Statistical modeling of physical activity based on accelerometer data

Dem Fachbereich 03 Mathematik und Informatik der
Universität Bremen
zur Erlangung des akademischen Grades eines
Dr. rer. nat.

eingereichte Dissertation

von
Dipl.-Math. Norman Wirsik
aus
Aurich

Datum der Einreichung: 2016/01/18

Erstgutachterin: Prof. Dr. rer. nat. Iris Pigeot-Kübler
Zweitgutachter: Prof. Dr. rer. nat. Wolfgang Ahrens

Tag der mündlichen Prüfung: 2016/03/21

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich die letzten Jahre auf meinem Weg zur Promotion begleitet und unterstützt haben.
Mein größter Dank gilt meiner Doktormutter Iris Pigeot. Erst ihre beharrliche Unterstützung und Motivation hat diese Arbeit zustande kommen lassen. Egal wie voll ihr Terminplan ist, sie findet immer Zeit für ihre Doktoranden, ist Kritikerin, Ideenlieferantin und versetzt Berge, um ihren Doktoranden zu helfen. Korrektur lesen erledigt sie sprichwörtlich über Nacht. Tausend Dank!
Weiterhin gilt mein herzlicher Dank Wolfgang Ahrens, der nicht nur mit innovativen Fragen und Anregungen unsere Paper verbessert hat, sondern auch bereit ist, diese Arbeit als Zweitgutachter zu bewerten.
Ein großes Dankeschön auch den Mitgliedern meiner Fachgruppe, Hermann Pohlabeln, Timm Intemann und Christoph Buck sowie Claudia Börnhorst. Mit ihnen konnte ich so manches Problem diskutieren und ihre Ratschläge waren sehr hilfreich. Besonderer Dank geht zudem an Timm Intemann, für seinen geduldigen Peer-Review und seine Hilfe bei all meinen R-Problemen.
Ole Sprengeler hat maßgeblich mein Verständnis von körperlicher Aktivität und Akzelerometern geprägt. Viele Ideen und Fragestellungen dieser Arbeit sind im Diskurs mit ihm entstanden; vielen Dank für die enge Zusammenarbeit.
Natürlich geht auch ein großer Dank an alle Kolleginnen und Kollegen im BIPS und in meiner Abteilung, insbesondere an die Fachgruppe IT und Datenmangement, ohne die es keine Daten zum Auswerten und kein HPC zum Rechnen geben würde. Auch möchte ich mich bei den WiMis der biometrischen Fachgruppen Sigrid Behr, Carmen Dering, Dirk Enders, Ronja Foraita, Moritz Hanke, Svenja Jacobs, Bianca Kollhorst, Tammo Reinders, Walter Schill und Marc Suling bedanken. Auch sie hatten immer ein offenes Ohr und haben meine Klagen geduldig ertragen.
Ebenso möchte ich mich bei allen Ko-Autoren der in dieser Arbeit verwendeten Publikationen bedanken, insbesondere der Oldenburg-Connection Mirko Brandes und Fabian Sobotka.
Nicht zuletzt gilt mein Dank meiner Familie und meinen Freunden, die mich in der Promotionszeit begleitet, unterstützt, ermutigt und ertragen haben. Euch allen danke!

Abstract

Physical activity is generally considered as being beneficial for many health outcomes. Lack of physical activity and increased sedentary behavior are regarded as major risk factors. Therefore physical activity has been in the focus of epidemiological research for a long time.

Subjective methods like standardized physical activity questionnaires are frequently used to assess physical activity in epidemiological research. In recent years, objective methods, like pedometers and accelerometers, have become more common. Accelerometers measure the body acceleration along up to three axes. The acceleration is stored as a numerical quantity, the counts for a certain period of the time (epochs). Counts are thought to be proportional to the intensity of the activity. Accelerometer measurements allow to derive the time a person spent in certain intensity ranges, like sedentary, light, moderate and vigorous.

After a motivation of the research presented in this thesis and a short outline, the concept of physical activity is introduced in Chapter 2, which particularly focuses on the description of the objective measurement of physical activity using accelerometers and pedometers in contrast to subjective measurements like physical activity questionnaires. Some methodological problems regarding the objective and subjective assessment of physical activity are identified and investigated in Chapter 5.

Chapter 3 presents more details on accelerometer measured physical activity. The intensity levels are commonly assigned using count thresholds, the so-called cutpoints. The time spent within one activity range without changing into another is called bout. The cutpoint method is a valid way to analyze accelerometer data under the quite unrealistic assumptions that the state of motion at a point in time is independent of the state of motion a person was in just before and that humans switch from sitting to running and back to sitting within a few seconds.

It is, however, more realistic to assume that human activity behavior consists of a sequence of non-overlapping distinguishable activities, like walking to work, sit at the desk and playing badminton after work that can be represented by a mean intensity level. The recorded accelerometer counts scatter around this mean level. If this holds true, then the cutpoint method leads to considerable

misclassification of the bouts into the wrong intensity levels and hence also to an incorrect estimation of the number of bouts.

In Chapter 4, two novel approaches to better capture physical activity under these assumptions are developed and implemented. The Hidden Markov models (HMM) are stochastic models that allow fitting a Markov chain with a predefined number of activities to the data. This new method is compared to the standard cutpoint method in a simulation study. HMMs require some a priori information that are not verifiable. Therefore, it is desirable to find a way to model physical activity data that does not need any other a priori information. Thus, a regression model is called for that allows to model accelerometer data as a sort of step function with each jump indicating the start of a new activity and the constant interval being the mean intensity level of that activity. Here, expectile regression utilizing the Whittaker smoother with an L_0-penalty is introduced as a second innovative approach, which allows the desired fit. The expectile regression is compared to the cutpoint method and the HMMs by means of Monte-Carlo experiments. Both methods, compared to the cutpoint method, reduce the misclassification rate of counts and the number of identified bouts and therefore present a substantial improvement for modeling accelerometer data to assess physical activity.

Chapter 5 presents the results of four studies on physical activity. In the large European IDEFICS study, accelerometer data were collected from several thousands children. These data are used to describe the physical activity behavior in European children using GAMLSS, which is also introduced in this chapter. A second study exploits the collected activity data of the IDEFICS study to investigate the influence of physical activity and sedentary behavior on high blood pressure in children. The PATREC study is a smaller study in German children and adolescents with a strong methodological focus. Data collected in this study are used to study some problems identified in Chapter 2 regarding objectively and subjectively measured physical activity in different domains of activity. In the fourth study an energy expenditure equation is derived for one pedometer model.

Chapter 6 summarizes and discusses the findings of the previous chapters and ends with an outlook on future research with respect to the assessment of physical activity data in epidemiological studies.

Physical activity, accelerometer data, hidden Markov models, expectile regression, L0-penalty, Whittaker smoother, pattern recognition, physical activity patterns, bout detection, GAMLSS, energy prediction equation

Zusammenfassung

Allgemein geht man davon aus, dass körperliche Aktivität einen positiven Einfluss auf viele Erkrankungen und respiratorische Fitness hat. Bewegungsmangel und sitzendes Verhalten gelten als Hauptrisikofaktoren. Daher steht körperliche Aktivität seit langer Zeit im Fokus epidemiologischer Forschung.
Typischerweise werden subjektive Methoden wie standardisierte Fragebögen zur Erfassung von körperlicher Aktivität großflächig eingesetzt. Seit einigen Jahren werden vermehrt Akzelerometer und Pedometer als objektive Methoden verwendet. Akzelerometer messen die Beschleunigung des Körpers entlang bis zu drei Achsen. Die Beschleunigung wird als natürliche Zahl, dem sogenannten Count, für eine bestimmte Zeitdauer (Epoche) im Gerät gespeichert. Es wird angenommen, dass diese Counts proportional zur Aktivitätsintensität sind. Mit Akzelerometermessungen kann die Zeit, die eine Person in den Intensitätsbereichen sitzend, leicht, moderat und stark verbracht hat, bestimmt werden.
Nach einer Motivation des Themas und einer kurzen Übersicht über die Arbeit wird in Kapitel 2 das Konzept von körperlicher Aktivität vorgestellt und objektiven Methoden zur Erfassung von körperlicher Aktivität werden subjektiven Methoden gegenübergestellt. Hieraus ergeben sich einige methodische Fragestellungen, die im weiteren Verlauf in Kapitel 5 untersucht werden.
In Kapitel 3 werden weitere Details zur Messung von körperlicher Aktivität mit Akzelerometern beschrieben. Intensitätsbereiche der Counts werden üblicherweise anhand von Schwellwerten zugeordnet. Dabei wird die Zeit, die eine Person in einem Intensitätsbereich verbringt, ohne in einen anderen zu wechseln, Bout genannt. Diese Schwellwertmethode ist nur unter den unrealistischen Annahmen, dass der Bewegungszustand zu einem bestimmten Zeitpunkt unabhängig vom vorangegangen Bewegungszustand ist und dass Menschen ihren Bewegungszustand üblicherweise innerhalb von Sekunden vom Sitzen zum Rennen und wieder zurück zum Sitzen wechseln, eine valide Möglichkeit, Akzelerometerdaten zu analysieren.
Dahingegen ist es wesentlich realistischer anzunehmen, dass körperliche Aktivität die diskrete Abfolge von unterscheidbaren Aktivitäten ist, wie zu Fuß zur Arbeit zu gehen, am Schreibtisch sitzen und nach der Arbeit Badminton

spielen. Die Aktivitäten können dabei durch ein mittleres Intensitätsniveau abgebildet werden und die gemessenen Akzelerometercounts streuen um dieses mittlere Niveau. Unter dieser Annahme führt die Schwellwertmethode zu erheblicher Missklassifikation der Counts in die falschen Intensitätsbereiche und damit schließlich zu einer verfälschten Schätzung der Anzahl von Bouts.

In Kapitel 4 werden zwei innovative Methoden, die diese Annahmen berücksichtigen, entwickelt und implementiert. Hidden Markov Modelle (HMM) sind stochastische Modelle, die es ermöglichen, eine Markovkette mit einer vorher definierten Anzahl von Aktivitäten an die Daten anzupassen. Diese neue Methode wird mit der üblichen Schwellwertmethode in einer Simulationsstudie verglichen. HMMs benötigen einige a priori Annahmen, die nicht überprüfbar sind. Daher ist es wünschenswert, einen Modellierung von körperlicher Aktivität zu finden, die ohne solche Annahmen auskommt. Es wird also ein Regressionsmodell gesucht, das es erlaubt, Akzelerometerdaten als eine Art Stufenfunktion zu modellieren, bei der jeder Sprung den Beginn einer neuen Aktivität anzeigt und das konstante Intervall das mittlere Intensitätsniveau der Aktivität darstellt. Hierzu wird Expektilregression unter Verwendung des Whittakerglätters mit L_0-Strafterm als zweite innovative Methode vorgestellt und ebenfalls mit der Schwellwertmethode und HMMs in einer weiteren Simulationsstudie verglichen. Beide Methoden reduzieren im Vergleich zur Schwellwertmethode die Missklassifikationrate der Counts und die Anzahl der erkannten Bouts und stellen somit eine substantielle Verbesserung der Modellierung von Akzelerometerdaten dar.

In Kapitel 5 werden die Ergebnisse von vier empirischen Studien zur körperlichen Aktivität vorgestellt. In der großen europäischen IDEFICS-Studie wurden von mehreren tausend Kindern Akzelerometerdaten gesammelt. Diese Daten werden genutzt, um das Bewegungsverhalten von europäischen Kindern mittels GAMLSS, das ebenfalls in diesem Kapitel eingeführt wird, zu beschreiben. Eine weitere Studie nutzt die Daten der IDEFICS-Studie, um den Einfluss von körperlicher Aktivität und sitzendem Verhalten auf kindlichen Bluthochdruck zu untersuchen. Die PATREC-Studie untersucht das Bewegungsverhalten deutscher Kinder und Jugendlicher mit einem besonderen methodischen Fokus. Die hier gesammelten Daten dienen zur Untersuchung der in Kapitel 2 aufgebrachten Fragen zu objektiven und subjektiven Erfassungsmethoden

in unterschiedlichen Aktivitätsdomänen. In der vierten Studie wird eine Gleichung zur Schätzung des Energieumsatzes für ein Pedometermodell bestimmt. In Kapitel 6 werden die Resultate zusammengefasst und diskutiert sowie ein Ausblick auf zukünftige Forschung im Bereich der Erfassung von körperlicher Aktivität in epidemiologischen Studien gegeben.

Körperliche Aktivität, Akzelerometerdaten, Hidden-Markov-Modelle, Expektilregression, L0-Strafterm, Whittaker-Glätter, Mustererkennung, körperliche Aktivitätsmuster, Bouterkennung, GAMLSS, Energievorhersagegleichung

Contents

List of Figures

List of Tables

List of Abbreviations

AEE	Activity-induced energy expenditure
AIC	Akaike information criterion
BIC	Bayesian information criterion
BCCG	Box-Cox Cole and Green distribution
BCPE	Box-Cox power exponential distribution
BEE	Basal energy expenditure
BF	Body fat
BLUP	Best linear unbiased predictor
BMI	Body mass index
CI	95% confidence interval
CO_2	Carbon dioxide
CPM	Counts per minute
CV	Coefficient of variation
DIT	Dietary induced thermogenesis
DLW	Doubly labeled water
FFM	Fat free mass
FMI	Fat mass index
GAM	Generalized additive models

GAMLSS	Generalized additive models for location scale and shape
GIS	Geographical information system
GPS	Global positioning system
GLM	Generalized linear model
GPS	Global positioning system
HBP	High blood pressure
HMM	Hidden Markov model
HPC	High performing computing cluster
IPAQ	International physical activity questionnaire
LAWS	Least asymmetrically weighted squares
LPA	Light physical activity
MAPE	Mean absolute percentage error
MCR	Misclassification rate
MET	Metabolic equivalent of task
MLM	Multi level model
MPA	Moderate physical activity
MVPA	Moderate-to-vigorous physical activity
OLS	Ordinary least squares
REE	Resting energy expenditure
RMR	Resting metabolic rate
RMSE	Root mean square error
RR	Relative risk
SD	Standard deviation

SED Sedentary behavior

SDS Standard deviation score

TEE Total energy expenditure

TEF Thermic effect of food

TPA Total physical activity

VM Vector magnitude

VPA Vigorous physical activity

VO_2 Oxygen intake

z-FMI z-score of the fat mass index

z-waist z-score of the waist circumference

Chapter 1

Introduction

1.1 Motivation

Physical activity is generally considered as being beneficial for many health outcomes. Lack of physical activity and increased sedentary behavior are regarded as major risk factors. Therefore physical activity has been in the focus of epidemiological research for a long time. Physical activity is typically described by the four *dimensions*, (1) frequency, (2) duration, (3) intensity and (4) type and is performed in so called *domains*, which typically include leisure time physical activity, occupational physical activity, transportation activity and activities performed at home.

In order to be able to properly investigate the association of physical activity with different outcomes, a good exposure measurement is required. In epidemiological research subjective methods like standardized physical activity questionnaires are broadly used. The advantages of the subjective assessment are its low costs, simple logistics and its broad application with an accompanying "validation".

In recent years, objective methods, like pedometers and accelerometers, have become more common. Accelerometers measure the body acceleration along up to three axes. The acceleration is stored as a numeric quantity, the *counts* for a certain period of the time (*epochs*). Counts are thought to be proportional to the intensity of the activity. Accelerometer measurements allow to derive the time a person spent in certain intensity ranges, like sedentary, light, moderate and vigorous. Physical activity is frequently summarized as minutes per day

spent in these activity ranges, hence dimensions (1) to (3) can be assessed.
The intensity levels are commonly assigned using count thresholds, the so-called *cutpoints*. The time spent within one activity range without changing into another is called *bout*. The cutpoint method is a valid way to analyze accelerometer data under the quite unrealistic assumptions that the state of motion at a point in time is independent of the state of motion a person was in just before and that humans switch from sitting to running and back to sitting within a few seconds.
It is, however, more realistic to assume that human activity behavior consists of a sequence of non-overlapping distinguishable activities, like walking to work, sit at the desk and playing badminton after work that can be represented by a mean intensity level. The recorded accelerometer counts scatter around this mean level. If this holds true, the application of the simple cutpoint method will lead to considerable misclassification of the counts and hence to an invalid exposure measurement. Additionally the number of bouts will be overestimated, as by misclassifying the count to the wrong intensity range, a new count is started by definition.
This thesis focuses on how to improve modeling accelerometer data to better reflect real-life behavior and also investigates methodological issues regarding the comparison of subjective and objective measurement of physical activity. The thesis also presents results from studies on physical activity that describe the physical activity behavior of European children and investigate the impact of sedentary behavior and physical activity on high blood pressure in children. An energy prediction equation for a pedometer model is also derived.

1.2 Outline

This thesis consists of six chapters based on six manuscripts, reprinted in the appendix. Chapter 2 gives an introduction to the concept of physical activity and its assessment in epidemiological studies. Chapter 3 presents more details on accelerometer measured physical activity, how it is commonly analyzed and what disadvantages may occur, given some assumptions on physical activity behavior. Chapter 4 describes and implements two novel approaches to reflect these assumptions, and Chapter 5 presents four papers on empirical studies related to physical activity. The thesis concludes with a detailed discussion in Chapter 6.

Chapter 2 introduces the concept of physical activity and particularly focuses on the description of the objective measurement of physical activity like accelerometers and pedometers in contrast to subjective measurements like physical activity questionnaires, especially with regard to their utilization in epidemiological studies. Some methodological problems regarding the objective and subjective assessment of physical activity are identified and further investigated in a paper on objectively and subjectively measured physical activity in different domains of activity.

Chapter 3 builds on the previous chapter and provides more details on how physical activity is objectively measured using accelerometers, which have become the method of choice in recent years. The data recorded by the devices is described, as well as the typical approach how they are analyzed, i.e. by applying the so-called *cutpoint method*. The assumptions underlying the cutpoint method are quite unrealistic. Under more realistic assumptions, namely that physical activity can be regarded as a sequence of non-overlapping activities with an distinguishable mean intensity, the simple cutpoint method has some serious drawbacks, leading to considerable misclassification. The assumptions are verified by the collection of labeled accelerometer data, where the performed activities are known.

In Chapter 4, two novel approaches to model accelerometer data under the assumptions introduced in Chapter 3 are developed. The hidden Markov models (HMM) are stochastic models that allow fitting a Markov chain to the data based on a predefined number of activities. In a methodological paper this new method is compared to the standard cutpoint method in a simulation study. Expectile regression utilizing the L_0-penalty and the Whittaker smoother are introduced as a second innovative approach. Fitting the 0.5-expectile curve to the data is basically a mean regression. Adding the Whittaker smoother with an L_0-penalty now allows the desired fit accounting for the above assumptions on physical activity behavior. In a second methodological paper the expectile regression is compared to the cutpoint method and the HMMs by means of Monte-Carlo experiments. In order to ensure using simulated data resembling real-life accelerometer data as closely as possible, the simulation was chosen to reflect the collected labeled data.

Chapter 5 presents the results of four studies on physical activity. In the large European IDEFICS study, accelerometer data were collected from several thousand children. These data are used to describe the physical activity behavior in European children using GAMLSS, which is also introduced in this chapter. A second paper based on the IDEFICS study exploits the collected activity data to investigate the influence of physical activity and sedentary behavior on high blood pressure in children. The PATREC study is a smaller study in German children and adolescents with a strong methodological focus. Data collected in this study are used to study some problems identified in Chapter 2 on objectively and subjectively measured physical activity in different domains of activity. In the fourth paper an energy expenditure equation is derived for one pedometer model. The data were collected by Oldenburg sports scientists combining this pedometer model with spirometry.

Chapter 6 summarizes and discusses the findings of the previous chapters and ends with an outlook on future research regarding the assessment of physical activity data in epidemiological studies.

The appendix of this thesis provides reprints of the published papers. The complete paper is presented in case the papers have been published in an open access journal, or if permission for reprint was obtained from the journal. In cases where papers have been just submitted and not yet published, the abstract will be presented.

Chapter 2

Methodological background

This chapter serves as an introduction to the wide spectrum of assessing physical activity in the context of modern epidemiological studies. Different assessment methods are presented and discussed with regard to their application in epidemiological studies. This chapter mostly summarizes results from Trost (2007), Beneke and Leithäuser (2008), Westerterp (2009), and Schmid and Leitzmann (2014).

2.1 Assessment of physical activity

Currently physical inactivity is considered as major risk factor for several health disorders like cancer (McTiernan, 2008), obesity (Kimm et al., 2005), cardiovascular disorders (Lee et al., 2012), muscular skeletal disorders (Janz et al., 2010), as well as mental disorders (Rethorst et al., 2009). "Valid and reliable measures of physical activity are therefore a necessity in studies designed to (1) document the frequency and distribution of physical activity in defined population groups, (2) determine the amount or dose of physical activity required to influence specific health parameters, (3) identify the psychosocial and environmental factors that influence physical activity behavior in youth, and (4) evaluate the efficacy or effectiveness of programs to increase habitual physical activity in youth." (Trost, 2007). "*Physical activity* is defined as any bodily movement produced by skeletal muscle that results in energy expenditure above resting" (Trost, 2007) and should not be confused with *exercise*, as "exercise is a specific type of physical activity that is defined as planned, structured, and repetitive bodily movement done to improve or maintain one or more components of physical fitness." (Trost, 2007). Studies show that the proportion of activity-induced energy expenditure (AEE) of total energy expenditure (TEE) varies between 5% in a subject with minimal activity level to about 45-50% in a subject with high activity level (Westerterp, 2009). Schmid and Leitzmann (2014) state that total energy expenditure typically consists of three components: (1) resting metabolic rate RMR, which is the minimal rate of energy that is required for basic bodily functions, (2) thermic effect of food (TEF) (also known as dietary induced thermogenesis (DIT)), which is the amount of energy required above RMR for processing food and (3) activity-induced energy expenditure. RMR is the main component with approximately 70% of TEE, TEF forms about 10% of TEE and AEE around 20%. Several measurement units are common when measuring physical activity. These include energy expenditure per time unit, e.g. kJ per hour per kg body mass, and metabolic equivalent of task (MET), as rate of oxygen (O_2) consumption. By definition

$$1 \text{ MET} = 3.5 \cdot \frac{\text{mL O}_2}{\text{kg} \cdot \text{min}},$$

which is equivalent to

$$1 \text{ MET} = 1 \frac{\text{kcal}}{\text{kg} \cdot \text{h}} = 4.184 \frac{\text{kJ}}{\text{kg} \cdot \text{h}}.$$

1 MET also roughly corresponds to the energy costs of sitting quietly. MET values range from 0.9 MET while sleeping to 23 MET for running at 22.5km/h. METs are often used to assign activities to *activity ranges*. Consequently 1 - 1.5 METs correspond to sedentary behavior, light intensity activities are those with 1.5 to $<$4 METs, moderate intensity activities are those with 4-6 METs and activities with $>$6 METs are called vigorous intensity activities (Trost et al., 2011). In the case of objective instruments, physical activity is commonly reported as time spent in these activity ranges. Physical activity can be described by four *dimensions*, (1) frequency, (2) duration, (3) intensity and (4) type and is performed in so called *domains*, which typically include leisure time physical activity, occupational physical activity, transportation activity and activities performed at home. Depending on the context and study population, additional domains, like for example in the case of school students physical education, sports clubs or after-school programs, should be added.

The perfect measurement instrument would allow a reliable and valid measurement of physical activity in all dimensions and domains (Trost, 2007). There are many different instruments available, which can be assigned to three categories. Category 1 contains the *reference methods*, or gold-standard. *Objective measurements* and *subjective or self-report methods* form categories 2 and 3. Reference methods measure energy expenditure directly and are used to validate instruments of categories 2 and 3. Validated instruments of category 2 in turn are frequently used to validate methods of category 3 (Beneke and Leithäuser, 2008). All instruments have certain advantages and disadvantages that one has to consider with regard to the question of interest. Trost (2007) and Westerterp (2009) as well as Schmid and Leitzmann (2014) provide overviews and ratings of the different methods, which are now discussed in detail.

2.1.1 Reference methods

Direct observation, indirect calorimetry and *doubly labeled water (DLW)* are considered reference methods for measuring physical activity.

Direct observation Direct observation is one of the first methods to measure physical activity in free-living individuals and is the only method to observe all dimensions and domains of physical activity. Specially trained personnel observes the study subject for a continuous observation period, ranging from a single physical education lesson, to four hours during the course of the day. In pre-defined observation intervals of 3, 10, 15 or 60 seconds, physical activity is recorded either as intensity equivalent within three to eight predefined categories, or as standardized activity, like sitting, running, swimming etc. in combination with an intensity (Beneke and Leithäuser, 2008). On the one hand, direct observation has proven itself to be very flexible and is able to record contextual information like environmental conditions. On the other hand, this method is very labor intensive and observers have to be thoroughly trained. In addition, one can argue that their presence will influence the behavior of the subject (reactivity effect) and that judging activity intensity is highly subjective, although studies have shown high inter-observer reliability (Trost, 2007; Westerterp, 2009). Another point of criticism is the fact that observations are only done for a relatively short period of time compared to other measurement methods and that results are therefore only valid for the observed setting. This is particularly true, if, for example, only a single physical education lesson was used for the observation, as this lesson can be hardly regarded as representative for the general behavior of a student. This disadvantage in combination with the immense need for personnel and the accompanying huge costs prohibit using this instrument in large cohort studies (Trost, 2007; Beneke and Leithäuser, 2008; Westerterp, 2009).

Indirect calorimetry This method is based on the oxygen intake (VO_2) and carbon dioxide (CO_2) production and calculates the energy expenditure using the measured amounts of breathing gas. This method has been used since the 1920s, with first devices being bulky and hence stationary. Nowadays, portable devices (spirometers) are available, allowing vigorous physical activity without too much interference, although mouthpieces and masks do cause discomfort to a certain degree and might not be tolerated by the subject. Particularly when dealing with children, additional weight burden exceeding 6% of the body mass will influence movement economy negatively and will lead to considerably increased energy expenditure. This instrument is frequently used to validate methods of category 2 and 3 (Beneke and Leithäuser, 2008). Indirect calorimetry is relatively expensive and burdensome for the participants, especially for longer periods of time, which are needed for the assessment of habitual physical activity. Therefore this method is not a feasible option in large scale epidemiological cohorts (Schmid and Leitzmann, 2014).

Doubly labeled water This method is considered as the gold standard for measuring total energy expenditure (TEE) in free-living subjects over a period of one to four weeks. Water containing doses of two stable water isotopes, ^2H_2O (deuterium-labeled water) and $H_2^{18}O$ (oxygen-18-labeled water), is given to the subject at specific points in time. The isotopes are naturally occurring and have no known toxicity. The deuterium-labeled water is only released through the body's water pool (urine, sweat, evaporative losses), while the oxygen-18-labeled water is additionally lost via the bicarbonate pool. Dissolved CO_2, which is the end product of metabolism, enters the blood stream and is exhaled. Samples of body fluids (urine, blood, saliva) are analyzed by mass spectrometry and the rates for the disappearance of the isotopes are determined. At least three samples are required. One baseline sample before DLW application, one after the DLW has equilibrated with the body water and one after one to four weeks. The measured CO_2 production can be converted to TEE and if basel energy expenditure (BEE) is known, either by separate measurement or estimation, activity-induced energy expenditure (AEE) can be calculated as

$$AEE = 0.9 \times TEE - BEE.$$

Although gold-standard for measuring TEE, this method has some considerable disadvantages that prohibit its use in large cohort studies. This method requires exact adherence to the study protocol by the subject. Information on the pattern of physical activity, such as energy spent in light, moderate and vigorous physical activity, cannot be derived from this method. The most important limitation of the DLW method is its excessive costs. Therefore DLW is typically used only in relatively small samples and mostly to validate instruments of category 2 (Trost, 2007; Westerterp, 2009). For example, DLW was used in the IDEFICS study (see Section 5.1) to validate accelerometer devices (Ojiambo et al., 2012).

2.1.2 Objective measurements

Heart rate monitoring, pedometry and *accelerometry* are objective methods to measure physical activity in free-living subjects. These methods are validated using one of the above described reference methods and are in turn used to validate methods of category 3. In the literature, usually a so-called “validation coefficient” is reported to assess validity. Often this term refers to Pearson’s and Spearman’s correlation coefficient interchangeably (see Trost, 2007, Table 1). It is obvious that two measurements should be highly correlated, if they are supposed to assess the same dimension, however, this is not sufficient to show validity of one of these instruments. This is especially true, if correlation coefficients ≤ 0.5 that turn out to be significantly different from 0 lead to the conclusion that the investigated instrument is valid, see Bland and Altman (1986) for considerations on the validity of instruments. However, a discussion of the correct interpretation and investigation of validity is beyond the scope of this chapter.

Heart rate monitoring Heart rate monitoring was one of the first objective methods used for the assessment of physical activity. Equations are available, which can be used to estimate daily energy expenditure based on the monitored heart rate. Validation studies using DLW were conducted showing good agreement on group level, but individual differences were large. For this method individual calibration and measurement of VO_2 at rest are needed to determine the so called flex heart rate. It is well known that people with higher physical fitness can perform more intense activities at lower heart rates than persons with low levels of fitness. Other factors like age, body size or emotional stress may also influence the relationship between heart rate and VO_2, as do substances like caffeine and medications like e.g. beta-blockers. Additionally, heart rate lags behind changes in movement and stays elevated after some exhausting activity, although the body is already at rest. Hence it can be suspected that heart rate monitoring is not suitable for measuring sporadic activity patterns that are found e.g. in children (Trost, 2007; Westerterp, 2009).

Pedometry A pedometer is a relatively simple device that registers steps and is quite cost-effective compared to accelerometers and is often used in health promotion programs and in clinical settings where walking is the main type of activity. Pedometers are easy to administer, which allows their use also in large groups of virtually any age. The concept of a "step" is easy to comprehend, therefore pedometers have the potential to promote behavior change, like for example in the "10,000 steps Rockhampton project" (Schmid and Leitzmann, 2014). A major limitation is the inability of the pedometer to record the magnitude/intensity of the activity. Movement above a certain threshold is registered as a step, regardless whether the movement was walking, running or jumping, although, of course, the step frequency allows conclusions regarding speed and thereby intensity. Pedometers can only register walking activities, but do not capture activities like swimming, cycling or weight lifting. Thus, pedometers are supposed to provide valid measurement of the relative amount of physical activity, but they cannot provide information on type of activity, frequency, intensity, or duration (Trost, 2007; Schmid and Leitzmann, 2014). In Section 5.6 an energy prediction equation is derived for one pedometer model, allowing at least to capture the energy expenditure and hence the intensity for walking activities.

Accelerometry In contrast to pedometers, accelerometers are able to measure acceleration in up to three planes. Uni-axial devices register acceleration along a vertical axis, bi-axial devices additionally along the medio-lateral plane and tri-axial devices also along the anterior-posterior plane. These, in the meanwhile relatively inexpensive, devices collect information known as *(impulse-)counts* and provide information on intensity, frequency and duration of physical activity of an individual. Counts represent a device-specific numeric quantity which is generated by the accelerometer for a specific time unit *(epoch)* (e.g. 1 to 60 sec). This quantity is proportional to the intensity of the physical activity performed by the subject. Devices of the first generations had only limited memory. Therefore epochs around 15 seconds to one minute were common, as well as observation times of only a few days. Nowadays, devices have become small, light and robust and are very well tolerated by subjects. Their improved batteries and increased memory now allow high frequency measurements with epoch length of 1-5 seconds over a complete week or more. The sequence of activities during a day is stored as a time series of counts by the device. The most common approach to derive the pattern of physical activity and its energy expenditure is to map these counts to a certain number of sedentary and activity ranges, such as sedentary behavior (SED), light (LPA), moderate (MPA) and vigorous (VPA) physical activity. So the most common measurement unit is minutes (per day) in SED, LPA, MPA, or VPA respectively. The duration of physical activity within the same activity range is known as *bout* and can be easily extracted from a given sequence of counts. A bout is defined as the time period in which the subject remains within one activity range without changing to another. Activity ranges are separated by thresholds known as cutpoints. Cutpoints are available for children (e.g. Evenson et al., 2008; Freedson et al., 2005; Guinhouya et al., 2009a; Pate et al., 2006; Puyau et al., 2002; Treuth et al., 2004; Trost, 2007) and adults (e.g. Freedson et al., 1998; Sasaki et al., 2011; Troiano et al., 2008) to assess the overall time spent in these ranges of physical activity. Alternatively, energy prediction equations (e.g. Crouter et al., 2012) can be used to derive energy expenditure from the accelerometer counts. Numerous validation studies have been performed to date (e.g. Ekelund et al., 2001; Hislop et al., 2012; Ojiambo et al., 2012; Plasqui and Westerterp, 2007) using DLW or indirect calorimetry, as well as direct observation as reference method. Due to these results, accelerometers can be regarded as a valid instrument to assess physical activity. Because of these features, accelerometry is now one of the most frequently used methods for assessing physical activity in free-living subjects. However, accelerometers are not able to register certain activities that are associated with increased energy costs like cycling, swimming, using stairs, carrying heavy objects, or walking uphill. Some people argue that these activities only make up a small proportion of the overall physical activity and therefore this disadvantage is neglectable. Modern tri-axial devices are more sensitive to activities of light intensity and provide better measurements of upper-body movement in activities like rowing and riding a bike. The Euclidean norm is then used to combine the counts along the axes to the *vector magnitude* (VM). Until now, only few cutpoints for VM are available. Another factor that may influence the results of an accelerometer measurement is the place where the device is attached, e.g. foot, hip or arm. Therefore standardization within one study is mandatory. Like other objective methods, accelerometers do not provide contextual information on the domains in which physical activity is performed. To overcome this shortage, oftentimes participants are asked to keep an activity diary, in which non-wearing periods, e.g. swimming, or other times when the accelerometer was not worn are recorded, as well as the beginning and end of certain domains like transportation, being in the work environment, at school etc. (Trost, 2007; Beneke and Leithäuser, 2008; Westerterp, 2009; Schmid and Leitzmann, 2014).

GPS The recent spreading of smart phones with GPS capability as well as standalone GPS trackers now allow to combine information on physical activity with the built environment using geographical information systems (GIS) and by this to investigate the interaction between people's physical activity and their environment. Many smart phones also have built in pedometers and more and more applications to monitor physical activity and exercises are introduced (Schmid and Leitzmann, 2014).

2.1.3 Subjective measurements

All presented methods above are objective in the sense that the subject is not forced to rate his or her own activity behavior. All methods of category 3, which are *proxy report, structured interview, questionnaire* and *activity diary*, require that the subject recalls physical activity from the past and rates/estimates duration and intensity. This is of course highly subjective and therefore all subjective methods are, to some degree, subject to recall bias and social desirability bias.
With the exception of the structured activity interview, all self-report methods are inexpensive and require only minimal personnel resources compared to other methods of assessment and are therefore widely applied in all kinds of (large) studies. A huge number of different physical activity questionnaires exist and their validity and reliability is subject to discussion. Particularly for self-report instruments, the criticism concerning the use and interpretation of correlation coefficients from above holds true.

Proxy report Proxy reports are used when the subject is considered to be unable to understand and/or answer questions concerning his or her physical activity due to e.g. age as it is the case for young children. These proxy reports are based on the assumption that parents or teachers know enough of the behavior pattern of the child to sufficiently answer questions on its behalf. Studies on the validity of this instrument only showed disappointing results (Beneke and Leithäuser, 2008; Verbestel et al., 2015).

Activity diary Subjects are asked to retrospectively indicate their activity performed and its intensity every few minutes (e.g. 1-15 minutes). Resulting estimates are quite good compared to objective measurements, yet some subjects have difficulties to rate their own intensity level and a diary can impose a considerable burden to the subject, especially, if reporting intervals are short (Beneke and Leithäuser, 2008). Recently electronic activity diaries using smart phones have been introduced. After a certain time interval participants are reminded by a signal to record their past physical activity. Using voice recognition the participant's reply is converted to a text form and then assigned to an activity category (Schmid and Leitzmann, 2014).

Structured interview The presence of an interviewer can reduce misunderstandings and incomplete answers on the one hand. On the other hand, direct contact between the interviewer and the subject increases the chances of socially desirable answers. Compared to other self-report instruments higher personnel resources are necessary and it is not clear, whether avoidance of misunderstandings outweighs interviewer induced bias (Beneke and Leithäuser, 2008).

Questionnaire Physical activity questionnaires are probably the most frequently used instrument to assess physical activity. They are easy to use, cheap and many different questionnaires for different target groups and settings are available. In theory, a questionnaire can assess all dimensions and domains of physical activity. In fact, next to direct observation, self-report instruments are the only ones that can provide contextual information. However, questionnaires are subject to considerable recall bias for subjects of all ages, as especially habitual physical activity is challenging to recall and rate retrospectively. Questionnaires tend to underestimate LPA and to overestimate MVPA. This might be explained by the fact that MVPA, like swimming and jogging, are mostly planned exercises and occur in more structured settings like visits to the gym, while low intensity activities, like walking, occur throughout the day and are therefore difficult to assess. One of the most popular physical activity questionnaires is the International Physical Activity Questionnaire (IPAQ)(Craig et al., 2003). It was developed in 1996 and is considered as an established population surveillance tool for the assessment and comparison of physical activity across countries. A long and a short version for e.g. telephone interviews are available and have been translated into more than 20 languages (Schmid and Leitzmann, 2014). Some physical activity questionnaires may not be suitable for all age groups. E.g. they may be unsuitable for children who are younger than 10 years of age. Young children seem to have problems to fully understand the concept of physical activity and they have problems to differentiate between sedentary activities like playing a video game and non-sedentary activities like playing outside and doing household chores (Trost, 2007). Additionally children's activity behavior is characterized by short bouts of activity. In order to deal with this, one could either opt for proxy reports, as mentioned above, or use questionnaires that have been especially designed for children, like the MoMo questionnaire as part of the German KiGGS study (Schmid and Leitzmann, 2014). There are many other questionnaires available (see Trost, 2007, Table 1). This reflects the lack of comparability when trying to measure physical activity. Researcher tend to rather create their own questionnaires than to use existing "validated" ones. This causes problems if one tries to compare results between studies using different questionnaires. As there are many different physical activity questionnaires, there are also many reviews available investigating the validity of these instruments, with varying results. Some certify sufficient validity for self-report instruments, with the exception of younger children (Trost, 2007), while other see rather low validity and reliability when habitual physical activity is measured. Some studies report systematic underestimation, some report overestimation and others report agreement at group level with considerable error on individual level (Westerterp, 2009). Methodological problems of physical activity questionnaires applied to children and adolescents are further investigated in Section 5.5.1. Here, subjective measured physical activity in different domains is compared with accelerometer assessed physical activity.

2.1.4 Observation period

The answer to the question how many days a subject's physical activity should be monitored strongly depends on the research question and the preferred method of assessment. Other considerations may be financial limitations and researchers have to make sure not to choose a monitoring protocol that is overly burdensome to the subjects. As said above, direct observation can be used to measure one's physical activity for a couple of hours, maybe a day or two, due to its limitations. Doubly labeled water, on the contrary, can by design only be used to assess physical activity over the course of one to several weeks. Self-report instruments can be used for arbitrary periods. Because of technological improvements objective instruments can be used for a few weeks, if desired. Some studies tried to calculate wearing days that are necessary to reach a certain degree of reliability. However, the results of these studies are inconclusive. Keeping this in mind and considering the strong likelihood that physical activity behavior will vary between weekdays and weekend days, a

7-day monitor protocol seems to be a reasonable choice (Trost, 2007).

2.1.5 Discussion

When looking at the different options available for measuring physical activity in free-living subjects, it is quite obvious that there does not exist the one instrument of choice. Physical activity is a rather complex concept that consists of four dimensions and several domains. Reference methods like doubly labeled water and indirect calorimetry are excellent for measuring energy expenditure, but neither provide information on the dimensions frequency, intensity, duration and type nor on the domains. In the case of doubly labeled water, the costs of this method prohibit its large scale use in cohorts. If a monitoring period of a complete week is intended, indirect calorimetry also seems to be inappropriate, although the spirometer and its mask/mouthpiece may be tolerated for a couple of hours under "laboratory conditions", it is certainly not feasible to wear this device for a complete week.

