the limitations of the previous chapter, as mentioned above, into account.

3.3 Methods

The study design was split up in an algorithm development section and a free-living experiment for algorithm evaluation. The experiments done for the algorithm development part were done by collaborators in Bethesda, US. The free-living experiments were done in Cambridge, organised by me.

3.3.1 Algorithm development

A mechanical shaker device as described by Rothney et al. [24] was used to empirically assess the relationship between the response of the Actigraph and raw accelerometer output. This process assisted in optimizing the parameters of the conversion algorithm. The mechanical shaker device applied a translational oscillation in a two-dimensional (horizontal) plane at the frequencies 20-250 RPM with steps of 10 RPM, respectively 0.33-4.17 Hz with steps of 0.17 Hz, and based on seven radii settings 0.50, 0.75, 1.00, 1.25, 1.50, 1.75 and 2.00 inch resulting in 7 x 24 = 168 experimental conditions. Each experimental condition lasted for five minutes.

Accelerometer

The Actigraph GT3X is a tri-axial accelerometer which can be set up to collect raw acceleration data (30 Hz) for a maximum duration of 25 hours. The raw data can then offline be converted to the proprietary Actigraph counts by using the manufacturer's software tool "ActiLife". It is not possible to collect both counts and raw data with one Actigraph device at the same time, but the software indirectly facilitates this by post-experiment data conversion; this was done in the current study. This route has the advantage that Actigraph counts and raw accelerometer data are both captured from the same sensing position in the same device. Actilife generated counts, counts generated onboard of the Actigraph GT3X and counts generated by the Actigraph GT1M are supposedly compatible with each other according to the manufacturer. Vanhelst recently showed that both devices are compatible in their estimation of time spent in intensity levels with an average difference of less than 0.56% in a study population of 25 adult participants [71]. No other evidence on the compatibility between the GT1M and the GT3X for the assessment of daily physical activity has been published in the scientific literature.

Algorithm design and optimization

The conversion algorithm was designed according to the processing blocks as outlined in Figure 3-1, which is a modified version of the scheme as shown in Chapter 2. First, the dynamic range of the raw acceleration data was computationally limited (truncated) to an upper limit (±2.5 g or ±5g) and a lower limit (0g or ±0.05g), after which a digital band-pass frequency filter was applied. Three types of filter were evaluated as available within R (http://cran.r-project.org/): Butterworth, Chebyshev Type 1, and Elliptical or Cauer. After the filtering stage, time series corresponding to the three accelerometer sensors were combined by taking the Euclidean norm (vector magnitude) for each time stamp. Both Actigraph counts and estimated Actigraph counts were averaged over the middle three minutes of each experimental condition. Finally, a regression model was derived to make the conversion to Actigraph counts.

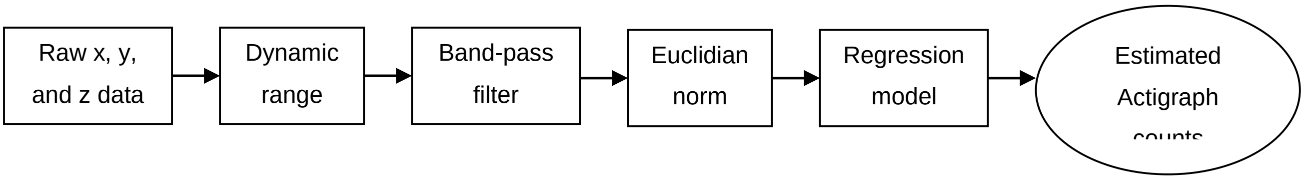

Figure 3-1: Signal processing scheme

Optimization of the algorithm parameters was organized in four stages: Stage 1 involved the evaluation of a wide range of plausible algorithm parameter values. The parameter combination that resulted in the highest correlation coefficient between Actigraph counts and estimated Actigraph counts was then used for further analysis. The parameters as evaluated are presented in Table 3-1. In stage 2, a more refined sub-range of parameter values close to the best performing parameters from stage 1 were explored. Stage 3 was identical to stage 1, but included a change to the signal processing scheme. Here, the dynamic range limitation is applied to both the individual signals and to the Euclidian norm of the three signals. If the best performing algorithm from stage 3 outperformed the best performing algorithm from stage 1 an additional stage, stage 4, was performed. Stage 4 would then include a refined exploration of algorithm parameters close to the optimal parameters from stage 3. Thus, the correlation coefficient was used as primary selection criteria. The standard error of the estimate was used as secondary criteria if correlation coefficient could not discriminate.

Table 3-1: Algorithm parameter values as evaluated for stage 1

Parameter	Values
Dynamic range (lower boundary)	0.00, 0.05
Dynamic range (upper boundary)	2.50, 5.00
Filter type	Butterworth*, Chebyshev Type 1*, Elliptical filter*
Filter Order	1, 2, 3, 4
ω_0 (lower boundary)	0.05 to 0.85 with steps of 0.10
ω_0 (upper boundary)	1.30 to 3.50 with steps of 0.20
pass band attenuation (dB)	3, 6
stop band attenuation (dB)	20, 40, 60, 80
Regression order	1, 2, 3, 4

[* as available within R (http://cran.r-project.org/)]

3.3.2 Free-living evaluation

Eighteen healthy adult volunteers took part in this study segment, characteristics of whom are described in

Table 3-2. Participants were informed about the objective and activity monitoring protocol to be used, in both written and oral forms. All participants gave written informed consent. The Research Ethics Committee of Cambridgeshire approved the study.

Table 3-2: Participant characteristics

	All (N = 18)	Males (n =9)	Females (N = 9)
Age	32 ± 10 (23 – 58)	31 ± 8 (24 – 49)	34 ± 11 (23 – 58)
Height	171 ± 11 (146 – 192)	178 ± 6 (171 – 192)	163 ± 8 (146 – 172)*
Weight	68.6 ± 12.1 (44.4 – 98.8)	75.5 ± 12.2 (59.2 – 98.8)	61.7 ± 7.5 (44.4 – 69.7)*
BMI	23.4 ± 2.6 (18.2 – 28.3)	23.7 ± 3.2 (18.2 – 28.3)	23.2 ± 1.9 (19.7 – 26.2)

[* significant difference between males and females, $p < 0.01$]

Experimental design

Participants were asked to wear a belt holding two packages of one Actigraph GT3X and one raw accelerometer taped together on their left and on the their right hip for all waking hours during a 48 hour period in their daily life. Experiments were only done on weekdays.

Accelerometers

The raw accelerometer device used (GENEA, Unilever Discover, UK) was the same as used in the pilot-study, see chapter 2. The Actigraph GT3X as described in paragraph 3.3 was used with firmware version 2.2.0 in recording mode 45 (no raw data), and an epoch length of 1 second. The raw recording mode of the Actigraph GT3X was not used for this study segment because raw data collection in the GT3X comes with maximum measurement duration of 25 hours (sample frequency: 30 Hz). Eight different raw accelerometer devices (GENEA) and six Actigraph GT3X devices were used randomly during the free-living experiments to increase the external validity of the results for each type of accelerometer. None of the accelerometers as used for the free-living experiment were used for algorithm development.

Data analyses

Raw accelerometer data was converted into Actigraph counts using the processing method presented in Figure 3-1 and optimized algorithm parameters from the shaker device experiment as described in Paragraph 3.3.1 on page 23. Only the first 47 hrs of the data were used due to premature ending of a few experiments. The mimicked Actigraph counts (ACmimicked) were calculated over 1 second non-overlapping windows (epoch) as for Actigraph data. Inspection of the data showed that timestamps were insufficient for synchronisation of time series. Time synchronisation was performed based on cross-correlation using the model [$t_{new} = c1 + t_{old} \cdot c2$], where coefficient c1 was varied over a ± 30 sec range with steps of 1 sec and coefficient c2 was varied from 0.998 to 1.002 with steps of 0.0005. For each combination of c1 and c2 the correlation coefficient was calculated between the two time-series and the values resulting in the highest correlation were used for further analysis. However, if the optimal correlation coefficient was lower than r = 0.80 the search range was expanded to ± 300 sec for coefficient c1. If this did not result in a correlation coefficient above 0.80 the range was increased to ± 2000 sec range. Finally, if the latter range did not result in a correlation higher than 0.80 the coefficients leading to the maximum correlation coefficients were chosen. After the synchronisation of time-series, data points were down sampled to 1 minute averages to further optimize overlap between time windows. Additionally, the average counts over the 47 hr period were calculated for each individual.

Explained variance in AC by ACmimicked was assessed through the calculation of the Pearson's correlation coefficient. To quantify the magnitude of disagreement, the root mean square of the error (RMSE) was computed. Bland & Altman plots were used to assess absolute agreement between the 47 hr average ACmimicked and AC. A paired sample t-test was applied to test for a significant difference between 47 hr averages. Linear regression analysis was then used to assess where the slope and y-intercept of the regression line through average individual ACmimicked and average individual AC in daily life differed significantly from respectively one and zero. Additionally, the algorithm's sensitivity and specificity to classify data in four count-based intensity intervals was assessed, defined as: sedentary (0 – 100 counts $\cdot$ min^{-1}), light (100 – 2000 counts $\cdot$ min^{-1}), moderate (2000 – 4000 counts $\cdot$ min^{-1}), and vigorous (> 4000 counts $\cdot$ min^{-1}) intensity. Finally, the errors in the estimation of time spent in count-based intensity levels were calculated and stratified to the three most frequently used Actigraph monitors in the free-living part of the current study. All analysis were performed in the open-source tool R using 'signal', 'Hmisc', and 'stat' packages[1]. P < .05 was denoted statistical significant.

[1] http://cran.r-project.org/

3.4 Results

3.4.1 Algorithm parameters

The best performing algorithm based on shaker data for stage 1 resulted in a squared correlation coefficient between AC and ACmimicked of $r^2 = 0.998$ ($p < 0.001$, SE = 3.39 counts · sec-1), see Table 3-3. The corresponding regression model is shown in Equation 3-1. Here, 'A' represents the output of the processing stage 'Euclidean norm'.

Equation 3-1

$$Counts \cdot \sec^{-1} = -2.21 + 76.49 \cdot A + 3556.22 \cdot A^2 - 10180.41 \cdot A^3 + 9981.65 \cdot A^4$$

Based on the optimal parameters from stage 1 the range of parameter values to be evaluated in stage 2 were chosen, see Table 3-4. The best performing parameter values for stage 2 had an r^2 of 0.998 ($p < 0.001$, SE = 3.20 counts · sec-1), see Table 3-3 and Figure 3-2. The regression model corresponding to the best algorithm parameters for stage 2 is shown in Equation 3-2.

Equation 3-2

$$Counts \cdot \sec^{-1} = -2.22 + 90.04 \cdot A + 3210.24 \cdot A^2 - 8929.35 \cdot A^3 + 8506.38 \cdot A^4$$

Algorithm performance for stage 3 was less compared with stage 1 and stage 2, see Table 3-3. Therefore, the optimal algorithm parameters resulting from stage 2 were chosen for evaluation in free-living participants, see Table 3-4.