Among the objective instruments, accelerometers and pedometers imply the least burden to subjects and are methodologically robust and well tested. Devices have become affordable to allow their use in large-scale field studies. Accelerometers can, in contrast to pedometers, additionally provide information on the dimensions of physical activity. No objective measurement can provide contextual information. This is the special advantage of self-report instruments like questionnaires. They are easy to use, cheap, widely used and, compared to accelerometers, do not require sophisticated logistics. Yet, there are substantial doubts regarding validity, reliability and comparability. These doubts are less pronounced for accelerometers. An accelerometer is a "heartless machinery", that is not tempted to record socially desirable physical activity behavior, it does not forget to register motion and it can judge intensity rather precise, although some activities cannot be registered, which in turn questionnaires can. Accelerometers can also be used in children of all ages, an area of application in which questionnaires reach their limits. As a matter of fact, accelerometers are becoming more and more broadly used in field studies and large cohorts, as their advantages are obvious.

The question is, however, whether accelerometers should be the only measurement of physical activity in such studies. As discussed, contextual information is not recorded. But this might be of particular interest. For example, it may be of interest to learn in which domain most of the physical activity is performed. It would be, of course, interesting to know, whether most physical activity is accumulated during regular activities during the day, or whether there are certain domains, like organized sports activities (e.g. physical education or school programs), or transportation activities with high intensity levels. Such information will be helpful when developing intervention programs that aim to increase physical activity. Thus, it seems reasonable to combine both instruments and their strengths by adding an activity diary, in which date and time of the domains of interest are recorded. This way highly validated and reliable objective measurements can be put into contextual settings of physical activity.

This approach was implemented in the PATREC study described in Section 5.5. Some results that can be obtained from the combined use of objective and subjective measurements are presented in Section 5.5.1.

Chapter 3

Accelerometer measured physical activity

This chapter provides further details on accelerometer measured physical activity. Counts, which are recorded by the accelerometer, are introduced, as well as the commonly used method to analyze them, the cutpoint method. This method is only valid under quite unrealistic assumptions. More realistic assumptions about human physical activity behavior are formulated, which, if true, lead to some serious drawbacks of the cutpoint method. In order to verify these assumptions, labeled accelerometer data were collected. In Chapter 4, two novel approaches will be presented that allow to model accelerometer data taking these assumptions into account.

3.1 Accelerometer counts

As described in Section 2.1.2, accelerometers as an objective measurement of physical activity have become the method of choice to access physical activity in recent years. Modern devices allow high frequency measurements for extended periods of time. The information is stored as a natural number, the so-called (impulse-)*counts* which provide information on intensity and duration of an individual's physical activity. Counts are a device-specific numeric quantity which is recorded for a specific time unit, the *epoch*, which ranges from 1 second in modern devices to 60 seconds in older ones. Counts are thought to be proportional to the intensity of the physical activity performed by the

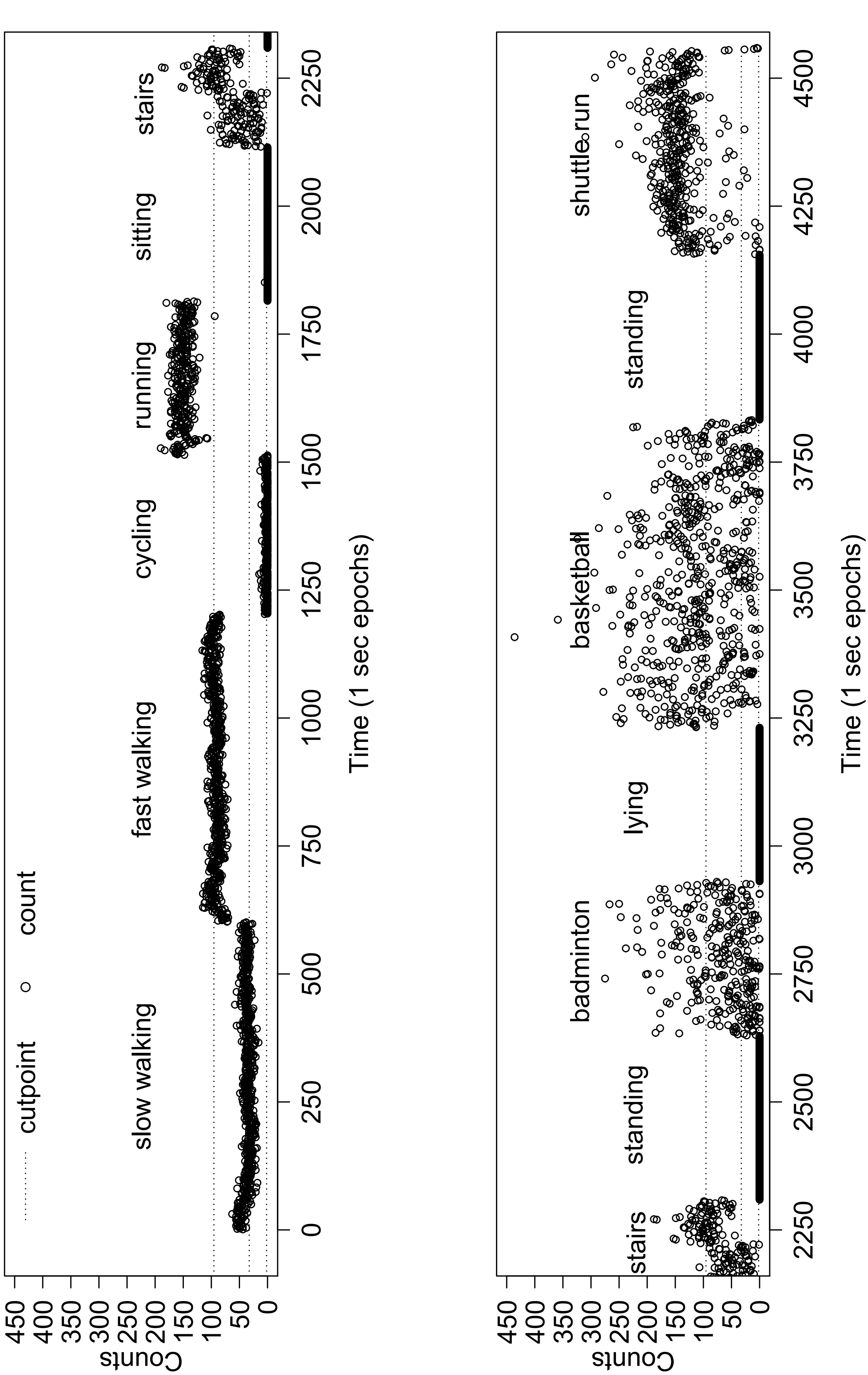

Figure 3.1: Example of collected labeled accelerometer data (1 second epochs).

subject. The sequence of activities during a day is stored as a time series of counts by the accelerometer, see Figure 3.1 for an example of collected labeled data, for which the underlying activity is known, with 1 second epochs.

3.2 Cutpoint method: choice of cutpoints and epoch length

The most common approach to derive the pattern of physical activity and its energy expenditure is to map these counts to a certain number of sedentary and activity ranges, such as sedentary behavior, light, moderate and vigorous physical activity. Activity ranges are separated by thresholds known as *cutpoints*. Cutpoints for different age groups are available for children (Evenson et al., 2008; Freedson et al., 2005; Guinhouya et al., 2009a; Pate et al., 2006; Puyau et al., 2002; Treuth et al., 2004; Trost, 2007) and adults (Freedson et al., 1998; Sasaki et al., 2011; Troiano et al., 2008) allowing to assess the overall time spent in these ranges of physical activity. The duration of physical activity within the same activity range is called a *bout* and is defined as the time period in which the subject remains within one activity range without changing to another.

Cutpoints according to Freedson et al. (1998) are included in Figure 3.1. In this example all epochs with $\leq$ 99 counts/min are classified as SED, epochs with 100-1951 counts/min as LPA which corresponds to $<$ 3 metabolic equivalent of task (METs). Epochs with 1952-5724 counts/min are assigned to MPA with 3-5.99 METs and epochs with 5725-9498 counts/min to HARD with 6.00-8.99 METS and epochs $>$ 9498 counts/min to VERY HARD with $>$ 9 METs. Commonly epochs with $\geq$ 1952 counts/min are characterized as moderate-to-vigorous physical activity (MVPA).

Apparently the choice of the cutpoints has a direct effect on the derived amounts of time spent in SED, LPA and MVPA. A discussion of the advantages and disadvantages of different cutpoints is beyond the scope of this chapter, but in the literature cutpoints according to Evenson et al. (2008) are frequently used for children and adolescents and cutpoints according to Freedson et al. (1998) are commonly used for adults.

Another influencing factor with regard to the identified intensities is the choice

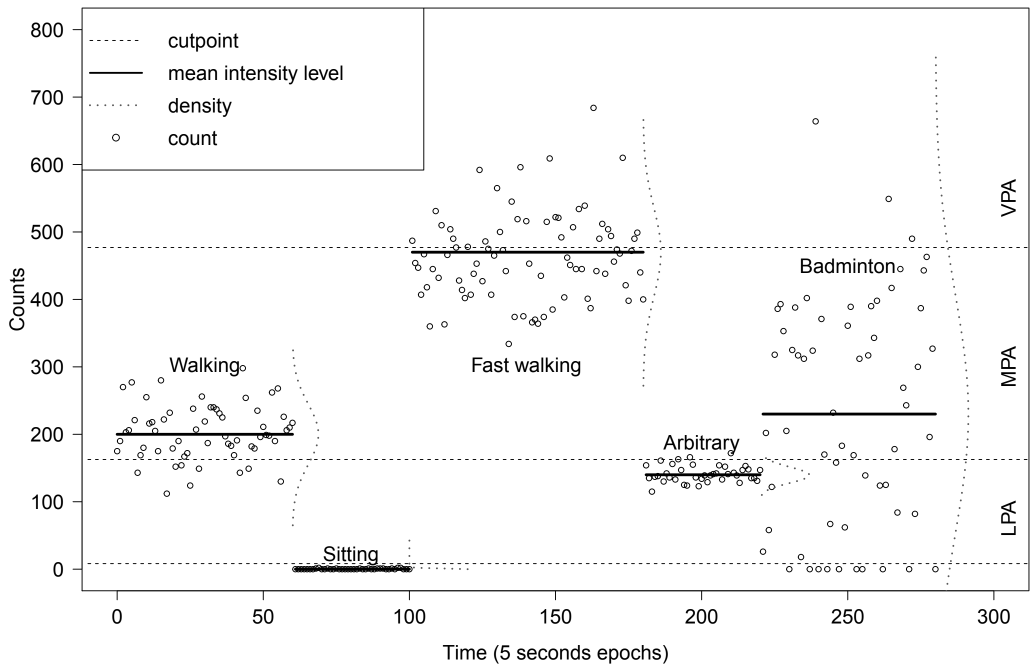

Figure 3.2: Assumed physical activity model: The figure shows five distinct activities: walking, sitting, fast walking, an arbitrary activity and playing badminton with mean activity levels represented by the solid line. The observed accelerometer counts scatter around them following a certain distribution depicted as dotted line (adopted from Witowski et al. (2014) and to be shown in the forthcoming paper presented in Appendix B).

of the epoch length. As will be seen later on, an increase in the chosen epoch length results in a reduction of the variation of counts and hence fewer counts will be at the extreme ends of the intensity range. This leads to an underestimation of time spent in SED/LPA and particularly MVPA. In the past, determining epoch lengths was a trade off between battery endurance and available memory, resulting in epoch lengths of 15 seconds to 1 minute. With the technological advances of accelerometer devices, nowadays epoch lengths of 1, 3 or 5 seconds are considered as sensible choices.

3.3 Assumption about physical activity behavior

The cutpoint method is very easily implemented and therefore widely used. It is a valid way to classify accelerometer data, if one assumes that the count at point in time t is independent of the count at $t-1$ and human beings are able to switch instantly from one mode of activity to the other. These assumptions are, however, quite unrealistic. Assuming a more realistic physical activity behavior may lead to serious flaws of the cutpoint method.

Let us assume that a person's daily activities are composed of a non-overlapping series of bouts of different activities. For example riding a bike to work, working at a desk, walking to lunch and so on. Let us further assume that all these activities have a certain intensity, which is represented by a true, mean count level. The registered counts by the accelerometer then scatter around this true intensity level. This assumption is depicted in Figure 3.2. The person first takes a short walk, after which she/he is sitting, maybe watching TV, followed by some fast walking, an arbitrary activity (see Section 4.4.1) and a game of badminton. The solid black lines represent the "true" average count level for each of these activities, which can be understood as the true intensity level. The counts registered by the accelerometer scatter around this true level, following a certain distribution (dotted gray line). So activities depicted in Figure 3.2 consist of five separate bouts, with five distinct activity levels.

If this assumption holds true, then the cutpoint method has some serious drawbacks. As long as the variation around the true intensity level is small and the true level is not close to a cutpoint the complete mode of activity can be correctly assigned to its corresponding activity range. However, in real-life applications there are activities showing large variation of counts, resulting in large scattering, as for example games such as basketball or badminton. Counts are then assigned to the wrong activity range, leading to considerable misclassification. The erroneous classification of counts may also lead to an overestimation of the number of activity bouts. As a bout is defined as the time a person spends within one activity range without switching to another range, misclassifying the count into a different activity range starts a new bout by definition. The subject seems to switch from one activity range to another

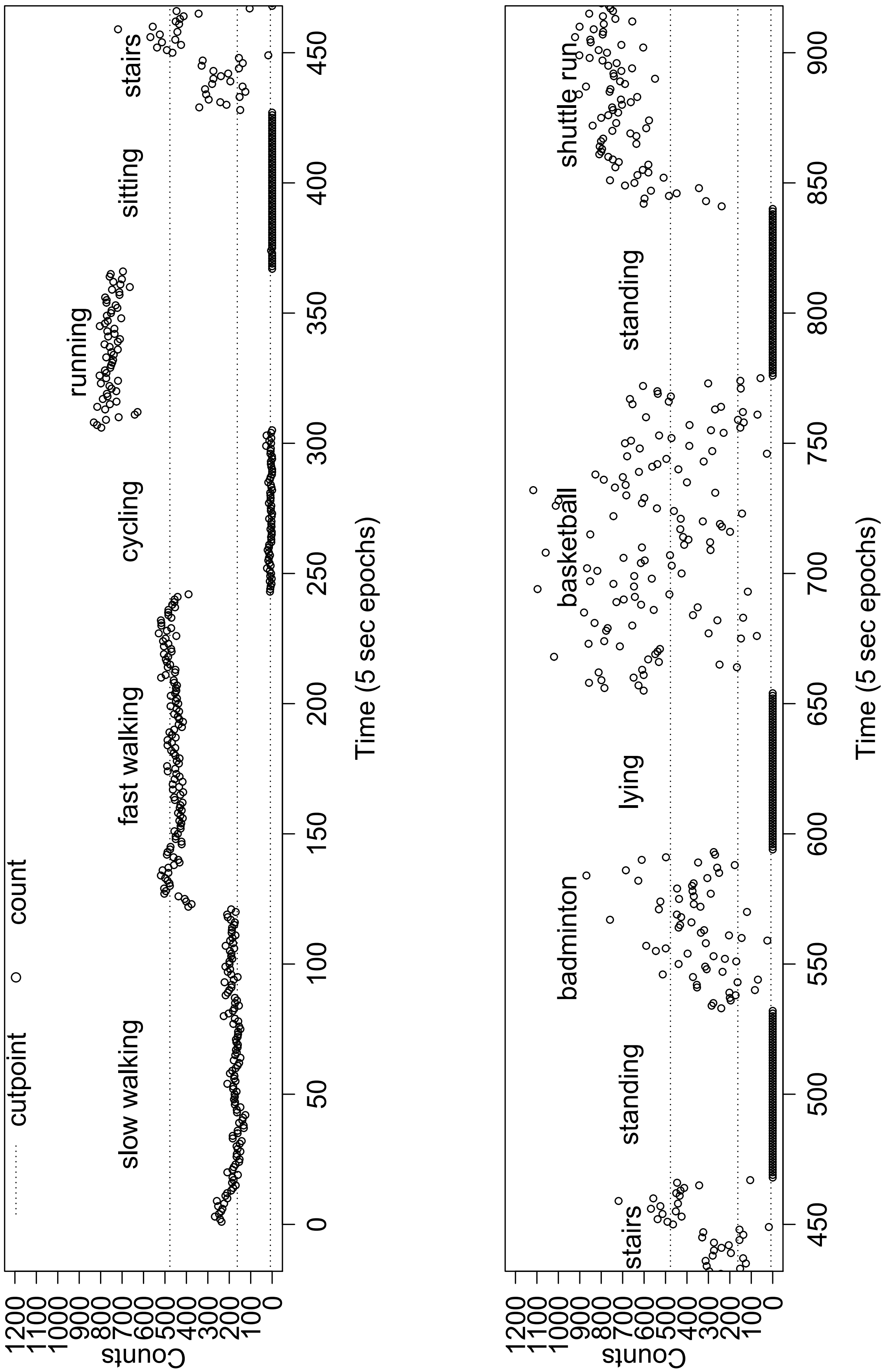

Figure 3.3: Example of collected labeled accelerometer data (5 seconds epochs).

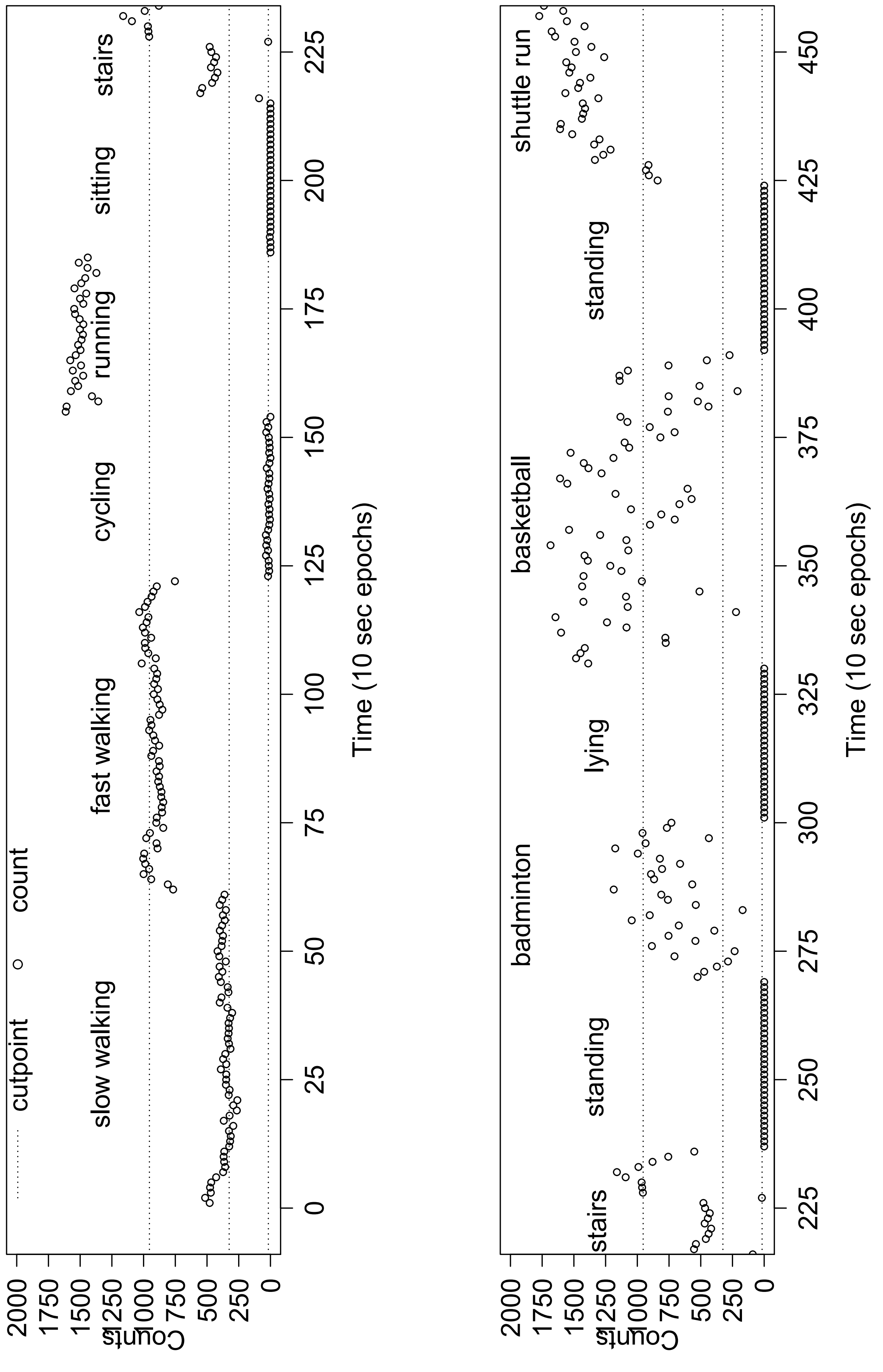

Figure 3.4: Example of collected labeled accelerometer data (10 seconds epochs).

Activity	Duration (min)	Speed ($\frac{\text{m}}{\text{s}}$)	Intensity
Standing still	5	0	SED
Lying on the ground	5	0	SED
Sitting	5	0	SED
Slow walking	10	1.08	LPA
Fast walking	10	1.67	LPA
Riding a bike	5	5.33	LPA
Climbing stairs of a five story building	≈ 4	N/A	LPA/MVPA
Jogging	5	2.83	MVPA
Badminton	5	N/A	MVPA
Basketball	10	N/A	MVPA
Shuttle run test	≈ 6	N/A	LPA to MVPA

N/A = not applicable

Table 3.1: List of activities performed for generating labeled data.

and back again within a few epochs.

3.4 Labeled data

Chapter 4 will present novel approaches to assign intensity levels to accelerometer counts to cope with the drawbacks of the cutpoint method mentioned above. These methodological approaches for modeling accelerometer data have to be evaluated. For this purpose accelerometer data are needed in which the underlying truth for each observation (count) is known. This includes the activity, which generated the measured count, as well as its intensity. These requirements are met by simulated data (see Section 4.2.1 and Section 4.4.1).
In order to simulate accelerometer data that resemble real life data as closely as possible, we collected labeled accelerometer data in a small sample. Five female and four male participants were asked to perform a sequence of predefined activities, covering the whole range of intensities. The participants wore GT3X+ Actigraph accelerometers (Pensacola, Florida, USA). The device was attached to the right hip using an elastic belt. The devices were initialized

using the ActiLife 6 software. Data were downloaded using the same software and counts were computed at 1, 5, 10 and 15 seconds epochs. Table 3.1 lists the performed activities, their duration and intensity. The specific activities were chosen to cover rather monotonic ones, like walking and cycling, resulting in a count series with little variation, as well as activities like badminton and basketball, which show considerably more variation. Figures 3.1, 3.3 and 3.4 show the collected labeled data for one participant displayed in 1, 5 and 10 seconds epochs. In this example the effects of increasing epoch lengths become obvious. The variation of the counts is reduced and hence less counts are found below the LPA cutpoint and above the VPA cutpoint, this is especially true for activities with high variation like badminton and basketball. Consequently less time spent in LPA and MVPA is identified by the cutpoint method.

Chapter 4

New approaches for assigning intensity levels

This chapter investigates two innovative approaches to model accelerometer data under more realistic assumptions then those underlying the cutpoint method. The theoretical background of both methods, namely hidden Markov models and expectile regression using a Whittaker smoother with L_0-penalty, will be introduced. In addition, their performance will be investigated by means of Monte Carlo experiments.
It will be shown that hidden Markov models are a promising improvement over the cutpoint method. Hence, this method will be compared with expectile regression utilizing a Whittaker smoother with an L_0-penalty, where we will see that the latter even outperforms hidden Markov models albeit at the cost of computational simplicity.

4.1 Hidden Markov models

Assumptions on the true physical activity behavior in human beings were formulated in Section 3.3 and the resulting drawbacks of the simple cutpoint method were described. As one solution to this problem the hidden Markov models (HMM) can be combined with the traditional cutpoint method. The idea is to identify the correct average intensity levels and map the counts to them. Afterwards the identified activities are then assigned to an intensity level via the regular cutpoints. The result of this proposal is depicted in Figure 4.1.

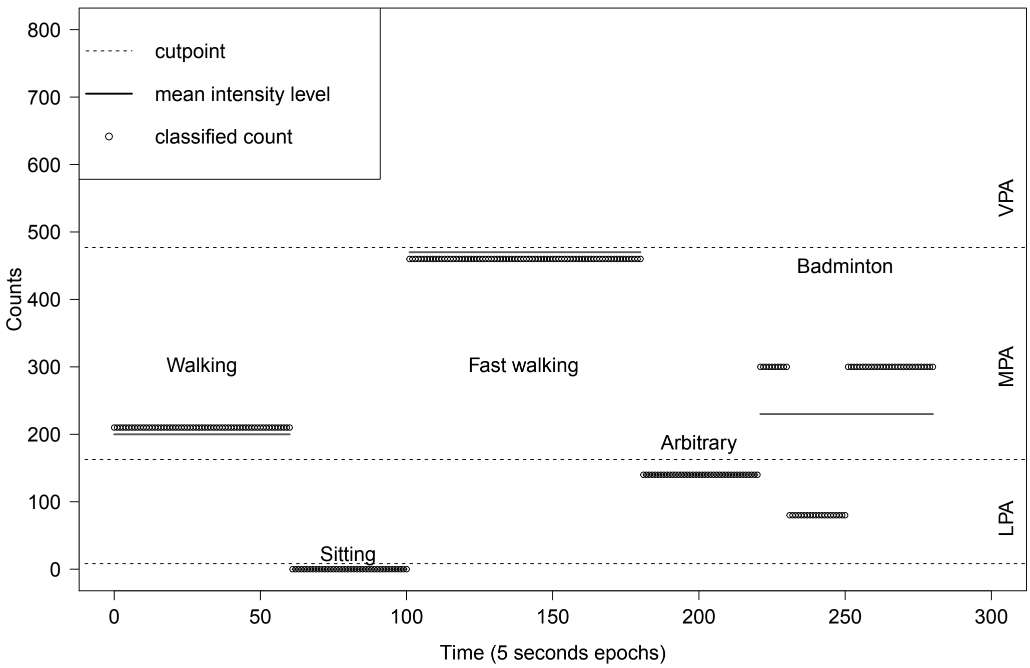

Figure 4.1: Identified activities and their intensities after the application of the HMM-method (adapted from Witowski et al. (2014))

The theory of hidden Markov models will be introduced in this section. In Section 4.2, HMMs will be applied to simulated accelerometer data to assess their performance.

4.1.1 Definition of hidden Markov models

This section follows the description of the mathematical background provided in Zucchini and MacDonald (2009) and Fink (2003). Let us now assume that the activities performed during the day can be represented as a time series of true activity states can be mathematically described as a stochastic process. The idea is that the observed time series, the counts registered by the accelerometer, have been generated by an underlying unobservable, time and value discrete, stochastic process whose random variables Z_t are *hidden*.

Definition 4.1. (Stochastic process) Let $(\Omega, \mathscr{A}, P)$ be a probability space. Let further $\mathscr{I}$ be an index set and $\mathscr{Z}$ a space with a σ-algebra. Then a *stochastic* *process* is a function

$$Z : \Omega \times \mathscr{I} \to \mathscr{Z}, \quad (\omega, t) \mapsto Z_t(\omega), \tag{4.1}$$

where the function

$$Z_t : \Omega \to \mathscr{Z} \tag{4.2}$$

is a random variable on $(\Omega, \mathscr{A}, P)$ for each $t \in \mathscr{I}$.

The range $\mathscr{Z}$ of random variables Z_t is the set of possible states. In the case of activities performed during the day, $\mathscr{Z} = \{1, \ldots, m\}$ is finite with $i \in \mathscr{Z}$ symbolizing a specific activity, e.g. walking, running or sitting at the desk. Correspondingly $\mathscr{I}$ is a countable set. The stochastic process $\{Z_t, t \in \mathbb{N}\}$ is called time and value discrete.

Definition 4.2. (Time series) A *times series* $\{z_1, \ldots, z_T\} = z_{1:T}$ is a finite realization of a stochastic process $\{Z_t, t \in \mathbb{N}\}$ with length $T \in \mathbb{N}$.
$z_t = i, i \in \mathscr{Z}$, is the realization of Z_t at point in time t.

The *true* time series of length T of activities $\{z_1, \ldots, z_T\}$ is thought to be hidden and can therefore only be observed indirectly via the recorded accelerometer counts $\{x_1, \ldots, x_T\}$, which are the observed realizations of random variables X_t. The underlying, unobservable and hence hidden stochastic process satisfies the Markov property and is therefore called *Markov chain*.

Definition 4.3. (Markov property) A time and value discrete stochastic process $\{Z_t, t \in \mathbb{N}\}$ is called *Markov chain*, if it satisfies the following *Markov property*:

$$P(Z_t = z_t | Z_1 = z_1, \ldots, Z_{t-1} = z_{t-1}) = P(Z_t = z_t | Z_{t-1} = z_{t-1}). \tag{4.3}$$

Let us now assume that the *transition probability* to switch from one state to another at point in time t only depends on the state a person is currently in and is independent of all states prior to t.

Definition 4.4. (Transition probability) The probability of a Markov chain to switch from state i to state j is given by the *transition probability*

$$\gamma_{ij} = P(Z_t = j | Z_{t-1} = i). \tag{4.4}$$

A Markov chain is called *homogeneous*, if the transition probability γ_{ij} is independent of t for all pairs of i and $j \in \mathscr{Z}$. The transition probability of a homogeneous Markov chain with finite $\mathscr{Z} = 1, \ldots, m$ can be summarized in an $(m \times m)$ *transition matrix*

$$\mathbf{\Gamma} = (\gamma_{ij})_{1 \leq i,j \leq m} \tag{4.5}$$

with

$$\gamma_{ij} \in [0,1], \quad i,j \in \mathscr{Z}, \tag{4.6}$$

$$\sum_{j \in \mathscr{Z}} \gamma_{ij} = 1, \quad i \in \mathscr{Z}. \tag{4.7}$$

A Markov chain is fully defined by its transition matrix $\mathbf{\Gamma}$ and a vector containing the *initial probabilities* $\boldsymbol{\pi_0} = (\pi_{01}, \ldots, \pi_{0m}) = (P(Z_1 = 1), \ldots, P(Z_1 = m))$ with $\sum_{i=1}^{m} \pi_{0i} = 1$ for the first state. Under the assumption described above, each state $i = 1, \ldots, m$ is linked with the mean activity count μ_i of the corresponding activity, which the state represents. Let μ_i denote the mean activity level of the i-th physical activity. Furthermore, the variable X_t is assumed to be conditionally independent of all remaining variables given its unobservable activity Z_t:

$$P(X_t = x_t | Z_1, \ldots, Z_t, X_1, \ldots, X_{t-1}) = P(X_t = x_t | Z_t = z_t). \tag{4.8}$$

At each point in time t, the observed accelerometer count x_t is assumed to be generated by a certain distribution, which depends on the activity state $z_t = i$ with the corresponding activity level μ_i as mean of this distribution. This assumption is depicted in Figure 3.2, in which the distribution is drawn as dotted grey line.

Definition 4.5. (Observation distribution) The probability that X_t takes a value x_t under the condition that $Z_t = i$ is given by the *observation distribution*

$$p_i(x_t) = P(X_t = x_t | Z_t = i). \tag{4.9}$$

In case of continuous distributions, $p_i(x_t)$ is the value of the density function at x_t.

The observation distributions are assumed to be a subset of a whole class of distributions to be specified in advance. Each observation distribution p_i is determined by $k \in \mathbb{N}$ parameters with parameter vector $\boldsymbol{\theta_i} = (\theta_{1i}, \ldots, \theta_{ki}) \in \mathbb{R}^k$. The $m \cdot k$ parameters in turn form the matrix $\boldsymbol{\theta} = (\theta_{li})_{1 \leq l \leq k; 1 \leq i \leq m}$. An HMM is fully defined by its model-specific parameter $\mathbf{\Theta} = (\boldsymbol{\pi_0}, \mathbf{\Gamma}, \boldsymbol{\theta})$.

4.1.2 Applying HMMs to accelerometer data

The application of HMMs can be subdivided into the following three steps.

Step 1: Building an HMM for an observed time series of counts

The model-specific parameter $\mathbf{\Theta}$ of an HMM is estimated based on an observed time series of counts $\{x_1, \ldots, x_T\}$. This first step is referred to as *training of the HMM.*

Definition 4.6. (Production probability) Let the hidden Markov model be defined by $\mathbf{\Theta}$, then the *production probability* of a certain observed series $\{x_1, \ldots, x_T\} = x_{1:T}$ is given by the probability

$$L(\mathbf{\Theta}) = P(X_{1:T} = x_{1:T}|\mathbf{\Theta}). \tag{4.10}$$

The *likelihood* of the model is finally the probability that a certain observed series $x_{1:T}$ as well as a certain series of activities $z_{1:T}$ have been generated by an HMM defined by $\mathbf{\Theta}$ summed over all possible series of activities $z_{1:T} \in \mathscr{Z}_{1:T}$:

$$L(\mathbf{\Theta}) = \sum_{z_{1:T} \in \mathscr{Z}_{1:T}} P(X_{1:T} = x_{1:T}, Z_{1:T} = z_{1:T}|\mathbf{\Theta}) \tag{4.11}$$

$$= \sum_{z_{1:T} \in \mathscr{Z}_{1:T}} [\pi_{0_{z_1}} p_{z_1}(x_1) \prod_{t=2}^{T} \gamma_{z_{t-1},z_t} p_{z_t}(x_t)] \tag{4.12}$$

The likelihood of the model with respect to $\mathbf{\Theta}$ can be either numerically maximized or by utilizing the so-called Baum-Welch algorithm (Baum et al., 1970) which is commonly used to fit HMMs. In real-life applications the number of underlying activities m given the observed accelerometer counts $x_{1:T}$ is unknown. Therefore several HMMs with different numbers of states m are trained and their goodnesses of fit are compared using the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). If both criteria suggest a different number of states, then one may opt for fewer states to have a more simplistic model or for a larger number of states if this better reflects the underlying practical situation.

Step 2: Decoding the hidden sequence of PA-levels

After the model parameter $\mathbf{\Theta}$ and an appropriate number of physical activities m have been estimated, the resulting HMM is used to link each observed count

x_t to an estimated activity level $\hat{\mu}_i$, $i = 1, \ldots, m$.

Step 2.1 First, the Viterbi algorithm (Forney Jr, 1973; Viterbi, 1967) is used to decode the globally most likely sequence of hidden activities denoted by $z_1^*, \ldots, z_T^*$ for the trained HMM and the same time series of counts $x_{1:T}$ that was used to train the HMM in Step 1 by comparing the joint probability of all T hidden states and the observed accelerometer counts.

Step 2.2 Second, each accelerometer count x_t is assigned to the estimated activity level $\hat{\mu}_{z_t^*}$ that corresponds to the decoded state z_t^* at this point in time.

Step 3: Extension of the cutpoint method

In the last step the results of the HMM-based method is combined with the traditional cutpoint approach. Now, each accelerometer count x_t is assigned to an activity range a_t via its corresponding (most likely) mean activity level $\hat{\mu}_{z_t^*}$.

In the example illustrated in Figure 4.1, the trained HMM identifies five activity levels $\hat{\mu}_1, \ldots, \hat{\mu}_5$, which leads to a misclassification of parts of the state 'badminton' into three instead of one bout, with two bouts being assigned to MPA and one to LPA. Even with this overestimation of six identified activity levels instead of five, the HMM-based method assigns most counts correctly to their actual activity range. The high number of bouts typically obtained from the cutpoint method is reduced by the HMM-based approach because a Markov chain is assumed to underlie the performed activities at each point in time. The present example consists of five bouts: the first is defined by the activity 'walking', which corresponds to the activity range LPA, the second bout is defined by 'sitting' in SED, the third by 'fast walking' in MPA, close to the cutpoint for VPA. The fourth bout is defined by a arbitrary activity and the last by 'badminton' in MPA. Due to the assumed Markov chain, the HMM-based approach detects seven bouts, which is an overestimation of the true value of five, but results are more precise than those obtained from the traditional cutpoint method, which identifies over 40 bouts.
Figures 4.4 and 4.5 show the HMMs (gray dashed line) fitted to the labeled

accelerometer presented in Section 3.4.

4.2 Modeling accelerometer data with HMMs

The hidden Markov models introduced in Section 4.1 have the potential to improve the analysis of accelerometer data, especially compared to the traditional cutpoint approach, described in Section 3.2. In order to investigate the general feasibility and the advantages over the cutpoint approach, we conducted a simulation study (Witowski et al., 2014), see Appendix A for a reprint. The HMMs were compared with the traditional cutpoint method in terms of (1) the misclassification rate (MCR), calculated as the percentage of how many of the counts were assigned incorrectly to any other activity range than their true activity range, (2) number of bouts correctly identified, (3) number of activity levels correctly identified, and (4) runtime.

4.2.1 Simulation study

In the simulation study, 1,000 days of labeled accelerometer data consisting of $T = 1,440$ counts at 15 seconds epochs were simulated. So each simulated time series represented a six-hour day. For labeled data, the true sequence of activities and their actual activity level and also the activity range of each count are known. Counts per day were randomly generated using the negative binomial distribution (with parameters $r = 1$ and $p = 0.0009$, resulting in the lowest activity level $\mu_1 = 111.11$) and the Gaussian distribution (with the parameters $\mu_2 = 400$, $\mu_3 = 600$ and $\mu_4 = 900$ as well as $\sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 10,000$) around three or four pre-defined activity levels (depending on the random time series generated by a Markov chain). For the simulation study cutpoints from Pate et al. (2006) were used. The lowest activity level of 400 of the simulated data was intentionally chosen to be very close to the lower cutpoint of 420 to investigate the performance of the HMMs close to a cutpoint. In the context of modeling accelerometer counts, three distributions are of particular interest: The first HMM is based on the Poisson distribution, which is typically used to model counts. The second model uses the generalized Poisson distribution (Joe and Zhu, 2005) that includes a further variance parameter to allow for a larger or smaller variation than the one assumed for a standard Poisson distribution. Real-life accelerometer data typically show larger variability than a simple Poisson distribution can accommodate. For the third HMM, a Gaussian distribution is assumed to capture the random scattering of the counts around the presumed activity level. For the purpose of the present analysis, the Poisson-based HMM is referred to as HMM[Pois], the HMM based on the generalized Poisson distribution as HMM[GenPois] and the Gaussian-based HMM as HMM[Gauss].

4.2.2 Results

The results of the simulation study clearly show the superiority of the HMM-based method over the traditional cutpoint approach. Among the different distributions used for the hidden Markov models, HMM[Pois] showed the weakest performance with regard to MCR, bout and activity detection. The results for HMM[GenPois] and HMM[Gauss] were similar. HMM[Gauss] led to a slightly better MCR, while HMM[GenPois] was better in terms of bout detection. HMM[GenPois] outperformed HMM[Gauss] with a considerably higher activity detection rate. This outperformance came at a price, namely runtime and problems with numerical stability. So depending on the particular research question one has to weigh the advantages and drawbacks of HMM[GenPois] and HMM[Gauss] to decide which model is best suited for the situation. For detailed results and an extensive discussion of the results see Witowski et al. (2014).

4.3 Expectiles and expectile regression

As was seen in the previous two sections, HMMs are a promising improvement to model accelerometer data compared to the cutpoint method. Yet, quite a lot of a priori information is required for a proper fit. That is the number of modeled activity levels has to be defined in advance, as well as the class of distributions. Both assumptions are difficult to verify. It is virtually impossible to know how many activities a participant performs during any given day and according to which distribution the observed counts scatter around the mean level in advance without inspecting the data. Visual inspections may provide a good guess, but this is unmanageable for thousands of accelerometer days.

So we are searching for an approach that takes the assumptions made on physical activity behavior in Section 3.3 into account, by modeling accelerometer data as a sort of step function with each jump indicating the start of a new activity and the constant interval being the mean intensity level of that activity. This should be accomplished without any a priori assumptions on the number of activities or any distribution.

Here, we propose expectile regression using a Whittaker smoother with L_0-penalty as a solution that allows the desired modeling of accelerometer data. In Section 4.4, the general performance and the advantages of this novel approach are investigated in a simulation study, in which it is compared to HMMs and the cutpoint method.

In this section, first, a short introduction to univariate expectiles is given and the concept of penalized regression is briefly explained. This introduction will be a little broader than is needed for our application to accelerometer measured physical activity to enable the reader to better understand the underlying idea. Building on this, expectile regression utilizing a Whittaker smoother with L_0-penalty will be presented, which allows to model a step curve with the desired properties, as described above.

4.3.1 Univariate expectiles

Usually regression models focus only on one quantity of the response distribution, the mean. However, there are situations in which it is needed to also model the extreme parts of the data. This can be, for example, done by using GAMLSS as presented in Section 5.3.1. *Quantile regression* as introduced by Koenker and Bassett Jr (1978) is frequently used, as quantiles have a natural interpretation. Another option is *expectile regression* introduced by Newey and Powell (1987) as an alternative to quantile estimation. Expectile regression, as well as quantile regression can be used to characterize the complete conditional distribution of a response. An overview of models beyond mean regression can be found in Kneib (2013). In recent years, expectile regression has been found to be a reasonable generalization of mean regression and an alternative to median regression.

For given observations $y_1, \ldots, y_n$ of independent identically distributed random variables $Y_1, \ldots, Y_n$ the τ-quantile q_τ can be estimated by minimizing the weighted absolute residuals

$$\hat{q}_\tau = \arg\min_{q_\tau} \sum_{i=1}^{n} w_\tau(y_i, q_\tau)|y_i - q_\tau| \tag{4.13}$$

with weights

$$w_\tau(y_i, q_\tau) = \begin{cases} \tau & \text{if } y_i > q_\tau \\ 1-\tau & \text{if } y_i \leq q_\tau \end{cases} \tag{4.14}$$

as introduced by Koenker and Bassett Jr (1978). The basic idea is to asymmetrically "punish" the residuals. As the sum in (4.13) is not differentiable, linear programming is used to obtain estimates (Koenker, 2005). An `R` package is available to calculate q_τ for $\tau \in (0,1)$ (Sobotka and Kneib, 2012).
In 1987 Newey and Paul extended the asymmetric weights used for quantile regression to what they called *asymmetric least squares*, which more recently was replaced by *least asymmetrically weighted squares* (LAWS). Instead of partitioning the data by a proportion τ being below the estimate, as quantiles do, the weight of a partial first moment with proportion τ is located below the estimate. τ is often referred to as *asymmetry parameter*, it specifies the strength of a specific interest in either the upper or lower tail of the distribution.
Newey and Powell (1987) replaced the L_1 distance in (4.13) by the L_2 distance, which makes the sum in (4.15) differentiable and allows an easy solution. An expectile estimate can then be calculated by fulfilling the LAWS criterion

$$\hat{\zeta}_\tau = \arg\min_{\zeta_\tau} \sum_{i=1}^{n} w_\tau(y_i, \zeta_\tau)(y_i - \zeta_\tau)^2 \tag{4.15}$$

with weight function $w_\tau(y_i, \zeta_\tau)$, as defined in Equation (4.14), and $\tau \in (0,1)$. As stated above, the sum in (4.15) is differentiable, but depends on the weights $w_\tau(y_i, \zeta_\tau)$, which in turn depend nonlinearly on ζ_τ. Therefore, estimates are obtained through an iteratively weighted least squares process. LAWS can be understood as a weighted generalization of the well-known ordinary least squares (OLS) estimation, which is the special case of LAWS for $\tau = 0.5$.
Unlike quantiles, expectiles lack an easy interpretation, with the exception of $\zeta_{0.5}$, which is the mean. Jones (1994) showed that the expectiles are in fact quantiles uniquely related to the distribution of Y. Yao and Tong (1996) showed that there exists a unique bijective function $h : (0,1) \to (0,1)$ such

that $q_\tau = \zeta_{h(\tau)}$ where

$$h(\tau) = \frac{-\tau q_\tau + G(q_\tau)}{-\zeta_{0.5} + 2G(q_\tau) + (1-2\tau)q_\tau} \tag{4.16}$$

with $G(q) = \int\limits_{-\infty}^{q} y\,dF(y)$ as the partial moment function and $F(y)$ as the cumulative distribution function. Here, $G(\infty) = \zeta_{0.5} = \mu$ is the expectation of Y.

This implies that quantiles can be calculated from a dense set of expectiles. Schulze Waltrup et al. (2015) used (4.16) to compare expectile-based quantile estimates with quantile estimates regarding efficiency and proposed a method to estimate non-crossing expectile curves based on splines.

All theoretical τ-expectiles can be calculated for a given distribution with cumulative distribution function F and finite expectation. The `R` package `expectreg` provides expectiles for various distributions and the necessary programs to calculate expectiles for given distributions (Sobotka et al., 2014).

4.3.2 Expectile regression

In the following, we will extend the above approach to a regression model with covariates x_i, $i = 1, \ldots, r$. Let us first consider the simple parametric model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}_\tau + \boldsymbol{\varepsilon}_\tau$$

with response $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$, design matrix $\boldsymbol{X} = (\mathbb{1}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ with $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nr})^T$, $j = 1, \ldots, r$ and errors $\boldsymbol{\varepsilon}_\tau = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Here $\zeta_\tau = \boldsymbol{X}\boldsymbol{\beta}_\tau$ is the expectile and the regression coefficient that minimizes (4.15) is estimated by iteratively reweighted least squares updates

$$\hat{\boldsymbol{\beta}}_\tau^{[b]} = (\boldsymbol{X}^T \boldsymbol{W}_\tau^{b-1} \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{W}_\tau^{b-1} \boldsymbol{y} \tag{4.17}$$

where $\hat{\boldsymbol{\beta}}_\tau^{[b]}$ is the estimated regression coefficient vector in the bth iteration step. As stated above, the estimation has to be iterated, since the weights in the *weight matrix* $\boldsymbol{W}_\tau^b = \operatorname{diag}(w_\tau(y_1, \boldsymbol{X}\hat{\boldsymbol{\beta}}_\tau^{[b]}), \ldots, w_\tau(y_n, \boldsymbol{X}\hat{\boldsymbol{\beta}}_\tau^{[b]}))$ also depend on the current estimates (Schulze Waltrup, 2014).

Now let us define a more flexible nonlinear model

$$Y_i = f_\tau(x_i) + \varepsilon_{\tau i}.$$

Several choices for the functional form for the expectile curve f_τ are possible. Newey and Powell (1987) originally proposed a linear model. Schnabel and Eilers (2009) favored *P(enalized)-splines* to model expectile curves.
The basisc idea is to approximate $f(x)$ by polynomial B-splines of degree l. Let us assume the domain is divided into $\tilde{M}-1$ equal intervals by $\tilde{M}$ *knots*. Then $f(x)$ can be approximated by $M = \tilde{M} + l - 1$ *basis functions* $B_m^l(x)$ of degree l as

$$f(x) = \sum_{m=1}^{M} \beta_m B_m^l(x)$$

with β_m denoting the coefficient (so-called amplitude) of basis function B_m^l. B-splines are described in detail by Eilers and Marx (1996). The aim is to construct a smooth function by joining polynomial pieces, resulting in a $(l-1)$-times continuously differentiable function f.
B-splines are defined recursively, with basis function of degree 0 defined as

$$B_m^0(x) = \begin{cases} 1, & \text{if } \kappa_m \leq x < \kappa_{m-1} \\ 0, & \text{otherwise} \end{cases}$$

for $m = 1, \ldots, M-1$ and knots κ_m. A general B-spline of degree $l \geq 1$ can now be defined as

$$B_m^l(x) = \frac{x - \kappa_{m-l}}{\kappa_m - \kappa_{m-l}} B_{m-1}^{l-1}(x) + \frac{\kappa_{m+1} - x}{\kappa_{m+1} - \kappa_{m+1-l}} B_m^{l-1}(x).$$

Thus, a B-spline of degree l can be constructed from a B-spline of degree $(l-1)$ and can be traced back to a B-spline of degree 0. A basis function of degree l consists of $l+1$ polynomial pieces, which are defined by $l+2$ knots, of which l are inner knots. A design matrix $\boldsymbol{B}$ containing the basis functions can be constructed as

$$\boldsymbol{B} = \begin{pmatrix} B_1(x_1) & \cdots & B_M(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \cdots & B_M(x_n) \end{pmatrix}.$$

For more details on B-Splines we refer to Eilers and Marx (1996) and Schulze Waltrup (2014).
The correct choice of the number of knots and their positions is a problem in B-spline regression, as it has an impact on the flexibility of the fitted curve. In order to correct for too much flexibility Eilers and Marx (1996) created

P-splines by using equidistant knots and by introducing a penalty term. Let $\boldsymbol{K}$ denote a symmetric *penalty matrix*, then

$$(\boldsymbol{y} - \boldsymbol{B\beta})^T(\boldsymbol{y} - \boldsymbol{B\beta}) + \delta\boldsymbol{\beta}^T\boldsymbol{K\beta}$$

with $\delta \geq 0$ is the penalized least squares criterion, which we minimize with respect to $\boldsymbol{\beta}$. With *smoothing parameter* δ the smoothness of the fitted curve can be tuned between a polynomial spline regression without penalty ($\delta \to 0$) or a polynomial fit of order $d-1$ ($\delta \to \infty$). Schnabel and Eilers (2009) investigated methods for choosing δ. In Section 4.3.4 two options for selecting δ will be presented.

Eilers and Marx (1996) proposed to use second order differences of adjacent coefficients of B-splines, which is accomplished with the difference operator $\triangle^2(\beta_m) = \triangle\triangle(\beta_m) = \beta_m - 2\beta_{m-1} + \beta_{m-2}$ for $m \geq 3$. This leads to the definition of the $(M-2) \times M$ dimensional *difference matrix* $\boldsymbol{D}_2$ as

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix},$$

which leads to the desired penalization and $\boldsymbol{K} = \boldsymbol{D}_2^T\boldsymbol{D}_2$.

In the context of expectile regression, the regression coefficients are estimated by iteratively reweighted least squares updates,

$$\hat{\boldsymbol{\beta}}_\tau^{[b]} = (\boldsymbol{B}^T\boldsymbol{W}_\tau^{b-1}\boldsymbol{B} + \delta\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{B}^T\boldsymbol{W}_\tau^{b-1}\boldsymbol{y}$$

similar to 4.17 (Schulze Waltrup, 2014). Expectile smoothing can also be achieved by other smoothers. For example, Sobotka and Kneib (2012) used bivariate P-splines and Markov fields for spatial smoothing in combination with expectiles. See also Kneib (2013) for an overview.

4.3.3 Modification for use in accelerometer data

The basic assumptions related to physical activity behavior were presented in Section 3.3. As a consequence, the best fitting regression model should be a) constant during the performance of one specific activity, b) close to the mean

intensity level of that activity and c) “jump” right to the next mean intensity level.
The concept of penalized smoothing introduced in the sections above is now used to modify expectile regression by applying the Whittaker smoother with L_0-penalty to ensure that the fitted expectile curve has the desired properties.

Model

As described in Section 4.1 the recorded accelerometer counts $y_1, \ldots, y_T$ are the observed realizations of random variables Y_t. Now, we consider the regression model

$$f(x) = \sum_{t=1}^{T} \beta_t B_t(x)$$

with regression coefficients β_t and basis elements B_t. The flexibility is ensured by construction, as there is one regression coefficient β_t for each observed point in time t, $t = 1, \ldots, T$. The support of each basis element is also only one point in time and hence the design matrix simplifies to $\boldsymbol{B} = \boldsymbol{I}$.

L_0-penalty

Whittaker (1923) introduced the smoother described in (4.18), see also Eilers (2003) for details on theory, implementation and applications. Rippe et al. (2012) presented a modification of the Whittaker smoother as signal smoother for segmented genetic data. In certain types of tumor tissue, segmentation can be observed and a visual representation fulfilling the same requirements as a regression curve for accelerometer data is required (see Figure 4.2). The authors proposed to use a smoother based on the L_0 norm.
Let the data consist of T data pairs (x_i, y_i) for which a smooth series $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_T)$ is fitted. The authors defined the so called *objective function* as

$$S_2 = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2 + \delta \sum_{i=2}^{T} (\hat{y}_i - \hat{y}_{i-1})^2. \tag{4.18}$$

The first term, the squared residuals, measure the fidelity of the fitted curve $\hat{\boldsymbol{y}}$ to data $\boldsymbol{y}$. The second term is the penalty on roughness with smoothing parameter δ. The larger δ is chosen, the smoother the curve will be (see top panel of Figure 4.2; please note that in this figure the original notation of

Rippe et al. (2012) is used with λ instead of δ). In quantile smoothing the L_2 norm (sum of squared values) in the penalty is replaced by the L_1 norm (sum of absolute values) with objective function

$$S_1 = \sum_{i=1}^{T} |y_i - \hat{y}_i| + \delta \sum_{i=2}^{T} |\hat{y}_i - \hat{y}_{i-1}|.$$

This modification results in a better visualization of the segmented data, as can be seen in the middle panel of Figure 4.2, although there is still a number of undesirable small jumps. As further improvement the use of the L_0 norm is proposed, resulting in

$$S_0 = \sum_{i=1}^{T} (y_i - \hat{y}_i)^2 + \delta \sum_{i=2}^{T} |\hat{y}_i - \hat{y}_{i-1}|^0. \tag{4.19}$$

This penalizes basically non-zero differences between neighboring points of $\hat{\boldsymbol{y}}$, that is jumps. Positive numbers raised to the power of 0 result in 1 and $0^0 = 0$ by convention. Therefore only jumps result in a penalty. The penalty is always 1, regardless of the magnitude of the jump. The result can be seen in the lower panel of Figure 4.2.

The optimal choice of the smoothing parameter δ will be discussed in Section 4.3.4.

Penalty matrix

As we look at differences of neighboring regression coefficients $\beta_t - \beta_{t-1}$, the penalty matrix $\boldsymbol{K}$ can be constructed as $\boldsymbol{K} = \boldsymbol{D}_1^T \boldsymbol{P} \boldsymbol{D}_1$ with $\boldsymbol{D}_1$ a $(T-1 \times T)$ dimensional difference matrix with

$$\boldsymbol{D}_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix},$$

and weight matrix $\boldsymbol{P} = \operatorname{diag}\left(\left(\frac{1}{(\boldsymbol{D}_1 \boldsymbol{\beta})^2 + \xi} \right)^T \right)$ with $\xi > 0$ added for computational stability. With this definition, the total penalty adds up to $\delta \boldsymbol{\beta}^T \boldsymbol{D}_1^T \boldsymbol{P} \boldsymbol{D}_1 \boldsymbol{\beta} = \delta \sum_{t=2}^{T} \frac{(\beta_t - \beta_{t-1})^2}{(\beta_t - \beta_{t-1})^2 + \xi}$. Typical choices for ξ are $10,000^{-2}$. This way, the summand is 0, if $\beta_{k-1} = \beta_k$, i.e. constant. For $\beta_{k-1} \neq \beta_k$, i.e. a jump in the regression

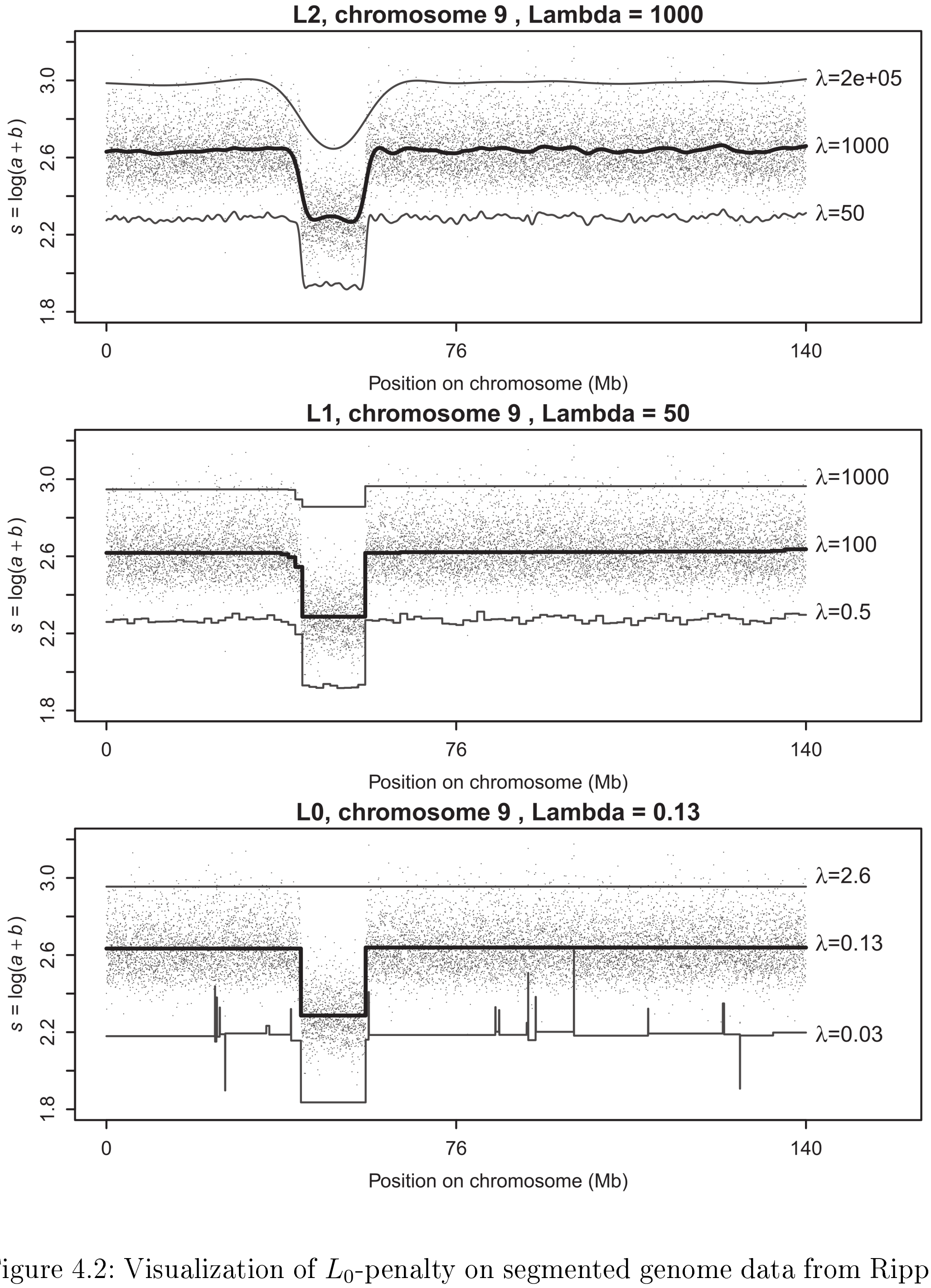

Figure 4.2: Visualization of L_0-penalty on segmented genome data from Rippe et al. (2012)

.

curve, each summand is about 1. Thus, the penalty punishes the number of jumps in the regression curve.
With these modifications, the expectile regression can be used to model accelerometer data while reflecting the underlying assumptions. As described previously, expectile curves have no natural interpretation, except for $\tau = 0.5$ as the mean. As we are interested in the mean intensity levels of activities, analyses should therefore focus on fitting the 0.5-expectile curve to accelerometer data, although additionally estimating lower and upper expectiles may provide further insights into the type of activity and its accompanying distribution of counts. Lower expectiles could be interpreted as stops or less active periods during an activity as e.g badminton, and higher expectiles represent periods of highest intensity.

4.3.4 Choice of smoothing parameter δ

With the roughness penalty being properly defined we can search for a proper smoothing parameter δ. It is obvious that for $\delta \to 0$ the regression model will be just an interpolation of the observed data, while for $\delta \to \infty$ it will become a constant. In the literature, two approaches to select the "optimal" δ can be found that have proven to work in practice.

Cross-validation

Rippe et al. (2012) proposed to use odd/even cross-validation when working with the L_0 smoother. All even observations are left out by setting a weight $\tilde{w}_i$ to 0. $\tilde{w}_i$ is set to 1 for all odd observations. For a series of different δs the value

$$\tilde{CV} = \sqrt{\sum_i (1 - \tilde{w}_i)(y_i - \hat{y}_i)^2}$$

is calculated. Then the value δ that minimizes $\tilde{CV}$ is determined. It should, however, be doubled when modeling the complete dataset. See Figure 4.3 for a visualization of odd/even cross-validation.
We adapt the odd/even cross-validation in a way that we are able to use all available observations. We separate the data into two folds, one containing all even observations, the other all odd observations. For each fold $h = 1, 2$ we

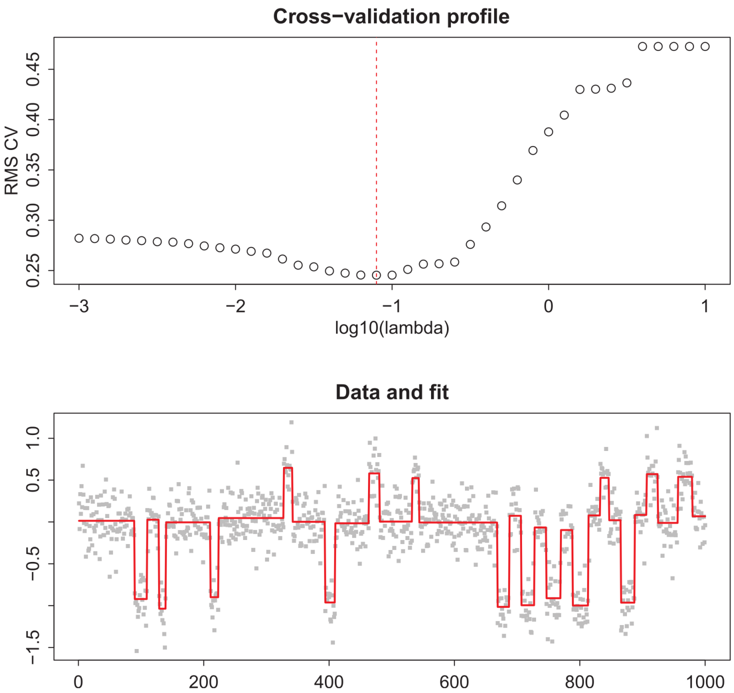

Figure 4.3: Example of odd/even cross-validation to select optimal smoothing parameter from (Rippe et al., 2012)

estimate

$$\hat{\boldsymbol{\beta}}^h = \arg\min_{\boldsymbol{\beta}} \left[\sum_{t=1}^{T} w_{t,h}^* (y_t - \beta_t)^2 + \delta \boldsymbol{\beta}^T \boldsymbol{D}_1^T \boldsymbol{P} \boldsymbol{D}_1 \boldsymbol{\beta} \right]$$

with additional weights $w_{t,1}^* = 0$ for $h = 1$ and t even. Thus, although only odd observations are used, we also obtain predictions for the omitted even observations. In the second fold, $h = 2$, $w_{t,2}^* = 0$ and t odd. The two predictions are combined to $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^1 + \hat{\boldsymbol{\beta}}^2$ and the cross-validation score $CV = \sum_{t=1}^{T} (y_t - \hat{\beta}_t)^2$ is calculated. δ that minimizes CV is then determined by a grid search.

L-curve

Alternatively, the so-called L-curve may be used to select an adequate value for δ. Hansen (1992) suggested to consider the two major components of every smoothing procedure that is godness of fit and smoothness of the final estimate. For this purpose, the logarithm of the magnitude of the penalty term of the regression model ($\Xi = \log_{10}(\boldsymbol{\beta}^T \boldsymbol{D}_1^T \boldsymbol{P} \boldsymbol{D}_1 \boldsymbol{\beta})^2)$ is plotted against the logarithm of the sum of squared residuals ($\Psi = \log_{10} \sum_{i=1}^{T} (y_i - \boldsymbol{\beta}_i)^2)$ parameterized by

δ resulting in the so-called L-curve. The "elbow" of the L-shaped curve is characterized by intermediate values of Ψ, Ξ and δ. Hansen (1992) opted to select as appropriate δ the value that corresponds to the point of maximum curvature, that is the elbow of this curve.

The L-curves were originally introduced for the selection of a regularization parameter in ill-proposed inverse problems. Frasso and Eilers (2015) showed that the L-curves can be applied to a wide variety of smoothing problems, even in data with correlated noise. See Hansen (1992) and Frasso and Eilers (2015) for computational details and graphical visualization of selecting an optimal δ by this approach.

Both methods for selecting the smoothing parameter δ suffer from the disadvantage that a grid search from values close to zero with almost no penalization to values implying a constant estimate has to be performed. This leads to a tremendous computational effort, as the time needed for the grid search is multiplied by the number of tested δ.

4.4 Comparison of HMM- and expectile-modeled accelerometer data

In Section 4.1 HMMs were introduced as a novel approach to allow modeling physical activity behavior as described in Section 3.3. Section 4.2.1 describes a simulation study conducted by Witowski et al. (2014) in which the performance of HMMs based on the Poisson, generalized Poisson and Gaussian distribution was compared with the cutpoint method, concluding that HMMs based on the Gaussian distribution, denoted HMM[Gauss], are a suitable new approach to model accelerometer data. Expectile regression using the L_0 norm penalty and a Whittaker smoother have been introduced in Section 4.3.3 as a second innovative approach to model accelerometer data accounting for the assumptions made for physical activity behavior. The new methods are compared with each other and with the cutpoint method in a second simulation study. The expectile regression is compared with the traditional cutpoint approach (Section 3.2) and with HMM[Gauss] with regard to (1) misclassification rate (MCR), (2) number of identified bouts and (3) identified levels, (4) the proportion of the estimated curve being in the range of $\pm 10\%$ of the true mean

No.	Resembled activity	1 second epochs Mean level	Distribution	% during the day
1	Sitting	-1	$\mathrm{N}(\mu = -1, \sigma = 1)$	40
2	Arbitrary	28	$\mathrm{Pois}(\lambda = 28)$	24
3	Slow walking	36	$\mathrm{N}(\mu = 36, \sigma = 11)$	24
4	Arbitrary	55	$\mathrm{N}(\mu = 55, \sigma = 13)$	2
5	Badminton	60	$\mathrm{N}(\mu = 60, \sigma = 55)$	2
6	Arbitrary	68	$\mathrm{Pois}(\lambda = 68)$	2
7	Fast walking	90	$\mathrm{N}(\mu = 90, \sigma = 15)$	2
8	Basketball	110	$\mathrm{N}(\mu = 110, \sigma = 65)$	2
9	Running	160	$\mathrm{N}(\mu = 160, \sigma = 30)$	1
10	Arbitrary	190	$\mathrm{N}(\mu = 190, \sigma = 40)$	1

Table 4.1: Characteristics of simulated activities (1 second epochs).

and (5) runtime.

4.4.1 Simulation study

The major advantage of simulated accelerometer data is that for each count the originating intensity level is known. In order to obtain plausible results it is essential to simulate data that resemble real-life accelerometer data as closely as possible. We collected labeled accelerometer data, for which the performed activities are known in a sample of nine adults, as described in Section 3.4. Based on these data 6 activities presented in Tables 4.1 and 4.2 were chosen for the simulation, covering the range from monotonic activities like walking with little variation of counts to ball games like basketball with large variation. Another 4 arbitrary activities were defined to introduce further activities with smaller or larger variation. For most simulated activities the Gaussian distribution was used, some arbitrary activities were assumed to be Poisson distributed. Additionally some mean activity levels were deliberately chosen to be close to a cutpoint defined by Freedson et al. (1998). In total 1,000 accelerometer days were simulated for 1 second epochs ($T = 43,200$) and 5 seconds epochs ($T = 8,640$) each. For each day five to ten activities were

No.	Resembled activity	5 seconds epochs Mean level	Distribution	% during the day
1	Sitting	-1	N($\mu = -1, \sigma = 1$)	40
2	Arbitrary	140	Pois($\lambda = 140$)	24
3	Slow walking	200	N($\mu = 200, \sigma = 45$)	24
4	Arbitrary	275	N($\mu = 275, \sigma = 75$)	2
5	Badminton	230	N($\mu = 230, \sigma = 177$)	2
6	Arbitrary	340	Pois($\lambda = 340$)	2
7	Fast walking	470	N($\mu = 470, \sigma = 66$)	2
8	Basketball	500	N($\mu = 500, \sigma = 240$)	2
9	Running	860	N($\mu = 860, \sigma = 125$)	1
10	Arbitrary	950	N($\mu = 950, \sigma = 200$)	1

Table 4.2: Characteristics of simulated activities (5 seconds epochs).

randomly chosen, the average percentage of time the activities were performed during the day can be found in Table 4.1 for 1 second epochs days and Table 4.2 for 5 seconds epochs, respectively. Minimum bout length was set to 240 seconds. Negative counts were set to 0. The number of activities and bouts for the simulated days can be found in Table 4.3.

4.4.2 Statistical analyses

As described in Section 4.3.4 there are basically two ways to determine a good smoothing parameter δ for the expectile regression. Odd/even cross-validation as well as L-curves are very computationally intensive and therefore it was not feasible to automatically determine an optimal δ for each individual day. Instead, the optimal choice was made in a subsample of accelerometer days, leading to $\delta = 630$ to be used in the analyses.
The main analyses were performed on a high performance computing cluster (HPC), providing two master servers and 28 computing nodes, each consisting of 12 CPU cores (2.53 GHz each) and 96 Gb RAM, using `R` version 3.2.0 (R Core Team, 2015). For the cutpoint method and HMM[Gauss] the `R` package `HMMpa` (Witowski and Foraita, 2014) was used. Expectile curves were calculated

No.	1 second epochs		5 seconds epochs	
	Activities	Bouts	Activities	Bouts
Min	6	30	5	27
Median	8	42	9	41
Mean	8.45	41.64	8.50	41.40
Max	10	56	10	54

Table 4.3: Characteristics of simulated accelerometer days.

with the `R` package `expectreg` (Sobotka et al., 2014). Expectile curves turned out to be very computationally intensive. In order to fully exploit the potential of the HPC, simulated 1 second epochs days were divided into five equal sized parts, which were sent to five different cores and results were recombined after analyses.

4.4.3 Results

As was previously known, the cutpoint method performed worst, with the obvious exception of runtime. Expectile regression performed considerately better than HMM[Gauss] with regard to MCR and number of identified bouts. Results for HMM[Gauss] were closer to the correct number of activities. As HMMs by construction are based on a pre-defined number m of levels this is not surprising, as expectile regression estimates a curve. Therefore we introduced the proportion of the estimated curve being in the range of $\pm 10\%$ of the true mean as a measure of "closeness of fit". Using this as a criterion, the expectile regression was by far better than the other two techniques. Expectile regression needed on average about seven times the runtime of HMM[Gauss], but given the increased performance, this is a reasonable price to pay. Both methods show improved results for data aggregated to 5 seconds epochs, mainly due to the accompanying reduction of variance relative to the mean as described in Section 3.2. Especially HMM[Gauss] benefited from this reduction with regard to all considered quality criteria. Expectile regression showed improvements regarding MCR, runtime and proportion of the estimated curve being in the range of $\pm 10\%$ of the true mean. Compared to 1 second epochs data, the number of identified bouts increased, but was still in the magnitude of and

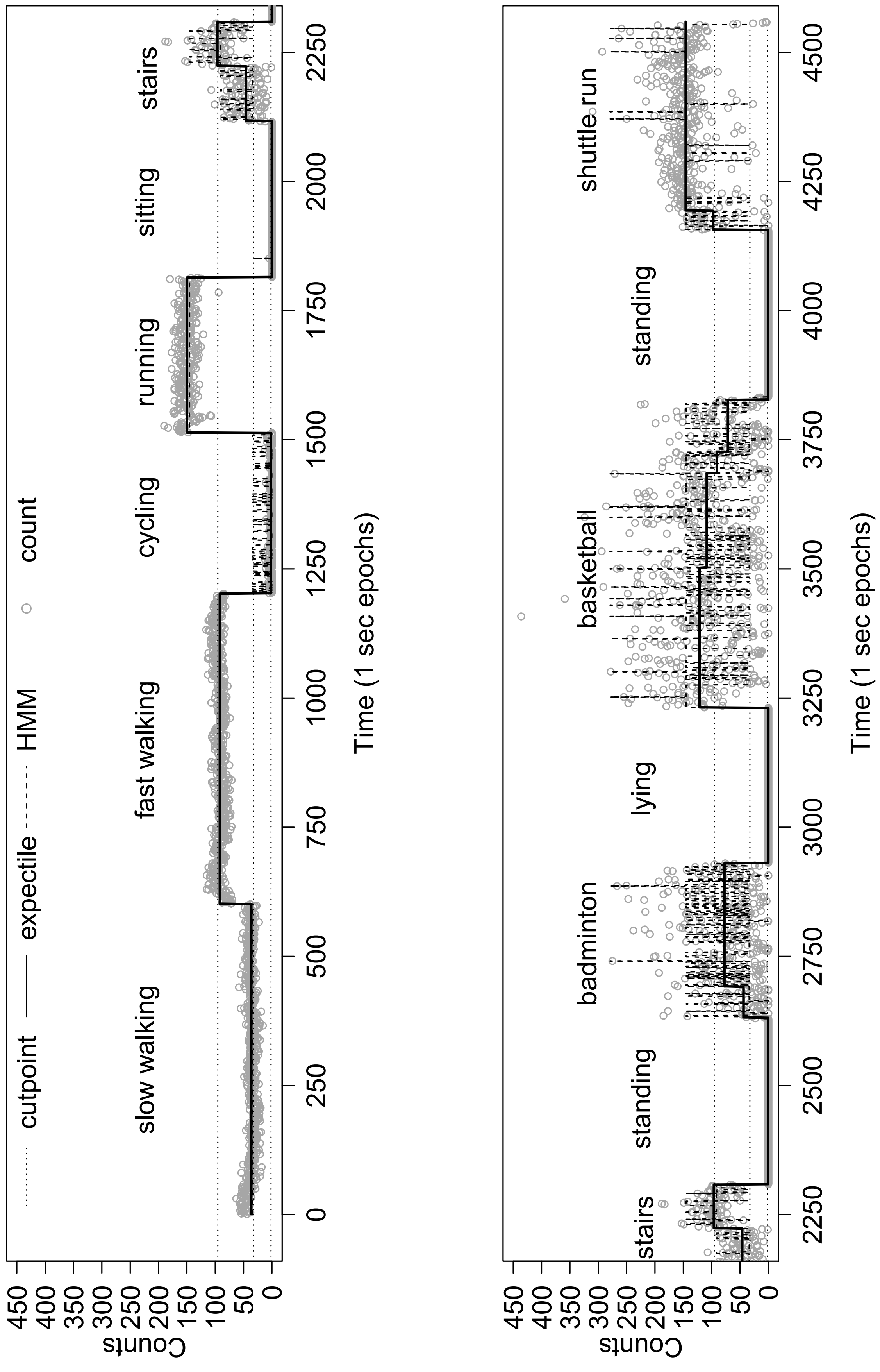

Figure 4.4: Example of collected labeled accelerometer data (1 second epochs) and the results of HMM[Gauss] and expectile regression applied to them (forthcoming paper presented in Appendix B).

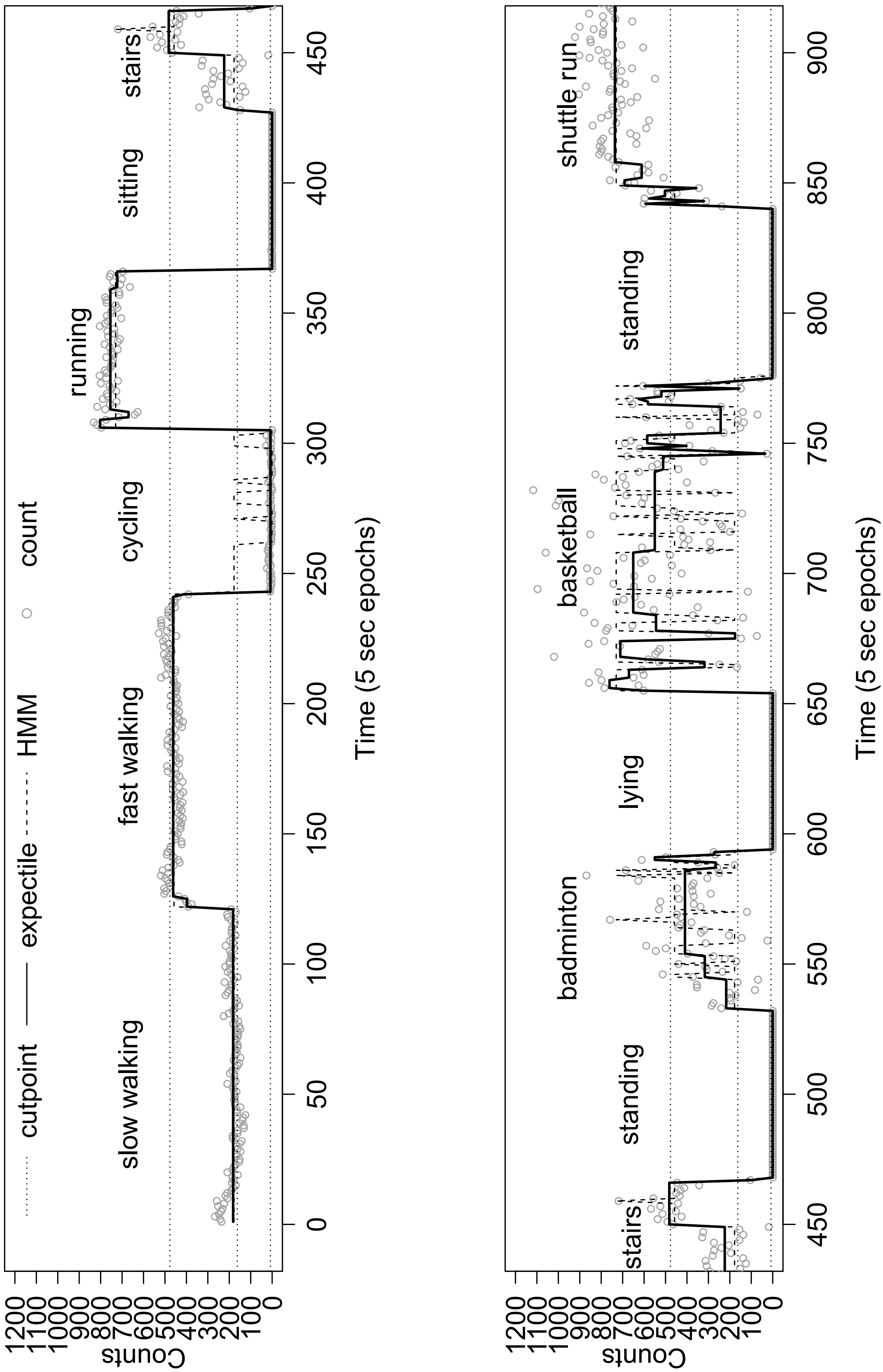

Figure 4.5: Example of collected labeled accelerometer data (5 seconds epochs) and the results of HMM[Gauss] and expectile regression applied to them (forthcoming paper presented in Appendix B).

considerably closer to the true number than for HMM[Gauss]. The number of identified levels also increased. It seemed that the expectile regression became more sensitive to (numerically) extreme values and added levels to compensate these values.

Figures 4.4 and 4.5 show the HMMs (gray dashed line) and expectile curve (solid black line) fitted to the labeled accelerometer presented in Section 3.4.