Table 3-3: Best performing algorithm parameter values for each stage of the development

Parameter	Stage 1	Stage 2	Stage 3
Pass-band filter	Elliptic/Cauer	Elliptic/Cauer	Elliptic/Cauer
Dynamic range	0.05 – 5	0.05 – 5	0.05 – 5
Filter Order	1	1	1
ω_0 (Hz)	0.35 – 2.5	0.36 – 2.55	0.35 – 2.50
pass band attenuation (dB)	3	3	3
stop band attenuation (dB)	20	30	20
Regression order	4	4	4
Performance			
R^2 ($p < .001$ for all)	0.998	0.998	0.998
SE (counts · sec^{-1})	3.39	3.20	3.41

Table 3-4: Algorithm parameter values as evaluated for stage 2

Parameter	**Values**	**Filter**
Dynamic range (lower boundary)	0.00, 0.05	BF, CF, EF
Dynamic range (upper boundary)	2.50, 5.00	BF, CF, EF
Filter Order	1, 2, 3, 4	BF, CF, EF
ω_0 (lower boundary)	0.30 to 0.40 with steps of 0.02	BF & EF
	0.03 to 0.07 with steps of 0.01	CF
ω_0 (upper boundary)	1.80 to 2.00 with steps of 0.05	BF
	2.00 to 2.20 with steps of 0.05	CF
	2.35 to 2.75 with steps of 0.05	EF
pass band attenuation (dB)	2, 3, 4, 5	BF, CF, EF
stop band attenuation (dB)	10, 15, 20, 25, 30	BF & EF
	20, 25, 30, 35, 40	CF
Regression order	1, 2, 3, 4	BF, CF, EF

[BF: Butterworth filter; CF: Chebyshev Type 1 filter; EF: Ellliptical filter]

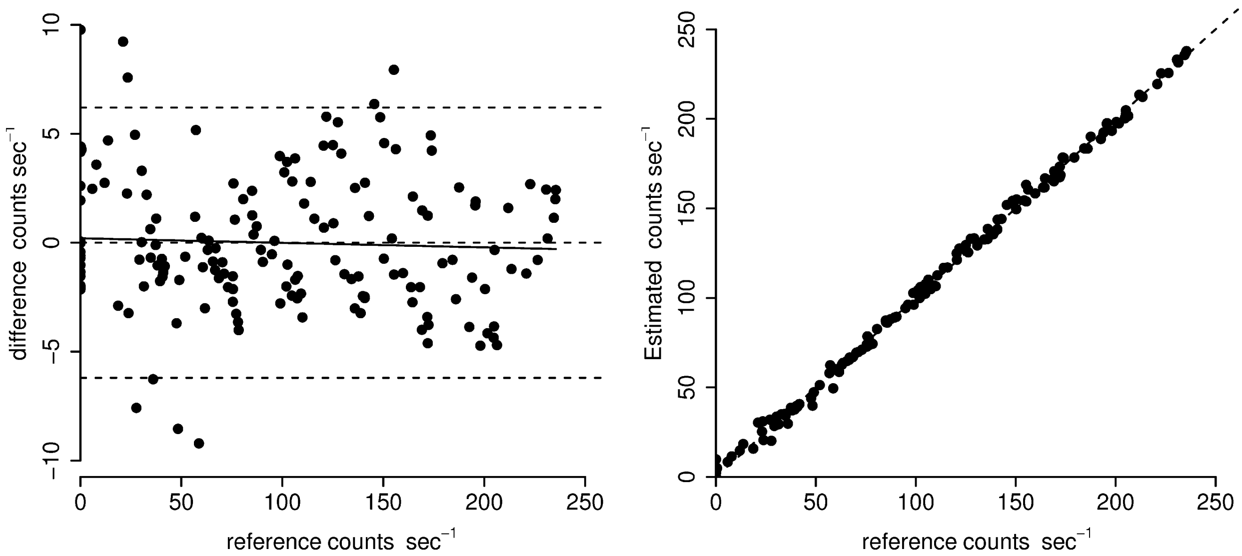

Figure 3-2: Algorithm performance on shaker device based on optimal algorithm parameters from phase 2. The dashed lines represents 95% limits of agreement and average difference (left), straight line represents best fit regression line (left) and line of identity (right).

3.4.2 Free-living evaluation

All GENEA measurements were successful but two out of thirty-six measurements could not be used because of malfunction of the Actigraph accelerometers resulting in 34 successful paired measurements with the Actigraph and the raw accelerometer. Signals were visually inspected for irregularities, which did not result in further data exclusion. All mimicked AC values were truncated to non-negative values.

Average 47 hr $AC_{mimicked}$ explained 95.5 % of the between-individual variance in 47hr period averages of AC ($p < 0.001$). Minute by-minute variance within participant (per hip) explained 94.9 ± 3.6% (mean ± SD), ranging from 83% to 98%. Average 47 hr $AC_{mimicked}$ underestimated AC by -72.3 counts · min^{-1} ($p < 0.001$, t = -9.96, df = 31, 95% confidence interval -87 to -57 counts · min^{-1}). The 95 % limits of agreement (representing the range in which 95% of the differences lie) were -150.2 and 6.94 counts · min-1, see Figure 3-3. Further the root mean square error of the 47 hr average was 82.8 counts · min-1 which amounts to 19.2% when expressed relative to the average.

The y-intercept of the linear regression line through the average 47hr values was -16.2 counts · min-1 and not significantly different from zero ($p = 0.28$). The slope of the regression line was 0.87 and did significantly differ from 1.00 ($p < 0.001$), 95% C.I: 0.81 – 0.93, see Figure 3-4. The distribution of the measurement value for the free-living data showed that the conversion algorithm assigned fewer minutes to the range 100-2000 and the range 2000-4000 compared to the Actigraph, see Table 3-5 and Table 3-6..Sensitivity of ACmimicked to discriminate between four intensity levels as measured by AC was > 56.6% and specificity was > 84.9%, see Table 3-5.. Systematic error of ACmimicked was highest in the 2000-4000 range (-28.40%) and lowest in the 0-100 range (4.72%). Random error was relatively low ranging from 0.62% to 2.36%, see Table 3-6.

Table 3-5: Confusion matrix, sensitivity, and specificity for 1 min epoch estimates in free-living evaluation (34 x 47 hrs)

Estimated intensity level (counts · min^{-1})	True intensity level (counts · min^{-1}) 0 – 100	100 – 2000	2000 – 4000	> 4000
0 – 100	68680	3947	6	0
100 – 2000	905	15253	1833	22
2000 – 4000	4	148	2502	411
> 4000	0	0	80	2055
Sensitivity (%)	98.7	78.8	56.6	82.6
Specificity (%)	84.9	96.4	99.4	99.9

Table 3-6: Relative (%) error (mean ± SD) for time spent in each intensity level (34 x 47 hrs)

Intensity level (counts · min^{-1})	All (n = 32)	Monitor 1 SN09090099 (n = 7)	Monitor 2 SN29099884 (n = 10)	Monitor 3 SN29099992 (n = 6)
0 – 100	4.72 ± 0.62	5.03 ± 0.96	4.72 ± 0.62	4.43 ± 0.10
100 – 2000	-6.25 ± 1.80	-6.90 ± 0.81	-7.00 ± 1.45	-6.60 ± 0.34
2000 – 4000	-28.40 ± 2.27	-28.84 ± 3.21	-28.31 ± 1.83	-30.08 ± 0.87
> 4000	-12.84 ± 2.36	-11.94 ± 3.89	-13.66 ± 1.60	-13.13 ± 1.42

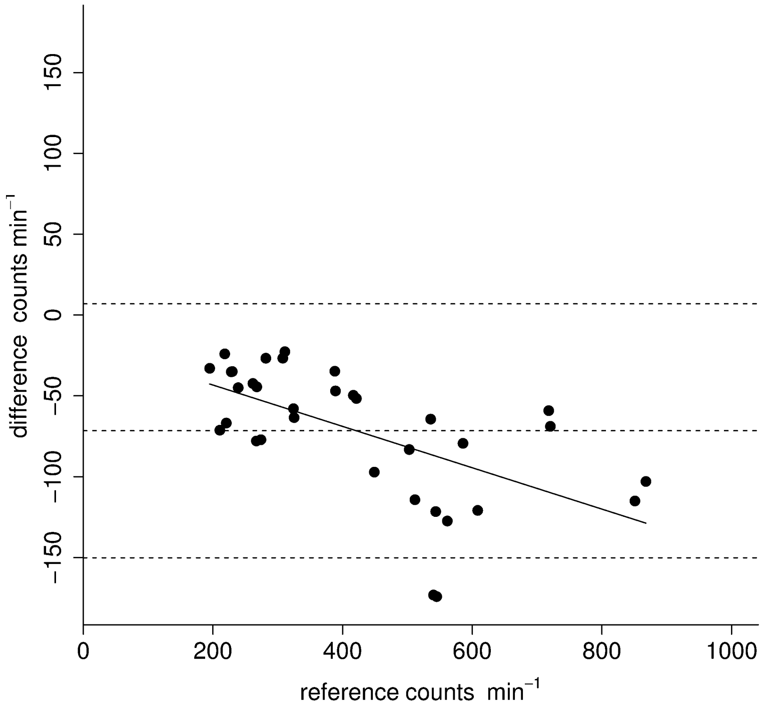

Figure 3-3: Agreement between 47 hr average of $AC_{mimicked}$ and AC. [dashed lines represent limits of agreement and the mean difference]

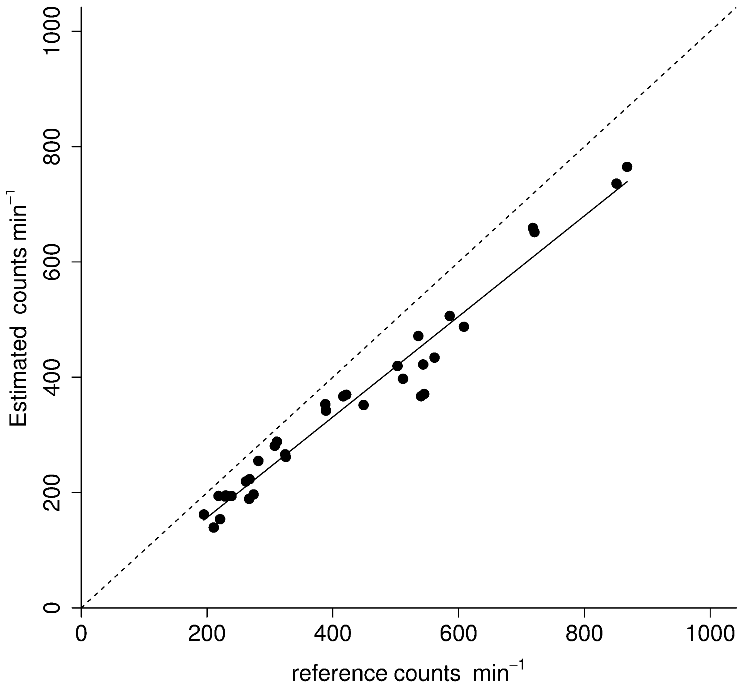

Figure 3-4: Comparison between 47 hr average of AC and $AC_{mimicked}$ [dashed line represents line of identity]

3.5 Discussion

The Actigraph accelerometer is one of the most widely used activity monitors in large-scale field studies. This chapter is a follow-up on the previous attempt to develop a method for converting raw acceleration data into Actigraph counts as described Chapter 2 [72]. Various changes were applied to the study design with the intention to improve the algorithm and make the evaluation under free-living conditions more accurate.

The algorithm performed better on the shaker device compared with Chapter 2 ($r = 0.998$ vs. 0.975). This may largely be explained by the thorough search for optimal algorithm parameters instead of solely relying on information specified by the manufacturer or the literature, and secondly by the approach of using one device for collecting both raw data and Actigraph counts instead of two separate accelerometers.

The experimental design for the evaluation under free-living conditions was improved by attaching both types of accelerometer to each hip. The explained variance in both minute-by-minute data and 47hr averages was high. However, the sensitivity to classify data in count-based intensity categories was limited (> 56.6%). Systematic measurement error for time spent in intensity categories was large (28.4%) in the intensity level 2000-4000 counts $\cdot$ min^{-1} while the standard deviation in the measurement error was low, < 2.36% for all intensity levels. Stratifying measurement error for the three most frequently used Actigraph devices in the present study showed that the systematic error per intensity level was almost constant across devices, which suggests that measurement error is predictable and can be addressed by further optimization of the algorithm.