In summary, expectile regression with an L_0 norm penalty and a Whittaker smoother showed superior results compared to HMMs and the cutpoint method and is hence promising approach to analyze accelerometer data. For more results and a more detailed discussion we refer to the forthcoming paper presented in Appendix B.

Chapter 5

Studies of physical activity in various age groups

Whereas the previous chapter introduced and compared two novel approaches to model accelerometer data, this chapter describes five empirical studies on physical activity. First, in the European IDEFICS study accelerometer data were collected in over 12,000 children. These data are used to describe the physical activity behavior of European children using GAMLSS, which is also introduced in this chapter. The IDEFICS data will also be used to investigate the association between physical activity and high blood pressure in children. Adding the follow-up data collected in the I.Family study, which continues the IDEFICS cohort, the data are used to assess the longitudinal associations between physical activity and obesity markers like BMI and fat mass.
The smaller PATREC study conducted in Bremen, Germany, collected accelerometer data in combination with an activity diary and activity questionnaire to investigate some methodical issues in the assessment of physical activity in adolescents for different domains by subjective and objective methods. Finally, an energy prediction equation for a pedometer model is derived based on a small study conducted in Oldenburg, Germany.

5.1 The IDEFICS study

The European IDEFICS (Identification and prevention of dietary- and lifestyle induced health effects in children and infants) study is a prospective cohort study, which started in 2006. The study aims at investigating the etiology of overweight, obesity and related disorders in children. A baseline survey (T0) was conducted in 2007-2008 and a follow-up survey (T1) in 2010-2011 in eight European countries (Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain and Sweden) (Ahrens et al., 2006; Bammann et al., 2006; Ahrens et al., 2014). Between T0 and T1 a primary prevention program was implemented in selected intervention regions in each country, each to be compared with a control region (Henauw et al., 2011; Pigeot et al., 2015). 16,228 children aged 2-9 years participated in the baseline survey. Parents were asked to report data on sociodemographic characteristics as well as on medical, nutritional and other lifestyle factors. Children also participated in an extensive examination protocol, which included anthropometry, accelerometry, blood pressure and a fitness test, as well as the collection of biological samples, including saliva for DNA extraction, blood and urine (Figure 5.1, left section). Additional protocols to collect information on the built environment, sensory taste perception and mechanisms of food choice and consumer behavior were implemented in subgroups. Signed written informed consent was obtained from children's parents in addition to verbal permission from each child before examination. The study protocol was approved by the local ethics committees. See Ahrens et al. (2011) for more details on the study design, the used instruments and a description of the study population. Free-living physical activity was assessed using GT1M or Actitrainer uniaxial devices from Actigraph, LLC, Pensacola, FL, USA. Both devices use identical sensor units. The general survey manual required the accelerometers to be set to 15 seconds epochs. The devices were attached to the right hip by means of an elastic belt. The children were asked to wear the devices from the getting up in the morning until bed time in the evening. Participants were asked to wear the accelerometer for at least three consecutive days including one weekend day. Parents completed a daily activity or non-wearing diary, in which wearing periods and periods during which the accelerometers were not worn should be recorded. A total of 18,745 children participated in both IDEFICS surveys. Of these, 12,014 provided data on physical activity. The remaining 6,731 either refused to wear the devices or the assessment was not completed due to other reasons, like lack of devices at the time of assessment. Children with musculoskeletal or orthopaedic diseases

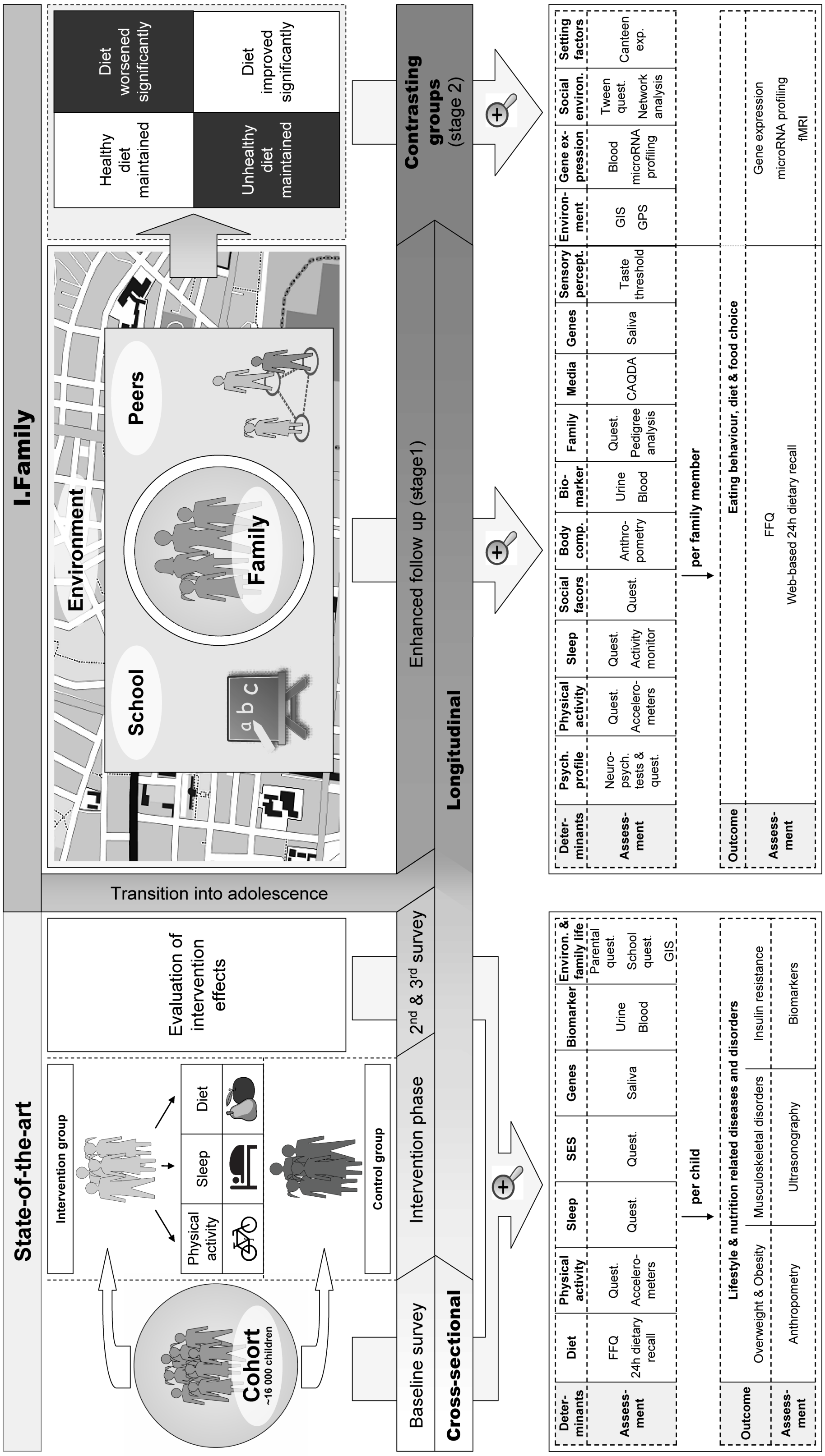

Figure 5.1: Longitudinal design of the IDEFICS study, its concatenation with the I.Family study and overview of all examination modules (left part presented in Ahrens et al. (2011)).

(n=332) were excluded from this assessment.

5.2 The I.Family study

The I.Family study (Determinants of eating behaviour in European children, adolescents and their parents) builds on the IDEFICS study. In this follow-up study dietary behavior and food choice within whole families and their lifestyle are investigated. As in the IDEFICS study, data on health and nutrition behavior were collected and complemented with family data by including siblings and parents.

A second follow-up survey (T3) was conducted in the year 2013/2014. All *index children* that is all children, who participated in T0 and/or T1 were invited to participate, as well as their siblings and parents. The examination program of T3 covered the majority of the modules employed during T0 and T1. Modules on family life, peers and kinship structure were introduced in T3 (Figure 5.1, right section). Additional assessment modules were implemented in a subgroup of participants, the so-called *contrasting groups*, these were defined as children who showed divergent developmental trajectories in their weight status. One of these modules investigated the built environment of the families using GIS and GPS trackers. Families were asked to wear an accelerometer together with a GPS tracking device. Like in the IDEFICS study, signed written informed consent was obtained from children's parents in addition to verbal permission from each participant before examination. The study protocol was approved by the local ethics committees.

In total 17,540 persons participated in I.Family of which 7,083 were index children, 2,548 were newly recruited siblings and 6,851 biological parents and 1,057 other adults. The 6,162 families in I.Family had on average 2 children (data status as of December 2015). More details on the study design, the used instruments and a description of the study population will soon be presented in a forthcoming publication.

Participants in T3 were asked to wear an Actigraph accelerometer for seven consecutive days. For 4,841 children (50.3% of children and adolescents participating in T3) and 1,427 adults (18.1%) data on physical activity were downloaded from the devices. Since also the raw data files were collected, the data can be derived for arbitrary epoch lengths like e.g. 15 or 60 seconds as used in the IDEFICS study, or even shorter epochs like 1 or 5 seconds.

5.3 Descriptive results of physical activity: application of GAMLSS to accelerometer data

In Konstabel et al. (2014) we used the Generalized Additive Models for Location Scale and Shape (GAMLSS) presented below to derive percentile curves for levels of physical activity. The publication aims at describing physical activity levels of European children and the provision of sex- and age-specific reference standards in children aged 2-10 years.

5.3.1 Generalized Additive Models for Location Scale and Shape (GAMLSS)

In clinical practice, especially for diagnostic purposes, reference ranges are needed for various clinical parameters to classify measurements as pathologic, unusual or usual. If the measurement depends on a covariate, for example age, so that the reference ranges change with the covariate, then this should be reflected in so-called percentile curves. Percentile curves show the percentiles of the distribution of a medical parameter depending on an additional covariate, typically depending on age. These percentile curves are thus of particular interest for measurements in children, as their bodies pass through dramatic changes during childhood and adolescence, so that one single reference range might be sufficient for adults, but certainly not for children and adolescents. In 1988 Cole introduced the *LMS* method, which summarizes the changing distribution by three separate curves which are estimated for the a) median, b) the variation coefficient and c) the skewness. This method has been extended in recent years to also allow for more than one covariate and for additional modeling of the kurtosis of a distribution. For this purpose the so-called Generalized Additive Models for Location Scale and Shape (GAMLSS) are used. Both techniques will be briefly introduced in this section.

LMS method for smooth reference percentile curves

The LMS method was firstly introduced by Cole in 1988 and further improved in 1992. The basic idea is to fit what was later on called a Box-Cox Cole and Green (BCCG) distribution to the empirical distribution of the variable of interest. The BCCG distribution is described by three parameters, 1) the Box-Cox power ψ, 2) the mean μ and 3) the coefficient of variation σ. The following paragraphs are based on Cole and Green (1992) and give a formal introduction to this technique. Let us denote e.g. the medical parameter of interest with Y. Let us further assume that Y is a positive random variable with median μ and that Y^{ψ} is normally distributed. For $\psi = 0$ let $\ln Y$ follow a normal distribution. Based on the family of transformations proposed by Box and Cox (1964) (therefore ψ is referred to as $Box - Cox\ power$) the following transformation

$$X = \frac{(Y/\mu)^{\psi} - 1}{\psi}, \quad \psi \neq 0 \tag{5.1}$$

or

$$X = \ln(Y/\mu), \quad \psi = 0$$

will map the median μ of Y to a median of 0 for X and is continuous at $\psi = 0$. Let σ denote the standard deviation (SD) of X. For $\psi = 1$, σ is the coefficient of variation (CV) of Y. Thus, X follows a normal distribution with mean μ and variance σ^2. The standard deviation score (SDS) or z-score of X and hence of Y is given by

$$\begin{aligned} Z &= X/\sigma \\ &= \frac{(Y/\mu)^{\psi} - 1}{\psi\sigma}, \quad \psi \neq 0, \sigma \neq 0 \end{aligned} \tag{5.2}$$

or

$$Z = \frac{\ln(Y/\mu)}{\sigma}, \quad \psi = 0,$$

respectively. Thus, Z follows a truncated standard normal distribution, as the condition $0 < Y < \infty$ leads to the condition $-1/(\sigma\psi) < Z < \infty$ if $\psi > 0$ and

$-\infty < Z < -1/(\sigma\psi)$ if $\psi < 0$. It follows that percentiles can be calculated as

$$P_{100\alpha} = \mu(1 + \psi\sigma z_\alpha)^{1/\psi}, \quad \psi \neq 0 \tag{5.3}$$

or

$$P_{100\alpha} = \mu e^{(\sigma z_\alpha)}, \quad \psi = 0,$$

where z_α is the α-quantile of the standard normal distribution. Now let us assume that the distribution of Y varies with some covariate t, i.e. age or height, then also the three parameters ψ, μ and σ vary with this covariate. The name giving idea of the LMS method is now to estimate three smooth curves $L(t)$, $M(t)$ and $S(t)$ for the parameters ψ, μ and σ. It follows that

$$Z = \frac{[Y/M(t)]^{L(t)} - 1}{L(t)S(t)}, \quad L(t) \neq 0,\ S(t) \neq 0, \tag{5.4}$$

or

$$Z = \frac{\ln[Y/M(t)]}{S(t)}, \quad L(t) = 0.$$

Analogously to (5.3) percentiles can be calculated as

$$P_{100\alpha}(t) = M(t)(1 + L(t)S(t)z_\alpha)^{1/L(t)}, \quad L(t) \neq 0 \tag{5.5}$$

or

$$P_{100\alpha}(t) = M(t)\exp[S(t)z_\alpha], \quad L(t) = 0$$

The probability density function for Y is given by

$$f_Y(y) = \frac{y^{\psi-1}\exp(-\frac{1}{2}z^2)}{\mu^\psi\sigma\sqrt{2\pi}\Phi(\frac{1}{\sigma|\psi|})}, \tag{5.6}$$

where z is given by (5.2) and $\Phi(\frac{1}{\sigma|\psi|})$ is the cumulative distribution function of a standard normal distribution. Cole and Green (1992) assumed a standard normal distribution for Z and that the truncation probability is negligible. With these assumptions and inserting $L(t)$, $M(t)$ and $S(t)$ for the parameters ψ, μ and σ, the log-likelihood function can be derived as:

$$l = l(L, M, S) = \sum_{i=1}^{n} \left(L(t_i) \ln \frac{y_i}{M(t_i)} - \ln S(t_i) - \frac{1}{2} z_i^2 \right)$$

for the case of independent random variables Y_i with observations y_i at corresponding covariate levels t_i with z_i being the SDS corresponding to y_i. If the L, M and S curves are smooth, then so are the percentile curves. In order to assure the smoothness of $L(t)$, $M(t)$ and $S(t)$ Cole and Green suggest to subtract penalties from the likelihood (penalized likelihood). $L(t), M(t)$ and $S(t)$ are thus estimated by maximizing the *penalized likelihood*

$$l_p = l - \frac{1}{2}\alpha_\psi \int (L''(t))^2 dt - \frac{1}{2}\alpha_\mu \int (M''(t))^2 dt - \frac{1}{2}\alpha_\sigma \int (S''(t))^2 dt$$

where α_ψ, α_μ and α_σ are smoothing parameters. Cubical splines are used for the estimation. The Fisher-score is used for iterative optimization; see Cole and Green (1992) for details on the numerical implementation and optimal choice of the smoothing parameters.

So in summary the LMS method can be used to derive percentile curves for a medical parameter that depends on one covariate, if the assumption that the variable follows a normal distribution after a suitable power transformation is fulfilled. The LMS method, however, is not suitable to model a medical parameter that depends on more than one covariate, i.e. age and height or height and weight, or if one wants to explicitly model kurtosis, which can only be indirectly modeled by the LMS method via the shape and scale parameters. In these cases one may want to opt for GAMLSS, which is a generalization of the presented LMS method.

GAMLSS

Generalized additive models for location scale and shape (GAMLSS) are very flexible regression models. GAMLSS are not restricted to response variables whose distribution belongs to an exponential family as e.g. the Generalized Additive Models (GAM) or Generalized Linear Models (GLM). A large number of distributions can be modeled, including distributions that are highly skew and/or kurtotic. The following paragraphs provide a short introduction and are based on Stasinopoulos and Rigby (2007).

Let us assume for $i = 1, 2, \ldots, n$ observations y_i to originate from independent random variables Y_i with probability (density) function $f(y_i|\boldsymbol{\theta}_i)$ conditional on a vector of four distribution parameters $\boldsymbol{\theta}_i^T = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4}) = (\mu_i, \sigma_i, \nu_i, \rho_n)$. Although GAMLSS is not restricted to distributions defined by up to four parameters, for most applications up to four distribution parameters will be sufficient, where μ_i and σ_i are usually characterized as location and scale parameters and ν_i, ρ_i as shape parameters, e.g., skewness and kurtosis. Following the definition of GAMLSS by Rigby and Stasinopoulos (2005), let $\mathbf{y}^T = (y_1, y_2, \ldots, y_n)$ be the vector of the response variable of length n. Let us further assume that the distribution parameters can be considered as functions of the potential covariates. Then, $g_k(.)$ for $k = 1, 2, 3, 4$, denote known monotonic link functions, which describe the functional relationship between the distribution parameters $(\mu_i, \sigma_i, \nu_i, \rho_i)$ and the J_k covariates and random effects by

$$g_k(\boldsymbol{\theta_k}) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk} \boldsymbol{\varphi}_{jk}, \tag{5.7}$$

where for $j = 1, \ldots, J_k$ and $k = 1, 2, 3, 4$ let us denote with

1. $\boldsymbol{\beta}_k^T = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{J_k' k}) \in \mathbb{R}^{J_k'}$, $J_k' \in \mathbb{N}$, the parameter vector,
2. $\mathbf{X}_k \in \mathbb{R}^{n \times J_k'}$ a fixed known design matrix,
3. $\boldsymbol{\varphi}_{jk}$ a q_{jk} dimensional random variable,
4. $\mathbf{Z}_{jk}$ a fixed known $n \times q_{jk}$ design matrix,
5. $\boldsymbol{\eta}_k$ the linear predictor.

With $\boldsymbol{\theta_1} = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T, \boldsymbol{\theta_2} = \boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^T, \boldsymbol{\theta_3} = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^T, \boldsymbol{\theta_4} = \boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^T$ we get

$$\begin{aligned}
g_1(\boldsymbol{\mu}) &= \boldsymbol{\eta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \boldsymbol{\varphi}_{j1}, \\
g_2(\boldsymbol{\sigma}) &= \boldsymbol{\eta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \boldsymbol{\varphi}_{j2}, \\
g_3(\boldsymbol{\nu}) &= \boldsymbol{\eta}_3 = \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \boldsymbol{\varphi}_{j3}, \\
g_4(\boldsymbol{\rho}) &= \boldsymbol{\eta}_4 = \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \boldsymbol{\varphi}_{j4},
\end{aligned}$$

where $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\rho}$ and $\boldsymbol{\eta}_k$ are vectors of length n. $\boldsymbol{\varphi}_{jk}$ is assumed to be distributed as $\boldsymbol{\varphi}_{jk} \sim N_{q_{jk}}(\mathbf{0}, \mathbf{G}_{jk}^{-1})$ with $\mathbf{G}_{jk}^{-1}$ being the (generalized) inverse of a $q_{jk} \times q_{jk}$ symmetric matrix $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\chi}_{jk})$. $\mathbf{G}_{jk} = \mathbf{G}_{jk}(\boldsymbol{\chi}_{jk})$ may depend on a vector of hyperparameters $\boldsymbol{\chi}_{jk}$ and if $\mathbf{G}_{jk}$ is singular then $\boldsymbol{\varphi}_{jk}$ is understood to have an improper prior density function proportional to $\exp\left(-\frac{1}{2}\boldsymbol{\varphi}_{jk}^T \mathbf{G}_{jk} \boldsymbol{\varphi}_{jk}\right)$.
For (5.7) every distribution parameter can be modeled as a linear function of explanatory variables and/or as linear functions of stochastic variables (random effects). An important special case of this very general definition gives the semi-parametric additive formulation of GAMLSS. Let $\mathbf{Z}_{jk} = \mathbf{I}_n$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\boldsymbol{\varphi}_{jk} = \mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ for all combinations of $j = 1, \ldots, J_k$ and $k = 1, \ldots, 4$. Then (5.7) is simplified as follows to the *semi-parametric additive* formulation of GAMLSS

$$g_k(\boldsymbol{\theta_k}) = \boldsymbol{\eta}_k = \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}(\mathbf{x}_{jk}), \tag{5.8}$$

where

1. $\mathbf{x}_{jk}$ for $1, 2, \ldots, J_k$ are vectors of length n,
2. h_{jk} is an unknown function of the explanatory variable X_{jk}
3. $\mathbf{h}_{jk} = h_{jk}(\mathbf{x}_{jk})$ is the vector which evaluates the function h_{jk} at $\mathbf{x}_{jk}$.

With $\boldsymbol{\theta_1} = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T, \boldsymbol{\theta_2} = \boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^T, \boldsymbol{\theta_3} = \boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^T, \boldsymbol{\theta_4} = \boldsymbol{\rho} = (\rho_1, \ldots, \rho_n)^T$ we get

$$\begin{aligned}
g_1(\boldsymbol{\mu}) &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} h_{j1}(\mathbf{x}_{j1}), \\
g_2(\boldsymbol{\sigma}) &= \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_1} h_{j2}(\mathbf{x}_{j2}), \\
g_3(\boldsymbol{\nu}) &= \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_2} h_{j3}(\mathbf{x}_{j3}), \\
g_4(\boldsymbol{\rho}) &= \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} h_{j4}(\mathbf{x}_{j4}).
\end{aligned} \tag{5.9}$$

The semi-parametric additive formulation of the GAMLSS (5.9) is most common and is implemented in the *R* package *gamlss*, in addition to the more general definition of (5.7). With (5.9) $h_{jk}(\mathbf{x}_{jk})$ can be modeled as natural cubic spline(s) of the covariate(s). Similar to the LMS method smoothness is ensured by maximizing the penalized likelihood, which is given by

$$l_p = l - \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{J_k} \boldsymbol{\varphi}_{jk}^T \mathbf{G}_{jk} \boldsymbol{\varphi}_{jk} \tag{5.10}$$

where $l = \sum_{i=1}^{n} \log(f(y_i|\boldsymbol{\theta}_i))$ is the log-likelihood function of the data given $\boldsymbol{\theta}_i$ for $i = 1, \ldots, n$; see Rigby and Stasinopoulos (2005) and Stasinopoulos and Rigby (2007) for details. GAMLSS is able to cope with a large number of distributions ranging from discrete one parameter distributions, like the binomial or Poisson distributions, to continuous distributions with four model parameters like the *Box-Cox power exponential* (BCPE), which was introduced by Rigby and Stasinopoulos (2004). A (non-conclusive) list can be found in Stasinopoulos and Rigby (2007). The only restriction on the distribution is that $f(y|\boldsymbol{\theta})$ and its first and second (and cross-) derivatives with respect to each element of $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \theta_3, \theta_4)$ need to exist, either in explicit form or as numerical derivatives.

Remark 5.1. For our application in the section below, we assume that Y_n are independent identically distributed with distribution parameters $(\mu, \sigma, \nu, \rho) \in \mathbb{R}^4$. Following the notation as introduced by Rigby and Stasinopoulos (2005), we will denote a model where the response variable Y follows a BCPE-distribution with (1) the location parameter μ modeled using the identity link as a cubic smoothing spline ($\mathrm{cs}(x, 3)$) with three degrees of freedom in x, i.e. age, combined with the linear term in x and (2) the scale parameter σ being modeled by a log-linear model in x and (3) ν and (4) ρ being modeled using a constant model expressed as 1 (in case of ρ on the log-scale), in short, with

$$Y \sim \mathrm{BCPE}\{\mu = \mathrm{cs}(x, 3), \log(\sigma) = x, \nu = 1, \log(\rho) = 1\}.$$

5.3.2 Objectively measured physical activity in European children

During the IDEFICS surveys T0 and T1 free-living physical activity was assessed using Actigraph GT1M and Actitrainer devices set to 15s epochs (see Section 5.1). However, in some of the IDEFICS centers 60s epochs were used.

Therefore, it was decided to reintegrate the data collected at 15s epochs to 60s epochs. Non-wearing time was defined as 20 minutes or more of consecutive zero counts and was removed for analysis. Minimum wearing time was set to at least 8 hours of valid time per day. In order to be included into the analysis at least one valid weekday and one valid weekend day were required. In total 7,684 children met the inclusion criteria (3,842 boys and 3,842 girls). Activity ranges were assigned to the accelerometer counts using the cutpoint method (Section 3.2) with Evenson cutpoints (Evenson et al., 2008). For the following dependent variables percentile curves were derived: 1) average counts per minute (CPM) that is sum of daily counts divided by valid time, 2) time spent in at least moderate activity (MVPA), 3) light activity (LPA) and 4) sedentary time (SED). For MVPA, LPA and SED the unadjusted as well as adjusted minutes were analyzed. To obtain adjusted minutes, unadjusted (raw) minutes were divided by wearing time for each wearing day and the resulting fraction was multiplied by the average wearing time across all valid days.

Statistical analysis

We used the *gamlss package* (version 4.2-6) of the statistical software R (version 3.0.1) (R Core Team, 2015). Different distributions were fitted to the observed distribution of physical activity variables as the Box-Cox power exponential (BCPE), the Box-Cox Cole and Green (BCCG), the Box-Cox t, the normal, the power exponential and the t family distribution. Age was modeled either as a constant, as a linear function, or as a cubic spline. Goodness of fit was assessed by the Bayesian Information Criterion (BIC) and Q-Q plots. As a result, percentile curves for the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles were calculated based on the model that showed the best goodness of fit (Cole et al., 2009; Stasinopoulos and Rigby, 2007). For comparative purposes, it is, however, beneficial, if the same distribution is used for all dependent variables. The BCCG distribution (Cole and Green, 1992) turned out to be the most appropriate distribution according to the BIC in most cases. In all other cases, the difference from the best fitting distribution in terms of BIC was negligible. Following the notation introduced in Remark 5.1, Table 5.1 presents the fitted GAMLSS for the physical activity data.

Variable	Distribution	$\log(\mu)$	$\log(\sigma)$	ν	Sex
CPM	BCCG	$\text{cs}(age, 3)$	$\text{cs}(age, 3)$	1	both
Adjusted MVPA	BCCG	$\text{cs}(age, 3)$	$\text{cs}(age, 3)$	1	boys
Adjusted MVPA	BCCG	$\text{cs}(age, 3)$	1	1	girls
Unadjusted MVPA	BCCG	$\text{cs}(age, 3)$	1	1	both
Adjusted LPA	BCCG	$\text{cs}(age, 3)$	1	1	both
Unadjusted LPA	BCCG	$\text{cs}(age, 3)$	1	1	both
Adjusted SED	BCCG	$\text{cs}(age, 3)$	1	1	both
Unadjusted SED	BCCG	$\text{cs}(age, 3)$	1	1	both

Table 5.1: Fitted GAMLSS for physical activities and sedentary behavior

As described in Section 5.3.1 one major advantage of GAMLSS is that distributions can be fitted, which allow the explicit modeling of kurtosis, like e.g. the BCPE distribution. However, here the BCCG distribution showed the best fit for the physical activity data. So in retrospective the LMS method by Cole, as presented in the beginning of Section 5.3.1, would have been sufficient to capture the structure of our data. But this was a priori unknown and was only confirmed by the more sophisticated analysis exploiting GAMLSS.

Results

Based on the fitted distributions listed in Table 5.1 smoothed percentile curves $P_5, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}$ and P_{95} were calculated as reference ranges for physical activity in European children (see Figures 5.2 to 5.4). In general boys show higher values for CPM and MVPA, while sedentary time was higher for girls. No differences can be seen with regard to light activities. The percentile curves show similar trends with increasing age for both sexes. Average sedentary behavior increases with age from about 240 minutes per day (min/day) at age 3 to about 380 min/day at age 10. At the same time an decrease from 410 min/day to 360 min/day can be seen in LPA. MVPA increases with age. Starting with 24 min/day on average in boys and girls, the time spent in MVPA doubles in boys until the age of 10. For girls an increase can be seen as well, but only to 35 min/day at the age of 10.

Children are recommended to perform MVPA activities for at least 60 minutes a day. To investigate how many children follow this recommendation, we

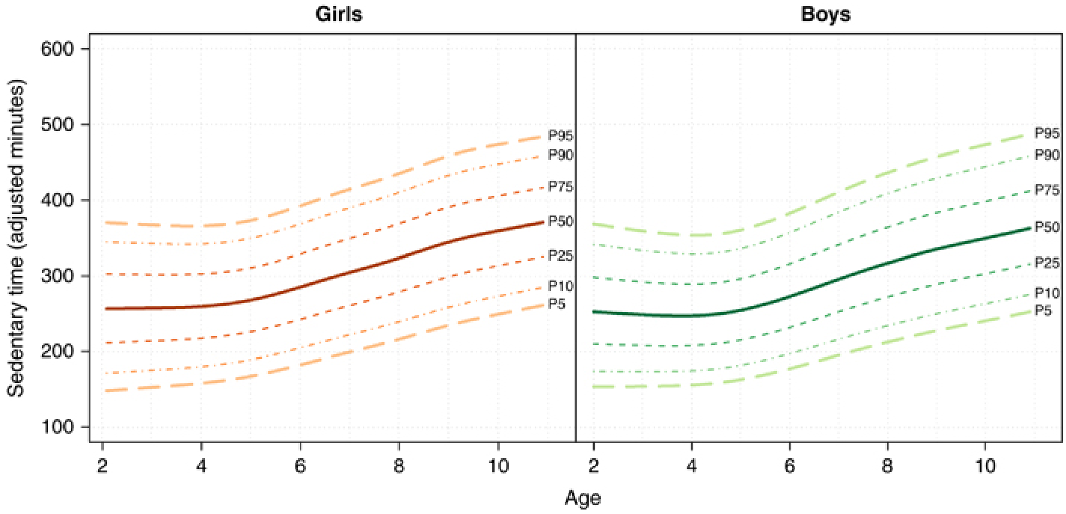

Figure 5.2: Percentile curves: adjusted SED for European boys and girls (Konstabel et al., 2014).

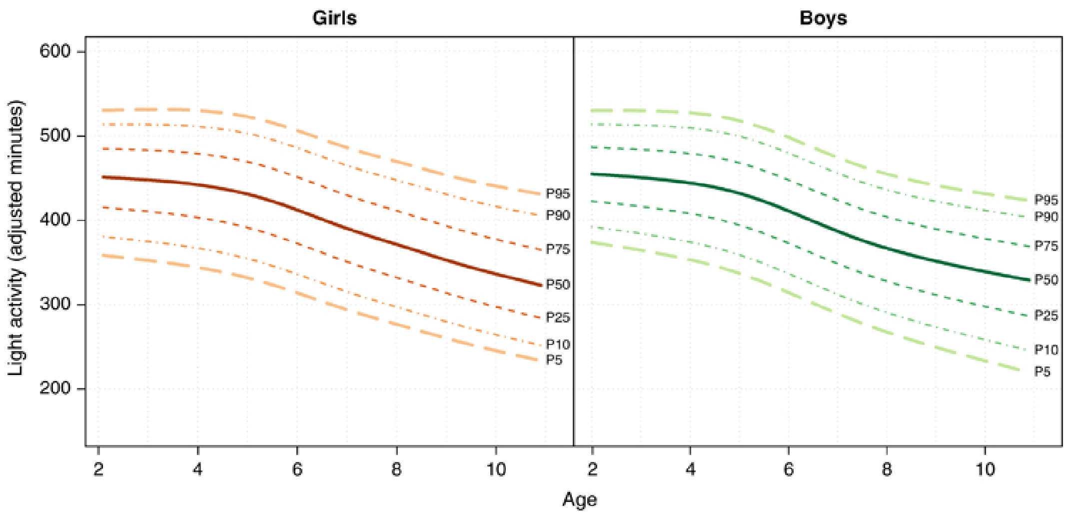

Figure 5.3: Percentile curves: adjusted LPA for European boys and girls (Konstabel et al., 2014).

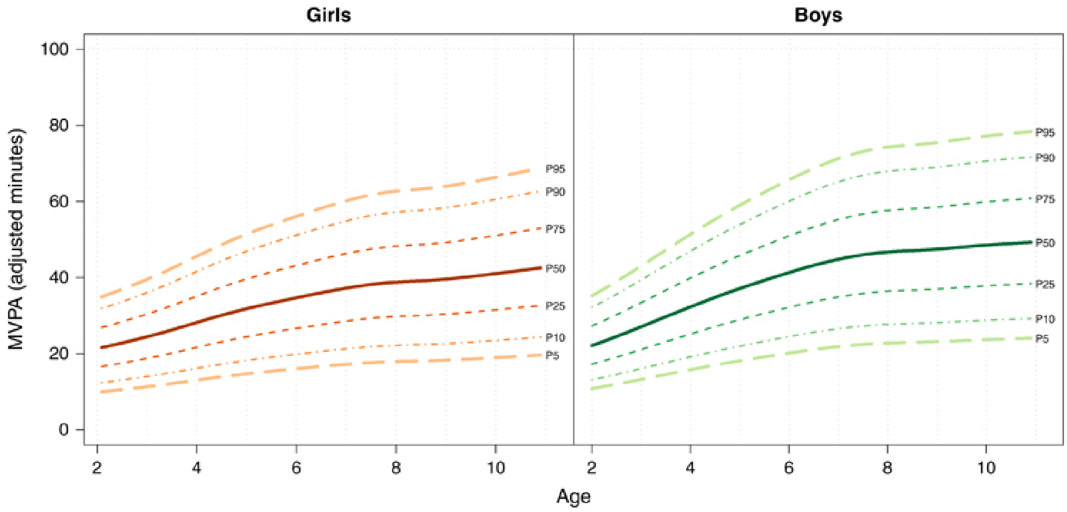

Figure 5.4: Percentile curves: adjusted MVPA for European boys and girls (Konstabel et al., 2014).

calculated the corresponding percentage of children in the IDEFICS study. General compliance was low with proportions ranging from 2.0% (Cyprus) to 14.7% (Sweden) in girls and from 9.5% (Italy) to 34.1% (Belgium) in boys. For detailed results and an extensive discussion see Konstabel et al. (2014).

5.4 Association of physical activity with specific endpoints

5.4.1 Longitudinal association of objectively measured physical activity behavior and obesity in European children

Physical activity is generally considered as being beneficial for body composition, that is high levels of physical activity lead to low body fat mass and a healthy body mass index (BMI) in general. Numerous studies, mostly cross-sectional, investigated the relationship between obesity and physical activity (Jimenez-Pavon et al., 2010; Rauner et al., 2013). There is general agreement that physical activity is negatively associated with obesity (Jimenez-Pavon et al., 2010).

Cross-sectional studies can only study associations rather than causality. To the best of our knowledge, only few longitudinal studies with small sample sizes investigated the interaction of objectively measured physical activity levels, fat mass and fat free mass (FFM) (Jimenez-Pavon et al., 2013). Until today it remains unclear whether physical activity leads to a reduction of fat mass or if fat mass hinders being physically active (Metcalf et al., 2010; Ekelund et al., 2014). Recently, the Ballabeina (Bürgi et al., 2011) and EarlyBird (Metcalf et al., 2010) studies investigated the interaction of objectively measured physical activity and fat mass.

Metcalf et al. (2010) investigated in a sample of about 200 children whether inactivity is the cause of fatness or fatness the cause of inactivity. This research was part of the EarlyBird study in which children were visited yearly from age 7 to age 10. Physical activity was assessed using Actigraph accelerometers on seven consecutive days. Total physical activity (TPA) as counts per week and minutes spent in MVPA were analyzed. Body fat per cent (BF) was measured by dual x-ray absorptiometry. The authors used partial correlation coefficients to compare baseline versus change to follow-up associations in order to examine the direction of association. First, the authors looked at cross-sectional associations for the four surveys, adjusted for age and sex (e.g. $TPA_{7y} = sex + age_{7y} + BF_{y7}$). Second, so-called time-lagged associations of physical activity on future BF measured 1, 2 and 3 years later adjusted for the earlier measurement were modeled, as well as the reverse association, i.e. the influence of BF on future physical activity (e.g. $TPA_{10y} = sex + age_{7y} + BF_{y7}$). Third, changes in physical activity and BF were calculated for each 1-, 2- and 3-year period. Partial correlation coefficients were then calculated for the predictor at a single point in time and the change in the outcome variable from that point in time to a 1-, 2- or 3-year follow-up. The authors adjusted for the outcome measure at the earlier point in time (e.g. $TPA_{10y} - TPA_{7y} = sex + age_{7y} + TPA_{7y} + BF_{y7}$). In this study, BF was predictive of changes in physical activity, but physical activity levels were not predictive of changes in BF. The authors concluded that physical inactivity seems to be the result of fatness rather than its cause. See Metcalf et al. (2010) for the complete study and results.

This result was also confirmed by the Ballabeina study: while children with higher body fat at baseline were observed to be less active at follow-up, baseline

	Cross-sectional			
	T0		T1	
	r	(95%-CI)	r	(95%-CI)
CPM	-0.05	(-0.09,-0.01)	-0.07	(-0.11,-0.04)
MVPA	-0.05	(-0.09,-0.01)	-0.08	(-0.12,-0.04)
VPA	-0.11	(-0.15,-0.08)	-0.16	(-0.2,-0.13)
	Time-lagged			
	PA T0 vs. z-FMI T1		z-FMI T0 vs. PA T1	
	r	(95%-CI)	r	(95%-CI)
CPM	-0.06	(-0.1,-0.02)	-0.07	(-0.11,-0.03)
MVPA	-0.05	(-0.09,-0.01)	-0.07	(-0.11,-0.04)
VPA	-0.1	(-0.14,-0.06)	-0.15	(-0.18,-0.11)
	Change in outcome			
	PA T0 vs Δz-FMI		z-FMI T0 vs. ΔPA	
	r	(95%-CI)	r	(95%-CI)
CPM	-0.02	(-0.06, 0.01)	-0.06	(-0.1,-0.03)
MVPA	-0.01	(-0.04, 0.03)	-0.06	(-0.1,-0.02)
VPA	-0.01	(-0.04, 0.03)	-0.11	(-0.14,-0.07)

r = Spearman's partial correlation coefficient, adjusted for sex and age.
CI = confidence interval; Δ = change;

Table 5.2: Preliminary results on the association between physical activity and BF based on follow-up data of the IDEFICS study.

physical activity did not reduce fat mass at follow-up (Bürgi et al., 2011).
Using the data collected in the IDEFICS study (Section 5.1), we tried to reproduce these findings. Body composition was assessed using the z-score of the fat mass index (FMI), which is the fat mass in kg divided by the squared height in m, abbreviated as z-FMI, the z-score of waist circumference (z-waist) (Nagy et al., 2014, corrected version to be published in 2016) and FFM. Physical activity was measured using Actigraph accelerometers for at least three days of at least eight hours wearing time of which at least one day was a weekend day. We used CPM and time spent in MVPA and VPA as components of physical activity. About 3,000 children had valid accelerometer measurements at T0 and T1, which is a requirement to calculate the changes between points in time. Using the same methods as Metcalf et al. (2010), we were able to derive similar results based on our data. Table 5.2 presents the resulting partial correlation coefficient for z-FMI. Like Metcalf et al. (2010) we observe statistically significant cross-sectional and time-lagged associations. When looking at the correlations of the changes, only z-FMI versus change in physical activity is significant. Hence, our findings support the results of Metcalf et al. (2010). But as a limitation, in the IDEFICS study only two points in time can be considered.

In order to increase the number of observations, data from the I.Family study (Section 5.2) are used to augment the IDEFICS study data. Both studies combined include about 3,600 participants with two valid accelerometer measurements and nearly 1,000 participants with three valid measurements. In order to use as many observations as possible, we will use multi-level models (MLM) to assess the direction of the association.

In a first step MLMs will be used to derive a random slope over time for each participant's exposure, e.g. MVPA. In a second step this result will be inserted into another MLM as predictor for the outcome, e.g. z-FMI at a later point in time, preferably T3. A paper is currently being prepared.

One could also argue that comparing the change in z-scores with the "raw" minutes spent in an intensity level is incorrect. The z-scores are by construction adjusted for age and sex, that means that a participant with a change of 0 maintains his/her position relative to the reference population, even if his/her "raw" FMI value changes. This is not true for e.g. minutes spent in MVPA.

As can be seen in Figure 5.4 MVPA increases over time, as does the variance. So a participant who is at P_{50} at age 4 can still be at P_{50} at age 6, hence his/her activity level has not changed relative to the population, yet he/she has a change of 10 minutes MVPA. A participant at P_5 can also remain at this percentile and has only a change of about 5 minutes MVPA.
In order to address these concerns, one may, aside from calculating the physical activity z-scores, use MLMs to consider the daily accelerometer measurements. So accelerometer measurements nested within individuals nested within countries may be modeled, rather than simply using the mean of the daily measured physical activity. Another alternative would be to use the best linear unbiased predictor (BLUP) to combine the daily measurements to one value as suggested by Olive et al. (2012) and Stanek 3rd et al. (1999) as the BLUP allows to consider the inter- as well as the intra-individual variability of the measurements.
Another idea is to use path models to further investigate the association between physical activity and body composition. If fatness leads to inactivity, which increases fatness further reducing physical activity, then this would be a classic vicious circle. Here, the path model may discover the best starting point to break this circle.

5.4.2 Incidence of high blood pressure in children - Effects of physical activity and sedentary behaviors

High blood pressure (HBP) is known to be one of the most important risk factors for cardiovascular diseases. De Moraes et al. (2015) determined the incidence of pre-HBP and HBP and analyzed the effect of physical activity and sedentary behavior on pre-HBP and HBP. Of 16,228 children participating in the baseline survey T0 of the IDEFICS study only a subset of 5,221 children provided information on the primary outcome HBP and the main exposures physical activity and sedentary behavior, as well as potential confounders. As this information were required for T0 and T1, a total of 5,061 participants were included in the analyses. Pre-HBP and HBP were defined according to the National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents (NHBPEP, 2004). Minimum accelerom-

eter wearing time was at least 6 h/d for at least 3 days (2 weekdays and 1 day of weekend/holiday). The sampling interval (epoch) was set to 15 seconds. Total volume of activity was expressed as the sum of recorded counts divided by total daily registered time expressed in minutes (counts/min; cpm). All children were categorized as fulfilling the current recommendation of $\geq$ 60 min MVPA per day or not. Another variable categorizing the change in physical activity from T0 to T1 was created with values (1) meeting the recommendation in T0 and T1; (2) meeting the recommendations in T0 but not T1; (3) meeting recommendation only in T1 and (4) meeting the recommendation neither in T0 nor T1. Sedentary behavior was assessed using a proxy report on activity behavior that was completed by the children's parents. Reported daily TV/DVD/video and computer/games-console use were summed up to obtain the total screen time per day. Finally the children were categorized according to the American Academy of Pediatrics: Committee on public education (AAoPCoPE, 2001) as having either $\leq$ 120 min screen time per day or more. In addition, a variable describing the change in sedentary behavior from T0 to T1 was derived.
Cumulative incidence was calculated for the two years of follow-up with 95% confidence intervals (CI) for both outcomes: Pre-HBP and HBP for total physical activity (fulfilling the recommendation of $\geq$ 60 min MVPA per day), categorized screen time and categorized change in physical activity. The magnitude of these associations was subsequently expressed as unadjusted and adjusted relative risk (RR) and 95% CI. Multinomial multilevel regression models using random intercepts were applied to estimate the effect of physical activity and sedentary behavior on pre-HBP and HBP incidence. The survey center was used as level in this model.
The incidence of pre-HBP per year was found to be 121/1000 children and 110/1000 children per year for HBP. Children, who had a reported screen time $>$ 120 min per day at T0 and T1, showed an RR of having HBP of 1.28 (1.03-1.60). For T1 an elevated RR of 1.53 (1.12-2.09) for having HBP can be seen for children not meeting the physical activity recommendation of at least 60 min MVPA per day. No association between pre-HBP and the considered behaviors was found. In conclusion, it can be stated that the incidence of pre-HBP and HBP is high in European children and that maintaining sedentary behaviors

during childhood increases the risk of developing HBP. For the detailed results and their discussion see Moraes et al. (2015).

5.5 The PATREC study

The PATREC study is a cross-sectional study which was carried out during the school year 2012/2013 (September 2012 - February 2013). The aim of this study was to evaluate the comparability of different instruments used to objectively and subjectively measure physical activity. 542 students from two primary and two secondary schools in the city of Bremen were invited to participate in four different modules: (1) wearing an Actigraph accelerometer (GT3X+, GT1M, ActiTrainer; Pensacola, Florida, USA) for seven consecutive days during waking hours, except when taking a bath and swimming, and to complete an activity diary. The students were asked to record sports clubs and physical education time frames in this diary, as well as non-wearing periods and their reasons. The accelerometers were attached with an elastic belt on the right hip. Accelerometer data were stored at 3-second-epochs and computed with the ActiLife 6 software. Non-wearing periods were defined according to Choi et al. (2011) also using the ActiLife 6 software. Cutpoints from Evenson et al. (2008) for SED, LPA and MVPA were applied; (2) a 7-day-recall questionnaire at the last day of the accelerometer-wearing-period with 12 items covering different domains and dimensions of physical activity and sedentary behavior; (3) a questionnaire on the habitual physical activity; and (4) a fitness test. The body mass index (BMI) was calculated by self-reported height and weight.

Participating students were required to ask for written parental consent. In addition, older students (11-17 years) had to give their written consent, while oral consent was obtained from the younger students (6-10 years). The study was approved by the ethical committee of the University of Bremen.

5.5.1 Domain-specific self-reported and objectively measured physical activity in children.

As was discussed in Section 2.1.5, the agreement of subjectively and objectively measured physical activity is generally low or moderate at best. Little is known about the extent different domains contribute to total physical activity and sedentary behavior. According to the SLOTH (sleep, leisure, occupation, transportation, household) model (Pratt et al., 2004), opportunities for children and adolescents to accumulate physical activity can be assigned to five domains: sleep, transportation, school time, leisure time and home. Among these domains school can be considered as a very important domain, as children and adolescents spend about half of their waking hours there (Bailey et al., 2012; Escalante et al., 2014; Guinhouya et al., 2009b).
Epidemiological studies commonly use questionnaires and proxy reports to assess physical activity and sedentary behavior by asking children or their parents to report the duration of outdoor playing time, organized sports activities or electronic media consumption. These domain-specific activity variables are often thought to sufficiently describe the physical activity behavior (Pratt et al., 2004; Trost, 2007).
Sprengeler et al. in a forthcoming publication presented in Appendix E studied the agreement of self-reported and objectively measured physical activity and sedentary behavior in specific domains (transportation, school time, physical education, leisure time and organized sports activities). Additionally, the contribution of these domains to total SED, LPA and MVPA was investigated by combining accelerometry (and an activity diary) with a questionnaire.

Methods

A modified version of the validated German MoMo-(Motoric Module) physical activity questionnaire (PAQ) was used, which is designed to assess habitual physical activity (Jekauc et al., 2013). The modified PAQ consists of 12 questions assessing five domains of physical activity (transport, school time, physical education, leisure time and organized sports activities). Children up to the age of 10 completed the questionnaire with their parents (proxy-reported), while the older students completed the questionnaire by themselves. For each domain, frequency, duration and intensity of physical activity were assessed. Self-reported SED was defined as the duration of sedentary activities (e.g. TV viewing, sitting during school hours). The intensity of physical activity was assessed as participants' self-perceived, typical intensity of breathing and sweating during physical activity.

Accelerometer data were used for analyses, if at least ten hours per day of valid wearing time for at least three days were available. Non-wearing time was removed according to Choi et al. (2011) and intensity levels assigned using cutpoints from Evenson et al. (2008). Using the information provided in the accelerometer diary and the PAQ, objectively measured SED, LPA and MVPA were each assigned to the five domains transport, school time, physical education, leisure time and organized sports activities. Dates and time of physical education and organized sports activities were linked to the accelerometer counts based on the information in the activity diary. School time was defined as the interval between the start and end of school. The time of physical education classes was excluded from school time and assessed separately. Transport was defined as the provided self-reported transport duration plus five additional minutes before start and after end of school. All other accelerometer counts were assigned to leisure time.

542 children and adolescents were invited to participate in the study and written consent was obtained from 390 (72%). At baseline, accelerometer and questionnaire data were available from 371 participants. Of these 298 provided valid questionnaire and accelerometer data (adequate valid wearing time on at least two weekdays and one weekend day) for the comparison of subjective vs. objective measurements. Information on school time was only available for children attending the primary schools, aged 6-10 years. Of these, 207 students had at least three valid accelerometer weekdays available for the domain-specific analyses.

Statistical analyses

Accelerometers and PAQs are supposed to be two instruments measuring essentially the same. In this case, it is to be expected that the corresponding measurements are highly correlated, however, a high positive correlation is not sufficient to show that two instruments measure the same (Bland and Altman, 1986). Self-reported and objectively measured minutes of total and domain-specific SED, LPA and MVPA were compared using the Spearman rank correlation coefficient r_S. The correlation coefficient ranges from -1 to 1, with 1 meaning perfect monotone correlation and 0 no correlation. The correlation is considered as weak if $r_S \leq 0.39$, as moderate if $0.40 \leq r_S \leq 0.59$, as strong if $0.60 \leq r_S \leq 0.79$ and as very strong to perfect if $r_S \geq 0.80$.

Results

Self-reported physical activity was generally over-reported compared to accelerometer measured physical activity. The agreement of self-reported and objectively measured physical activity was low for total LPA ($r_S = 0.09$, 95%-confidence interval (CI) = (-0.03, 0.20)), total MVPA ($r_S = 0.21$, CI = (0.10, 0.32)). Moderate agreement was found for total SED ($r_S = 0.44$, CI = (0.34, 0.53)).

Among the domain-specific correlations, moderate agreement could be seen for LPA during transport ($r_S = 0.59$, CI = (0.49, 0.67)) and MVPA during organized sports activities ($r_S = 0.54$, CI = (0.38, 0.67)). About half of total objectively measured SED, LPA and MVPA (55%, 53% and 46%, respectively) occurred during school time, while organized sports activities contributed 24% to total MVPA.

In conclusion, the school setting is the most important domain, contributing about half of total SED, LPA and MVPA in children aged 6-10 years. Accelerometers should be preferred over questionnaires to measure duration and intensity of physical activity. As was known, domain-specific data still require self-reported information. Further results and a discussion will be included in the forthcoming publication presented in Appendix E.

5.6 Energy expenditure using pedometers

Pedometer data were collected and combined with spirometer measurements of the activity energy expenditure (AEE) in 207 participants (110 females) in the age of 8 to 74 years. These were recruited by newspaper announcements and telephone calls. All participants (and their parents/legal guardians if applicable) were carefully instructed and gave written informant consent. All

procedures of the study were approved by the local ethics committee. Participants had to be healthy and free of impairments with an BMI ≤ 27.5 kg m^{-2}
.

Resting energy expenditure (REE) was measured by a portable oxygen analyzer system (MetaMax 3b, Cortex Biophysik, Leipzig, Germany) for 30 minutes under controlled conditions. The same device was used to record AEE for three different walking speeds/intensities. Participants first walked at their preferred speed for eight minutes, which was considered moderate walking. Then participants were asked to walk slowly for eight minutes and afterwards to walk fast, without running for three minutes. Participants younger than 18 years only walked moderately and fast. AEE was computed as the average metabolic equivalent of task (MET) from data measured between minute 5:00 and minute 7:00 for the first two intensities and minutes 1:30 to 3:00 for the last intensity.

Gait cycles per minute for the relevant time frames were recorded using a step activity monitor (StepWatch 3.0, Orthocare Innovations, WA, USA), which was attached to the right ankle of participants. The step activity monitor was previously adjusted to the participant's gait characteristics and steps were stored in 15 seconds epochs by the device (Brandes et al., 2012).

5.6.1 Estimating energy expenditure from gait intensity

Brandes et al. used the collected data in a forthcoming publication presented in Appendix F to derive an energy expenditure equation for the step activity monitor StepWatch 3.0. Here, $\log(AEE)$ was used as dependent variable and gait cycles, body weight, height, age and sex as independent variables. In order to account for the repeated measurements (up to three measurements per participant) mixed linear models were fitted. Several different models based on the predictor variables were calculated. As a way to assess the accuracy of the energy prediction and to compare the different models, we used leave-one-out cross-validation to calculate the root-mean-square-error (RMSE) and mean absolute-percentage-error (MAPE).

Model	Independent variables	Walking		
		r2	RMSE	MAPE
1	Weight	0.183	8.54	39.20
2	Weight + SAM	0.582	6.19	20.40
3	**Weight + height + SAM**	0.644	5.69	18.92
4	**Weight + height + age + SAM**	0.651	5.64	18.76
5	Weight + height + height · SAM + age + age · SAM + SAM	0.651	5.60	19.11
6	Weight + weight · SAM + height + age + age · SAM + SAM · SAM + SAM	0.648	5.66	18.68

RMSE = root-mean-square-error; MAPE = mean-absolute-percentage-error; SAM = gait cycles

Table 5.3: Overview of models for predicting AEE during walking.

k-fold cross-validation

Let us assume for $n = 1, 2, \ldots, N$ observations y_n to originate from independent random variables Y_n. Let $k \in \mathbb{N}$, then the original sample of N observations is divided into k equal sized subsamples. One subsample is kept as the *validation dataset* and the remaining $k-1$ samples (*training data*) are used to derive the energy prediction equation. This in turn is used to calculate a prediction $\hat{y}$ based on the validation dataset.

This cross-validation is repeated k times (the *folds*). This way all data are used for both training and validation, while each single observation is used only once for validation. The RMSE is calculated as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2},$$

being a measure for the accuracy of the prediction. Different models can now be compared by their RMSE, with a smaller RMSE meaning a better prediction. MAPE is also commonly used as another accuracy measure, which can be interpreted as the percentage the prediction differs from the observation on average, calculated as

		Walking			
Model	Independent	Coefficient	(95% CI)	SE	p-value
3	Intercept	-2.9522	(-3.5498, -2.3547)	0.30	<0.0001
	Weight	0.01129	(0.0077, 0.0148)	0.002	<0.0001
	Height	0.01359	(0.0095, 0.0177)	0.002	<0.0001
	SAM	0.04597	(0.0425, 0.0494)	0.002	<0.0001
4	Intercept	-3.1430	(-3.7549, -2.5310)	0.31	<0.0001
	Weight	0.009068	(0.0051, 0.0193)	0.002	<0.0001
	Height	0.01507	(0.0108, 0.0193)	0.002	<0.0001
	Age	0.001891	(0.0003, 0.0035)	0.001	0.021
	SAM	0.04621	(0.0428, 0.0497)	0.002	<0.0001

SAM = gait cycles; CI = confidence interval

Table 5.4: Regression coefficients for selected models of walking AEE.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\%.$$

Please note, that $k = 10$ is a common choice in cross-validation and $k = N$ leads to the special case of leave-one-out cross-validation.

Statistical analyses

A priori the best prediction model to be used is unknown. Various models starting with only gait cycles as predictor for AEE up to complex model with quadratic terms, interactions terms and random effects are possible. In a first step over 160 models representing many of the possible combinations of the predictor variables and e.g. containing indicators for fast walking, age groups rather than age as a continuous variable, about half of the models also allowing for an individual random intercept were considered.

RMSE and MAPE were then calculated using 10-fold cross-validation. Models were compared with regard to RMSE and MAPE and a set of 15 candidate models was selected and then leave-one-out cross-valdiation was conducted to improve RSME and MAPE estimation. As a last step non-significant fixed effects were removed from the models leaving the ones presented in Table 5.3.

Results

Of the 207 participants originally participating in the study 167 provided 452 observations, which were included in the analyses. Table 5.3 shows the selected models after cross-validation, with the simple models 1 and 2 added for comparison. RMSE and MAPE were calculated after transforming the prediction back to the original scale of AEE. As it turned out, the more complex models 5 and 6 only slightly reduced RMSE and MAPE and therefore the simpler models 3 and 4 were selected.

The parameter estimates for model 3 and 4 were obtained using the complete dataset and are shown in Table 5.4 and can be applied to predict AEE given gait cycles measured by the StepWatch 3.0 pedometer.

Chapter 6

Discussion and conclusion

In Chapter 4, hidden Markov models based on the Gaussian and generalized Poisson distribution and the expectile regression using the Whittaker smoother in combination with an L_0-penalty were introduced as new approaches for modeling accelerometer data going beyond the commonly used cutpoint method. Both methods were investigated using simulated accelerometer data in which the true intensity level underlying the observed accelerometer count is known. As has been shown by Witowski et al. (2014) and in Section 4.4 both methods are superior to the cutpoint method with regard to all considered quality measures, except runtime.

In a direct comparison HMMs and expectile regression perform similarly well with a slight advantage of the expectile regression. Expectile regression shows the smallest misclassification rate of all investigated methods, which is the most important quality measure, especially in epidemiological studies, where the total amount of physical activity is of particular interest. Additionally, the expectile regression does not need any prior assumption on the distribution, which the counts follow around the mean intensity level. Even the smoothing parameter δ can be automatically selected. So in summary the expectile regression has proven to be the most flexible method, which needs no prior information for modeling the data. However, there may still be some situations, where HMMs based on a certain distribution are advantageous as discussed in Witowski et al. (2014), like e.g. if the number of different activities performed during the day is of particular interest.

It has to be pointed out that both methods still use cutpoints at the end of

	Mean proportion of physical activity intensity levels (1 second epochs)			
Model	SED (%)	LPA (%)	MVPA (%)	VPA (%)
Simulation	38.0	23.0	39.0	2.8
Cutpoint	38.4	28.2	33.4	4.0
HMM[Gauss]	38.3	35.2	26.5	4.0
Expectile	36.0	27.6	36.4	2.8

SED = sedentary behavior; LPA = light physical activity;
MVPA = moderate-to-vigorous physical activity;
VPA = vigorous physical activity

Table 6.1: Mean proportions of assigned physical activity intensity levels by the cutpoint method, HMMs and expectile regression, compared to the simulated data for 1 second epochs.

the day. The next step to a convincing method for modeling physical activity should consider an approach that does not need to use established cutpoints in order to assign intensity levels and hence AEE to the counts. It is hard to imagine how this can be accomplished just using accelerometer counts without external information on the actual energy expenditure. For example, even if a dense set of expectile or percentile curves is fit to the data, how should a threshold for e.g. MVPA be selected for each participant? Such an approach would imply that each participant spends a certain amount of time in the MVPA range. This assumption can be hardly justified, hence a method to account for externally measured (individual) energy expenditure will be required for the near future.

Both methods suffer from a considerably longer runtime than the cutpoint method, but both have the potential to improve the modeling of physical activity based on accelerometer data, especially in large epidemiological studies, where good estimates of the total time spent in different intensity levels and of the physical activity episodes (bouts) are particularly important.

	Mean proportion of physical activity intensity levels (5 seconds epochs)			
Model	SED (%)	LPA (%)	MVPA (%)	VPA (%)
Simulation	38.1	22.9	39.0	2.9
Cutpoint	38.6	28.0	33.4	3.6
HMM[Gauss]	38.4	27.6	34.0	3.1
Expectile	38.1	22.9	39.0	2.8

SED = sedentary behavior; LPA = light physical activity;
MVPA = moderate-to-vigorous physical activity;
VPA = vigorous physical activity

Table 6.2: Mean proportions of assigned physical activity intensity levels by the cutpoint method, HMMs and expectile regression, compared to the simulated data for 5 seconds epochs.

6.1 Future applications

HMMs and the expectile regression have only been applied to simulated data. The next step will be to apply both methods to the accelerometer data collected during the IDEFICS and I.Family studies (Sections 5.1 and 5.2). Both methods are computationally intensive, so applying them to ten thousands of accelerometer days will take some time, especially since the collected free-living physical activity data have to be cleaned first, that is any non-wearing time has to be removed before processing the data any further. As the assessment of the times spent in different intensity levels should be improved by both methods, the results of the HMM- and expectile-analyzed physical activity has to be compared to the results obtained by the cutpoint method. Tables 6.1 and 6.2 show the mean proportions of physical activity intensity levels as they were simulated and as identified by the cutpoint method, the HMMs and the expectile regression for 1 and 5 seconds epochs in the simulated data from Section 4.4.1. As can be seen, the cutpoint method tends to underestimate the time spent in MVPA, while at the same time overestimating time spent in VPA. This is crucial, because even in very intensive, complex activities like e.g. playing badminton, there will always be counts equal or close to 0 (see Figures 3.1-3.3), which will be assigned to SED by the cutpoint method instead of MVPA, while HMMs and expectile regression will find the underlying mean activity level, so counts even "far away" from then mean intensity level will be assigned to MVPA rather than to SED. The same principle applies to extremely large counts, which are assigned to VPA by the cutpoint method, but to MPA by the HMMs and expectile regression. So applying HMMs and expectile regression to the collected data should result in an increased estimate of the time spent in MVPA and a reduction of time spent in VPA.

This in turn will most likely have an effect on the associations between physical activity and any outcome that can be seen in the IDEFICS and I.Family data. Therefore it may be reasonable to perform statistical analyses for physical activity based on the improved physical activity assessment and to compare the obtained results to the ones derived by the "conventional" cutpoint method. The analysis of longitudinal associations of physical activity and body composition as described in Section 5.4.1 is an obvious candidate for such a comparison. As the cutpoint method tends to underestimate time spent in MVPA, any associations (if any) seen for MVPA should be more pronounced when using HMM- and expectile regression-assessed physical activity. However, associations identified for VPA should become smaller, if just VPA is considered, since it tends to be overestimated by the cutpoint method.

6.2 Going beyond accelerometer assessed physical activity

Recently so-called activity or fitness trackers have become an emerging technology, which address a broad community of leisure time athletes promising to improve a healthy lifestyle by tracking physical activity and various other health parameters. The devices range from simple step counters to fully fledged smartwatches with built in heart rate monitors, GPS and even a pulse oximeter. The concept "step" and the daily goal to reach 10,000 steps are simple to understand and to interpret. All devices allow an instant feedback to the wearer and a wireless download to his/her smartphone and online account. Depending on the producer and the features of the device, summary statistics on the physical activity behavior are portrayed, like steps and distances walked that day and calories burnt. Some devices register the floors climbed during the day and give a detailed time trend of the heart rate over time and during exercise. Sleep duration and quality can be monitored. Some providers created software applications for sports such as swimming, which supposedly registers the swum laps and the required strokes, aside from burnt calories and distances. Another application is supposed to improve training results for cycling using also GPS information. All collected data can be displayed by the smartphone application and is stored for longer periods of time allowing to monitor the training progress over weeks or even months. Some smartphone apps allow the entry of further information regarding weight and body composition as well as consumed foods. Training results can be shared with others via social media, others can also be challenged to certain activities. All devices are designed to be worn in daily life, either clipped to the clothes or worn as wristwatch, hence special attention is given to the design.

These activity trackers have great potential for epidemiological research, although some questions have to be clarified first. Activity trackers could be particularly useful in intervention studies and long-term monitoring. The devices give instant feedback and the smartphone app provides a detailed graphical summary and additional features and advice on how to improve towards an active lifestyle. The devices are made for constant wearing, so it is likely that participants will tolerate the trackers and include them in their daily life, if participants perceive them as providing (instant) benefit. The recorded information is automatically downloaded from the device and linked to an online account, which in turn can be accessed by a researcher given informed consent and that all required data protection regulations are met. This way a lot of information on the physical activity behavior can be collected with a very limited, cost effective logistical effort for very long periods of time.

Accelerometers typically used in epidemiological studies do not offer these advantages. Although modern devices are very small and light they are commonly worn on the hip with an additional elastic belt, which adds some burden to the participants, as it is hard to disguise the device. It is not waterproof, such that is has to be taken off when showering or swimming and may provide discomfort in some situations, like e.g. lying on the side the device is attached to. Accelerometers also do not allow instant feedback to the participants, as the devices have to be collected and the stored data need to be downloaded by special software and have to be processed before the feedback can be generated and mailed to participants. Thus, in epidemiological studies the feedback on physical activity behavior of a participant comes with some delay in an often very simple form providing only limited information. All these issues regularly result in a relatively short observation time with sometimes only moderate compliance.

Activity trackers may lead to a higher compliance and as they are relatively inexpensive they could serve as an incentive for participants. It has to be kept in mind that activity trackers might not be suitable for all age groups. Participants need to be technologically knowledgeable, which might prohibit their use in very young children and the elderly.

But before activity trackers can be widely applied, it has to be investigated, if the quality of the recorded data is similar to that of accelerometer data. It has to be checked, if the data are similarly structured, that is a time series of observations, e.g. steps and heart rate recorded along with time and date and possibly even GPS coordinates. If so, the question remains, if they can be used to derive AEE and hence the intensity of the activity and the burnt calories. Apparently energy prediction equations are used to estimate the burnt calories, but their validity needs to be reviewed. One of the imminent weaknesses of accelerometers is their inability to reliably assess climbing stairs, cycling and swimming. Some manufacturers of activity trackers claim that their devices are able to correctly assess physical activity in such situations, which has to be verified.

If the recorded data can be retrieved as a time series similar to accelerometer data, HMMs and expectile regression may be useful approaches to improve modeling these data. As the information is supposedly very similar to regular pedometers an energy prediction equation could be derived as done in Section 5.6.

Appendix A

Paper: Using Hidden Markov models to improve quantifying physical activity in accelerometer data - A simulation study

Contribution to the manuscript I herewith certify that I conceived and designed the experiments with the other authors, interpreted the results, drafted parts of the manuscript and revised it critically for important intellectual content.
The presented study was published in the peer-reviewed open access journal *PLOS ONE* (Witowski et al., 2014).

PLOS ONE

RESEARCH ARTICLE

Using Hidden Markov Models to Improve Quantifying Physical Activity in Accelerometer Data – A Simulation Study

Vitali Witowski[1,2]*, Ronja Foraita[1], Yannis Pitsiladis[3], Iris Pigeot[1,2], Norman Wirsik[1]

1. Department Biometry and Data Management, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany, **2.** Department of Mathematics and Computer Science, University of Bremen, Bremen, Germany, **3.** School of Sport and Service Management, University of Brighton, Eastbourne, United Kingdom

*wirsik@bips.uni-bremen.de

OPEN ACCESS

Citation: Witowski V, Foraita R, Pitsiladis Y, Pigeot I, Wirsik N (2014) Using Hidden Markov Models to Improve Quantifying Physical Activity in Accelerometer Data – A Simulation Study. PLoS ONE 9(12): e114089. doi:10.1371/journal.pone.0114089

Editor: Jérémie Bourdon, Université de Nantes, France

Received: April 24, 2014

Accepted: November 4, 2014

Published: December 2, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All data files are available from the Dryad database (accession number doi:10.5061/dryad.tq0gt).

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Introduction: The use of accelerometers to objectively measure physical activity (PA) has become the most preferred method of choice in recent years. Traditionally, cutpoints are used to assign impulse counts recorded by the devices to sedentary and activity ranges. Here, hidden Markov models (HMM) are used to improve the cutpoint method to achieve a more accurate identification of the sequence of modes of PA.

Methods: 1,000 days of labeled accelerometer data have been simulated. For the simulated data the actual sedentary behavior and activity range of each count is known. The cutpoint method is compared with HMMs based on the Poisson distribution (HMM[Pois]), the generalized Poisson distribution (HMM[GenPois]) and the Gaussian distribution (HMM[Gauss]) with regard to misclassification rate (MCR), bout detection, detection of the number of activities performed during the day and runtime.

Results: The cutpoint method had a misclassification rate (MCR) of 11% followed by HMM[Pois] with 8%, HMM[GenPois] with 3% and HMM[Gauss] having the best MCR with less than 2%. HMM[Gauss] detected the correct number of bouts in 12.8% of the days, HMM[GenPois] in 16.1%, HMM[Pois] and the cutpoint method in none. HMM[GenPois] identified the correct number of activities in 61.3% of the days, whereas HMM[Gauss] only in 26.8%. HMM[Pois] did not identify the correct number at all and seemed to overestimate the number of activities. Runtime varied between 0.01 seconds (cutpoint), 2.0 minutes (HMM[Gauss]) and 14.2 minutes (HMM[GenPois]).

Conclusions: Using simulated data, HMM-based methods were superior in activity classification when compared to the traditional cutpoint method and seem to be appropriate to model accelerometer data. Of the HMM-based methods, HMM[Gauss] seemed to be the most appropriate choice to assess real-life accelerometer data.

Introduction

Currently physical inactivity is considered a major risk factor for several health disorders like cancer [1], obesity [2], cardiovascular disorders [3], muscular skeletal disorders [4], as well as mental disorders [5]. An appropriate assessment of physical activity (PA) is therefore essential in disciplines like medicine and epidemiology to improve the existing evidence-base. The use of accelerometers as an objective measurement of PA has become the most preferred method of choice in recent years, as modern devices allow high frequency measurements for extended periods of time. These now relatively inexpensive devices collect information known as (impulse-)*counts* and provide information on intensity and duration of PA in an individual.

Counts represent a device-specific numeric quantity which is generated by an accelerometer for a specific time unit (*epoch*) (e.g. 1 to 60 sec). This quantity is proportional to the intensity of the physical activity performed by the subject. The sequence of activities during a day is stored as a time series of counts by the device. The most common approach to derive the pattern of PA and its energy expenditure is to map these counts to a certain number of sedentary and activity ranges, such as sedentary, light, moderate and vigorous activity. The duration of PA within the same activity range is known as *bout* and can be easily extracted from a given sequence of counts. A bout is defined as the time period in which the subject remains within one activity range without changing to another. Activity ranges are separated by thresholds known as *cutpoints*. Cutpoints for different age groups are available for children [6, 7, 8, 9, 10, 11, 12] and adults [13, 14, 15] allowing to assess the overall time spent in these ranges of PA.

While the ease of implementation of this *cutpoint method* is an obvious advantage, this method has certain important disadvantages. Counts are being incorrectly assigned to the wrong activity range, leading to misclassification and thereby to an increase of bouts. In the following we assume that the PA of an individual is composed of a sequence of non-overlapping bouts, i.e. each bout being a discrete activity, which is performed over a period of time. Furthermore, the modes of activity can be represented by a ‘true’ average count level. This assumption is depicted in Figure 1. The person first takes a short walk, after which she/he watches TV, followed by a game of basketball and running afterwards. The solid black lines represent the ‘true’ average count level for each of these activities. For example the short walk at the beginning has a true count level of about $\mu_{2(walking)} = 300$ counts per epoch, which can be understood as the true intensity level of this walk. The counts registered by the accelerometer scatter around this

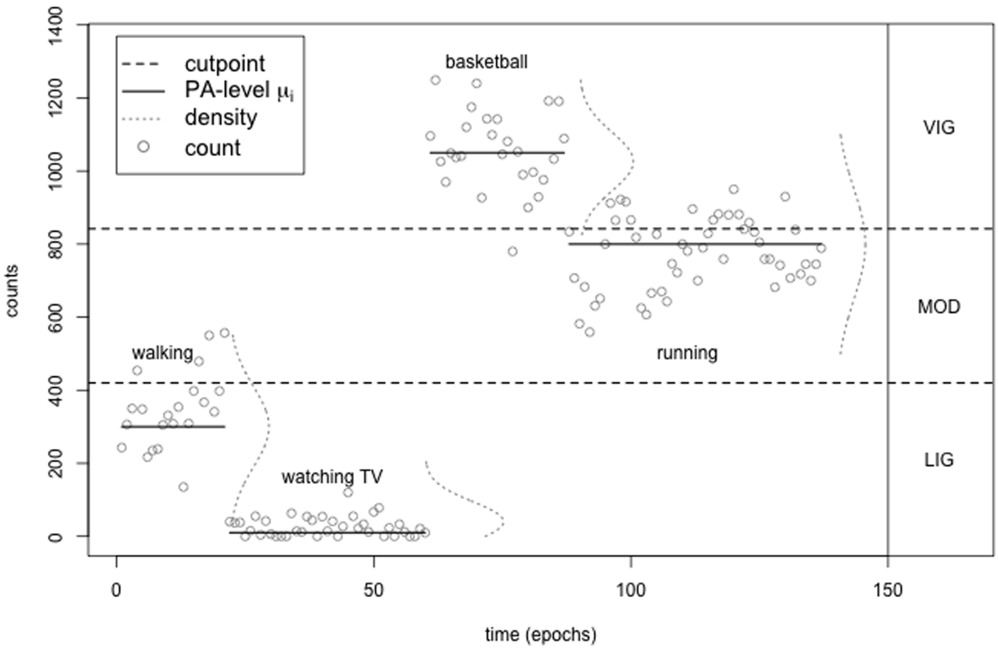

Figure 1. Modeling of accelerometer counts using HMMs. The figure shows the three activity ranges LIG, MOD, VIG, separated by the cutpoints at 420 counts and 842 counts. The accelerometer counts x_t scatter around four different activity states ("watching TV", "walking", "running" and "playing basketball") following a state dependent distribution $X_t|Z_t \sim N(\mu_i, \sigma_i^2)$ with $\theta_i = (\mu_i, \sigma_i^2)$ and fictitious PA-levels $\mu_{1(TV)} = 10, \mu_{2(walking)} = 300, \mu_{3(running)} = 800, \mu_{4(basketball)} = 1050$ respectively.

doi:10.1371/journal.pone.0114089.g001

true level, following a certain distribution (dotted grey line). So the PA depicted in Figure 1 consists of four separate bouts, with four distinct PA-levels μ_1 to μ_4.

As long as the variation around the true intensity level is small and the true level is not close to a cutpoint the complete mode of activity can be correctly assigned to its corresponding activity range. However, in real-life applications the variation of counts and the resulting scattering is large, leading to considerable misclassification of the registered counts into erroneous activity ranges. As a consequence the number of bouts is dramatically increased, as a subject seems to switch from one activity range to another and back again within a few epochs. Therefore, the duration a subject spends in one activity range can be significantly under- or overestimated (cf. numerous validation studies performed to date e.g. [16, 17, 18, 19]).

There has been a number of attempts to resolve the misclassification issue. For example, Pober et al. [20] proposed stochastic models to allow the identification of modes of activity like working at a computer, walking, walking uphill and vacuuming in accelerometer data. A *hidden Markov model* (HMM) was successfully trained to identify these activities. The model correctly identified the activity mode in 80.8% of the data. Vacuuming was correctly identified most frequently in 98.8% of all cases, and walking/walking uphill in about 62%. This approach requires annotated data for training the HMM and activity mode identification is therefore limited to the modes used during training. In order to use this method in free living environments, one would need to train the HMM with all possible activity modes.

As a solution to the misclassification problem caused by large variation of the counts registered by accelerometers, we suggest a new approach that combines the HMM-based method with the traditional cutpoint method. The aim is to provide a better estimate of the activity modes that generated the sequence of counts during the day and by this to decrease the misclassification error, which is inevitably introduced by the cutpoint approach. In order to achieve this, an HMM-based approach was developed to model accelerometer data. 1,000 days of labeled accelerometer data were simulated. HMM models based on the Gaussian, the Poisson and the generalized Poisson distribution were compared with the cutpoint method with regard to misclassification rate (MCR), bout detection, detection of the number of activities performed during the day and runtime.

Methods

Traditional cutpoint approach

The cutpoint method assigns an activity range to each epoch. There are various cutpoints available in the literature. Any of these cutpoints could have been used for our simulation study where we decided to use cutpoints from Pate et al. [9]. According to [9] epochs with <420 counts/15 sec are assigned to light physical activity (*LIG*) with 0–3 *metabolic equivalent of task* (*METs*), epochs with 420–841 counts/15 sec to moderate physical activity (*MOD*) with 3–6 METs and epochs >841 counts/15 sec to vigorous physical activity (*VIG*) with more than 6 METs.

HMM-based approach

An HMM is a stochastic model based on the idea that an observed time series has been generated by an underlying unobservable, time and value discrete, stochastic process whose random variables Z_t are hidden. This sequence of hidden states satisfies the Markov property and forms a Markov chain, i.e. the transition probability to switch from one state to another only depends on the state of interest and is independent of all states prior to *t*. The hidden Markov chain represents a sequence of unobserved random variables Z_t with a finite number of states m. Let $\mathcal{Z} = \{1,...,m\}$ denote the set of possible states and represent the realization of Z_t at point in time t. Each state $i \in \mathcal{Z}$ symbolizes different activities that change from one activity $z_{t+1} = i, i \in \mathcal{Z}$, to another $z_{t+1} = j, j \in \mathcal{Z}$, over time.

The states, however, cannot be observed directly, but they generate a state-dependent output according to a known or presumed probability distribution (see [21, 22] for further details on HMMs). For the purpose of this analysis, the hidden sequence of states is the true, but unknown sequence of PA each subject performed in a free living-environment, while the recorded accelerometer counts x_t are the observed realizations of random variables X_t. The sequence of hidden states satisfies the *Markov property*:

$$P(Z_t = z_t \mid Z_1,...,Z_{t-1},X_1,\ldots,X_{t-1}) = P(Z_t = z_t \mid Z_{t-1} = z_{t-1}),$$

i.e. each activity solely depends on its predecessor. The resulting time series of length $T$of performed physical activities $z_1,...,z_T$ can only be observed indirectly via a parallel time series represented by the assessed accelerometer counts.

The probability of a Markov chain to switch from state i to state j is given by the *transition probability* $\gamma_{i,j} = P(Z_t = j|Z_{t-1} = i)$. A Markov chain is called *homogeneous*, if the transition probability $\gamma_{i,j}$ is independent of t for all pairs of i and $j \epsilon \mathcal{Z}$. The transition probability of a homogeneous Markov chain with finite $\mathcal{Z} = \{1,...,m\}$ can be summarized in an $(m \times m)$ *transition matrix* $\mathbf{\Gamma} = (\gamma_{i,j})_{1 \le i,j \le m}$, with elements of $\mathbf{\Gamma}$ being probabilities and therefore the following conditions have to be fulfilled: and A Markov chain is fully defined by this transition matrix $\mathbf{\Gamma}$ and a vector containing the *initial probabilities* $\pi_0 = (\pi_{01},\ldots,\pi_{0m}) = (P(Z_1 = 1),\ldots,P(Z_1 = m))$ with $\sum_{i=1}^{m} \pi_{0i} = 1$ for the first state.

In the HMM-based approach, each state $i = 1,...,m$ is linked with the mean activity count μ_i of the PA, which the state represents. μ_i denotes the *PA-level* of the PA i. Furthermore, the variable X_t is assumed to be conditionally independent of all remaining variables given its hidden PAZ_t :

$$P(X_t = x_t \mid Z_1,...,Z_t,X_1,\ldots,X_{t-1}) = P(X_t = x_t \mid Z_t = z_t).$$

This means, at each point in time t, the count x_t is assumed to be generated by a certain distribution, which depends on the activity state $z_t = i$, with the corresponding PA-level μ_i as mean of this distribution. The *observation distribution* is the probability that X_t takes a value x_t under the condition that $Z_t = i$. The observation distributions are assumed to be a subset of a whole class of distributions to be specified in advance. Depending on the class of distributions, p_i is determined by $k\epsilon\mathbb{N}$ parameters. These form the *parameter vector* $\theta_i = (\theta_{1i},\ldots,\theta_{ki}) \epsilon \mathbb{R}^k$. The $m \cdot k$ parameters in turn form the matrix $\boldsymbol{\theta} = (\theta_{li})_{1 \le l \le k; 1 \le i \le m}$. An HMM is fully described by its *model-specific parameter* $\mathbf{\Theta} = (\pi_0,\mathbf{\Gamma},\boldsymbol{\theta})$. This setting is illustrated in Figure 1 and depicts the fictitious output of an accelerometer, while the subject performed four activities 'walking', 'watching TV', 'running' and 'playing basketball'. The sequence of activities is assumed to follow a Markov chain, and the accelerometer counts are assumed to be generated by four activity-state-dependent Gaussian distributions with the corresponding PA-levels $\mu_{1(TV)}$, $\mu_{2(walking)}$, $\mu_{3(running)}$ and $\mu_{4(basketball)}$ as their means.