The large underestimation of time spent in the range 2000-4000 counts $\cdot$ min^{-1} was not observed during the shaker experiment, see Figure 3-2, which may partly be explained by additional sources of measurement error other then a limited accuracy of the conversion algorithm alone. Possibly one or more of the underlying assumptions about accelerometer reliability are incorrect, including: (a) raw acceleration is measured the same by both Actigraph in raw storage mode, Actigraph in count storage mode (before onboard conversion into count) and Genea; (b) The way in which the Actilife software converts raw data into counts is compatible with the way the Actigraph generates counts based on onboard signal processing, and; (c) Batches of the same accelerometer type (Actigraph GT3X) record data in the same way (high inter-instrument reliability).

The current study design does not allow for standardised evaluation of the possible sources of error as mentioned above. It would require the combined application of various batches of all three accelerometer versions – i.e. Genea, Actigraph raw and Actigraph count – both on a shaker device and in free-living experiments.

A possible additional way to gain insight in the Actigraph signal processing is by use of machine learning techniques to search for signal features that help to explain the variance in counts. Further, the free-living evaluation may be improved by addressing the difficulty of synchronising Actigraph counts with genea data either electronically or by the application of easy detectable movement to both devices at various time points during the measurement.

The attempts to aim for methodological consistency as described in the current and previous chapter are not intended to facilitate methodological consistency between traditional accelerometer attachment to the hip and the recent trend in the epidemiological field to apply acceleration sensors to the wrist, for example the Pelotas 2004 birth cohort (Pelotas) and the Whitehall III study (London). The relation between wrist and hip movement may not necessarily be linear or reproducible, which complicates methodological consistency. Therefore, it seems impossible to account for this from a methodological end. Epidemiologists may however, want to verify whether wrist and hip accelerometer output rank participants similarly in the study population of interest.

In conclusion, perfect emulation of Actigraph counts has not been achieved, although insight has been gained in how a conversion algorithm could be developed. The recent trend to attach accelerometers to the wrist instead of the traditional hip location complicates methodological consistency even if there would be a perfect conversion algorithm. Therefore, combined application of traditional accelerometers attached on the hip and raw accelerometers attached on the wrist may be the only way to facilitate historical comparisons of data.

4 Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer

This chapter was published on 29th of July 2011 and freely accessible: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0022922

5 Separating movement and gravity components in an acceleration Signal and implications for the assessment of human daily physical activity

This chapter was published on 23rd of April 2013 and freely accessible::

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0061691

6 Wearable sensor-based classification of activity types: literature review

6.1 Introduction

The previous chapters focussed on the extraction of simple summary measures to quantify body acceleration and the use of this summary measure to estimate daily PAEE. Another area of physical activity assessment is the classification of activity types. This chapter aims to provide a narrative review of what has been published on activity type classification based on wearable sensor technology in general. This chapter was not submitted to any journal as it was mainly aimed at supporting the work in the next chapter.

6.2 Methods

The literature review as described in this chapter was restricted to studies using wearable sensors and aiming for the classification of activity types in people's daily life. The PubMed Central database (U.S. NIH free digital archive of biomedical and life sciences journal literature) was searched for the following key words: activity type or activity mode in combination with classification, detection, annotation, assessment. Most of the literature in the field of activity classification was published in conference proceedings related to the fields of computer science and engineering. These conference proceedings are not all indexed by the PubMed Central database. Therefore, the literature search was extended by a search in Google (www.google.com) using the same key words. Reference lists within articles were screened for extra articles.

In total over 120 publications were identified. Based on abstracts 33 were selected for this narrative review. In addition, ten literature reviews were found covering the issue of activity classification. However, most of these literature reviews also provide experimental data. Topics of the reviews included: context-awareness systems [108], promotion of the use of body-fixed sensors for human movement analysis over the use of laboratory based measurement equipment [109], activity recognition from user annotated data [48], general usage of accelerometers [110, 111].

health care systems [112], telemetric monitoring in the behavioural sciences [113], interpretation of accelerometer data [114], data processing techniques used in activity classification [115] and most recently Hasley at al. reviewed the use of raw accelerometry in animals with a paragraph on the potential value for activity type classification [116].

Most publications focus on classification accuracy. However, a standardized comparison between accuracy rates from different studies is difficult as differences in study design affect the absolute accuracy. For example: the detection of walking out of sedentary activities is likely to be much easier than the detection of walking out of a dataset containing Nordic walking, running, stair walking and uphill walking. This review will only refer to accuracy rates to report on comparisons as made within individual studies and will not attempt to systematically compare study results.

6.3 Results

6.3.1 History

The field of activity classification was pioneered by Veltink and colleagues in the 1990s [35, 117, 118] and was fuelled by innovations in electronics in those times. These innovations include the introduction of acceleration sensors sensitive to gravitational acceleration as described in paragraph 1.4, and feasible to be worn on the human body. Further, the development of pocket-sized digital data recorders, increased battery capacity and increased computer capacity facilitated the development of activity type classifiers [45].

6.3.2 Sensor type

Out of all types of wearable sensor technology the acceleration sensor is most commonly applied for activity type classification. The three main types of acceleration sensors are: piezo-electric, piezo-resistive, and capacitive. Piezo-resistive and capacitive acceleration sensors are sensitive to the static component of acceleration and are referred to as “seismic” or “inertial” sensors [37]. Most accelerometers of newer date are seismic sensors, as for example the latest versions of Actigraph which is the most commonly used brand of accelerometer in the epidemiology field. In most published studies sensor type is not specified. Seven of papers applied capacitive acceleration sensors [48, 56, 111, 112, 119-121] and another seven studies applied piezo-resistive acceleration sensors [35, 45, 118, 122-125]. Piezo-electric sensors have been applied in the earlier version of the Actigraph [21, 22, 126]. The advantage of seismic acceleration sensors is that they measure the gravitational component, which can be used for posture detection [35].

When only one acceleration sensor is used, it is called a uni-axial accelerometer. The sensitivity of a seismic acceleration sensor is lowest when it is parallel to gravity, which impacts the accuracy for angle assessment under static conditions. The inaccuracy can be reduced by mounting two acceleration sensors perpendicularly, which is called a bi-axial accelerometer [45]. The earliest papers were based on uni-axial or biaxial acceleration sensors [35]. Nowadays tri-axial acceleration sensors are most commonly applied for activity classification.

Combining accelerometers with gyroscopes and magnetometers has been considered to improve the ability for angle assessment [101, 109, 127]. A disadvantage of the addition of additional sensors is an increase in the amount of data to be processed and stored, and the increased power consumption by the electronic components. At the moment no measurement tools are available that allow for feasible application of gyroscopes and accelerometers for the assessment of daily physical activity over seven day periods in epidemiology on a single charge.

The sensor system proposed by Intille includes audio and video recording [108]. The data processing for such system is likely to put high demands on computational power, which may make it less applicable for epidemiological research. Physiological sensors like heart rate monitors or temperature sensors have also been used. Physiological sensors are characterized by a natural time delay and inter-individual differences. Physiological signals contain information not derivable from movement sensors but so far, only very few studies have even attempted to utilise physiological information for activity type classification [44, 128, 129]. More advanced signal analysis may be required to successfully integrate physiological signals in models for activity classification. The recently introduced GENEActive accelerometer collects near-skin temperature and light [130].

Makikawa et al. was one of the first who presented ideas on combining an accelerometer with ECG and GPS for activity classification, but did not report on any experiments or prototypes [124]. A more recent study showed that GPS may improve activity type classification for outdoor activities [126]. However, no definitive conclusions can be made as literature on this topic is still limited.

6.3.3 Sensor location

Most studies have attached the accelerometer solely to the waist [21, 56, 110, 120, 123-126, 131]. Other locations for sensor attachment are wrist and chest [44], chest and upper leg [118], upper leg [121], chest and tights [35, 45, 111], forearm, thigh and chest [43], chest [109] or a combination of five or more locations [48, 119].

Boa and Intille concluded that a limitation in most studies is that sensors are connected by wires [48]. In this review the use of wires will not be discussed since new wireless technologies are likely to cope with this limitation as observed in the past. Bao and Intille concluded that the combination thigh-wrist or hip-wrist are the most feasible alternatives to a five sensor approach with less than 5% decrease in classification accuracy [48]. Kern showed from experiments in only one participant that sensors mounted on the upper body part are preferable above sensors attached to the lower body part [119]. It was also suggested that for complex activities at least one sensor on the lower and one on the upper body is a desirable configuration [119]. It has been argued that accelerometers attached to the wrist and chest are not useful to discriminate sitting from standing [44]. Further it has been concluded that attachment on the outer side of the upper leg is the most accurate way to distinguish sitting from standing [35, 121]. One study aimed to develop a model that is robust for a random variation in accelerometer positioning on the waist. The accelerometer was removed every night [56]. Foerster et al. compared a five-sensor configuration to a two-sensor configuration but did not provide evidence on the difference [45] and concludes that a four-sensor configuration should be applied without providing results for a four-sensor configuration. Foerster argued that if more sensors are applied more types of activity can be classified, but did not provide evidence for this statement [45]. Zhang et al. compared classification accuracy of wrist and waist attachment and concluded that each location provides acceptable classification accuracy [58]. Additionally, Zhang highlighted that measurement feasibility is also important [58]. In chapter 4 I found that perceived acceptability of wrist attachment when worn for 24 hours by the participant is equal or better compared with waist attachment when worn during waking hours only. Therefore, wrist attachment is likely to allow for capturing a large fraction of the intended measurement period, which is relevant to both the estimation of PAEE and the classification of activity types. Most studies, however, do not acknowledge this need for feasibility and focus on the classification accuracy only.

Finally it was observed that a detailed description of sensor attachment for each measurement condition is important to allow for reproduction of study design but is not always provided in the literature (e.g. Nawab [128] and Veltink [35]).

6.3.4 Sensor specifications

Sample frequencies as applied are: 5 Hz [120], 10 Hz [111], 32 Hz [21, 45, 123, 126], 45 Hz [56, 110], 75 Hz [131], 100 Hz [35, 118, 119], 128 Hz [43] and 256 Hz [125], although some then collapse the data before analysis, e.g. to 1Hz [21, 126]. Dynamic range as applied are: ± 2 g [45, 120], ± 2.7 g [131], ± 4 g [123], ± 5 g [35, 118], ± 10 g [110, 111].

The required sample frequency has not been investigated in detail. Several papers advise a sample frequency of 20 Hz or higher [48, 112] and incorrectly refer to Bouten et al. [122] assuming that Bouten and colleagues investigated the required sample frequency, while they only cited other papers who did not apply acceleration sensors [122]. The Shannon-Niquist sampling theorem states that the sampling frequency should at least be twice as high as the frequency of interest [132]. It is however undefined what the frequency of interest in acceleration signal should be, this may be application dependent. Selecting a high sample frequency has the advantage that all relevant information is collected but at the same time results in more data points putting higher requirements on storage space and computational capacity. The choice of dynamic range is likely to be dependent on sensor positioning on the body and the type of activity to be classified. A wider dynamic range has the disadvantage of a loss in amplitude resolution.

6.3.5 Experimental design

The experimental design aims to facilitate the empirical assessment of the relation between sensor recordings, related features and types of physical activity. Most studies use an exercise laboratory or the near vicinity of a research institute to collect a training dataset for the development of an activity type classifier [22, 58, 133].

A laboratory setting increases standardisation but may not be representative for real life applications. Recent reviews concluded that the development of classification algorithms should therefore take place under conditions highly comparable to free-living conditions [21, 45, 48]. However, no literature was found proposing a study design that allows for standardised model development under free-living conditions. Most of the studies in literature are performed in a laboratory setting, and limited by a small sample size, a small number of activity types, and a small number of sensors [45, 48].