The HMM approach developed for such situations can be subdivided into the following three steps.

Step 1: Building an HMM for an observed time series of counts. The model specific parameters $\mathbf{\Theta}$ of the HMM given an observed time series of counts $x_1,...,x_T$ are estimated. This is referred to as *training of the HMM*. Parameter estimation can either be performed by numerical maximization of the likelihood of the model with respect to $\mathbf{\Theta}$ or by utilizing the so-called *Baum-Welch algorithm* [23] which is commonly used to fit HMMs.

Typically the number of hidden states m (respectively the number of hidden activities) given the counts $x_1,...,x_T$ is unknown. In this case the basic idea is to train several HMMs with different numbers of states m and to evaluate the goodness of fit of the model by the *Bayesian Information Criterion (BIC)* and *Akaike's Information Criterion (AIC)*. If both criteria suggest a different number of states, then one may opt for fewer states to have a more simplistic model or for a larger number of states if this better reflects the underlying practical situation.

Step 2: Decoding the hidden sequence of PA-levels. After the model parameters $\mathbf{\Theta}$ and an appropriate number of physical activities m have been estimated, the resulting HMM is used to link each count x_t to an estimated PA-level $\hat{\mu}_i(i=1,\ldots,m)$.

Step 2.1 First, the *Viterbi algorithm* [24, 25] decodes the globally most likely sequence of hidden activities denoted by $z_1^*,...,z_T^*$ for the trained HMM and the same time series of counts $x_1,\ldots,x_T$ that was used to train the HMM in Step 1 by comparing the joint probability of all T hidden states and the observed accelerometer counts. Alternatively, a local method can be used to decode the most likely hidden activity z_t^*, given all accelerometer counts $x_1,\ldots,x_T$, for each single $t=1,...,T$ by comparing the joint probability of the hidden state at point in time t and the observed accelerometer counts.

Step 2.2 Second, each count x_t is assigned to the estimated PA-level $\hat{\mu}_{z_t^*}$ that corresponds to the decoded state z_t^* at this point in time. Step 2.2 is demonstrated in Figure 2. In this example, the trained HMM with $m=5$ leads to an overfitting of the four activities performed, where the state 'running' is mistakenly split into two different PA-levels by the decoding step.

Step 3: Extension of the cutpoint method. In the last step of our approach, which combines the HMM-based method with the traditional cutpoint approach, each accelerometer count x_t will be assigned via its corresponding (most likely) PA-level $\hat{\mu}_{z_t^*}$ to an activity range a_t, using the traditional cutpoint method.

Overall, the procedure of the new HMM-based cutpoint approach can be summarized as follows:

Step 1: Train the HMM parameters assuming a probability distribution for the counts for each (hidden) PA.

Step 1.1: (optional): Estimate the number of different states m.

Step 2: Decode the hidden sequence.

Step 2.1: Estimate the most likely sequence of states (HMM-decoding): $x_1,\ldots,x_T \rightarrow z_1^*,\ldots,z_T^*$.

Step 2.2: Assign a PA-level (HMM-decoding): $z_t^* \rightarrow \hat{\mu}_{z_t^*}$

Step 3: Assign an activity range (cutpoint method): $\hat{\mu}_{z_t^*} \rightarrow a_t$.

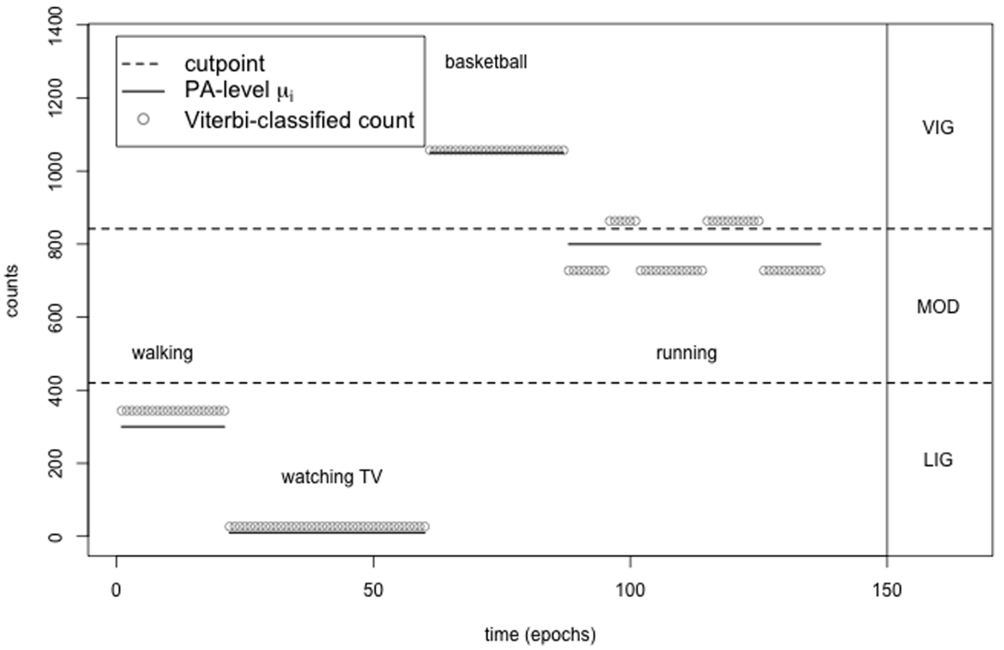

Figure 2. HMM-decoding using the Viterbi algorithm to extract the most likely sequence of physical activity.

doi:10.1371/journal.pone.0114089.g002

In the example illustrated in Figure 2, the trained HMM identifies five PA-levels $\hat{\mu}_1,\ldots,\hat{\mu}_5$, which leads to a misclassification of parts of the state 'running' into five instead of one bout, with two bouts being assigned to the highest activity range. Even with this overestimation of five identified PA-levels instead of four, the HMM-based method assigns most counts correctly to their actual activity range. The high number of bouts typically obtained from the cutpoint method is reduced by the HMM-based approach because a Markov chain is assumed to underlie the performed activities at each point in time. The present example consists of three bouts: the first is defined by the two activities 'walking' and 'watching TV' that correspond to the activity range LIG; the second bout is defined by 'running' in MOD and the third by 'playing basketball' in VIG. Due to the assumed Markov chain, the HMM-based approach detects eight bouts, which is an overestimation of the true value of three, but results are more precise than those obtained from the traditional cutpoint method, which identifies 25 bouts.

The underlying distributions of the states which generate the observed time series are a priori unknown. In the context of modeling accelerometer counts, three distributions are of particular interest: The first HMM is based on the Poisson distribution, which is typically used to model counts. The second model uses the generalized Poisson distribution [26] that includes a further variance parameter to allow for a larger or smaller variation than the one assumed for a standard Poisson distribution. Real-life accelerometer data typically show larger variability than a simple Poisson distribution can accommodate. For the third HMM, a Gaussian distribution is assumed to capture the random scattering of the counts around the presumed PA-level. For the purpose of the present analysis, the Poisson-based HMM is referred to as HMM[Pois], the HMM based on the generalized Poisson distribution as HMM[GenPois] and the Gaussian-based HMM as HMM[Gauss].

Simulation Study

The performance of the traditional cutpoint method is assessed by comparing it with the extended cutpoint method using HMMs in terms of (1) the misclassification rate (MCR), calculated as the percentage of how many of the counts were assigned incorrectly to any other activity range than their true activity range, (2) number of bouts correctly identified, (3) number of activity levels correctly identified, and (4) runtime. For this purpose, *labeled* accelerometer count time series, for which the correct activity range of each count is known, with the length of $T = 1{,}440$ and an epoch length of 15 seconds have been simulated. This particular epoch length and length of T were chosen to reflect typical situations in population-based epidemiological studies [27]. The HMMs does also work with shorter epoch lengths and larger T. Please note that for our simulated, labeled data, the true sequence of activities and therefore the actual PA-level and also the activity range of each count are known. In total, 1,000 different time series were simulated, each representing 6 hours of counts per day (data available under doi:10.5061/dryad.tq0gt). Counts per day were randomly generated using the negative binomial distribution (with parameters $r = 1$ and $p = 0.0009$, resulting in the lowest PA-level $\mu_1 = 111.11$) and the Gaussian distribution (with the parameters and $\mu_4 = 900$, with $\sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 10{,}000$) around three or four pre-defined PA-levels (depending on the random time series generated by a Markov chain), with the lowest PA-level (400) chosen to be very close to the lower cutpoint of 420. To create random activity modes that are time periods spent on the same PA-level, e.g. walking or running, the sequence of PA-levels has been generated using a Markov chain. The simulations were designed to reflect free living-environment observations obtained for children (see Table 1). The simulated data were specifically designed to accommodate cutpoints proposed by [9]. As a large amount of misclassification is expected to occur in activity modes close to a cutpoint, the lowest PA-level (400) was intentionally chosen to be close to one cutpoint, in order to demonstrate the advantage of this method. Any other cutpoints available in the literature could have been chosen, since the application of HMMs does not depend on the choice of the cutpoints. On average, one 6 hour day comprised of 23.66 bouts and 3.97 activities during the day. For data simulation and analysis the R package *HMMpa* [28] was used.

Table 1. Statistical characteristics of the simulated 1,000 data sets (SD=standard deviation).

	Mean	SD	Min	Median	Max
b [bouts]	23.66	7.03	5.00	23.00	47.00
m [activities]	3.97	0.17	3.00	4.00	4.00

doi:10.1371/journal.pone.0114089.t001

Results

Table 2 displays the MCR of all considered methods based on 1,000 simulated days. The cutpoint method shows the highest misclassification rate with about 11% followed by HMM[Pois] with about 8%. HMM[GenPois] and HMM[Gauss] correctly assign 97% and 98% of all counts to their true activity range, respectively. HMM[GenPois] and HMM[Gauss] outcomes are very close to the simulated data of 23.7 bouts with a mean of 31.2 and 32.5 detected bouts, respectively, while HMM[Pois] detects five times and the cutpoint method ten times as many bouts (Table 2). HMM[Gauss] detects the correct number of bouts in 12.8% and HMM[GenPois] in 16.1% of the days.

The proposed methods do not need a priori information on the number of different activities performed during the day. The algorithms identify the most appropriate number *m* by minimizing AIC and BIC. On average, the simulated days have 3.97 different activities. HMM[GenPois] identifies on average 4.19 followed by HMM[Gauss] with 5.18 (Table 2). HMM[GenPois] identifies the correct number of activities in 61.3% of the days, whereas HMM[Gauss] only in 26.8%. HMM[Pois] does not identify the correct number at all.

Mathematical models often have the disadvantage of being numerically instable or having a long runtime. With the exception of the HMM[GenPois], which was numerically instable in 6.9% of the simulated days, all other presented methods converged for the simulated days. Runtime varied from 0.01 seconds (cutpoint) to

Table 2. Misclassification rate, number of identified bouts and identified activities for the traditional cutpoint method and the HMM-based method with different state-dependent observation distributions (SD=standard deviation).

Measure	Method	Mean	SD	Min	Median	Max	Correctly identified [%]
Misclassification rate	Cutpoint	11.14	2.16	5.35	11.18	19.31	88.86
	HMM[Gauss]	1.77	3.53	0.00	0.90	31.94	98.23
	HMM[Pois]	8.21	5.97	1.53	5.56	32.64	91.79
	HMM[GenPois]	3.03	5.58	0.14	1.18	23.06	96.97
Number of identified bouts	Cutpoint	229.55	38.52	129.00	229.00	345.00	0.00
	HMM[Gauss]	32.52	12.84	1.00	31.00	125.00	12.8
	HMM[Pois]	136.43	46.75	37.00	131.00	283.00	0.00
	HMM[GenPois]	31.16	9.86	13.00	31.00	51.00	16.1
Number of identified activities	Cutpoint	–	–	–	–	–	–
	HMM[Gauss]	5.18	0.96	3.00	6.00	6.00	26.8
	HMM[Pois]	5.66	0.47	5.00	6.00	6.00	0.00
	HMM[GenPois]	4.19	0.60	3.00	6.00	6.00	61.8

doi:10.1371/journal.pone.0114089.t002

2.0 minutes (HMM[Gauss]) to 14.2 minutes (HMM[GenPois]) on a regular Windows workstation.

Discussion

This paper investigated the feasibility and the potential advantages of the HMM-based method over the cutpoint approach in identifying the sequence of modes of PA. The results of the simulation study clearly showed the inferiority of the cutpoint method compared to HMM-based approaches. By default the cutpoint method was not able to identify the number of activities performed by a subject. Depending on the specific research question, this may, however, be of particular interest in addition to the correct identification of bouts.

For example, typical recommendations on how much PA children and adolescents need each day suggest 60 minutes (1 hour) or more of physical activity that are age-appropriate, enjoyable and offer variety [29]. While moderate to vigorous activity can be adequately assessed using the cutpoint method, this method leads to a rather rough classification, if one wishes to distinguish the intensities of activities within one activity level, as e.g. between fast walking and slow jogging. Both would be simply assigned to a moderate activity level, whereas HMM-based methods have proven to be much more appropriate for this purpose. The HMM-based method can distinguish those activities and therefore our findings have important implications for the measurement of PA in individuals in free-living conditions for monitoring and surveillance purposes.

Among the HMM-based methods, HMM[Pois] revealed the weakest performance in terms of MCR, bout and activity detection. As anticipated, a simple Poisson distribution cannot accommodate the variance seen in accelerometer data. The results for HMM[GenPois] and HMM[Gauss] were very similar to each other. HMM[Gauss] had a slightly better MCR (1.77% vs. 3.03%), while HMM[GenPois] was better in terms of bout detection (16.1% vs. 12.8%). This is a considerable improvement compared to the cutpoint method, especially if one keeps in mind that a bout is considered as incorrectly identified if the detected bout is just one epoch shorter or longer than the true one. This situation can easily occur at the 'point of discontinuity' when the person switches from one activity to the other. HMM[GenPois] also performed better than HMM[Gauss] and HMM[Pois] with regard to the number of correctly identified activities, which may be particularly relevant for the analysis of accelerometer data. According to the present results, HMM[GenPois] outperforms HMM[Gauss] in this respect as reflected by the considerably higher activity detection rate of 61.3% for HMM[GenPois] compared to only 26.8% for HMM[Gauss]. HMM[Pois] did not identify the correct number at all. As the mean of identified activities is greater than the mean number of simulated activities, combined with the fact, that HMM[Pois] was not able to detect the correct number of activities at all, it can be concluded that HMM[Pois] in general overestimates the number of activities.

However, this outperformance of HMM[GenPois] comes at a price, namely runtime and problems with numerical stability. Even with 6 hour days and 15 seconds epochs ($T = 1{,}440$), HMM[GenPois] needed seven times more runtime than HMM[Gauss]. As modern accelerometers are becoming increasingly powerful, subjects can be monitored for up to 24 hours a day for 7 or more days at 1–5 seconds epochs. This results in a time series of more than $T = 24 \times 7 \times 3600 = 604{,}800$ counts, which will dramatically increase the runtime. For small sample sizes this fact can be disregarded and HMM[GenPois] can be used, but in large cohort studies with more than 10,000 subjects, runtime can become an issue and hence HMM[Gauss] may be preferred.

A simulation study was used here to explore the general feasibility and the potential advantages of the presented HMM-based method, as simulated data have the advantage that the 'truth' is known for every individual count. That means for example that the activity that generated this count and its true intensity level are known for comparative purposes. Using real-life accelerometer data this information would not be available, even if annotated data with measured oxygen consumption would be at hand. In the present study, the data were generated such that one simulated PA-level was close to a cutpoint to investigate whether HMM-based methods are able to correctly identify PA-levels in such situations. Although the simulation study was especially designed for the comparison of methods when using cutpoints from [9], this does not constitute a limitation to the HMM-based approach presented here since it can be easily adapted to any cutpoints proposed in the literature.

Nevertheless, in a next step, the HMM-based methods have to be applied to real-life accelerometer data, where it will be especially interesting to apply these models to annotated data, where the energy expenditure is known. Another promising future application of the presented method is to use HMMs to estimate $\bar{\mu}_{z_t^*}$ as mean PA-level and use the resulting count estimate in energy prediction equations, as e.g. provided by [30]. The HMM-based methods may lead to improved energy expenditure estimates based on better count estimates.

Conclusion

HMM-based methods for modeling accelerometer data are a promising extension of the traditional cutpoint method and on the basis of data presented here ought to improve the analysis of PA. While both HMM[GenPois] and HMM[Gauss] methods seem superior to current cutpoint methods, HMM[Gauss] may be more suitable for real-life applications and if estimation of activity levels is not the main focus. HMM[GenPois] should be used if a better activity and bout detection is desired and runtime is not an issue. Despite these encouraging results, both models will have to be applied to real accelerometer data in future studies in order to prove their superiority over traditional cutpoint method in practice.

Author Contributions

Conceived and designed the experiments: VW RF IP NW. Performed the experiments: VW. Analyzed the data: VW RF YP IP NW. Wrote the paper: VW RF YP IP NW.

References

1. **McTiernan A** (2008) Mechanisms linking physical activity with cancer. Nat Rev Cancer 8(3): 205–211.
2. **Kimm SY, Glynn NW, Obarzanek E, Kriska AM, Daniels SR, et al.** (2005) Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. Lancet 366(9482): 301–307.
3. **Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, et al.** (2012) Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 380(9838): 219–229.
4. **Janz KF, Letuchy EM, Eichenberger Gilmore JM, Burns TL, Torner JC, et al.** (2010) Early physical activity provides sustained bone health benefits later in childhood. Med Sci Sports Exerc 42(6): 1072–1078.
5. **Rethorst CD, Wipfli BM, Landers DM** (2009) The antidepressive effects of exercise: a meta-analysis of randomized trials. Sports Med 39(6): 491–511.
6. **Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG** (2008) Calibration of two objective measures of physical activity for children. J Sports Sci 26(14): 1557–1565.
7. **Freedson PS, Pober D, Janz KF** (2005) Calibration of accelerometer output for children. Med Sci Sports Exerc 37(11): 523–530.
8. **Guinhouya CB, Lemdani M, Vilhelm C, Durocher A, Hubert H** (2009) Actigraph-defined moderate-to-vigorous physical activity cut-off points among children: statistical and biobehavioural relevance. Acta Paediatrica 98(4): 708–714.
9. **Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M** (2006) Validation and calibration of an accelerometer in preschool children. Obesity 14(11): 2000–2006.
10. **Puyau MR, Adolph AL, Vohra FA, Butte NF** (2002) Validation and calibration of physical activity monitors in children. Obes Res 10(3): 150–157.
11. **Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, et al.** (2004) Defining accelerometer thresholds for activity intensities in adolescent girls. Med Sci Sports Exerc 36(7): 1259–1266.
12. **Trost SG** (2007) Measurement of physical activity in children and adolescents. Am J Lifestyle Med 1(4): 299–314.
13. **Freedson PS, Melanson E, Sirard J** (1998) Calibration of the computer science and applications, Inc. accelerometer. Med Sci Sports Exerc 30(5): 777–781.
14. **Sasaki JE, John D, Freedson PS** (2011) Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 14(5): 411–416.
15. **Troiano R, Berrigan D, Dodd K, Masse L, Tilert T, et al.** (2008) Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 40(1): 181–188.
16. **Ekelund U, Sjöström M, Yngve A, Poortvliet E, Nilsson A, et al.** (2001) Physical activity assessed by activity monitor and doubly labeled water in children. Med Sci Sports Exerc 33(2): 275–281.
17. **Hislop JF, Bulley C, Mercer TH, Reilly JJ** (2012) Comparison of epoch and uniaxial versus triaxial accelerometers in the measurement of physical activity in preschool children: a validation study. Pediatr Exer Sci 24(3): 450–460.
18. **Ojiambo R, Konstabel K, Veidebaum T, Reilly J, Verbestel V, et al.**; IDEFICS Consortium (2012) Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: the IDEFICS Validation Study. J Appl Physiol 113(10): 1530–1536.

19. **Plasqui G, Westerterp K** (2007) Physical Activity Assessment with accelerometers: An Evaluation against doubly labeled water. Obesity 15(10): 2371–2379.

20. **Pober DM, Staudenmayer J, Raphael C, Freedson PS** (2006) Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc 38(9): 1626–1634.

21. **Rabiner L** (1989) A tutorial on hidden markov models and selected applications in speech recognition. Proc IEEE 77(2): 257–286.

22. **Zucchini W, MacDonald IL** (2009) Hidden Markov Models for Time Series: An Introduction Using R. Monographs on Statistics and Applied Probability 110. Boca Raton (FL): Chapman and Hall/CRC.

23. **Baum L, Petrie T, Soules G, Weiss N** (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. Ann Math Statist 41(1): 164–171.

24. **Forney GD** (1973) The Viterbi algorithm. Proc IEEE 61(3): 268–278.

25. **Viterbi AJ** (1967) Error Bounds for convolutional codes and an asymptotically optimal decoding algorithm. IEEE T Inform Theory 13(2): 260–269.

26. **Joe H, Zhu R** (2005) Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution. Biometrical 47(2): 219–229.

27. **Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, et al.**; IDEFICS Consortium (2011) The IDEFICS cohort: design, characteristics and participation in the baseline survey. Int J Obes (Suppl 1): 3–15.

28. **Witowski V, Foraita R** (2013) HMMpa: Analysing accelerometer data using hidden markov models. R package version 1.0.

29. **Centers for Disease Control and Prevention (09 Nov 2011) How much physical activity do children need? Available:** http://www.cdc.gov/physicalactivity/everyone/guidelines/children.html**.** Accessed 2014 Oct 20.

30. **Crouter SE, Horton M, Basset DR Jr** (2012) Use of a 2-regression model for estimating energy expenditure in children. Med Sci Sports Exerc 44(6): 1177–1185.

Appendix B

Paper: Modeling physical activity data using L_0-penalized expectile regression

Contribution to the manuscript I herewith certify that I conceived and designed the experiments, performed large parts of the statistical analyses, interpreted the results and drafted parts of the manuscript and revised it critically for important intellectual content.
This manuscript will be submitted shortly after submission of this thesis. Only the abstract is presented here. The complete manuscript can be obtained from the author upon request.

Modeling physical activity data using L_0-penalized expectile regression

N. Wirsik[1], F. Sobotka[2], I. Pigeot[1,3]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[2]Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[3]University of Bremen, Bremen, Germany

Correspondence
Prof. Dr. Iris Pigeot
Department of Epidemiological Methods and Etiologic Research
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Achterstr. 30, 28359 Bremen, Germany
Phone: +49-421-218 56942
Email: pigeot@bips.uni-bremen.de

Abtract

In recent years accelerometers have become widely used for the objective assessment of physical activity. Usually intensity ranges are assigned to the measured accelerometer counts by simple cutpoints, disregarding the underlying activity pattern. Under the assumption that physical activity can be seen as distinct sequence of distinguishable activities, the use of hidden Markov models (HMM) has been proposed to improve the modeling of accelerometer data. As a potential further improvement we propose to use expectile regression utilizing a Whittaker smoother with an L_0-penalty to better capture the mean intensity levels underlying the observed accelerometer counts. The performance of this new approach is investigated in a simulation study, where we simulated 1,000 days of accelerometer data with 1 second epochs and 5 seconds epochs, based on collected labeled data to resemble real-life data as closely as possible. The expectile regression is compared to HMMs and the commonly used cutpoint method with regard to misclassification rate (MCR), number of identified bouts and identified levels, the proportion of the estimate being in the range of $\pm 10\%$ of the true mean. In summary, expectile regression utilizing a Whittaker smoother with an L_0-penalty seems to outperform HMMs and the cutpoint method and is hence a promising approach to model accelerometer data.

Appendix C

Paper: Objectively measured physical activity in European children: the IDEFICS study

Contribution to the manuscript I herewith certify that I consulted on the statistical analyses and critically revised the manuscript for important intellectual content.
The presented study was published in the peer-reviewed *International Journal of Obesity* (Konstabel et al., 2014) under a creative commons license.

OPEN

International Journal of Obesity (2014) **38,** S135–S143

www.nature.com/ijo

ORIGINAL ARTICLE

Objectively measured physical activity in European children: the IDEFICS study

K Konstabel[1,2], T Veidebaum[1], V Verbestel[3], LA Moreno[4], K Bammann[5,6], M Tornaritis[7], G Eiben[8], D Molnár[9], A Siani[10], O Sprengeler[6], N Wirsik[6], W Ahrens[6,11] and Y Pitsiladis[12] on behalf of the IDEFICS consortium

OBJECTIVES: To provide sex- and age-specific percentile values for levels of physical activity (PA) and sedentary time of European children aged 2.0–10.9 years from eight European countries (Sweden, Germany, Hungary, Italy, Cyprus, Spain, Belgium and Estonia).
METHODS: Free-living PA and sedentary time were objectively assessed using ActiGraph GT1M or ActiTrainer activity monitors in all children who had at least 3 days' worth of valid accelerometer data, with at least 8 h of valid recording time each day. The General Additive Model for Location Scale and Shape was used for calculating percentile curves.
RESULTS: Reference values for PA and sedentary time in the European children according to sex and age are displayed using smoothed percentile curves for 7684 children (3842 boys and 3842 girls). The figures show similar trends in boys and girls. The percentage of children complying with recommendations regarding moderate-to-vigorous physical activity (MVPA) is also presented and varied considerably between sexes and country. For example, the percentage of study participants who were physically active (as assessed by MVPA) for 60 or more minutes per day ranged from 2.0% (Cyprus) to 14.7% (Sweden) in girls and from 9.5% (Italy) to 34.1% (Belgium) in boys.
CONCLUSION: This study provides the most up-to-date sex- and age-specific reference data on PA in young children in Europe. The percentage compliance to MVPA recommendations for these European children varied considerably between sexes and country and was generally low. These results may have important implications for public health policy and PA counselling.

International Journal of Obesity (2014) **38,** S135–S143; doi:10.1038/ijo.2014.144

INTRODUCTION

Physical activity (PA) has assumed an increasingly prominent role in health promotion efforts given the childhood obesity epidemic. Consequently, assessing levels of regular PA and sedentary behaviour among children has become an important public health surveillance activity. Currently, the most important official reports about PA levels in European children have been based on data obtained through questionnaires.[1–3] However, there is no established PA questionnaire for use in children and even the most widely used PA questionnaires, such as the International Physical Activity Questionnaire[4,5] developed as an international standard questionnaire to monitor PA across diverse adult populations may be subject to recall bias, social desirability and misinterpretation. As questionnaires are too imprecise in the assessment of PA and given the advances made in this area over the last 10–15 years, objective methods should be favoured in large-scale studies or for surveillance purposes.[6–8] For purposes of monitoring and surveillance, accelerometers provide a reasonable compromise between validity, reliability, ease of administration and cost.[9] As such accelerometry is considered as the reference method for measuring PA and sedentary behaviour of children in free-living conditions.[10] Until recently, no large representative set of accelerometer data existed to describe PA behaviour especially in young European children. In response, the International Children's Accelerometry Database (http://www.mrc-epid.cam.ac.uk/research/studies/icad/) was established to pool data from 14 studies collected between 1998 and 2009 comprising 20 871 children (4–18 years). Time spent in moderate and vigorous PA and sedentary time was measured using accelerometry after re-analysing raw data files and independent associations between accelerometry outcomes and measures such as waist circumference, systolic blood pressure, fasting triglycerides, high-density lipoprotein-cholesterol and insulin were examined using meta-analysis.[11] In the results, reported by Ekelund *et al.*,[12] children accumulated a modest 30 ± 21 min per day of moderate-to-vigorous physical activity (MVPA) and this activity was associated with all cardio-metabolic outcomes independent of sex, age, monitor wear time, time spent sedentary and waist circumference. The IDEFICS study (Identification and prevention of dietary and lifestyle-induced health effects in children and infants)[13] provides an excellent opportunity to augment current data by reporting the objectively measured PA data from this large sample of children from a wider range of European countries. It will extend available data as well as allow comparison with other important European studies of childhood PA.[14,15] Therefore, the

[1]National Institute for Health Development, Tallinn, Estonia; [2]Institute of Psychology, University of Tartu, Tartu, Estonia; [3]Department of Movement and Sport Sciences, Ghent University, Ghent, Belgium; [4]GENUD (Growth, Exercise, Nutrition and Development) research group, Faculty of Health Sciences, University of Zaragoza, Zaragoza , Spain; [5]Institute for Public Health and Nursing Research, University of Bremen, Bremen, Germany; [6]Leibniz Institute for Prevention Research and Epidemiology—BIPS, Bremen, Germany; [7]Research and Education Institute of Child Health, Strovolos, Cyprus; [8]Department of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden; [9]Department of Paediatrics, University of Pécs, Pécs, Hungary; [10]Unit of Epidemiology & Population Genetics, Institute of Food Sciences, National Research Council, Avellino, Italy; [11]Institute of Statistics, Faculty of Mathematics and Computer Science, University Bremen, Bremen, Germany and [12]Centre for Sport and Exercise Science and Medicine (SESAME), University of Brighton, Eastbourne, UK. Correspondence: Professor Dr W Ahrens, Department of Epidemiological Methods and Etiologic Research, Leibniz Institute for Prevention Research and Epidemiology —BIPS, Achterstr. 30, D-28359 Bremen, Germany or Professor YP Pitsiladis, The Brighton Centre for Regenerative Medicine (BCRM), Hillbrow, Denton Road, Eastbourne BN20 7SP, UK.
E-mail: ahrens@bips.uni-bremen.de or Y.Pitsiladis@brighton.ac.uk

aim of the present analysis is to describe PA levels and to provide sex- and- age-specific PA reference standards in European children aged 2–10 years. A General Additive Model for Location Scale and Shape developed by Rigby and Stasinopoulos[16] was used to calculate percentile curves. This method is an extension of the LMS method to model the distribution of PA depending on multiple covariates while accounting for dispersion, skewness, and particularly the kurtosis of this distribution.[16]

MATERIAL AND METHODS

Data collection

The IDEFICS study[13] is a multi-centre, prospective cohort study on lifestyle and nutrition among children aged 2.0–10.9 years from eight European countries (Sweden, Germany, Hungary, Italy, Cyprus, Spain, Belgium and Estonia) where an intervention was embedded. Data collection for T0 (baseline) took place from September 2007 to June 2008 and for T1 (follow-up) from September 2010 to May 2011. Detailed description of IDEFICS sampling and recruitment approaches, standardisation and harmonisation process, data collection, analysis strategies, quality control activity and inclusion/exclusion criteria have been published elswhere.[17] Free-living PA and sedentary time were objectively assessed using Actigraph uniaxial accelerometers (either ActiTrainer or GT1M; Actigraph, LLC, Pensacola, FL, USA). The sensor units of both models are identical. The monitor was set to record PA in a 15 s epoch. Children were instructed to wear the accelerometer for at least 3 days (including 1 weekend day). Accelerometers were mounted on the right hip of each child by means of an elastic belt and adjusted to ensure close contact with the body. Parents were also asked to complete a daily activity or 'non-wear' diary during the monitoring period with instructions to record the time the accelerometer belt was attached and removed. Children were encouraged to wear the accelerometer from the moment they woke up in the morning until bedtime in the evening so that a full day of PA and sedentary time could be assessed. The Research Ethics Committees of each survey region involved approved the study and informed consent for participation in the study was signed by each parent/guardian.

From the starting-point database (18 745 children), 12 014 had some data on PA; the others (6731) either refused or the assessment was not completed for other reasons (mostly, lack of devices at the time that the child was measured). In addition, subjects were excluded from analysis if the child had chronic orthopaedic, bone or joint problems, chronic rheumatic disease, or musculoskeletal diseases, as indicated in the health and medical history questionnaire (total number of children excluded for these reasons was 332).

According to the protocol, a 15-s sampling interval ('epoch') was to be used in PA data collection; however, a 60-s sampling interval (the default in Actigraph software) was inadvertently used in some centres for a considerable proportion of data. Three options were considered to rectify this problem: (1) include only data recorded at 15 s (or, in a few cases, sub-multiple of 15 s) epoch or re-integrate all data to 60 s, (2) use 6, 8, or 10 h as minimum requirement of valid day, and (3) the numbers of valid weekdays (1 or 2) required in addition to at least 1 weekend day. The sample sizes following each option are shown in Table 1. Of the total of 12 014 children having at least some accelerometer data, the number of children to be included in the analysis varied from 8193 according to the most liberal criterion (at least 6 h daily, at least 1 weekend day+1 weekday, using 60 s epoch), to 3390 according to the most stringent criterion considered (at least 10 h of data, at least 1 weekend day and 2 weekdays; using 15 s epoch). First, all data were re-integrated to 60 s epoch in order to incorporate the data from Italy. Second, the decision was made to use 8 h as the lower threshold of daily wearing time as previously adopted in the HELENA study,[17] thus avoiding the loss of at least 1335 participants had a 10 h rule been adopted. Finally, the requirements of at least 1 vs 2 valid weekdays (in addition to at least 1 weekend day) were contrasted. Requiring 2 weekdays would have resulted in a total sample size of 5047 (42% of all children having some accelerometer data), as opposed to 7684 (64%) with the alternative criterion. The requirement was thus set to at least 1 weekday.

Table 1. Number of subjects included in the final analysis of PA (highlighted in bold) and number of subjects according to alternative criteria of epoch length and minimum wearing time

	Epoch = 60 s			*Epoch = 15 s*		
Required wearing time (hours per day)	6	8	10	6	8	10
At least 1 weekend day and 2 weekdays						
Included in the analysis	5721	5047	3667	5316	4680	3390
Incomplete data	5961	6635	8015	6366	7002	8292
At least 1 weekend day and 1 weekday						
Included in the analysis	8193	**7684**	6349	6941	6516	5365
Incomplete data	3489	**3998**	5333	4741	5166	6317
Excluded children						
No accelerometer data was collected		6731				
Musculoskeletal diseases		26				
Orthopaedic and musculoskeletal diseases		7				
Orthopaedic diseases		280				
Rheumatic diseases		5				
Age > 10.9		14				

The possibility of sample bias induced by an additional restriction of the sample was examined by conducting an analysis of covariance using body mass index (BMI *Z*-scores[18]) as a dependent variable. The sample was therefore divided into three categories: (1) no accelerometer data; (2) incomplete data (that is, < 1 weekday and 1 weekend day of at least 8 h of recording); and (3) complete data. The effect of this categorisation, adjusted for region (two regions that later became 'intervention' and 'control' regions in each country), sex and age was negligible in size ($\eta^2 = 0.0002$) and not statistically significant ($P = 0.14$). Moreover, the differences were not in the expected direction: the mean BMI *Z*-score was 0.342 in the group of children who had complete accelerometer data and 0.282 in the group who had no data (unadjusted standardised difference, Cohen's $d = 0.051$). Therefore, the children having complete data had, on the average, higher BMI Z-scores than those having no accelerometry data. Similarly, children having incomplete data tended to have higher BMI, than those who had complete data (respective means: 0.383 and 0.342; Cohen's $d = 0.035$). Note that these effect sizes are not adjusted for possible confounders.

Given PA levels tended to be higher on weekends and for ease of presentation, PA data for weekdays and weekend days were analysed separately and then combined so that weekdays were weighted by five, weekend days by two, and the result was divided by seven as previously recommended.[19] To be able to compare the results of the present study (that is, re-integrated into 60 s before analysis) with data recorded at lower epoch, a formula for converting MVPA recorded at 15 s epoch to 60 s epoch is presented in the Supplementary Information. The percentage of children complying with the 60 min MVPA recommendations was determined using the weighted-average day (that is, weekend average weighted by two and weekday weighted by five).

Table 2. Sample characteristics (sample size by country, age group, and sex)

Age	*Italy*	*Estonia*	*Cyprus*	*Belgium*	*Sweden*	*Germany*	*Hungary*	*Spain*	*Total*
Boys									
2.0–2.9	4	14		3	19	11		20	71
3.0–3.9	30	112	5	17	37	60	30	97	388
4.0–4.9	43	153	28	31	32	85	31	133	536
5.0–5.9	85	64	60	38	33	45	49	62	436
6.0–6.9	133	29	54	111	57	85	57	93	619
7.0–7.9	128	106	50	128	57	108	113	126	816
8.0–8.9	112	89	45	68	48	78	89	60	589
9.0–9.9	19	80	28	60	14	29	30	9	269
10.0–10.9	14	43	8	16	3	15	13	6	118
Total	568	690	278	472	300	516	412	606	3842
Mean age	6.81	6.36	6.89	7.20	6.19	6.38	7.00	5.78	
s.d. (age)	1.61	2.34	1.63	1.64	2.00	1.95	1.75	1.85	
Girls									
2.0–2.9	2	12		1	8	5	3	29	60
3.0–3.9	35	113	4	22	24	52	25	79	354
4.0–4.9	44	144	22	36	37	69	52	103	507
5.0–5.9	66	57	64	49	31	45	44	40	396
6.0–6.9	109	32	47	119	37	72	66	105	587
7.0–7.9	101	119	57	115	61	145	141	148	887
8.0–8.9	121	68	52	70	54	104	80	57	606
9.0–9.9	38	96	42	57	19	24	12	17	305
10.0–10.9	22	67	9	15	7	5	15		140
Total	538	708	297	484	278	521	438	578	3842
Mean age	6.98	6.59	7.11	7.05	6.60	6.62	6.82	5.98	
s.d. (age)	1.77	2.39	1.69	1.69	2.00	1.80	1.72	1.87	
T0	822	898	402	430	486	907	693	1109	5747
T1	284	500	173	526	92	130	157	75	1937
All	1106	1398	575	956	578	1037	850	1184	7684

Abbreviation: s.d. = standard deviation.

Accelerometer data reduction

For comparability, data reduction criteria were chosen similar to those used in the HELENA study (www.helenastudy.com).[17] Namely, non-wearing time was defined as 20 min or more of consecutive zero counts and at least 8 h of wear time was necessary to constitute a valid day and be included in the final analysis. Using more restrictive criteria like those in the European Youth Heart Study[15] (that is, counting >10 consecutive minutes of zero recording as non-wearing time and requiring at least 10 h daily wear time) would have resulted in a loss of data from >1000 children in the present study.

Accelerometer data were analysed using algorithms developed in R (version R 3.0.1; R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org[20]). A set of add-on functions to R was developed that allowed R to automatically read in the accelerometer raw files and where necessary re-integrate any data collected to standardise epoch settings, edit the data for excluding the likely non-wearing periods and compute daily summary statistics. Two rules were used for excluding data: (1) all negative counts were replaced by missing data code and (2) periods of 20 min or more consecutive zero counts were replaced by missing data code before further analysis. The output generated by R included accelerometer counts per minute (CPM), total monitoring time and time spent sedentary and in physical activities of different intensities based on Evenson cutoffs[21] (sedentary: 0–100, light: 101–2295, moderate: 2296–4011, vigorous: 4012 and more CPM).

The following dependent variables were used for the analysis: average CPM (that is, sum of daily counts divided by valid time, averaged over all 'valid days'), MVPA (minutes spent in at least moderate activity according to Evenson cutoffs), light activity, and sedentary time. The last two were used in two versions: (1) 'raw'

Table 3. Descriptive statistics (mean ± s.d.) of physical activity in 2–10-year-old European children

Age	*CPM*		*Sedentary (minutes per day)*		*Light (minutes per day)*		*MVPA (minutes per day)*	
	Mean	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*	*Mean*	*s.d.*
Boys								
2.0–2.9	567	141	237	68	410	63	24	14
3.0–3.9	625	159	241	70	416	62	34	18
4.0–4.9	662	160	241	70	422	57	42	21
5.0–5.9	657	168	262	81	420	66	46	22
6.0–6.9	631	185	296	98	397	64	50	26
7.0–7.9	610	175	321	94	381	66	52	27
8.0–8.9	581	186	342	101	368	68	49	28
9.0–9.9	551	159	366	98	364	75	48	26
10.0–10.9	540	162	378	91	360	71	48	25
Girls								
2.0–2.9	567	114	245	70	414	54	24	12
3.0–3.9	570	145	243	69	410	62	27	15
4.0–4.9	600	156	250	71	412	64	33	18
5.0–5.9	587	154	274	90	419	69	35	17
6.0–6.9	559	163	301	92	398	73	37	20
7.0–7.9	549	148	320	90	381	66	39	19
8.0–8.9	518	172	339	99	373	75	36	22
9.0–9.9	481	139	370	89	352	71	36	20
10.0–10.9	471	138	380	87	357	60	35	21

Abbreviations: CPM = counts per minute; MVPA = moderate-to-vigorous physical activity (using Evenson cutoffs); s.d. = standard deviation.

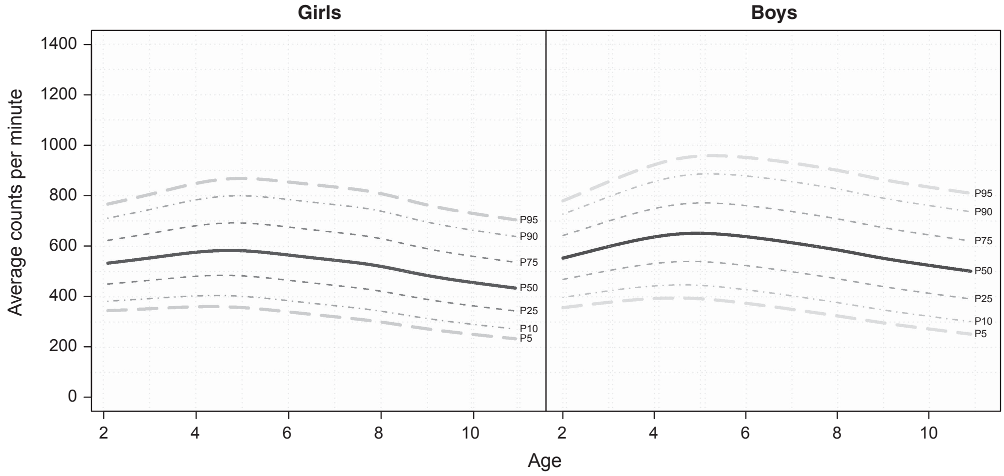

Figure 1. Reference values for average counts per minute.

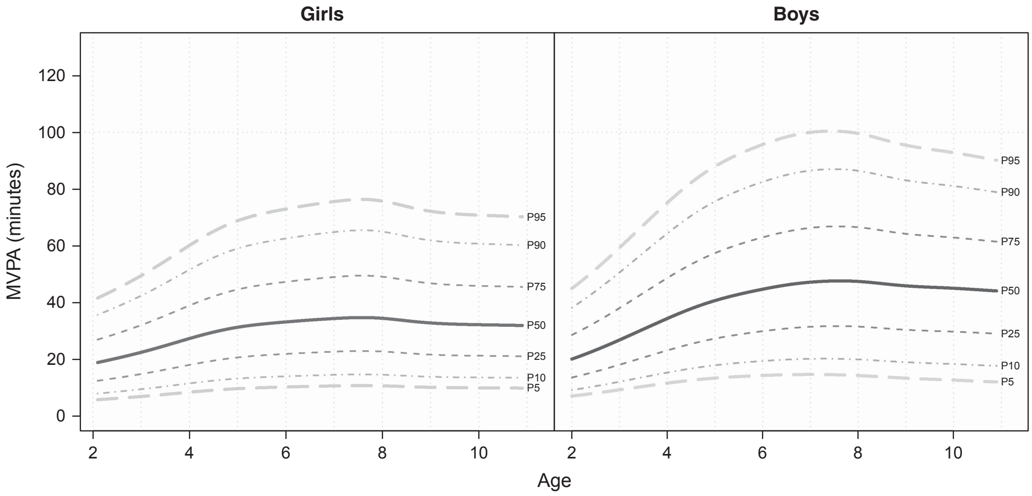

Figure 2. Reference values for moderate-to-vigorous physical activity (MVPA) according to Evenson cutoffs (minutes per day).

minutes, and (2) 'adjusted' minutes. Adjusted minutes were computed by dividing the raw minutes by wearing time and multiplying the resulting fraction by the average wearing time. Average wearing time of the final sample was 740 ± 100 min (mean ± s.d.). Data have also been analysed using other common cutoffs ie Sirard,[22] Pate,[23] Puyau,[24] and Van Cauwenberghe.[25] Results are presented in Supplementary Tables A–C.

Statistical analysis

The General Additive Model for Location Scale and Shape method is able to particularly model the kurtosis using other distributions and to include more than one covariate. We used the *gamlss package* (version 4.2–6) of the statistical software *R* (version 3.0.1).[20] Different distributions, that is, the Box-Cox power exponential, the Box-Cox *t* or the Box-Cox Cole and Green, the normal, the power exponential and the *t* family distribution were fitted to the observed distribution of PA variables. Moreover, the influence of age on parameters of the considered distributions was modelled either as a constant, as a linear function, or as a cubic spline of the covariates. Goodness of fit was assessed by the Bayesian Information Criterion and Q–Q plots to select the final model including the fitted distribution of PA variables and the influence of covariates on distribution parameters. Finally, percentile curves for the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles were calculated based on the model that showed the best goodness of fit.[16,26]

In order to use the same distribution for all dependent variables, it was decided to use the Box-Cox Cole and Green distribution as it was in most cases the best fitting model according to the Bayesian Information Criterion; or the difference in Bayesian Information Criterion from the best fitting model was negligible. For average CPM, the final model for both sexes considered Box-Cox Cole and Green distribution modelling log μ and log σ as a cubic spline of age, and ν as a constant. For unadjusted MVPA and light activity (both unadjusted and adjusted minutes), as well as for sedentary time (adjusted and unadjusted minutes), the final model for both sexes considered the Box-Cox Cole and Green distribution modelling log μ as a cubic spline of age, and log σ and ν as a constant. The form of the model for adjusted MVPA was different for boys and girls: for both sexes, log μ was modelled as a cubic spline of age and ν as a constant, but for boys, log σ was also a cubic spline of age but a constant for girls.

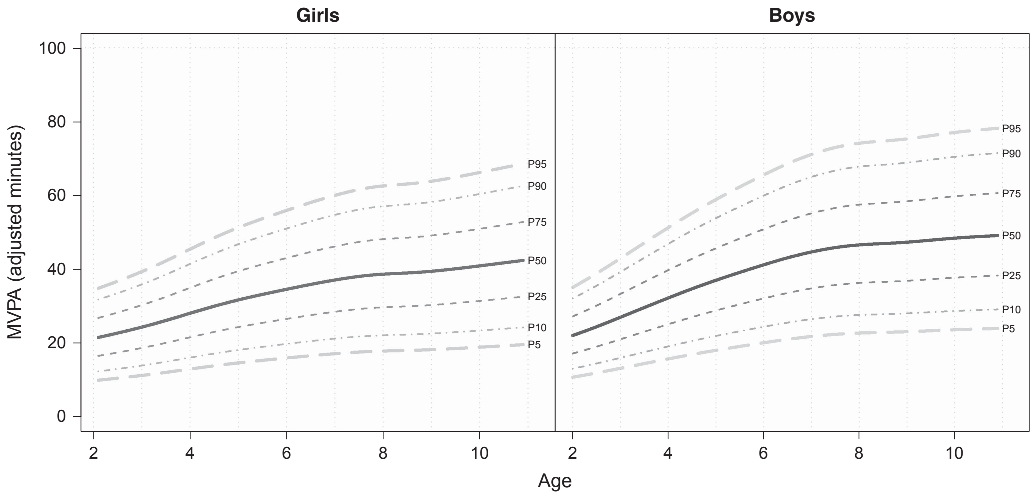

Figure 3. Reference values for moderate-to-vigorous physical activity (MVPA) according to Evenson cutoffs (adjusted minutes per day). Adjusted minutes were computed by dividing the raw minutes by wearing time and multiplying the resulting fraction by the average wearing time.

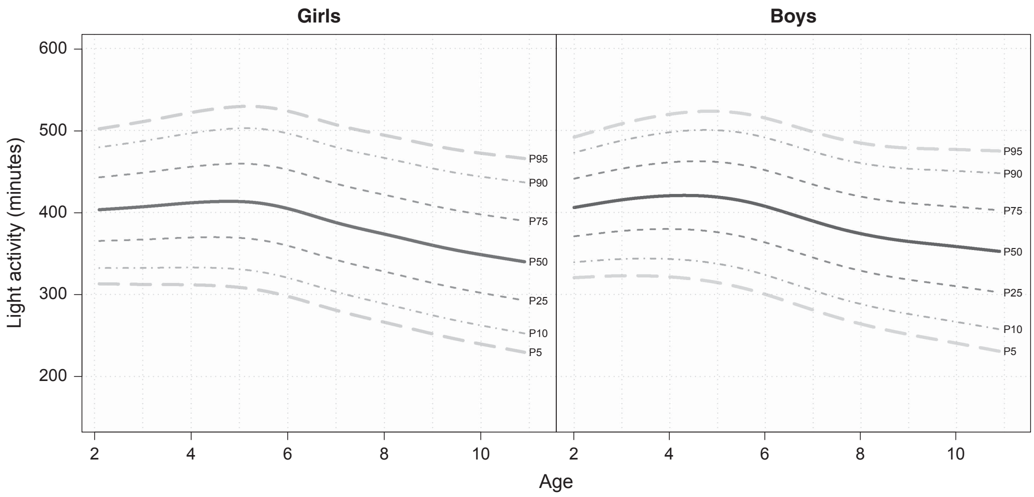

Figure 4. Reference values for light physical activity according to Evenson cutoffs (minutes per day).

RESULTS

In total, data from 7684 children (3842 boys and 3842 girls) from T0 and T1 are presented in Table 2. No longitudinal data were included: for most children ($N=5747$), T0 values were used; the T1 values were included only if the child was newly recruited in T1, or if their PA was not measured in T0 ($N=1937$). The descriptive statistics of PA and sedentary time in these 2.0–10.9-year-old European children are presented in Table 3. Both CPM and MVPA were higher in boys, whereas sedentary time was higher for girls; there was no overall difference in light activity. Reference values for PA and sedentary time (Evenson cutoffs) in the European children according to sex and age are expressed using smoothed centile curves (P_5, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}, P_{95}) and are illustrated in Figures 1, 2, 3, 4, 5, 6 and 7. These figures show similar trends in boys and girls. Tabulated percentiles of these data are presented in the Supplementary Table A. Centile curves for PA and sedentary time were similar irrespective of whether data was adjusted or not by dividing the raw minutes by wearing time and multiplying the resulting fraction by the average wearing time (see Figures 2, 3, 4, 5, 6 and 7). The percentage of children complying with recommendations regarding MVPA is shown in Figure 8. The percentage compliance to MVPA recommendations for these European children varied considerably between sexes and country and was generally low with the percentage of study participants who were physically active for 60 or more minutes per day ranging from 2.0% (Cyprus) to 14.7% (Sweden) in girls and from 9.5% (Italy) to 34.1% (Belgium) in boys (Figure 8). Percentile curves for average CPM obtained from the whole sample were similar to those obtained after excluding overweight and obese children (Supplementary Tables A–C).

DISCUSSION

The present study provides the most up-to-date sex- and age-specific data on PA in young children in Europe with important implications for public health policy and PA counselling. Although the present data should not be used as reference standards, scientists, medical and biomedical personnel and related stake-holders, children and parents will be able to compare the

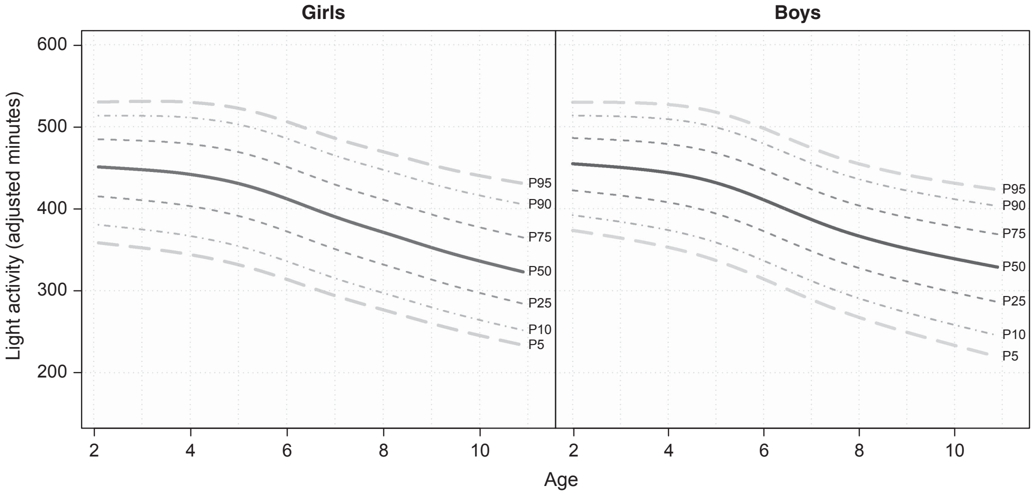

Figure 5. Reference values for light physical activity according to Evenson cutoffs (adjusted minutes per day). Adjusted minutes were computed by dividing the raw minutes by wearing time and multiplying the resulting fraction by the average wearing time.

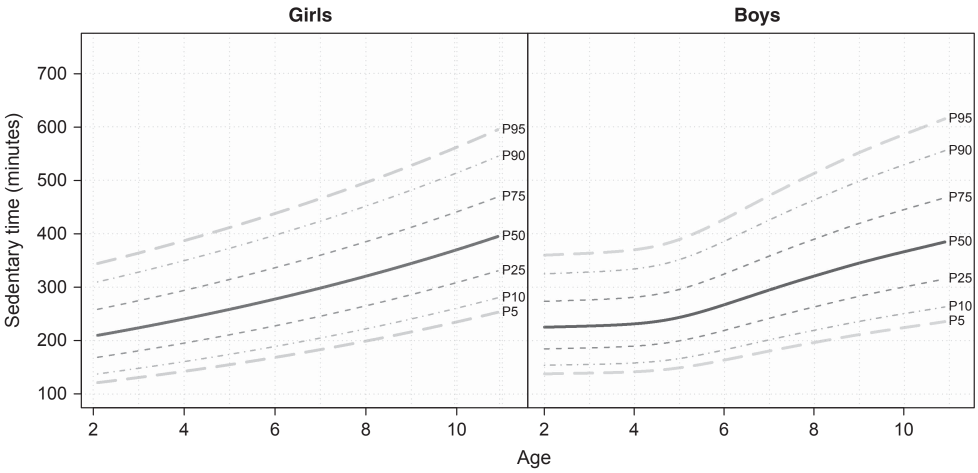

Figure 6. Reference values for sedentary time according to Evenson cutoffs (minutes per day).

obtained scores with these sex- and age-specific objective measures of PA and sedentary time of European children from eight different countries. These data complement a growing literature of comparative data across a range of different health-related fitness measures such as body mass index, physical fitness, bone health and blood lipids from this unique cohort of European children (see Nagy *et al.*, De Henauw *et al.*, Peplies *et al.*, De Miguel-Etayo *et al.*; this issue). The present analysis quantifies the magnitude and direction of sex- and age-related differences in children's PA patterns and shows boys having higher MVPA than girls on average from the age of 5 years onwards and a larger range in MVPA. These differences in MVPA are in line with previous studies and could be largely explained by socio-cultural reasons such as more vigorous exercise outside of school, during school physical education and more participation in sports teams in boys.[27]

Despite the importance of PA assessment, most previous reports and fact sheets about PA levels in European youth used data obtained through questionnaires.[1–3] For example, PA patterns were estimated in a cross-sectional survey of 137 593 youth (10–16 years) from the 34 (primarily European) participating countries of the 2001–2002 Health Behaviour in School-Aged Children Study (HBSC survey) using self-completed questionnaires administered in the classroom.[3] Subjects were asked how many days in the past week and in a typical week they were physically active (cumulative activity including sports, school activities, playing with friends, and walking to school) for 60 min or more. Using this approach, the percentage of study participants who were physically active for 60 or more minutes on 5 or more days per week ranged from 19.3% in France to 49.5% in the United States. Notwithstanding the usefulness of the data generated using questionnaires (for example, useful to identify, which activities were performed), such subjective approaches are heavily prone to recall bias, social desirability and misinterpretation.

Until the present analysis, no large representative set of accelerometer data existed to provide normative values especially for young European children. Notable attempts to review objectively measured PA data obtained by accelerometry from children and adolescents in Europe have been published.[11,12,28] In the first attempt, data were pooled from 14 studies collected between 1998 and 2009 comprising 20 871 children (4–18 years).[11] Using a meta-analysis, Ekelund *et al.*[12] found children

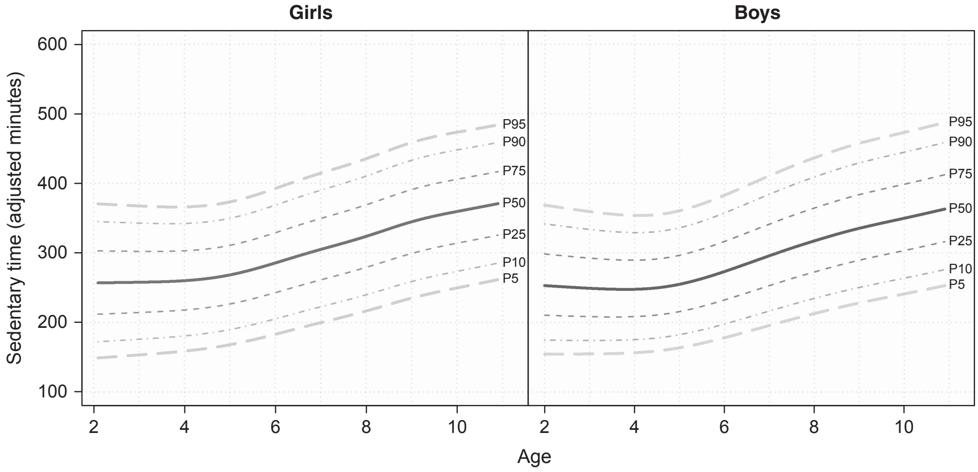

Figure 7. Reference values for sedentary time according to Evenson cutoffs (adjusted minutes per day). Adjusted minutes were computed by dividing the raw minutes by wearing time and multiplying the resulting fraction by the average wearing time.

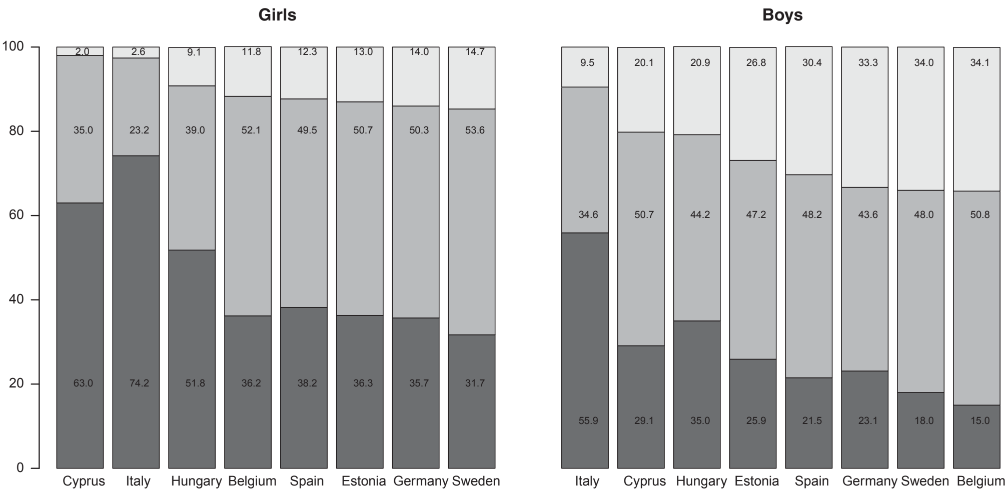

Figure 8. Percentage compliance to MVPA recommendations: lowest part: 0–30 min, middle part: 30–60 min, upper part: over 60 min.

accumulated a modest 30 ± 21 min per day of MVPA and this magnitude is very much in line with the data presented here from the IDEFICS study. MVPA was associated with all measured cardio-metabolic outcomes such as waist circumference, systolic blood pressure, fasting triglycerides, high-density lipoprotein-cholesterol and insulin independent of sex, age, monitor wear time, time spent sedentary and waist circumference (when waist circumference was not the outcome).[12] The subsequent study explored the proportion of European children who were assessed, based on objective assessment using accelerometry, as sufficiently active according to PA recommendations.[28] In order to do this, the authors conducted systematic electronic search of studies involving European youth published up to March 2012. The analysis and interpretation of the data is difficult owing to the use of different PA thresholds lying between 1000 and 4000 counts $\min^{-1}$ to define MVPA. For example, up to 100% of youth may be assessed as sufficiently active when using a threshold of $\sim > 1000$–1500 CPM. Applying > 2000 CPM as the cutoff, which is most frequently used, up to 87% of European children and adolescents would be considered physically active according to the current recommendations that children aged 5–18 years should engage in MVPA for a minimum of 60 min on a daily basis.[29] If the cutoff would be raised to > 3000 CPM, only 3–5% of the children would reach these recommendations.

The large discrepancy in outcomes released by accelerometer data is mainly due to the variety of cutoff points for MVPA among children and adolescents, hindering the definition of a clear goal towards PA promotion in Europe and beyond; standardisation of these methods is, therefore, urgently required. The results of the study by Guinhouya *et al.*[28] illustrate the significant impact of methodological decisions on accelerometer outcome variables and consequently the observed compliance to public health guidelines in young children. For example, choosing a low cutoff can wrongly classify inactive children as active and vice versa.[23] Trost *et al.*[30] have compared an array of varying cutoffs for children and recommend Evenson's cutoffs as the most accurate when assessing all levels of activity. Cutoffs developed for pre-school children by Van Cauwenberghe *et al.*[25] are similar to those by Evenson for MVPA counts (Evenson: $\geqslant 585$ counts per 15 s vs Van Cauwenberghe: $\geqslant 574$ counts per 15 s) but ~ 120 counts per 15 s lower by Van Cauwenberghe for vigorous activities. However, the validity of these recommendations remains to be determined conclusively and, therefore, consensus on this subject remains elusive, although essential.

Given the recommendation[30] that the Evenson's cutoffs are the most accurate when assessing all levels of activity in young children, these cutoffs have been preferred to generate the primary normative PA data for this cohort (albeit the data using the other commonly used cutoffs can also be found in the supplementary material). Using the Evenson cutoffs PA patterns were also generated for the different countries. One must exercise caution in extrapolating this data to each respective country as the IDEFICS study is not nationally representative because each survey centre only covered a delimited geographic area within a country. Furthermore and like all studies, the present study is not without limitations. Most important of these limitations relate to the variable use of accelerometers (that is, typically less than 4 days) and in only a proportion of the IDEFICS study cohort (see Table 2) due primarily due to the prohibitive cost of purchasing sufficient accelerometer devices for the testing of as many of the IDEFICS cohort possible and avoiding any negative impact on compliance given the already significant burden for parents, teachers and the children from the large amount of data being collected on numerous occasions. Nevertheless, some interesting patterns do emerge that tend to agree with other published data. For example, the proportion of children complying with MVPA recommendations is low in Italy and Cyprus and higher in Estonia, Hungary, Germany, Sweden, Belgium and Spain (see Figure 8). This trend is in almost complete agreement with data on physical fitness published previously.[31]

In conclusion, accelerometry is currently considered to be the most valid method of assessing PA among children in free-living conditions. In anticipation of an even greater reliance on accelerometry for measuring movement behaviours of children, the present study provides the most up-to-date and comprehensive set of sex- and age-specific reference data on PA in children and youth. These reference values may have important implications for public health policy and PA counselling to motivate young individuals with low levels of PA to set appropriate goals and monitor individual changes in PA.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work has been done as part of the IDEFICS Study (http://www.idefics.eu). We gratefully acknowledge the financial support of the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD). We thank the IDEFICS children and their parents who generously volunteered and participated in this project. We are also grateful to the Volkswagen Foundation that financially supported the production of this supplement.

DISCLAIMER

The information in this document reflects the authors' view and is provided as is.

STATEMENT OF ETHICS

We certify that all applicable institutional and governmental regulations concerning the ethical use of human volunteers were followed during this research. Approval by the appropriate Ethics Committees was obtained by each of the eight centres doing the fieldwork. Study children did not undergo any procedure before both they and their parents had given consent for examinations, collection of samples, subsequent analysis and storage of personal data and collected samples. Study subjects and their parents could consent to single components of the study while abstaining from others.

REFERENCES

1 Carroquino-Salto MJ. *Percentage of Physically Active Children and Adolescents. Fact Sheet 2.4. Report No.: RPG2_Hous_Ex1*. WHO Regional Office for Europe & the European Environment and Health Information System (ENHIS): Bonn, 2009.
2 Cavill N, Kahlmeier S, Racioppi F. *Physical Activity and Health in Europe: Evidence for Action*. WHO Regional Office for Europe: Copenhagen, 2006.
3 Janssen I, Katzmarzyk PT, Boyce WF, Vereecken C, Mulvihill C, Roberts C *et al*. The Health Behaviour in School-Aged Children Obesity Working Group. Comparison of overweight and obesity prevalence in school-aged youth from 34 countries and their relationships with physical activity and dietary patterns. *Obesity Reviews* 2005; **6**: 123–132.
4 Craig CL, Marshall AL, Sjostrom M, Bauman A, Booth ML, Ainsworth BE *et al*. International Physical Activity Questionnaire: 12-country reliability and validity. *Med Sci Sport Exerc* 2003; **35**: 1381–1395.
5 Paul HL, Macfarlane DJ, Lam TH, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. *Int J Behav Nutr Phys Act* 2011; **8**: 115.
6 Basterfield L, Adamson AJ, Parkinson KN, Maute U, Li P-X, Reilly JJ. Surveillance of physical activity in the UK is flawed: validation of the health survey for England physical activity questionnaire. *Arch Dis Child* 2008; **93**: 1054e8.
7 Adamo KB, Prince SA, Tricco AC, Connor-Gorber S, Tremblay M. A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review. *Int J Ped Obes* 2009; **4**: 2e27.
8 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? *Am J Clin Nutr* 2009; **89**: 862e70.
9 Troiano E. Large-scale applications of accelerometers: new frontier and new questions. *Med Sci Sport Exerc* 2007; **39**: 1501.
10 Bornstein DB, Beets MW, Byun W, McIver K. Accelerometer derived physical activity levels of preschoolers: a meta-analysis. *J Sci Med Sport* 2011; **14**: 504e11.
11 Sherar LB, Griew P, Esliger DW, Cooper AR, Ekelund U, Judge K *et al*. International Children's Accelerometry Database (ICAD): design and methods. *BMC Public Health* 2011; **11**: 485.
12 Ekelund U, Luan J, Sherar LB, Esliger DW, Griew P, Cooper C. Association of moderate to vigorous physical activity and sedentary time with cardiometabolic risk factors in children and adolescents. *JAMA* 2012; **307**: 704–712.
13 Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L *et al*. IDEFICS consortium. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond)* 2011; **35**: S3–S15.
14 Ruiz JR, Ortega FB, Martínez-Gómez D, Labayen I, Moreno LA, De Bourdeaudhuij I *et al*. HELENA Study Group. Objectively measured physical activity and sedentary time in European adolescents: the HELENA study. *Am J Epidemiol* 2011; **174**: 173–184.
15 Riddoch CJ, Bo Andersen L, Wedderkopp N, Harro M, Klasson-Heggebø L, Sardinha LB *et al*. Physical activity levels and patterns of 9- and 15-yr-old European children. *Med Sci Sports Exerc* 2004; **36**: 86–92.
16 Stasinopoulos DM, Rigby R. Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw* 2007; **23**: 1–46.
17 Martinez-Gomez D, Ruiz JR, Ortega FB, Veiga OL, Moliner-Urdiales D, Mauro B *et al*. HELENA Study Group. Recommended levels of physical activity to avoid an excess of body fat in European adolescents The HELENA Study. *Am J Prev Med* 2010; **39**: 203–211.
18 Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatr Obes* 2012; **7**: 284–294.
19 Ortega FB, Konstabel K, Pasquali E, Ruiz JR, Hurtig-Wennlöf A, Mäestu J *et al*. Objectively measured physical activity and sedentary time during childhood, adolescence and young adulthood: a cohort study. *PLoS One* 2013; **8**: e60871.
20 R Development Core Team. *R: a Language and Environment for Statistical Computing R Foundation for Statistical Computing (Online)*. R Development Core Team: Vienna, Austria, (ISBN 3–900051-07–0') http://www.R-project.org/, 2013.
21 Evenson KR, Cattellier D, Gill K, Ondrak K, McMurray RG. Calibration of two objective measures of PA for children. *J Sports Sci* 2008; **26**: 1557–1565.
22 Sirard J, Trost SG, Pfeiffer KA, Dowda M, Pate RR. Calibration and evaluation of an objective measure of physical activity in preschool children. *J Phys Act Health* 2005; **3**: 345–357.
23 Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. *Obesity (Silver Spring)* 2006; **14**: 2000–2006.
24 Puyau MR, Adolph AL, Vohra FA, Butte NF. Validation and calibration of physical activity monitors in children. *Obes Res* 2002; **10**: 150–157.
25 Van Cauwenberghe E, Labarque V, Trost SG, de Bourdeaudhuij I, Cardon G. Calibration and comparison of accelerometer cut points in preschool children. *Int J Pediatr Obes* 2011; **6**: 582–589.
26 Cole TJ, Stanokevic S, Stocks J, Coates A, Hankinson J, Wade A. Age-and size-related reference ranges: a case study of spirometry through childhood and adulthood. *Stat Med* 2009; **28**: 880–898.

27 Sallis JF, Zakarian JM, Hovell MF, Hofstetter CR. Ethnic, socioeconomic, and sex differences in physical activity among adolescents. *J Clin Epidemiol* 1996; **49**: 125–134.
28 Guinhouya BC, Samouda H, de Beaufort C. Level of physical activity among children and adolescents in Europe: a review of physical activity assessed objectively by accelerometry. *Public Health* 2013; **127**: 301–311.
29 Strong WB, Malina R, Blimkie CJ, Daniels SR, Dishman RK, Gutin B *et al.* Evidence based physical activity for school-age youth. *J Pediatr* 2005; **146**: 732e7.
30 Trost SG, Loprinzi PD, Moore R, Pfeiffer KA. Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc* 2011; **43**: 1360–1368.
31 Olds T, Tomkinson G, Leger L, Cazorla G. Worldwide variation in the performance of children and adolescents: an analysis of 109 studies of the 20-m shuttle run test in 37 countries. *J Sports Sci* 2006; **24**: 1025–1038.

Supplementary Information accompanies this paper on International Journal of Obesity website (http://www.nature.com/ijo)

Appendix D

Paper: Incidence of high blood pressure in children - Effects of physical activity and sedentary behaviors

Contribution to the manuscript I herewith certify that I consulted on the statistical analyses and critically revised the manuscript for important intellectual content.

The presented study was published in the peer-reviewed *International Journal of Cardiology* (Moraes et al., 2015).

International Journal of Cardiology 180 (2015) 165–170

Contents lists available at ScienceDirect

International Journal of Cardiology

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

Incidence of high blood pressure in children — Effects of physical activity and sedentary behaviors: The IDEFICS study High blood pressure, lifestyle and children

Augusto César Ferreira de Moraes [a,b,c,*], Heráclito Barbosa Carvalho [a,c], Alfonso Siani [d], Gianvincenzo Barba [d,1], Toomas Veidebaum [e], Michael Tornaritis [f], Denes Molnar [g], Wolfgang Ahrens [h], Norman Wirsik [h], Stefaan De Henauw [i], Staffan Mårild [j], Lauren Lissner [k], Kenn Konstabel [e], Yannis Pitsiladis [l], Luis A. Moreno [b,m], on behalf of the IDEFICS consortium

[a] School of Medicine of the University of São Paulo, FMUSP, São Paulo, Brazil
[b] Universidad de Zaragoza, Facultad de Ciencias de la Salud, GENUD Research Group (Growth, Exercise, Nutrition and Development) Zaragoza, Spain
[c] YCARE (Youth/Child and Cardiovascular Risk and Environmental) Research Group, FMUSP, Brazil
[d] Unit of Epidemiology and Population Genetics, Institute of Food Sciences, National Research Council, Avellino, Italy
[e] Department of Chronic Diseases, National Institute for Health Development, Tallinn, Estonia
[f] Research and Education Foundation of Child Health, Strovolos, Cyprus
[g] Department of Paediatrics, Medical Faculty, University of Pécs, Pécs, Hungary
[h] Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[i] Department of Public Health, Ghent University, Ghent, Belgium
[j] Department of Paediatrics, University of Gothenburg, Gothenburg, Sweden
[k] Department of Public Health and Community Medicine, University of Gothenburg, Sweden
[l] University of Glasgow, Faculty of Biomedical and Life Sciences, Glasgow, Scotland, United Kingdom
[m] School of Medicine of the University of São Paulo, Department of Preventive Medicine, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 29 October 2014
Accepted 23 November 2014
Available online 26 November 2014

Keywords:
Hypertension
Lifestyle behaviors
Cohort study
Multicenter study
Children

ABSTRACT

Background/objectives: High blood pressure (HBP) is one of the most important risk factors for cardiovascular diseases and it has a high prevalence in pediatric populations. However, the determinants of the incidence of Pre-HBP and HBP in children are not well known. i) To describe the incidence of HBP in European children; and ii) to evaluate the effect of physical activity (PA) and sedentary behavior (SB) on the Pre-HBP and HBP.
Methods: The IDEFICS cohort study. A total of 16,228 children 2–9 years at baseline were recruited by complex sampling population-based survey in eight European countries. At baseline (T0), 5221 children were selected for accelerometer measurements; 5061 children were re-examined 2 years later (T1). We estimated the incidence of Pre-HBP and HBP and evaluate the effect of PA and SB on the Pre-HBP and HBP, by computing relative risks and the corresponding 95% confidence intervals (RR, 95% CI).
Results: Incidences of Pre-HBP and HBP per year were: 121/1000 children and 110/1000 children, respectively. We found that children maintaining SB > 2 h/d during the two year follow-up showed a RR of having HBP of 1.28 (1.03–1.60). Children in T1 not performing the recommended amount of PA (<60 min/d) have a RR of HBP of 1.53 (1.12 to 2.09). We found no association between pre-HBP and the behaviors.
Conclusion: The incidence of pre-HBP and HBP is high in European children. Maintaining sedentary behaviors during childhood increases the risk of developing HBP after two years of follow-up.

1. Introduction

Chronic non-communicable diseases are the main source of disease burden worldwide and are thus a major public health problem [1]. Among non-communicable diseases, hypertension has been shown to have the highest prevalence in adults [2], and studies have shown that blood pressure (BP) levels in childhood and adolescence greatly impact the development of hypertension in adulthood [3].

Among the factors that may influence blood pressure levels (e.g. genetics, intrauterine development, socioeconomic status, tobacco use, total and abdominal obesity), physical activity (PA) and sedentary behaviors (SB) have been shown inverse [4] and direct associations [5,6], respectively, with blood pressure in children.

* Corresponding author at: Faculdade de Medicina da USP, Departamento de Medicina Preventiva, 2° andar, sala 2162, Secretaria de Pós-Graduação, Av. Dr. Arnaldo, 455, Cerqueira César, CEP: 01246-903 São Paulo, Brazil.
E-mail address: augustocesar.demoraes@usp.br (A.C.F. de Moraes).
[1] In memoriam.

http://dx.doi.org/10.1016/j.ijcard.2014.11.175

Although the effects of PA and SB on BP have mainly been examined in isolation, there are studies suggesting that these behaviors have an aggregate effect in children [7]; however, few studies have quantified the association between combined PA/SB levels and cardiovascular risk in children, like blood pressure. On the other hand, PA/SB levels are associated with sociodemographic and economic variables. The influence of sociodemographic factors on PA/SB has been described in a review [8]. There is no consensus in the literature regarding socioeconomic variables as determinants of these behaviors since such differences may be attributed to the demographic context and characteristics of the populations studied rather than the individual [9,10].

Reproducing the same results in different population groups with different characteristics would increase their biological plausibility and provide a higher level of scientific evidence. For this reason, we have included results from a multi-national European study in this report. We tested our hypothesis, in cohort studies conducted with children within the IDEFICS study (***I**dentification and prevention of **D**ietary- and lifestyle-induced health **EF**fects **I**n **C**hildren and infant**S***).

Thus, we hypothesized that low levels of PA and high levels of SB may contribute to the development of high blood pressure (HBP).

2. Methods

2.1. Study population

The IDEFICS study (***I**dentification and prevention of **D**ietary- and lifestyle-induced health **EF**fects **I**n **C**hildren and infant**S***) is an epidemiological multicenter European study, aiming to identify nutritional and lifestyle-associated etiological factors of childhood obesity and related morbidities. A cohort of 16,228 children aged 2–9 years at baseline (51% of eligible sample), recruited from eight European countries (Germany, Hungary, Italy, Cyprus, Spain, Estonia, Sweden and Belgium), was examined in the baseline survey (T0) and it was the starting point of the prospective study with the largest European children's cohort established to date. They were assessed between September 2007 and May 2008 according to a standardized protocol. Details of the procedures of the IDEFICS project have been previously published [11,12]. These children were followed up longitudinally to assess their development and to determine the etiological associations between baseline predictors and selected follow-up end points by a follow-up survey 2 years later at T1 (September 2009 to May 2010).

As accelerometry was measured only in a random subset of children from every center (due to availability of accelerometers), when the objective measurement of physical activity (PA) was included in the analyses, the sample size was reduced. The present analysis was performed in 5221 children (32.2% of the sample; boys = 51%; age = 6.1 ± 1.8 years; mean ± s.d.) at T0, with a complete set of data including: systolic blood pressure (SBP), diastolic blood pressure (DBP), height, exposures [PA intensities, sedentary behavior (SB)] and confounding variables (Fig. 1). Parents or legal guardians provided written informed consent to participate in the full program or in a selected set of examination modules. For each survey center, the approval of the local Ethical Committee was obtained.

2.2. Outcomes

Data collection procedures were described previously [13]. An arm BP oscilometric monitor device WelchAllyn 42008™, previously validated in this age group was used [14]. It was previously tested for reliability and reproducibility in the IDEFICS project [13]. Two BP readings were taken after 10-min rest, with a 5-min interval between them, and the lowest reading was recorded. The inter-observer coefficients of variation were below 5% for both BP levels.

The outcomes for this study are: Pre-HBP and HBP [15]. Pre-HBP was defined as SBP or DBP between 90th to 95th percentile for age and height; and HBP defined as SBP or DBP above the 95th percentile for age and height too.

2.3. Principal exposures

PA and SB levels were considered independent variables.