Experimental designs as presented by Intille [108] and van Laerhoven [121] rely on an effort made by the participant to annotate the activity type he/she performs, which may be affected by the time the participant has for making this annotation. It is questionable whether the annotations done by the participants are universal and match exactly the true pattern of activity types. Allen et al. asked their participants to perform one daily routine of 17 activities in their home environment. This routine was used for automatic training of the classification algorithm to be applied during the rest of the day [56]. The individual training adds 11.6 % accuracy compared to group training where all training data was used and 15.6% when combined with group training. However, it should be noticed that these improvements of user-specific training were only obtained after 21 or more iterations (21 days), which is not feasible for an epidemiological research setting [56]. Future research will have to evaluate whether re-designing the daily routine could reduce the required number of measurement days.

Another way to improve the experimental design may be by aiming for a better simulation of daily life in the laboratory setting. Bao and Intille introduced an obstacle course involving a list of activities [48]. In this obstacle course the participant wrote down the point in time a part of the course was completed and was not restricted in the performance of each activity.

The sample size was small in most of the studies and even limited to one participant in some [119, 121]. It is not always clear whether the participants in the validation group also participated in the evaluation group [45]. Keeping these groups separate, or at the very least maintaining the independence of training data and evaluation data, should be a good research practice to improve interpretability of classification accuracy measures. In Pober et al. the leave-one-particpant-out approach for model evaluation resulted in similar classification accuracy compared to a traditional comparison between a training group and a validation group, indicating that a leave-one-out approach is a sufficient way to evaluate models in studies with a limited sample size [21]. Nguyen proposed an unsupervised classification technique that does not require training and therefore avoids the problem of training group versus evaluation group [131]. However, the disadvantage of unsupervised classification is that the meaning of a classification is less clear. Additional information would then be needed to interpret the outcome of the classification. In line with this Allen et al. investigated a user adaptation method that relies on less user-specific information. It was stated that a variation of this method is currently widely accepted as a solution to the problem of limited training data in speaker verification [56]. Unfortunately no definition was provided on 'less user-specific information' which makes it difficult to interpret their results.

In conclusion, more research is needed to assess which study designs are best for developing activity type classifiers and for estimating classifier performance in daily life. Chapter 7 in this thesis describes an attempt to evaluate this particular issue, see chapter 7.

6.3.6 Definitions of the type of physical activity

Paragraph 1.2 on page 10 paid attention to the definition of 'physical activity'. Standardisation in the definition of what constitutes a type of physical activity becomes important when the collected data and corresponding algorithms are exchanged between people. Mathie et al. proposed the use of a flexible hierarchical tree that allows for structuring the type of physical activities towards the context of each specific study [110]. The variety in tree structures resulting from this flexibility will probably prevent comparability between studies as the performance of the algorithms is only known for one particular tree structure with a limited set of activity types. Further, an activity type within one tree structure may have another position in an extended tree structure, e.g. sitting in a bus may be annotated as "motorized transport" in one study and as "sitting" in another study.

The problem is that the design of a decision tree is potentially subjective. For example, the decision to classify a person who makes two steps as "walking" instead of a sub-activity of "standing" could have impact on the accuracy of the classification algorithm when there is no agreement on the border between those definitions. A similar problem was observed by Pober et al. who faced difficulty in distinguishing uphill walking from level walking [21]. It is questionable whether these two locomotion activities should be regarded as different types of activity or only as different intensity levels of the same broader type of activity. Nawab et al. proposed the partitioning of activity types in multiple hierarchically related scales: Principle activities, transitional activity, and co-activity [128]. Also here, the decision on principle activities and class seems to be subjective.

In most of the literature the type of activity was defined by one observer or annotator. In Nawab et al. the annotations was done by 20 participants, who were asked to view videotapes and annotate the activities as seen. The definition by these judges was then used as the final annotation [128]. It is unclear what the value is of the average opinion of 20 individuals, and whether it relates to cultural background or educational degree. Sherill encourages scientists to consider a good and [122] not too specific definition of a task to improve classification accuracy [43]. This strategy turned out to be more important compared to improving the number of sensor locations [43], a conclusion also supported by Mathie [110]. Conventions on the definition of physical activity types are needed to resolve the potential problem of a lack of agreement on what needs to be classified.

I recommend that the definition of activity types is split up in two definitions for respectively two stages: the definition at the stage of 'data collection and annotation' and its definition at the stage of 'classifier development'. The definition at the stage of 'data collection and annotation' needs to be as close to the reality as possible and not to be constrained by prescribed formats like hierarchies. Here, a video or picture illustrating the activity would be ideal as that minimizes the subjective classification by an observer. Next, the developer of the activity classifier will have to transform the detailed definition of each activity type into a simple finite number of classes that are practical and logical for the classifier development. This two-phase approach would leave maximal freedom to the developer of the classifier to decide on a logical definition especially when datasets from various research groups are to be merged. Such a procedure would also facilitate re-use of datasets as a high amount of details are captured and not summarized to abstract and potentially subjective definitions.

6.3.7 Feature extraction

Veltink and Bussman introduced the use of the DC component of a seismic acceleration sensor to detect the orientation of the human body with regard to gravity [35, 117, 118]. According to Sekine the changes in gait are mainly characterized by changes in posture [125]. Sekine concludes that all participants lean forwards when ascending stairs: this is detectable with wavelet analysis on the anterior-posterior signal derived with trunk accelerometer. Descending stairs could be detected with wavelet analysis on the vertical signal. Level walking has three peaks, and stairway walking has one or two peaks in a walking cycle. The lateral signal was not significantly different among walking patterns. Fast Fourier Transform analysis was regarded to not be sufficient by Sekine [125]. However, Zhang mainly relied on STFT (short-time Fourier transform) features and concluded that STFT-features are valuable for activity classification with only little beneficial value for discreet-wave transform (DWT) features [58]. Several papers suggest the incorporation of temporal information on the order and duration of activity types could improve classification accuracy for accurate movement classification and have the potential to reduce computational time and better detect short-duration movements [21, 56, 110]. It is questionable whether a time-sequence in types of activity is really of additional value. In daily life, movements related to types of activity may adhere to a certain pattern but behaviour in general may need to be regarded as a stochastic process.

Some studies are limited in their feature extraction due to pre-processing of the data by the manufacturer, in particular when this dramatically reduces the time resolution [21, 126]. The additional value of more complex features compared to simple features for the classification accuracy was small in the study by Allen [56]. However, this finding could be restricted to the specific features extracted, the classification algorithms used and the sensor wearing position in the particular study. The disadvantages of onboard feature extraction for data reduction are that it consumes power and that the possibilities of post-processing of the data are limited. This is especially the case when the classification algorithms have not been optimized. However, onboard data processing has also some advantage. Onboard data processing, if performed in near-realtime, allows for direct feedback to the participant [41] and can be used for alarm activation in case of an emergency, e.g. a fall [134].

Feature selection is another important step in the classification inference scheme. Most of the papers as reviewed do not report the criteria used for initial feature selection. Studies usually extract as many features as possible and then select features based on their performance, although some perform a dedicated dimensionality reduction first [135].

6.3.8 Classification framework

The classification framework refers to the techniques used to classify types of physical activity from non-annotated data. The techniques are usually developed from a separate learning dataset. It has been recognized that the important question in this field should be: "What classification approach should we use and what is the best sensor placement?" [45]. This paragraph includes a short overview of the techniques as used.

Several classification frameworks have been investigated including: quadratic discriminant analysis (QDA), Hidden Markov models (HMM), topic models, neural networks, k-nearest neighbour, classification trees, and heuristic models. No clear guidance has been found in the literature for ranking or grouping these classification approaches. Most of the discussion in conference proceedings and peer-reviewed journal publications focuses on the resulting classification accuracy rather than explaining the methods to the non-expert. The classification accuracy is likely to be affected by a number of factors, including the study design, the instrumentation, the definition of the activity types, and the classification framework, thus complicating the comparison between classification frameworks as presented in different papers. No discussion has been found in literature on why some classification frameworks perform better than others: classification frameworks are often approached as a black box. A detailed evaluation of classification approaches would help to improve insight into classification frameworks and allow for better justification when choosing a classification approach.

Some papers apply multiple classification frameworks and then select the best performing framework [58]. The role of chance in finding a good performing classification framework when evaluating multiple frameworks is usually not acknowledged.

A short overview will now follow of some of the findings by various papers on the topic of selecting a classification framework. Allen et al. investigated the advantages of moving away from a static rule-based classification approach towards a more general Gaussian mixture model (GMM)-based approach and concluded that a GMM model performs better than a heuristic model except for the classification of standing [56]. It is, however, highly unlikely that the heuristic model as used in their study is representative for all heuristic models, and also unclear whether the design of the heuristic model had in one way or the other been inspired by the same dataset as with which the GMM-based model was trained. Van Laerhoven applied a Kohonen

Self-Organizing Map and claims that this provides more insight in what is happening compared to other classification approaches. However, evidence for this claim is missing [121]. Mathie et al. introduced a classification structure based on a binary decision tree, which is claimed to be modular and flexible [110]. The tree is flexible from a visual perspective but no evidence or clarification is provided that a binary decision or a series of binary decisions is more flexible than a non-binary decision tree. Mathie et al. admitted that the accuracy of a whole branch relies on one binary decision based on one parameter. Therefore small adjustments to the top level of the three may have large impact on the accuracy of the overall classification.

Pober concludes that QDA and HMM are both applicable in physical activity classification [21]. The difference in performance is small and may be related to other differences in the data processing than the difference between QDA and HMM.

Zhang et al. compared a variety of classification frameworks, including decision trees, naïve Bayes, logistic regression, support vector machine and neural networks using the open source WEKA toolkit and concluded that a decision tree is the best candidate [58]. A neural network framework was considered by far the computational slowest [58]. Gyllensten and Bonomi also compared a variety of classification frameworks, including decision trees, neural networks, support vector machine and majority voting using the Matlab environment and concluded that differences in performance between the classification models are small, with the SVM and the majority voting scheme achieving slightly higher accuracy [136].

The problem when interpreting the recommendations as made by the studies as described above is that classifier frameworks were evaluated for a specific set of activity types, performed under specific conditions and using a specific set of signal features. The reported classification accuracy is specific to those conditions and may not be representative for other conditions. Therefore, caution should be taken when implementing recommendations in daily physical activity.

6.3.9 Facilitation of method implementation

The majority of publications focus on classifier accuracy and describing how the method was developed. Minor or no attention is commonly paid to facilitating the application of these classifiers. The following problems were identified:

1. The actual classification models including the model coefficients resulting from the classifier training stage are never published. These coefficients are essential for applying and independently evaluating a classifier.
2. Training datasets used for developing a classifier are commonly not published. The disadvantages of not publishing datasets are: a) resources are potentially not used to their full extent [137], publishing the data would allow other groups to come up with different and improved classifiers; b) data quality can not be verified; and; c) meta-analysis by merging multiple datasets is impossible.
3. Most studies have a low sample size (N < 30) predominantly recruited in a university environment possible causing selection bias, which limits representativeness for the wider population, even if a cross-validation would be performed. In relation to this, studies often fail to give clear guidance under what conditions a classifier can be used.
4. Most studies fail to provide a detailed description of the activity types as performed by the participants. The exact instructions to the study participant, the characteristics of the objects involved, and the environment in which the activity takes place may all influence the movements of the participant and by that sensor output. Providing a detailed description of these procedures may therefore be essential for understanding differences between signals and subsequent classifier performance. For example, various studies investigate the activity type "sitting", but most do not describe the object on which the participant sits. The dimensions, shape, and flexibility of the sitting object may determine the representativeness of 'sitting' for sitting in real life, e.g. sitting on a non-cushioned chair in a lecture hall and paying full attention to the lecture may well differ from sitting in a sofa at home in front of the television.
5. Most study protocols are restricted to a limited number of activity types often performed in a standardized fashion. Daily physical activity is likely to involve numerous variations on what we consider basic activity types. Testing a small number of standardized activity types may not cover the natural variation with which activities are performed and it is therefore likely that the classification task is easier in these standardised study environments, compared with real life situations.
6. Some studies use sensors that express their output in brand specific units often referred to as "counts". The underlying signal processing for a count is usually

not provided. If wearable sensor output is not expressed in SI units, a full description of the underlying signal processing or an exact copy of the same hardware would be required to replicate a dataset. Most studies lack such description.

7. Conflicts of interest are difficult to assess. Often, the authors are involved in the development of the hardware, the development of a classifier, and the evaluation of the method as a whole. In addition, some method developers are also method-user. In combination with the inaccessibility of data and signal processing scripts a subjective judgement or more problematically, a scoop, is easily made.

6.3.10 Data sharing

Openness and transparency of information would allow for classifier development in large study populations, with more natural variation in activity performance and the ability to complement each other's expertises and resources. However, collaboration may not be as easy as it sounds: A variety of scientific fields are interested in the development of methods for activity type classification, e.g.: metabolic research, occupational research, telecare, clinical research, sport, psychological research, army research, sleep analysis, fall detection, estimation of walking distance, and others. Each of these fields has different expectations about accuracy, precision, and feasibility, complicating a one method suits all approach. The introduction of standards on study design and data format to ensure consistency may constrain the freedom of the individual researcher to contribute and innovate. Other challenges related to experimental data sharing in a more general context have been described by Nelson [137]. Visscher and Weissman listed the advantages and challenges specific for fMRI data: Practical challenges and time issues were brought forward as reasons against sharing data. Further, the danger of having your potentially valuable data being described by others as invaluable could be a legitimate motivation to not share the data in the first place. A strong motivation against data sharing is the loss of hard work and exclusive potential for publications. Visscher and Weisman are positive about openness but do not provide a proposal on how to addresses the disadvantages as listed.

The innovation of method development by data sharing has been on the general agenda since the start of this PhD-studentship but has not materialized. A few initiatives by other groups were identified in the literature: An open-source network by MIT [108], the Palantir context library [44], and Jovanov proposed a composition of technologies including a research database to support home rehabilitation [112]. None of these projects are currently widely used or cited. More recently, other groups and companies have started the development of repositories and tools for sharing signal processing scripts [138, 139].

6.4 Concluding remarks

A variety of sensors types, sensor locations, study designs and signal processing techniques have been explored for activity type classification. Standardised comparisons on each of these areas are hard to make based on existing literature, which complicates making informed recommendations on the method to use. Additionally, published studies commonly fail to facilitate the implementation of the activity type classifier. Addressing the problems related to the implementation of activity classifiers may be an essential step for future development of methods for activity type classification. While others work on the development of repositories the next chapter will aim to gain insight in the validity of commonly used study designs for the development and evaluation of activity type classifiers which was identified in paragraph 6.3.5 as an unexplored area of research.

7 Impact of study design on development and evaluation of an activity-type classifier

This chapter was published on the 15th of April 2013 and freely accessible here:

http://jap.physiology.org/content/114/8/1042.long

8 Using activity type to improve the estimation of energy expenditure: three types of locomotion

This chapter was published in 2012 and available here:
https://www.researchgate.net/publication/229011489_Estimating_Energy_Expenditure_from_Raw_Accelerometry_in_Three_Types_of_Locomotion

9 General discussion

Raw accelerometry is a measurement technology that has recently become applicable in the field of physical activity epidemiology. Physical activity assessment based on accelerometry involves a wearable device that collects and stores measurements of acceleration and is worn by the study participant for usually seven days while pursuing their normal daily activities. In contrast to the accelerometers as traditionally used, raw accelerometers store data in manufacturer independent units at a sample frequency of at least twice the maximum frequency of interest. For example, if the frequencies of interests lie within the range 0 to 10 Hz then a sample frequency of at least 20 Hz would be needed. Raw accelerometery has the potential of enhancing the assessment of physical activity as it provides more detailed and more transparent information compared with traditional accelerometers. However, the absence of readily available signal processing techniques complicates immediate implementation of the technology. The objective of this thesis was to facilitate the implementation of raw accelerometry in the epidemiological field by developing signal processing techniques to extract meaningful outcome measures from the raw data. The next paragraph summarises the contributions of each chapter.

9.1 Chapter-specific contributions

In chapter 2 and 3 an algorithm for converting raw accelerometer output to traditional accelerometer output of the Actigraph accelerometer was developed. Although, no perfect conversion algorithm was achieved more than 94% of the variance in Actigraph counts was explained by the algorithm. Measurement error as reported may provide hints for further development. This work continues in collaboration with a group from Odense, Denmark.

In chapter 4, it was shown that a simple summary measure derived from a raw accelerometer attached to the wrist allows for explaining variance in daily PAEE. Further, it was found that perceived user-acceptability of an accelerometer attached to the wrist and worn for 24 hrs at least equals the perceived user-acceptability of a similar accelerometer attached to the hip worn for waking hours only.

In chapter 5, signal processing metrics were evaluated for their ability to remove the gravitational component from a raw acceleration signal and extract movement-related acceleration. The experiments were based on standardised motions of an industrial robot arm. A newly designed metric called $HFEN_+$ outperformed metrics that have been used in the literature for rotational frequencies above the frequency corresponding to the cut-off point of the high-pass frequency filter. Metric ENMO on the other hand outperformed all other metrics at lower rotational frequencies. In order to test the relevance of these findings for the assessment of daily physical activity the metrics were applied to the raw accelerometer data as collected on the wrist in real life, see chapter 5. Results show that the choice of metric affects the precision with which daily PAEE can be estimated and confirms the finding from chapter 5 that metric choice does matter.

I then moved on from basic summary measures to more sophisticated signal features to make inference on physical activity type. A literature review on this topic was presented in chapter 6. Main gaps in knowledge as identified by the literature review were: a limited insight in the validity of experimental designs for classifier evaluation and the lack of access to classifiers prohibiting the implementation and evaluation and error in the time spent in activities.

Chapter 7 then continued by investigating the effect of experimental design for classifier evaluation on the validity of classifier performance estimates. Results showed that experimental designs under confined conditions may provide biased estimates of classifier performance in real life scenarios, in particular classifier specificity.

The final chapter follows up on my MSc thesis as published in 2009 [49]. In chapter 8, activity type specific prediction equations for PAEE were developed for walking on level ground, cycling on level ground and stair walking based on accelerometer attachment to the lower back. The study is one of the first to develop and share activity-specific predictions equations for PAEE based on raw accelerometry. Accelerometer attachment to the lower-back is not common in the epidemiological field, which limits the contribution of this chapter to the general objective of this thesis. Nonetheless, the chapter could be seen as an example of the kind of study that needs to be done for wrist- and/or hip attached accelerometers.

In parallel to conducting the work as described in this thesis, signal processing scripts were developed for the (batch) analysis of epidemiological datasets. The scripts were written in R and made use of some of the techniques as described in the chapters.

So far, the R-scripts have been tested in: the 2004 Pelotas birth cohort (N = 3000) involving seven-day wrist recordings for each participant and a pilot trial in children with sever acute malnutrition in Ethiopia (N = 25) involving repeated five day recordings on wrist and hip. Additionally, data analysis is currently in preparation for various other datasets. The development of these R-scripts has not been described in a separate chapter as it was considered a practical rather than a scientific achievement. However, the experience of developing these R-scripts has been helpful in keeping method development connected with method implementation.

9.2 Mission accomplished?

The work as presented in this thesis facilitates the implementation of raw accelerometry in the epidemiological field as it provides insight in differences between metrics to summarize the data, insight in how daily PAEE can be predicted from the output from these metrics and insight in the potential weakness of study designs with which activity type classifiers are commonly evaluated.

A few recent journal publications indicate that colleagues have also been working on the implementation of raw accelerometry in epidemiology. One paper by Esliger et al. assessed the reliability and the validity of the Genea accelerometer [54]. The paper contained three parts: (i) the assessment of mechanical reliability and validity by using a shaker device, (ii) the development and evaluation of a method to use accelerometer data to classify intensity levels, and (iii) the comparison of the Genea against other widely used accelerometers [54]. The mechanical reliability as reported was slightly lower compared to the mechanical reliability of a raw accelerometer, the Dynaport, which I evaluated outside the context of this thesis [37]. The lower reliability of the Genea may be a consequence of its higher dynamic range compared with the dynamic range of the Dynaport (+/- 6g vs +/- 2g) resulting in a lower resolution to detect alternations in acceleration and therefore higher noise level although only a head-to-head comparison using the same design could firmly establish such a difference. Intensity level thresholds were derived from laboratory experiments involving 60 adult participants performing 10-12 activities and using the ROC-curve technique [54]. The laboratory experiments involved time proportions between activities that may not resemble real life scenarios. Therefore, based on my observations in chapter 7 it may be concluded that threshold performance estimates by Esliger et al. may have been biased by the experimental design of their study and not accurately reflect method performance in real life. Future studies are warranted to evaluate method performance in experimental designs that better reflect real life conditions. The main value of Esliger's paper is that it pitches the Genea as a new technology that is different compared with traditional accelerometers.

Zhang et al. re-used Esliger's data and developed activity type classifiers for both wrist and waist attachment and concluded that wrist attachment provides a good alternative to waist attachment for the classification of four activity categories based on features calculated over 12.8 second window lengths [58].

A third study by Rowlands and Stiles assessed the relation between accelerometer output and mechanical loading in order to demonstrate that accelerometers can be used to assess mechanical loading [168]. The associations as found for walking, running and jumping were predictable from the dynamics of mechanical movement in which more acceleration generally results from higher forces. The study does not provide insight in potential measurement error in real life involving a wider variety of activity types and predominantly static mechanical loading. Accelerometers are insensitive to mechanical loading by static forces, which is not acknowledged by Rowlands and Stiles [168] . Therefore, it is unclear whether raw accelerometry can be regarded as a valid measure for overall mechanical loading in daily physical activity independent from being a measure of (peak) acceleration. The use of pressure sensors in foot soles could be hypothesised to be a more accurate way of capturing mechanical loading, including static mechanical loading.

Despite the joined efforts of this thesis and recent work published by colleagues full robust implementation of raw accelerometry in the epidemiological field is challenged by the need for methodological consistency. I made an effort to preserve methodological consistency with traditional accelerometer output in chapters 2 and 3. However, achieving methodological consistency between studies that apply raw accelerometry is equally important but possibly even more difficult to achieve. The combination of the freedom that has been given to the user by access to raw data and the lack of consensus on how raw data ought to be processed may lead to a high diversity in signal processing techniques complicating comparison of journal publications. Obviously there is the opportunity of re-analysis but this does not guarantee that data will truly be re-analysed. Subsequently, differences in seemingly similar outcome measures between studies based on (slightly) different signal processing methods are likely to cause confusion.

It seems unlikely that consensus on how raw accelerometer data ought to be processed will be reached in the near future. Various candidate metrics for the extraction of a simple indicator of the magnitude of acceleration exist as demonstrated in chapter 4 and 5. Additional research may lead to different insights in the advantages and disadvantages of the metrics as evaluated. Other physical activity outcome like activity type, may require a similar quest for consensus. Also in terms of data quality checks more research is needed. Non-wear detection does leave some influence on the explained variance in average daily PAEE, see chapter 4 and 5. However, this does not mean that other outcome measures are not affected by the accuracy of the non-wear detection and other types of quality checks on the data; ongoing evaluation is warranted.

Consensus on signal processing may not be achievable by doing more methodological studies alone. The actual challenge underlying a lack of consensus may be twofold: (i) there is no gold standard for physical activity assessment in daily life for intensity and type and (ii) the urgent need for feasible methods.

Methods are usually developed and evaluated under confined laboratory-conditions, examples include: methods for step detection, methods for activity type detection and methods for estimation of minute-by-minute or even second-by-second energy expenditure. The uncertainties surrounding the representativeness of these experimental conditions for real life make it hard to prove that a method truly works in real life. The doubly labelled water method could be considered as the only method that can be applied in real life. The doubly labelled water method is considered the gold standard for the assessment of total energy expenditure over multiple days but in order to derive daily PAEE from TEE a measure of energy expenditure in rest is needed. The accuracy of such rest test in addition to the uncertainty about the representativeness of derived REE for the entire measurement period affects the accuracy of the REE assessment and indirectly the assessment of PAEE.

The urgent need for a feasible method causes methods to be used without thorough evaluation. Users are strongly dependent on claims made by developers in scientific journals and the quality of the peer-reviewed publication process or there may be publication bias which prohibits such evaluation to appear very often in the literature. Claims by methodological papers are rarely evaluated by independent scientists. The users of measurement tools are then forced to make a decision based on (existing) literature claims, the opinion of an authority in the field, or by running their own method development studies without any subsequent external evaluation. As a consequence certain techniques may become popular without being thoroughly investigated.

9.3 Recommendations for future work

In order to facilitate the full and successful implementation of raw accelerometry in the epidemiological field the following measures may be necessary:

- Continuation of the development of methods, where methods are not restricted to the signal processing but also include study design (incl. sensor location) and the means of facilitating the implementation of the method. The type of methods that are of interest are methods for classification of activity type, methods for estimation of average daily PAEE, methods for estimation of instantaneous (e.g. minute-by-minute) patterns in PAEE, methods for sleep analysis and methods for converting raw accelerometer data to traditional accelerometer output. Furthermore, methods need to be developed for infants, children and elderly who are often left out from original validation studies.
- Scientific journal editors and reviewers may need to encourage sharing of signal processing techniques [169]. This would increase transparency in the literature and ensure that methods can be replicated.
- Continued attention to the potential limitations of study designs as used for method evaluation studies.
- Scientific journals and/or method developers to play a more active role in educating method-users about the advantages and disadvantages of certain methods. Scientific literature reviews often expire within a year after publication when it comes to their ability to cover the most recent technological developments. An online platform with links to journal publications, sensor manufacturers, practical user-reviews and links to signal processing scripts may be a much more effective approach.
- Continued pilot implementation of methods in the epidemiological field in order to evaluate method feasibility.
- More insight is needed on how perceived user acceptability relates to selection bias and protocol adherence in order to make better informed decisions about sensor location.

- Introduction of data management tools that allow for repeated re-analyses of datasets to preserve methodological consistency both for the epidemiological field as well as for the methodological field.

9.4 Conclusion

The work as presented in this thesis facilitates the implementation of raw accelerometry in epidemiological research by providing a range of signal processing tools and by identifying potential sources of measurement error. An important challenge for the future is to preserve methodological consistency between studies based on raw accelerometry without prohibiting methodological innovation. Methodological consistency is challenged by a lack of consensus on which signal processing techniques to use, a lack of consensus on sensor location, a lack of clarity on how a classifier should be evaluated and the urgent need for signal processing tools for activity type classification and for profiling activity-related energy expenditure.

10 Literature

[1] Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? Int J Obes Relat Metab Disord. 1998 Aug;22 Suppl 2:S30-8.
[2] Corder K, Brage S, Ekelund U. Accelerometers and pedometers: methodology and clinical application. Current opinion in clinical nutrition and metabolic care. 2007 Sep;10(5):597-603.
[3] Corder K, Ekelund U, Steele RM, Wareham NJ, Brage S. Assessment of physical activity in youth. J Appl Physiol. 2008 Sep;105(3):977-87.
[4] Wong MY, Day NE, Luan JA, Chan KP, Wareham NJ. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? International journal of epidemiology. 2003 Feb;32(1):51-7.
[5] Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985 Mar-Apr;100(2):126-31.
[6] Neilson HK, Robson PJ, Friedenreich CM, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? The American journal of clinical nutrition. 2008 Feb;87(2):279-91.
[7] Wijndaele K, Brage S, Besson H, Khaw KT, Sharp SJ, Luben R, et al. Television viewing and incident cardiovascular disease: prospective associations and mediation analysis in the EPIC Norfolk Study. PloS one. 2011;6(5):e20058.
[8] Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. The international journal of behavioral nutrition and physical activity. 2008;5:56.
[9] Corder K, van Sluijs EM, Wright A, Whincup P, Wareham NJ, Ekelund U. Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? The American journal of clinical nutrition. 2009 Mar;89(3):862-70.
[10] Buchowski MS, Matthews CE, Cohen SS, Signorello LB, Fowke JH, Hargreaves MK, et al. Evaluation of a Questionnaire to Assess Sedentary and Active Behaviors in the Southern Community Cohort Study. Journal of physical activity & health. 2011 Aug 2.
[11] McNamara E, Hudson Z, Taylor SJ. Measuring activity levels of young people: the validity of pedometers. British medical bulletin. 2010;95:121-37.
[12] Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. Medicine and science in sports and exercise. 2011 May;43(5):867-76.
[13] Plasqui G, Westerterp KR. Physical activity assessment with accelerometers: an evaluation against doubly labeled water. Obesity (Silver Spring, Md. 2007 Oct;15(10):2371-9.
[14] Feito Y, Bassett DR, Thompson DL. Evaluation of Activity Monitors in Controlled and Free-Living Environments. Medicine and science in sports and exercise. 2011 Sep 7.
[15] Le Masurier GC, Lee SM, Tudor-Locke C. Motion sensor accuracy under controlled and free-living conditions. Medicine and science in sports and exercise. 2004 May;36(5):905-10.

[16] Pomeroy J, Brage S, Curtis JM, Swan PD, Knowler WC, Franks PW. Between-monitor differences in step counts are related to body size: implications for objective physical activity measurement. PloS one. 2011;6(4):e18942.

[17] Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Medicine and science in sports and exercise. 2008 Jan;40(1):181-8.

[18] Esliger DW, Tremblay MS, Copeland JL, Barnes JD, Huntington GE, Bassett DR, Jr. Physical activity profile of Old Order Amish, Mennonite, and contemporary children. Medicine and science in sports and exercise. 2010 Feb;42(2):296-303.

[19] Metcalf BS, Hosking J, Jeffery AN, Voss LD, Henley W, Wilkin TJ. Fatness leads to inactivity, but inactivity does not lead to fatness: a longitudinal study in children (EarlyBird 45). Archives of disease in childhood. 2011 Oct;96(10):942-7.

[20] Peters TM, Moore SC, Xiang YB, Yang G, Shu XO, Ekelund U, et al. Accelerometer-measured physical activity in Chinese adults. American journal of preventive medicine. 2010 Jun;38(6):583-91.

[21] Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. Medicine and science in sports and exercise. 2006 Sep;38(9):1626-34.

[22] Staudenmayer J, Pober D, Crouter SE, Bassett DR, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol. 2009 Jul 30.

[23] Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. Medicine and science in sports and exercise. 2007 Apr;39(4):716-27.

[24] Rothney MP, Apker GA, Song Y, Chen KY. Comparing the performance of three generations of ActiGraph accelerometers. J Appl Physiol. 2008 Oct;105(4):1091-7.

[25] Krasnoff JB, Kohn MA, Choy FK, Doyle J, Johansen K, Painter PL. Interunit and intraunit reliability of the RT3 triaxial accelerometer. Journal of physical activity & health. 2008 Jul;5(4):527-38.

[26] Westerterp KR. Assessment of physical activity: a critical appraisal. European journal of applied physiology. 2009 Apr;105(6):823-8.

[27] Assah FK, Ekelund U, Brage S, Corder K, Wright A, Mbanya JC, et al. Predicting physical activity energy expenditure using accelerometry in adults from sub-Sahara Africa. Obesity (Silver Spring, Md. 2009 Aug;17(8):1588-95.

[28] Rothney MP, Brychta RJ, Meade NN, Chen KY, Buchowski MS. Validation of the ActiGraph two-regression model for predicting energy expenditure. Medicine and science in sports and exercise. 2010 Sep;42(9):1785-92.

[29] Besson H, Brage S, Jakes RW, Ekelund U, Wareham NJ. Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. The American journal of clinical nutrition. 2010 Jan;91(1):106-14.

[30] Watkinson C, van Sluijs EM, Sutton S, Marteau T, Griffin SJ. Randomised controlled trial of the effects of physical activity feedback on awareness and behaviour in UK adults: the FAB study protocol [ISRCTN92551397]. BMC public health. 2010;10(1):144.

[31] Rennie K, Rowsell T, Jebb SA, Holburn D, Wareham NJ. A combined heart rate and movement sensor: proof of concept and preliminary testing study. European journal of clinical nutrition. 2000 May;54(5):409-14.

[32] Brage S, Brage N, Franks PW, Ekelund U, Wong MY, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring

improves estimate of directly measured physical activity energy expenditure. J Appl Physiol. 2004 Jan;96(1):343-51.
[33] Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor Actiheart. European journal of clinical nutrition. 2005 Apr;59(4):561-70.
[34] Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC, Wareham NJ. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. International journal of epidemiology. 2010 Feb;40(1):112-20.
[35] Veltink PH, Bussmann HB, de Vries W, Martens WL, Van Lummel RC. Detection of static and dynamic activities using uniaxial accelerometers. IEEE Trans Rehabil Eng. 1996 Dec;4(4):375-85.
[36] Using an accelerometer for inclunation sensing. 2012 [cited 2012 16th of May]; Available from: http://www.analog.com/static/imported-files/application_notes/AN-1057.pdf
[37] Van Hees VT, Slootmaker SM, De Groot G, Van Mechelen W, Van Lummel RC. Reproducibility of a triaxial seismic accelerometer (DynaPort). Medicine and science in sports and exercise. 2009 Apr;41(4):810-7.
[38] Uiterwaal M, Glerum EB, Busser HJ, van Lummel RC. Ambulatory monitoring of physical activity in working situations, a validation study. Journal of medical engineering & technology. 1998 Jul-Aug;22(4):168-72.
[39] Bussmann JBJ, Martens WLJ, Tulen JHM, Schasfoort FC, van den Berg-Emons HJG, Stam HJ. Measuring Daily Behavior Using Ambulatory Accelerometry: The Activity Monitor. Behavior Research Methods Instruments and Computers. 2001;33(3):349-56.
[40] Pitta F, Troosters T, Spruit MA, Probst VS, Decramer M, Gosselink R. Characteristics of physical activities in daily life in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2005 May 1;171(9):972-7.
[41] Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a Real-Time Human Movement Classifier Using a Triaxial Accelerometer for Ambulatory Monitoring. Ieee Transactions on Information Technology in Biomedicine. 2006 Numb 1;10:156-67.
[42] Preece SJ, Goulermas JY, Kenney LP, Howard D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE transactions on bio-medical engineering. 2009 Mar;56(3):871-9.
[43] Sherrill DM, Moy ML, Reilly JJ, Bonato P. Using hierarchical clustering methods to classify motor activities of COPD patients from wearable sensor data. Journal of NeuroEngineering and Rehabilitation. 2005;2:16.
[44] Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity Classification Using Realistic Data From Wearable Sensors. Ieee Transactions on Information Technology in Biomedicine. 2006 Numb 1;10:119-28.
[45] Foerster F, Fahrenberg J. Motion Pattern and Posture: Correctly Assessed by Calibrated Accelerometers. Behavior Research Methods Instruments and Computers. 2000;32(3):450-7.
[46] Brandes M, Zijlstra W, Heikens S, van Lummel R, Rosenbaum D. Accelerometry based assessment of gait parameters in children. Gait & posture. 2006 Dec;24(4):482-6.
[47] Zijlstra W, Hof AL. Assessment of spatio-temporal gait parameters from trunk accelerations during human walking. Gait & posture. 2003 Oct;18(2):1-10.

[48] Bao L, Intille SS. Activity Recognition from User-Annotated Acceleration Data. In: Ferscha A, Mattern F, editors. Pervasive Computing, Second International Conference, PERVASIVE; 2004; Vienna, Austria: Springer; 2004.
[49] van Hees VT, van Lummel RC, Westerterp KR. Estimating Activity-related Energy Expenditure Under Sedentary Conditions Using a Tri-axial Seismic Accelerometer. Obesity (Silver Spring, Md. 2009 Mar 12.
[50] Forouhi NG, Luan J, Hennings S, Wareham NJ. Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely study 1990-2000. Diabet Med. 2007 Feb;24(2):200-7.
[51] Butte NF, Wong WW, Adolph AL, Puyau MR, Vohra FA, Zakeri IF. Validation of cross-sectional time series and multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents using doubly labeled water. The Journal of nutrition. 2010 Aug;140(8):1516-23.
[52] Redmond DP, Hegge FW. Observations on the design and specification of a wrist-worn human activity monitoring system. Behavior Research Methods. 1985 Dec;17(6):11.
[53] Van Someren EJ, Lazeron RH, Vonk BF, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. Journal of neuroscience methods. 1996 Mar;65(1):55-62.
[54] Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Medicine and science in sports and exercise. 2011 Jun;43(6):1085-93.
[55] Chen KY, Acra SA, Majchrzak K, Donahue CL, Baker L, Clemens L, et al. Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. Diabetes technology & therapeutics. 2003;5(6):1023-33.
[56] Allen FR, Ambikairajah E, Lovell NH, Celler BG. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiological measurement. 2006 Oct;27(10):935-51.
[57] Schoeller DA, Jefford G. Determinants of the energy costs of light activities: inferences for interpreting doubly labeled water data. Int J Obes Relat Metab Disord. 2002 Jan;26(1):97-101.
[58] Zhang S, Rowlands AV, Murray P, Hurst T. Physical Activity Classification using the GENEA Wrist Worn Accelerometer. Medicine and science in sports and exercise. 2011 Oct 8.
[59] Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Improving the assessment of daily energy expenditure by identifying types of physical activity using a single accelerometer. J Appl Physiol. 2009.
[60] van Hees VT, Ekelund U. Novel daily energy expenditure estimation by using objective activity type classification: where do we go from here? J Appl Physiol. 2009 Sep;107(3):639-40.
[61] Klesges RC, Klesges LM, Swenson AM, Pheley AM. A validation of two motion sensors in the prediction of child and adult physical activity levels. American journal of epidemiology. 1985 Sep;122(3):400-10.
[62] Matthews CE, Freedson PS. Field trial of a three-dimensional activity monitor: comparison with self report. Medicine and science in sports and exercise. 1995 Jul;27(7):1071-8.
[63] Tryon WW, Williams R. Fully Proportional Actigraphy: A New Instrument. Behavior Research Methods Instruments and Computers. 1996;28(3):392-403.

[64] De Vries SI, Van Hirtum HW, Bakker I, Hopman-Rock M, Hirasing RA, Van Mechelen W. Validity and reproducibility of motion sensors in youth: a systematic update. Medicine and science in sports and exercise. 2009 Apr;41(4):818-27.
[65] Esliger DW, Tremblay MS. Physical activity and inactivity profiling: the next generation. Canadian journal of public health. 2007;98 Suppl 2:S195-207.
[66] Chen KY, Bassett DR, Jr. The technology of accelerometry-based activity monitors: current and future. Medicine and science in sports and exercise. 2005 Nov;37(11 Suppl):S490-500.
[67] Brage S, Brage N, Wedderkopp N, Froberg K. Reliability and Validity of the Computer Science and Applications Accelerometer in a Mechanical Setting. Measurement in Physical Education and Exercise Science. 2003;7(2):101-19.
[68] Esliger DW, Tremblay MS. Technical reliability assessment of three accelerometer models in a mechanical setup. Medicine and science in sports and exercise. 2006 Dec;38(12):2173-81.
[69] Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989 Mar;45(1):255-68.
[70] Edwards P, Roberts I. Population adiposity and climate change. International journal of epidemiology. 2009 Aug;38(4):1137-40.
[71] Vanhelst J, Mikulovic J, Bui-Xuan G, Dieu O, Blondeau T, Fardy P, et al. Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. BMC research notes. 2012 Apr 25;5(1):187.
[72] van Hees VT, Pias M, Taherian S, Brage S, Ekelund U. A method to compare new and traditional accelerometry data in physical activity monitoring. 2nd IEEE International WoWMoM Workshop on Interdisciplinary Research on E-Health Services and Systems (IREHSS); 2010 14-17 June 2010 Montreal, Canada; 2010. p. 1-6.
[73] Wareham NJ, Brage S, Franks PW, Abbott RA. Physical activity and insulin resistance. In: Kumar S, O'Rahilly S, eds. *Insulin Resistance: Insulin action and its disturbances in disease*: Wiley 2005:317-400.
[74] Hales CN, Barker DJ. The thrifty phenotype hypothesis. British medical bulletin. 2001;60:5-20.
[75] Poudevigne MS, O'Connor PJ. A review of physical activity patterns in pregnant women and their relationship to psychological health. Sports medicine (Auckland, NZ. 2006;36(1):19-38.
[76] Lof M, Forsum E. Activity pattern and energy expenditure due to physical activity before and during pregnancy in healthy Swedish women. The British journal of nutrition. 2006 Feb;95(2):296-302.
[77] Hayes C, Kriska A. Role of physical activity in diabetes management and prevention. Journal of the American Dietetic Association. 2008 Apr;108(4 Suppl 1):S19-23.
[78] Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Medicine and science in sports and exercise. 2005 Nov;37(11 Suppl):S531-43.
[79] Swartz AM, Strath SJ, Bassett DR, Jr., O'Brien WL, King GA, Ainsworth BE. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. Medicine and science in sports and exercise. 2000 Sep;32(9 Suppl):S450-6.
[80] Choi L, Chen KY, Acra SA, Buchowski MS. Distributed Lag and Spline Modeling for Predicting Energy Expenditure from Accelerometry in Youth. J Appl Physiol. 2009 Dec 3.

[81] Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA Accelerometer. Medicine and science in sports and exercise. (in press) Nov 11.
[82] Cook DG, Shaper AG, MacFarlane PW. Using the WHO (Rose) angina questionnaire in cardiovascular epidemiology. International journal of epidemiology. 1989 Sep;18(3):607-13.
[83] Phan DH, Bonnet S, Guillemaud R, Castelli E, Pham Thi NY. Estimation of respiratory waveform and heart rate using an accelerometer. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:4916-9.
[84] Hoffman DJ, Sawaya AL, Coward WA, Wright A, Martins PA, de Nascimento C, et al. Energy expenditure of stunted and nonstunted boys and girls living in the shantytowns of Sao Paulo, Brazil. The American journal of clinical nutrition. 2000 Oct;72(4):1025-31.
[85] Verweij PJ, Van Houwelingen HC. Cross-validation in survival analysis. Stat Med. 1993 Dec 30;12(24):2305-14.
[86] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986 Feb 8;1(8476):307-10.
[87] Prentice AM. Stable isotopic methods for measuring energy expenditure. Applications of the doubly-labelled-water (2H2(18)O) method in free-living adults. The Proceedings of the Nutrition Society. 1988 Sep;47(3):259-68.
[88] Plasqui G, Joosen AM, Kester AD, Goris AH, Westerterp KR. Measuring free-living energy expenditure and physical activity with triaxial accelerometry. Obesity research. 2005 Aug;13(8):1363-9.
[89] Masse LC, Fulton JE, Watson KL, Mahar MT, Meyers MC, Wong WW. Influence of body composition on physical activity validation studies using doubly labeled water. J Appl Physiol. 2004 Apr;96(4):1357-64.
[90] Moe-Nilssen R, Helbostad JL. Estimation of gait cycle characteristics by trunk accelerometry. Journal of biomechanics. 2004 Jan;37(1):121-6.
[91] Aminian K, Robert P, Buchser EE, Rutschmann B, Hayoz D, Depairon M. Physical activity monitoring based on accelerometry: validation and comparison with video observation. Medical & biological engineering & computing. 1999 May;37(3):304-8.
[92] Corder K, Brage S, Wright A, Ramachandran A, Snehalatha C, Yamuna A, et al. Physical activity energy expenditure of adolescents in India. Obesity (Silver Spring, Md. 2011 Nov;18(11):2212-9.
[93] Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. Obesity (Silver Spring, Md. 2010 Sep;18(9):1845-51.
[94] Adams SA, Matthews CE, Ebbeling CB, Moore CG, Cunningham JE, Fulton J, et al. The effect of social desirability and social approval on self-reports of physical activity. American journal of epidemiology. 2005 Feb 15;161(4):389-98.
[95] Corder K, Brage S, Wright A, Ramachandran A, Snehalatha C, Yamuna A, et al. Physical activity energy expenditure of adolescents in India. Obesity (Silver Spring, Md. (in press) Nov;18(11):2212-9.
[96] Hagstromer M, Troiano RP, Sjostrom M, Berrigan D. Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. American journal of epidemiology. 2010 May 15;171(10):1055-64.
[97] Colley RC, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian children and youth: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health reports / Statistics Canada, Canadian Centre

for Health Information = Rapports sur la sante / Statistique Canada, Centre canadien d'information sur la sante. 2010 Mar;22(1):15-23.
[98] Bouten CV, Sauren AA, Verduin M, Janssen JD. Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. Medical & biological engineering & computing. 1997 Jan;35(1):50-6.
[99] Shepard ELC, Wilson RP, Halsey LG, Quintana F, Gomez Laich A, Gleiss AC, et al. Derivation of body motion via appropriate smoothing of acceleration data. Aquatic Biology. 2008;4, Numb 3:235-41.
[100] Roetenberg D, Slycke PJ, Veltink PH. Ambulatory Position and Orientation Tracking Fusing Magnetic and Inertial Sensing. Ieee Transactions on Biomedical Engineering Bme. 2007 Numb 5;54:883-90.
[101] Sabatini AM. Quaternion-Based Extended Kalman Filter for Determining Orientation by Inertial and Magnetic Sensing. Ieee Transactions on Biomedical Engineering Bme. 2006 Numb 7;53:1346-56.
[102] Yun X, Lizarraga M, Bachmann ER, McGhee RB. An Improved Quaternion-Based Kalman Filter for Real-Time Tracking of Rigid Body Orientation. Intelligent robots and systems; 2003 IEEE/RSJ international conference on intelligent robots and systems (IROS 2003); 2003 Oct; Las Vegas, NV: Ieee; 2003. p. 1074-9.
[103] Piazzi A, Visioli A. An interval algorithm for minimum-jerk trajectory planning of robot manipulators. Decision and control; 1997 Dec; Diego; CA: Ieee; 1997. p. 1924-7.
[104] Kyriakopoulos KJ, Saridis GN. Minimum jerk path generation. IEEE International Conference on Robotics and Automation; 1988 24-29 Apr 1988 Philadelphia, PA , USA 1988.
[105] van Hees VT, Renstrom F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PloS one. 2011;6(7):e22922.
[106] Gradmark A, Pomeroy J, Renstrom F, Steiginga S, Persson M, Wright A, et al. Physical activity, sedentary behaviors, and estimated insulin sensitivity and secretion in pregnant and non-pregnant women. BMC pregnancy and childbirth. 2011;11:44.
[107] Rothney MP, Neumann M, Beziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. J Appl Physiol. 2007 Oct;103(4):1419-27.
[108] Intille SS, Tapia EM, Rondoni J, Beaudin J, Kukla C, Agarwal S, et al. Tools for Studying Behavior and Technology in Natural Settings. Lecture Notes in Computer Science. 2003(2864):157-74.
[109] Aminian K, Najafi B. Capturing human motion using body fixed sensors: outdoor measurement and clinical applications. Computer Animation and Virtual Worlds. 2004 Numb 2;15:79-94.
[110] Mathie MJ, Coster AC, Lovell NH, Celler BG. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiological measurement. 2004 Apr;25(2):R1-20.
[111] Knight JF, Bristow HW, Anastopoulou S, Baber C, Schwirtz A, Arvanitis TN. Uses of accelerometer data collected from a wearable system. Personal and Ubiquitous Computing. 2007;11(2):117-32.
[112] Jovanov E, Milenkovic A, Otto C, de Groen PC. A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. Journal of NeuroEngineering and Rehabilitation. 2005;2:6.
[113] Goodwin MS, Velicer WF, Intille SS. Telemetric monitoring in the behavior sciences. Behavior Research Methods. 2008 Numb 1;40:328-41.

[114] Godfrey A, Conway R, Meagher D, G OL. Direct measurement of human movement by accelerometry. Medical engineering & physics. 2008 Dec;30(10):1364-86.
[115] Preece SJ, Goulermas JY, Kenney LP, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors--a review of classification techniques. Physiological measurement. 2009 Apr;30(4):R1-33.
[116] Halsey LG, Shepard EL, Wilson RP. Assessing the development and application of the accelerometry technique for estimating energy expenditure. Comparative biochemistry and physiology. 2011 Mar;158(3):305-14.
[117] Veltink PH, Bussmann HBI, Koelma F, Franken HM. The Feasibility of Posture and Movement Detection by Accelerometry. In: Szeto AYJ, Rangayyan RM, editors. Pt 3; Biomechanics, rehabilitation, electrical phenomena, biomaterials; 1993 Oct; San Diego; CA: Ieee; 1993. p. 1230-1.
[118] Bussmann JBJ, Veltink PH, Koelma F, van Lummel RC, Stam HJ. Ambulatory monitoring of mobility-related activities: the initial phase of the development of an activity monitor. European Journal of Physical Medicine and Rehabilitation. 1995;5:2-7.
[119] Kern N, Schiele B, Schmidt A. Multi-sensor Activity Context Detection for Wearable Computing. Lecture Notes in Computer Science. 2003(2875):220-34.
[120] Randell C, Muller H. Context Awareness by Analysing Accelerometer Data. Wearable computers; 2000 Oct; Atlanta, GA: IEEE Computer Society; 2000. p. 175-6.
[121] Van Laerhoven K, Cakmakci O. What Shall We Teach Our Pants? Wearable computers; 2000 Oct; Atlanta, GA: IEEE Computer Society; 2000. p. 77-86.
[122] Bouten CV, Koekkoek KT, Verduin M, Kodde R, Janssen JD. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE transactions on bio-medical engineering. 1997 Mar;44(3):136-47.
[123] Park HD, Lee KJ, Lee SH, Hong SY, Kim YH. A Study on the Activity Classification Using a Triaxial Accelerometer. IEEE Engineering in Medicine and Biology Society; a new beginning for human health; 2003 Sep; Cancun, Mexico: Ieee; 2003. p. 2941-3.
[124] Makikawa M, Murakami D. Development of an Ambulatory Physical Activity and Behavior Map Monitoring System. IEEE: Engineering in Medicine and Biology Society; 1996 Oct; Amsterdam: Ieee; 1996. p. 71-2.
[125] Sekine M, Tamura T, Ogawa M, Togawa T. Classification of Acceleration Waveform in A Continous Walking Record. In: Chang HK, Zhang YT, editors. IEEE: Engineering in Medicine and Biology Society; 1998 Nov; Hong Kong: Piscataway NJ; 1998. p. 1523-6.
[126] Troped PJ, Oliveira MS, Matthews CE, Cromley EK, Melly SJ, Craig BA. Prediction of Activity Mode with Global Positioning System and Accelerometer Data. Medicine and science in sports and exercise. 2008 Numb 5;40:972-8.
[127] Yun X, Bachmann ER. Design, Implementation, and Experimental Results of a Quaternion-Based Kalman Filter for Human Body Motion Tracking. Ieee Transactions on Robotics. 2006 Numb 6;22:1216-27.
[128] Nawab SH, Roy SH, De Luca CJ. Functional Activity Monitoring From Wearable-Sensor Data. IEEE Engineering in Medicine and Biology Society; 2004 Sep; San Francisco, CA: Piscataway N.J.:; 2004. p. 979-82.
[129] Pawar T, Chaudhuri S, Duttagupta SP. Body Movement Activity Recognition for Ambulatory Cardiac Monitoring. Ieee Transactions on Biomedical Engineering Bme. 2007 Numb 5;54:874-82.

[130] GENEActive. 2012 [cited 2012 16th of May]; Available from: http://www.geneactiv.co.uk/
[131] Nguyen A, Moore D, McCowan I. Unsupervised clustering of free-living human activities using ambulatory accelerometry. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:4895-8.
[132] Shannon CE. Communication in the Presence of Noise (Classic Paper). Proceedings- Ieee. 1998;86(2):447-57.
[133] Bonomi AG, Goris AH, Yin B, Westerterp KR. Detection of type, duration, and intensity of physical activity using an accelerometer. Medicine and science in sports and exercise. 2009 Sep;41(9):1770-7.
[134] Yuwono M, Moulton BD, Su SW, Celler BG, Nguyen HT. Unsupervised machine-learning method for improving the performance of ambulatory fall-detection systems. Biomedical engineering online. 2012;11:9.
[135] Atallah L, Lo B, King R, Yang GZ. Sensor Positioning for Activity Recognition Using Wearable Accelerometers. Ieee Transactions on Biomedical Circuits and Systems. 2011;5, Numb 4:320-9.
[136] Gyllensten IC, Bonomi AG. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. IEEE transactions on bio-medical engineering. 2011 Sep;58(9):2656-63.
[137] Nelson B. Data sharing: Empty archives. Nature. 2009 Sep 10;461(7261):160-3.
[138] Actilife 6 data management solutations. [cited 2012 25th April]; Available from: www.theactigraph.com
[139] MOVEeCloud. [cited 2012 25th April]; Available from: http://moveecloud.com
[140] Mathie MJ, Coster ACF, Lovell NH, Celler BG. Detection of daily physical activities using a triaxial accelerometer. Medical and Biological Engineering and Computing. 2003;41(3):296-301.
[141] Freedson PS, Lyden K, Kozey-Keadle S, Staudenmayer J. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: validation on an independent sample. J Appl Physiol. 2011 Dec;111(6):1804-12.
[142] Ruch N, Rumo M, Mader U. Recognition of activities in children by two uniaxial accelerometers in free-living conditions. European journal of applied physiology. 2011 Aug;111(8):1917-27.
[143] Parkka J, Ermes M, Antila K, van Gils M, Manttari A, Nieminen H. Estimating intensity of physical activity: a comparison of wearable accelerometer and gyro sensors and 3 sensor locations. Conf Proc IEEE Eng Med Biol Soc. 2007;2007:1511-4.
[144] Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971 Mar;9(1):97-113.
[145] Swedish time use survey. 2010 [cited 2011 18th December]; Available from: http://www.scb.se/Pages/Product____12211.aspx
[146] American time use survey. 2010 [cited 2011 18th December]; Available from: http://www.bls.gov/tus/current/household.htm
[147] Time use of Canadians. 2010 [cited 2011 18th December]; Available from: http://publications.gc.ca/site/eng/393062/publication.html
[148] Ridgers ND, Salmon J, Ridley K, O'Connell E, Arundell L, Timperio A. Agreement between activPAL and ActiGraph for assessing children's sedentary time. The international journal of behavioral nutrition and physical activity. 2012;9:15.

[149] Zhang K, Werner P, Sun M, Pi-Sunyer FX, Boozer CN. Measurement of human daily physical activity. Obesity research. 2003 Jan;11(1):33-40.
[150] Biddle SJ, Gorely T, Stensel DJ. Health-enhancing physical activity and sedentary behaviour in children and adolescents. J Sports Sci. 2004 Aug;22(8):679-701.
[151] Levine JA, McCrady SK, Lanningham-Foster LM, Kane PH, Foster RC, Manohar CU. The role of free-living daily walking in human weight gain and obesity. Diabetes. 2008 Mar;57(3):548-54.
[152] Morgan PJ, Okely AD, Cliff DP, Jones RA, Baur LA. Correlates of objectively measured physical activity in obese children. Obesity (Silver Spring). 2008 Dec;16(12):2634-41.
[153] Aadahl M, Kjaer M, Jorgensen T. Associations between overall physical activity level and cardiovascular risk factors in an adult population. Eur J Epidemiol. 2007;22(6):369-78.
[154] Rothstein JM. Objective versus subjective: kudzu terminology. Phys Ther. 1996 Aug;76(8):814-6.
[155] Trost SG. Objective measurement of physical activity in youth: current issues, future directions. Exerc Sport Sci Rev. 2001;29(1):32-6.
[156] Schutz Y, Weinsier RL, Hunter GR. Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures. Obes Res. 2001 Jun;9(6):368-79.
[157] Tudor-Locke C, McClain JJ, Hart TL, Sisson SB, Washington TL. Pedometry methods for assessing free-living youth. Res Q Exerc Sport. 2009 Jun;80(2):175-84.
[158] Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. Arch Dis Child. 2008 Jul;93(7):614-9.
[159] Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans Inf Technol Biomed. 2006 Jan;10(1):156-67.
[160] Mathie MJ, Celler BG, Lovell NH, Coster AC. Classification of basic daily movements using a triaxial accelerometer. Med Biol Eng Comput. 2004 Sep;42(5):679-87.
[161] van Hees VT, van Lummel RC, Westerterp KR. Estimating activity-related energy expenditure under sedentary conditions using a tri-axial seismic accelerometer. Obesity (Silver Spring). 2009 Jun;17(6):1287-92.
[162] Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol. 1949 Aug;109(1-2):1-9.
[163] Calabro MA, Welk GJ, Eisenmann JC. Validation of the SenseWear Pro Armband algorithms in children. Med Sci Sports Exerc. 2009 Sep;41(9):1714-20.
[164] Arvidsson D, Slinde F, Larsson S, Hulthen L. Energy cost in children assessed by multisensor activity monitors. Med Sci Sports Exerc. 2009 Mar;41(3):603-11.
[165] Olds TS, Ferrar KE, Schranz NK, Maher CA. Obese adolescents are less active than their normal-weight peers, but wherein lies the difference? J Adolesc Health. 2011 Feb;48(2):189-95.
[166] Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Aspects of activity behavior as a determinant of the physical activity level. Scand J Med Sci Sports. [Epub ahead of print] Jun 1.
[167] Wu S, Cohen D, Shi Y, Pearson M, Sturm R. Economic analysis of physical activity interventions. Am J Prev Med. 2011 Feb;40(2):149-58.

[168] Rowlands AV, Stiles VH. Accelerometer counts and raw acceleration output in relation to mechanical loading. Journal of biomechanics. 2012 Feb 2;45(3):448-54.
[169] Ince DC, Hatton L, Graham-Cumming J. The case for open computer programs. Nature. 2012 Feb 23;482(7386):485-8.