Physical activity: was measured using a uniaxial accelerometer (Actigraph model GT1M GT1M or ActiTrainer. The sensor unit of both models is identical). Recordings were for at least 6 h/d for at least 3 days (2 weekdays and 1 day of weekend/holiday). The sampling interval (epoch) was set at 15 s. A measure of average total volume activity (hereafter called total PA) was expressed as the sum of recorded counts divided by total daily registered time expressed in minutes (counts/min; cpm). The cut-offs to define the PA intensity categories were derived from previously-validated cut-offs [16], with time spent in light PA (minutes) defined as the sum of time-per-day in which counts per epoch were 26 to 573 cpm. The time engaged in moderate PA was calculated based upon a cut-off of 574 to 1002 cpm per epoch. The time engaged in vigorous PA was

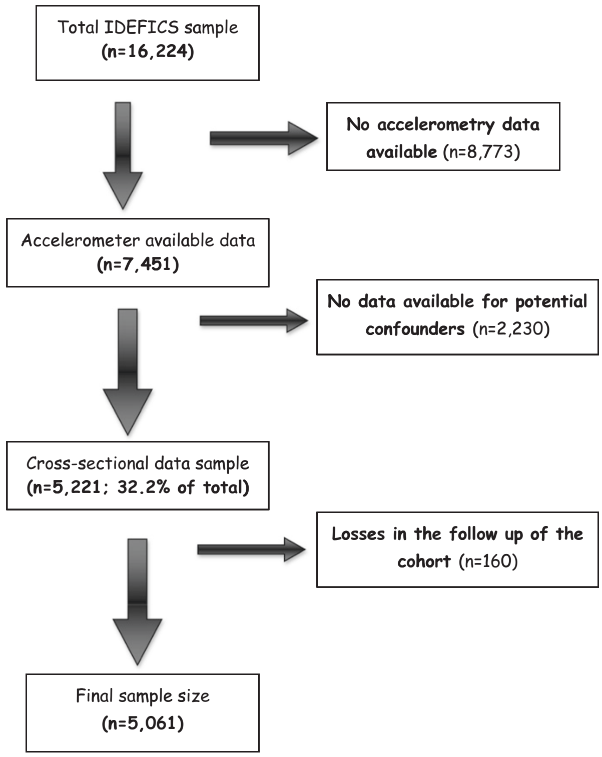

Fig. 1. Final sample size flowchart.

calculated based upon a cut-off of ≥1003 cpm per epoch. In addition, the time spent at the 'effective' intensity level was calculated as the sum of time spent in moderate + vigorous PA (MVPA).

Following current PA guidelines [17], subjects were classified in T0 and T1 as: meeting current PA recommendations when they accumulated at least 60 min/d of MVPA and not meeting current PA recommendations when MVPA was <60 min/d. We also established the variable change in PA based on the distribution in PA categories in T0 and T1; subjects were classified into the following categories: always ≥60 min/d (meeting current PA recommendations in both T0 and T1); ≥60 min/d → <60 min/d (meeting current PA recommendations in T0 to not meeting current PA recommendations in T1); <60 min/d → ≥60 min/d (not meeting current PA recommendations in T0 to meeting current PA recommendations T1); and always <60 min/d (not meeting current PA recommendations in T0 and T1).

Sedentary behaviors: The parental questionnaire was used to obtain information on children's sedentary behaviors'. Parents reported hours of TV/DVD/video viewing and computer/games-console use both for a typical weekday and weekend day. For the purpose of the current analysis, children's daily TV/DVD/video and computer/games-console use were summed to obtain the total screen time per day (the whole week). The used questionnaire had previously been tested for its reliability and validity in this population [18].

Thereafter, participants were classified into two groups according to the American Academy of Pediatrics (AAP's) guidelines on total screen time: ≤2 h/d and >2 h/d [19]. We also established the variable change in SB based on the distribution in SB categories in T0 and T1; subjects were classified into the following categories: always ≤2 h/d (meeting current SB recommendations in both T0 and T1); ≤2 h/d → >2 h/d (meeting current SB recommendations in T0 to not meeting current SB recommendations in T1); >2 h/d → ≤2 h/d (not meeting current SB recommendations in T0 to meeting current SB recommendations T1); and always >2 h/d (not meeting current SB recommendations in T0 and T1).

2.4. Potential confounders

The potential confounders for this study were divided into two groups: Contextual Factors and Individual Factors, and described bellow:

Contextual Factors

Centres in 8 European countries: Belgium, Cyprus, Estonia, Greece, Germany, Hungary, Italy, Spain and Sweden.

Seasonality: A variable was computed by recoding the original variable "blood drawing date" into "seasonality", as follows: winter (from 21st December to 20th March, coded as 1), autumn (from 21st September to 20th December, coded as 2), spring (from 21st March to 20th June, coded as 3), and summer (from 21st June to

20th September, coded as 4), as performed in previous studies. As the IDEFICS study was performed during the academic year, only a few children (n = 2%) were assessed in the first days of summer. They were included along with those assessed during spring. Therefore, the final variable was composed of three groups: winter (coded as 1), autumn (coded as 2), and spring (coded as 3).

Individual Factors

Study group: The children were divided into two groups: Intervention and control. The intervention group received a community-based intervention program with multicomponent education topics to promote healthy lifestyle. Children received nutrition education at school and community activities were performed to prevent obesity/overweight and metabolic syndrome components. The control group did not receive any intervention.

Age group: 2–5 years; 6–9 years and 10–12 years (only T1 analyses);

Parental education of the family provider: determined with a self-reported questionnaire and categorized according to the International Standard Classification of Education (ISCED) [20] in five levels: ISCED 1 = illiterate; ISCED 2 = up to 4th year of primary school; ISCED 3 = completed primary school; ISCED 4 = completed high school; and ISCED 5 = completed higher education.

Waist circumference: It was measured at the midpoint between the lowest rib cage and the top of the iliac crest with a non-elastic tape to the nearest 0.1 cm. The intraobserver technical errors of measurement were between 0.53 and 1.75 cm and interobserver reliability was greater than 94.9% [21], for this circumference.

2.5. Statistical analysis

The descriptive analyses were performed by mean (continuous variables) and percentage (categorical variables) and respective 95% confidence intervals (95% CI).

We calculated the cumulative incidence for the two years of follow-up and 95% CI of both outcomes: Pre-HBP and HBP for total and principal exposures. The magnitude of these associations was subsequently expressed in, unadjusted and adjusted, relative risk (RR) and 95% CI. Multinomial multilevel regression models using mixed effects intercept were applied to estimate the effect of PA and SB on Pre-HBP and HBP incidence [22,23]. The context variable used was the center.

For the adjusted analysis we developed a conceptual framework (Fig. 2) previously separated into five levels (the association of these levels is not shown): 1) center, seasonality; 2) sex; age (years); 3) parental education; 4) waist circumference; and 5) PA and SB. In this model, variables were controlled for those in the same level but also for those in the higher one [24]. P-values ≤ 0.20 were adopted in the univariate analysis [24] (as necessary to include a variable in the multivariate analysis and, then, it was entered through the hierarchical model method following the levels above) or when there was more than 10% modification in RR of any variable already in the model.

Multilevel analyses were performed with two objectives: 1st) to test the associations between BP categories and two separate measures of individual behaviors; and 2nd) to test the extent to which country-specific characteristics and contextual variables mediate the associations between BP categories and PA and SB.

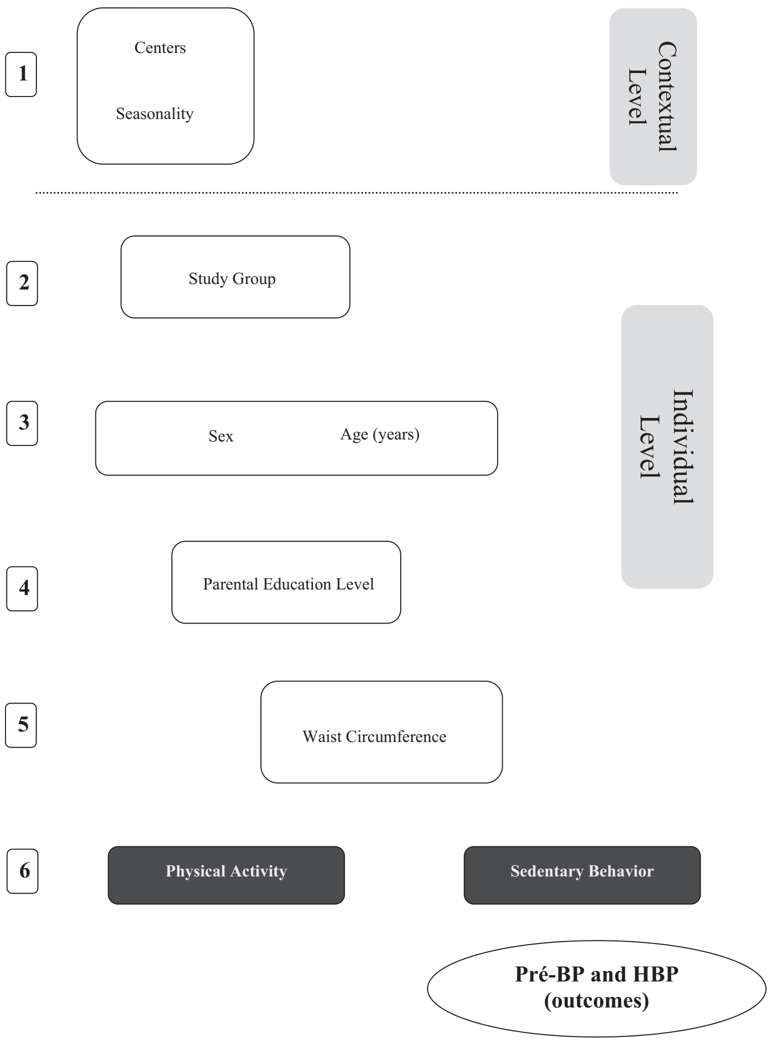

Fig. 2. Theoretical conceptual model of the association between contextual and individual variables on children's blood pressure categories. The effect of each variable on the outcome was adjusted for other variables in the same model or above in the hierarchical model. Variables with P > 0.2 were not included in the subsequent adjustment models. HBP = high blood pressure. The principal's independent variables are described in level 6.

Significance was set at P-values < 0.05 and the Stata 12 (Stata Corp., College Station, TX, USA) was used for all statistical calculations. All analyses were adjusted for the clustered nature of the sample using the "svy" set of commands.

3. Results

We performed sensitivity analyses in the sample by comparing the prevalence of the outcomes among children who had complete data (baseline = T0 and cohort = T1), and those who did not have complete data (only T0). We found no significant differences (P > 0.05) in prevalence in this analysis. Fig. 1 presents a description of the study sample, which represents 32.2% of the total IDEFICS study sample. A strong point of our study is that we only had 3.1% loss in the two years of follow-up (n = 5061).

Table 1 shows the characteristics of the sample at T0 and T1 of data collection. The only variable found showing significant differences between boys and girls was the increased prevalence of high sedentary behaviors (>2 h/d) both at T0 and T1, being higher in males than in females.

Cumulative incidences of both outcomes (Pre-HBP and HBP) according to the main exhibits examined are presented in Table 2. Significant effects were only found in the incidence of HBP for PA and SB at follow-up in cohort and in those maintaining high SB (>2 h/d) at follow-up.

Children who perform <60 min/d of PA at the time of the second data collection or maintained higher SB (>2 h/d) during the period of follow-up are at risk for HBP (Table 3).

4. Discussion

To the best of our knowledge, this is the first study examining the incidence of HBP in children, and exploring the effect of PA and SB on this incidence. We analyzed the incidence of pre-HBP and HBP in a large sample of European children from eight different countries, and also the effect of PA and SB in the incidences of these outcomes. We found a high incidence of both outcomes during the follow-up of two years. Children not meeting with the PA recommendation (≥60 min/d) in T1 or maintaining high SB (>2 h/d) at follow-up are at high risk of developing HBP. Our findings are relevant, because HBP is considered the risk factor with the highest attributable fractions for cardiovascular diseases (CVD) mortality (40.6%) [25] and some studies have shown that if HBP is developed in childhood and adolescence. Therefore, this could be crucial for developing CVD such as stroke and myocardial infarction in adulthood [3]. Including data from a cohort study adds consistency and temporality to our report as some of the results were similar in different populations and the risk factor came first than the outcome (Hill's principles) [26].

Our results corroborate studies that have evaluated, in cross-sectional studies, the association of PA on BP levels [27,28]. Several mechanisms can explain the positive effects PA induces on BP levels. There is strong evidence that the sheer stress caused by regular PA has a powerful effect on the release of vasodilator factors produced by the vascular endothelium [29], such as nitric oxide and endothelium-derived hyperpolarizing factor (EDHF) [30], and the children that

Table 1
Descriptive analysis of characteristics of the sample on cross-sectional and cohort, and their respective confidence intervals 95% (95% CI), according to independent variables.

Independents variables	Cross-sectional (n = 5221)				Cohort (n = 5061)			
	Males (n = 2638)		Female (n = 2583)		Males (n = 2485)		Female (n = 2576)	
	%	95% CI	%	95% CI	%	95% CI	%	95% CI
Age (years)								
2–5 years	42,8	(40,9–44,6)	39,6	(37,8–41,6)	16,5	(9,1–11,5)	15,5	(14,1–16,9)
6–9 years	57,2	(55,4–59,1)	60,4	(58,4–62,2)	69,3	(66,2–69,8)	70,3	(68,5–72,2)
10–12 years	–	–	–	–	14,2	(13,4–16,2)	14,3	(12,9–15,6)
Countries								
Belgium	6,3	(5,4–7,2)	7,5	(6,5–8,5)	14,2	(12,7–15,4)	13,8	(12,4–15,1)
Cyprus	7,6	(6,6–8,6)	7	(6,0–8,0)	5,1	(4,3–6,0)	6,7	(5,7–7,7)
Estonia	16,1	(14,7–17,5)	16,9	(15,5–18,4)	20,1	(18,5–21,6)	21,5	(19,9–23,1)
Germany	11,5	(10,3–12,7)	12,9	(11,6–14,2)	14,5	(13,2–15,9)	13,8	(12,4–15,1)
Hungary	16,3	(14,9–17,7)	16,4	(14,9–17,8)	5,7	(4,8–6,6)	5,5	(4,6–6,4)
Italy	12,2	(10,9–13,4)	11,4	(10,2–12,6)	12,6	((11,3–13,9)	10,4	(9,2–11,6)
Spain	20,1	(18,6–21,6)	19,2	(17,6–20,7)	20,1	(18,5–21,7)	19,4	(17,9–20,9)
Sweden	9,9	(8,8–11,1)	8,7	(7,6–9,8)	7,8	(6,8–8,9)	9	(7,9–10,1)
Parental education level (ISCED)								
ISCED 1	2,1	(1,5–2,6)	2,3	(1,7–2,9)	2,2	(1,6–2,8)	2,4	(1,8–3,0)
ISCED 2	6,4	(5,4–7,3)	6,2	(5,3–7,1)	7,5	(6,4–8,6)	6,4	(5,5–7,4)
ISCED 3	34,5	(32,7–36,3)	33,2	(31,3–35,0)	32,5	(30,6–34,4)	32,1	(30,3–34,0)
ISCED 4	17	(15,6–18,5)	18,8	(17,3–20,4)	21,2	(19,5–22,8)	20,1	(19,3–22,5)
ISCED 5	40	(38,1–41,9)	39,6	(37,6–41,4)	36,6	(34,7–38,6)	38,2	(36,3–40,1)
Physical activity								
<60 min/d	78,4	(76,8–79,9)	91,4	(90,3–92,5)	74,3	(7,26–76,0)	90,4	(89,2–91,5)
≥60 min/d	**21,6**	**(20,1–23,2)**	**8,6**	**(7,5–9,7)**	**25,7**	**(24,0–27,4)**	**9,6**	**(8,5–10,8)**
Sedentary behavior								
≤2 h/d	63,7	(61,9–65,6)	70,5	(68,8–72,3)	51,1	(49,1–53,2)	60,1	(58,2–62,0)
>2 h/d	**36,3**	**(34,4–38,1)**	**29,5**	**(27,7–31,2)**	**48,9**	**(46,9–50,9)**	**39,9**	**(38,0–41,8)**
Nutritional status by BMI								
Undernutrition	10,2	(9,0–11,3)	10,6	(9,4–11,8)	10,3	(9,1–11,5)	10,1	(8,9–11,3)
Eutrophic	72,5	(70,8–74,2)	69,5	(67,8–71,3)	68	(66,2–69,8)	67,8	(66,0–69,6)
Overweight	10,8	(9,7–12,0)	13,6	(12,3–15,0)	14,8	(13,4–16,2)	16,2	(14,8–17,6)
Obese	6,5	(5,6–7,5)	6,3	(5,3–7,2)	6,8	(5,8–7,8)	5,9	(5,0–6,8)
Blood Pressure								
Normal	78,8	(77,2–80,4)	75,5	(73,8–77,2)	84,5	(83,1–86,0)	83,7	(82,3–85,2)
Pre-HBP	12,1	(10,9–13,4)	13,2	(11,8–14,5)	8,8	(7,7–9,9)	8,4	(7,3–9,4)
HBP	9,1	(7,9–10,2)	11,3	(10,0–12,5)	6,7	(5,7–7,6)	7,9	(6,8–8,9)

Significant associations are in bold.
HBP = high blood pressure.

Table 2
Cumulative incidence and respective confidence interval 95% (95% CI) of the outcomes per 1000 individual lifestyle behaviors.

Independents variables	Incidence	
	Pre-HBP	HBP
Physical activity (PA) in cross-sectional		
≥60 min/d	**84 (56–112)**	**66 (41–91)**
<60 min/d	81 (69–92)	75 (64–86)
Physical activity (PA) in cohort		
≥60 min/d	**83 (61–104)**	**40 (25–55)**
<60 min/d	63 (54–72)	65 (56–74)
PA changes (cross-sectional == > cohort)		
Always ≥ 60 min/d	68 (43–93)	43 (23–63)
≥60 min/d == > <60 min/d	52 (40–63)	42 (31–52)
<60 min/d == > ≥60 min/d	105 (68–143)	35 (13–58)
Always < 60 min/d	74 (61–87)	86 (72–99)
Sedentary behavior (SB) in cross-sectional		
≤2 h/d	67 (56–78)	57 (47–67)
>2 h/d	66 (54–79)	65 (53–77)
Sedentary behavior (SB) in cohort		
≤2 h/d	66 (56–76)	55 (46–63)
>2 h/d	69 (54–83)	73 (58–88)
SB changes (cross-sectional == > cohort)		
Always ≤ 2 h/d	67 (56–79)	56 (45–67)
≤2 h/d == > >2 h/d	62 (45–80)	51 (35–67)
>2 h/d == > ≤2 h/d	64 (37–91)	61 (35–87)
Always > 2 h/d	70 (53–88)	78 (59–96)
Total	121 (53–188)	110 (93–162)

HBP = high blood pressure.

perform physical activity less than 60/min/d have lower vasodilation capacity of the endothelium and this could be the biological mechanism by which they develop HBP.

An important result we found was that children who maintained SB > 2 h/d during the two-year follow-up showed a high incidence of HBP. There are several possible physiological mechanisms by which SB may contribute to increased BP, and more research is needed to analyze the pathophysiological processes of increased BP due to high SB. One possible biological explanation is that SB changes the myokine response in the skeletal muscle and these alterations promote the endothelial dysfunction in the cardiovascular system by increase of the pro-inflammatory adipokines. Consequently, this increase could be the start of the pathological processes of atherosclerosis, and progressively develop into hypertension [31].

The results of this study are in agreement with some cross-sectional studies suggesting that lower PA level is associated with higher levels of BP [32]. Additionally, present results corroborate with previous survey observations [33] and a recent review [5] suggesting that SB is an independent risk factor for detrimental cardiovascular outcomes independent of PA level. Our results are of importance and highlight that the youth should be encouraged to engage in recommended levels of MVPA and reduce excessive time spent in screen-based sedentary behavior. In adolescents our group found that recommended levels of MVPA could attenuates the harmful effects of SB in increased blood pressure [6].

The behavioral patterns under consideration during childhood tend to continue into adulthood [34] and high levels of sedentary behaviors in adults increase the risk of mortality from cardiovascular diseases [35,36].

A limitation of this study is that it was not possible to adjust the analysis for other potentially BP-associated factors in either of the large children sample, such as genetics or intrauterine development, but we developed an adjusted analysis for a large set of potential of confounders. On the other hand, the diverse geographic origin of the samples, the cohort design consequently, temporal sequence between risk factor and outcome can be established, the use of objective measures

Table 3
Relative risk and respective confidence interval 95% (RR, 95% CI) by multilevel Poisson according lifestyle behaviors changes.

Independents variables	Pre-high blood pressure		High blood pressure	
	RR (95% CI) unadjusted	RR (95% CI) adjusted[a]	RR (95% CI) unadjusted	RR (95% CI) adjusted[a]
Random effects intercept	0,06 (0,03–0,10)	0,08 (0,04–0,15)	0,06 (0,03–0,10)	0,08 (0,04–0,15)
Physical activity (PA) in cross-sectional				
≥60 min/d	**1,00**	**1,00**	**1,00**	**1,00**
<60 min/d	0,93 (0,64–1,37)	0,89 (0,61–1,31)	1,10 (0,82–1,49)	0,93 (0,68–1,27)
Physical activity (PA) in cohort				
≥60 min/d	**1,00**	**1,00**	**1,00**	**1,00**
<60 min/d	0,85 (0,63–1,16)	0,83 (0,61–1,13)	**1,67 (1,25–2,24)**	**1,53 (1,12–2,09)**
PA changes (cross-sectional == > cohort)				
Always ≥ 60 min/d	1,00	1,00	1,00	1,00
≥60 min/d == ><60 min/d	0,84 (0,54–1,32)	0,80 (0,51–1,27)	1,05 (0,71–1,54)	0,97 (0,64–1,47)
<60 min/d == > ≥60 min/d	1,26 (0,73–2,16)	1,23 (0,72–2,12)	0,67 (0,38–1,19)	0,65 (0,36–1,18)
Always < 60 min/d	1,03 (0,67–1,57)	0,98 (0,64–1,51)	**1,66 (1,16–2,37)**	1,43 (0,97–2,12)
Sedentary behavior (SB) in cross-sectional				
≤2 h/d	1,00	1,00	1,00	1,00
>2 h/d	1,05 (0,82–1,34)	1,04 (0,81–1,33)	**1,24 (1,04–1,48)**	1,20 (0,99–1,44)
Sedentary behavior (SB) in cohort				
≤2 h/d	1,00	1,00	1,00	1,00
>2 h/d	1,03 (0,82–1,31)	1,01 (0,80–1,28)	1,17 (0,98–1,39)	1,16 (0,97–1,40)
SB changes (cross-sectional == > cohort)				
Always ≤ 2 h/d	1,00	1,00	1,00	1,00
≤2 h/d == > >2 h/d	0,96 (0,70–1,35)	0,95 (0,69–1,29)	1,07 (0,84–1,35)	1,06 (0,82–1,35)
>2 h/d == > ≤2 h/d	0,93 (0,61–1,43)	0,94 (0,61–1,44)	1,15 (0,85–1,56)	1,07 (0,78–1,46)
Always > 2 h/d	1,08 (0,81–1,44)	1,05 (0,78–1,44)	**1,32 (1,07–1,63)**	**1,28 (1,03–1,60)**
Random effects — countries	0,73 (0,42–1,29)	0,72 (0,41–1,27)	1,21 (0,71–2,06)	1,18 (0,70–2,02)
Akaike information criterion	2076,51	2065,9	3334,77	3098,58

Significant associations are in bold.
[a] This analysis was adjusted for potential confounders: country, seasonality, sex, age, parental education and waist circumference.

to assess PA and SB and multilevel adjusted analysis are some of the main strengths of our study.

5. Conclusions

According to our results, the incidence of pre-HBP and HBP is high in European children, low levels of PA are a risk factor for developing HBP and to maintain sedentary behaviors increases the risk of developing HBP after two years of follow-up. These results suggest that regular PA should be promoted and SB should be discouraged in children to prevent high blood pressure and its consequences in adulthood.

Funding and conflict of interest

Augusto César F. de Moraes is in receipt of a PhD scholarship from the São Paulo State Research Foundation (FAPESP: proc. 2011/11137-1 and 2011/20662-2); Luis A. Moreno was given scholarship of visiting professor from Brazilian government by Science without Borders Program by CNPq (National Counsel of Technological and Scientific Development) (proc. 007/2012) and CAPES (Coordination of Improvement of Higher Education Personnel) (proc. 007/2012). The GENUD Research Group was co-financed by the European Regional Development Fund (MICINN-FEDER). The IDEFICS Study was supported by the European Community Sixth Research, Technological Development and Demonstration Framework Programme (FOODCT-0 16181). The authors have no conflict to disclose and they certify that they comply with the Principles of Ethical Publishing in the International Journal of Cardiology.

Acknowledgments

We gratefully acknowledge all participating adolescents and their parents for their collaboration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijcard.2014.11.175.

References

[1] R. Beaglehole, R. Horton, Chronic diseases: global action must match global evidence, Lancet 376 (2010) 1619–1621.
[2] G. Danaei, M.M. Finucane, J.K. Lin, G.M. Singh, C.J. Paciorek, M.J. Cowan, et al., National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5 · 4 million participants, Lancet 377 (2011) 568–577.
[3] R.M. Lauer, W.R. Clarke, Childhood risk factors for high adult blood pressure: the Muscatine Study, Pediatrics 84 (1989) 633–641.
[4] D.G. Christofaro, R.M. Ritti-Dias, A. Chiolero, R.A. Fernandes, J. Casonatto, A.R. de Oliveira, Physical activity is inversely associated with high blood pressure independently of overweight in Brazilian adolescents, Scand. J. Med. Sci. Sports 23 (2013) 317–322.
[5] A. Fröberg, A. Raustorp, Objectively measured sedentary behaviour and cardio-metabolic risk in youth: a review of evidence, Eur. J. Pediatr. 173 (2014) 845–860.
[6] A.C. de Moraes, H.B. Carvalho, J.P. Rey-López, L. Gracia-Marco, L. Beghin, A. Kafatos, et al., Independent and combined effects of physical activity and sedentary behavior on blood pressure in adolescents: gender differences in two cross-sectional studies, PLoS One 8 (2013) e62006.
[7] S. Bel-Serrat, T. Mouratidou, A.M. Santaliestra-Pasías, L. Iacoviello, Y.A. Kourides, S. Marild, et al., Clustering of multiple lifestyle behaviours and its association to cardiovascular risk factors in children: the IDEFICS study, Eur. J. Clin. Nutr. 67 (2013) 848–854.
[8] C. Edwardson, T. Gorely, Parental influences on different types and intensities of physical activity in youth: a systematic review, Psychol. Sport Exerc. 11 (2010) 522–535.
[9] A. Seabra, D. Mendonça, M. Thomis, L. Anjos, J. Maia, Biological and socio-cultural determinants of physical activity in adolescents, Cad. Saude Publica 24 (2008) 721–736.
[10] G. Rose, Sick individuals and sick populations, Int. J. Epidemiol. 14 (1985) 32–38.
[11] M. Dello Russo, W. Ahrens, T. De Vriendt, S. Marild, D. Molnar, L.A. Moreno, et al., Gestational weight gain and adiposity, fat distribution, metabolic profile, and blood pressure in offspring: the IDEFICS project, Int. J. Obes. (Lond) 37 (2013) 914–919.
[12] W. Ahrens, K. Bammann, A. Siani, K. Buchecker, S. De Henauw, L. Iacoviello, et al., The IDEFICS cohort: design, characteristics and participation in the baseline survey, Int. J. Obes. (Lond) 35 (Suppl. 1) (2011) S3–S15.
[13] M. Suling, A. Hebestreit, J. Peplies, K. Bammann, A. Nappo, G. Eiben, et al., Design and results of the pretest of the IDEFICS study, Int. J. Obes. (Lond) 35 (Suppl. 1) (2011) S30–S44.
[14] J. Sorof, S. Daniels, Obesity hypertension in children: a problem of epidemic proportions, Hypertension 40 (2002) 441–447.
[15] NHBPEP. National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents, The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents, Pediatrics 114 (2004) 555–576.
[16] K.R. Evenson, D.J. Catellier, K. Gill, K.S. Ondrak, R.G. McMurray, Calibration of two objective measures of physical activity for children, J. Sports Sci. 26 (2008) 1557–1565.
[17] U.S. Department of Health and Human Services, Key Guidelines for Children and AdolescentsAccessed September 2011 ed: http://www.health.gov/PAGuidelines/ 2008.
[18] V. Verbestel, S. De Henauw, K. Bammann, G. Barba, C. Hadjigeorgiou, G. Eiben, et al., Are context-specific measures of parental-reported physical activity and sedentary behaviour associated with accelerometer data in 2–9-year-old European children? Public Health Nutr. 1–9 (2014).
[19] Education AAoPCoP, American Academy of Pediatrics: children, adolescents, and television, Pediatrics 107 (2001) 423–426.
[20] UNESCO. United Nations Educational Scientific and Cultural Organization, International Standard Classification of Education (ISCED), in: UIS (Ed.), , 2012, pp. 1–88.
[21] E. Nagy, G. Vicente-Rodriguez, Y. Manios, L. Béghin, C. Iliescu, L. Censi, et al., Harmonization process and reliability assessment of anthropometric measurements in a multicenter study in adolescents, Int. J. Obes. (Lond) 32 (Suppl. 5) (2008) S58–S65.
[22] A.V. Diez Roux, Next steps in understanding the multilevel determinants of health, J. Epidemiol. Community Health 62 (2008) 957–959.
[23] T. Snijders, R. Bosker, Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling, SAGE Publication, 1999.
[24] C. Victora, S. Huttly, S. Fuchs, M. Olinto, The role of conceptual frameworks in epidemiological analysis: a hierarchical approach, Int. J. Epidemiol. 26 (1997) 224–227.
[25] A.S. Go, D. Mozaffarian, V.L. Roger, E.J. Benjamin, J.D. Berry, W.B. Borden, et al., Executive summary: heart disease and stroke statistics—2013 update: a report from the American Heart Association, Circulation 127 (2013) 143–152.
[26] A.B. HILL, The environment and disease: association or causation? Proc. R. Soc. Med. 58 (1965) 295–300.
[27] U. Ekelund, S. Brage, K. Froberg, M. Harro, S. Anderssen, L. Sardinha, et al., TV viewing and physical activity are independently associated with metabolic risk in children: the European Youth Heart Study, PLoS Med. 3 (2006) 2449–2457.
[28] C. Lazarou, D.B. Panagiotakos, A.L. Matalas, Lifestyle factors are determinants of children's blood pressure levels: the CYKIDS study, J. Hum. Hypertens. 23 (2009) 456–463.
[29] A.S. Zago, A. Zanesco, Nitric oxide, cardiovascular disease and physical exercise, Arq. Bras. Cardiol. 87 (2006) e264–e270.
[30] B.A. Kingwell, Nitric oxide-mediated metabolic regulation during exercise: effects of training in health and cardiovascular disease, FASEB J. 14 (2000) 1685–1696.
[31] B.K. Pedersen, M.A. Febbraio, Muscles, exercise and obesity: skeletal muscle as a secretory organ, Nat. Rev. Endocrinol. 8 (2012) 457–465.
[32] A.R. Gaya, P. Silva, C. Martins, A. Gaya, J.C. Ribeiro, J. Mota, Association of leisure time physical activity and sports competition activities with high blood pressure levels: study carried out in a sample of Portuguese children and adolescents, Child Care Health Dev. 37 (2011) 329–334.
[33] J.P. Chaput, T.J. Saunders, M. Mathieu, M. Henderson, M.S. Tremblay, J. O'Loughlin, et al., Combined associations between moderate to vigorous physical activity and sedentary behaviour with cardiometabolic risk factors in children, Appl. Physiol. Nutr. Metab. 38 (2013) 477–483.
[34] V. Cleland, T. Dwyer, A. Venn, Which domains of childhood physical activity predict physical activity in adulthood? A 20-year prospective tracking study, Br. J. Sports Med. 46 (2012) 595–602.
[35] H.P. van der Ploeg, T. Chey, R.J. Korda, E. Banks, A. Bauman, Sitting time and all-cause mortality risk in 222 497 Australian adults, Arch. Intern. Med. 172 (2012) 494–500.
[36] K. Wijndaele, S. Brage, H. Besson, K.T. Khaw, S.J. Sharp, R. Luben, et al., Television viewing and incident cardiovascular disease: prospective associations and mediation analysis in the EPIC Norfolk Study, PLoS One 6 (2011) e20058.

Appendix E

Paper: Domain-specific self-reported and objectively measured physical activity in children

Contribution to the manuscript I herewith certify that I performed all statistical analyses, interpreted the results and drafted parts of the manuscript and revised it critically for important intellectual content.
The manuscript was still under review when this thesis was submitted. Only the abstract is presented here. The complete manuscript can be obtained from the author upon request.

Domain-specific self-reported and objectively measured physical activity in children

O. Sprengeler[1], N. Wirsik[1], A. Hebestreit[1], D. Herrmann[1], W. Ahrens[1]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

Correspondence
Prof. Dr. Wolfgang Ahrens
Department of Epidemiological Methods and Etiologic Research
Leibniz Institute for Prevention Research and Epidemiology - BIPS
Achterstr. 30, 28359 Bremen, Germany
Phone: +49-421-218 56822
Email: ahrens@bips.uni-bremen.de

Abtract

Purpose: Little is known about the extent different domains contribute to total sedentary (SED), light (LPA) and moderate-to-vigorous physical activity (MVPA). We compared domain-specific PA behaviour assessed by questionnaire and accelerometry during a typical school week.

Method: 298 German school children and adolescents aged 6-17 years wore an accelerometer for one week and completed a PA recall-questionnaire for the same period. Spearman coefficients were used to compare self-reported and objectively measured total and domain-specific PA.

Results: Self-reported PA was generally over-reported when compared to accelerometry. The agreement of self-reported and objectively measured PA was low for total LPA (r=0.09, 95%-CI:-0.03-0.20), total MVPA (r=0.21, 95%-CI: 0.10-0.32). In contrast, moderate agreement was found for total SED (r=0.44, 95%-CI: 0.34-0.53), LPA during transport (r= 0.59; 95%-CI: 0.49-0.67) and MVPA during organized sports activities (r= 0.54; 95%-CI: 0.38-0.67). About half of the objectively measured SED, LPA and MVPA (55%, 53% and 46%, respectively) occurred during school time, while organized sports activities contributed 24% to total MVPA.

Conclusions: The school setting is the domain contributing about half to total SED, LPA and MVPA in children. Accelerometry should be preferred over questionnaires to assess duration and intensity of PA in youth while domain-specific data require self-reported information.

Appendix F

Paper: Estimating energy expenditure from gait intensity

Contribution to the manuscript I herewith certify that I performed all statistical analyses, interpreted the results, and drafted parts of the manuscript and revised it critically for important intellectual content.
The manuscript was still under review when this thesis was submitted. Only the abstract is presented here. The complete manuscript can be obtained from the author upon request.

Estimating energy expenditure from gait intensity
M. Brandes[1], N. Wirsik[2], V. Hannöver[3], M. Fikus[3]

[1]Institute of Sports Science, Oldenburg University, Germany
[2]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[3]Institute of Sports Science, Bremen University, Germany

Correspondence
PD Dr. Mirko Brandes
Carl-von-Ossietzky University of Oldenburg
Institute of Sports Science
Ammerländer Heerstraße 114-118
Phone: +49-441-798 3177

Email: mirko.brandes@uni-oldenburg.de

Abtract

Purpose: A reliable estimation of activity-related energy expenditure (AEE) from gait intensity may increase the value of step counting in daily life. The current study aims to improve previous equations for calculating AEE based on gait intensity measurements, i.e. the number of steps taken in a one minute interval.

Methods: 167 participants (82 females, 85 males, 6-81 years old) were equipped with a step activity monitor (Step Watch 3.0, Orthocare Innovations, Washington, USA). Total energy expenditure (TEE) was measured with a mobile oxygen analyzer (MetaMax 3b, Cortex Biophysik, Leipzig, Germany). Resting energy expenditure (REE) was measured for 30 min, followed by three walking conditions with slow, moderate and fast pace on level ground in outdoor conditions. AEE was calculated as TEE - REE and expressed in kJ/min. Mixed linear models were used to derive an energy prediction equation. Leave-one-out cross validation was utilized to calculate the accuracy of the models.

Results: A model involving the number of gait cycles, weight and height performed best (r^2=0.644, RMSE=5.69, MAPE=18.92). The model improved only marginally with additional variables such as age, or cross-interactions. Furthermore, discriminating fast walking with a pronounced variation on AEE did not improve the model prediction.

Conclusions: Gait intensity measured by step counters in one minute intervals offers a reliable estimation of AEE across a wide age range. This approach is superior to the prediction of AEE by means of daily step counts, but does not reach the accuracy of accelerometers mounted on the lower back and using raw acceleration for AEE prediction.

Bibliography

Ahrens W, Bammann K, Henauw S de, Halford J, Palou A, Pigeot I, Siani A, Sjostrom M, European consortium of the IDEFICS project (2006). Understanding and preventing childhood obesity and related disorders – IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis* **16**(4): 302–308.

Ahrens W, Bammann K, Siani A, Buchecker K, Henauw S de, Iacoviello L, Hebestreit A, Krogh V, Lissner L, Marild S, Molnar D, Moreno LA, Pitsiladis YP, Reisch L, Tornaritis M, Veidebaum T, Pigeot I, IDEFICS consortium (2011). The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obesity* **35**(Suppl 1): S3–15.

Ahrens W, Pigeot I, Pohlabeln H, Henauw S de, Lissner L, Molnar D, Moreno LA, Tornaritis M, Veidebaum T, Siani A, IDEFICS consortium (2014). Prevalence of overweight and obesity in European children below the age of 10. *Int J Obesity* **38**(Suppl 2): S99–107.

American Academy of Pediatrics: Committee on public education (2001). Children, adolescents, and television. *Pediatrics* **107**(2): 423–426.

Bailey DP, Fairclough SJ, Savory LA, Denton SJ, Pang D, Deane CS, Kerr CJ (2012). Accelerometry-assessed sedentary behaviour and physical activity levels during the segmented school day in 10-14-year-old children: the HAPPY study. *Eur J Pediatr* **171**(12): 1805–1813.

Bammann K, Peplies J, Sjöström M, Lissner L, Henauw S de, Galli C, Iacoviello L, Krogh V, Marild S, Pigeot I, Pitsiladis Y, Pohlabeln H, Reisch L, Siani A, Ahrens W, on behalf of the IDEFICS consortium (2006). Assessment of diet, physical activity and biological, social and environmental factors in a multi-centre European project on diet- and lifestyle-related disorders in children (IDEFICS). *Public Health* **14**(5): 279–289.

Baum L, Petrie T, Soules G, Weiss N (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann Math Stat* **41**(1): 164–171.

Beneke R, Leithäuser R (2008). Körperliche Äktivitat im Kindesalter - Messverfahren. *Deut Z Sportmed* **10**(59): 215–221.

Bland JM, Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**(8476): 307–310.

Box GE, Cox DR (1964). An analysis of transformations. *J R Stat Soc Series B Methodol* **26**(2): 211–252.

Brandes M, van Hees VT, Hannöver V, Brage S (2012). Estimating energy expenditure from raw accelerometry in three types of locomotion. *Med Sci Sports Exerc* **44**(11): 2235–2242.

Bürgi F, Meyer U, Granacher U, Schindler C, Marques-Vidal P, Kriemler S, Puder JJ (2011). Relationship of physical activity with motor skills, aerobic fitness and body fat in preschool children: a cross-sectional and longitudinal study (Ballabeina). *Int J Obes* **35**(7): 937–944.

Choi L, Liu Z, Matthews CE, Buchowski MS (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Med Sci Sports Exerc* **43**(2): 357–364.

Cole TJ, Green PJ (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat Med* **11**(10): 1305–1319.

Cole TJ, Stanojevic S, Stocks J, Coates AL, Hankinson JL, Wade AM (2009). Age- and size-related reference ranges: a case study of spirometry through childhood and adulthood. *Stat Med* **28**(5): 880–898.

Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, Oja P (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* **35**(8): 1381–1395.

Crouter SE, Horton M, Bassett Jr DR (2012). Use of a two-regression model for estimating energy expenditure in children. *Med Sci Sports Exerc* **44**(6): 1177–1185.

Eilers PHC (2003). A perfect smoother. *Anal Chem* **75**: 3631–3636.

Eilers PHC, Marx BD (1996). Flexible smoothing with B -splines and penalties. *Stat Sci* **11**(2): 89–121.

Ekelund U, Sjostrom M, Yngve A, Poortvliet E, Nilsson A, Froberg K, Wedderkopp N, Westerterp K (2001). Physical activity assessed by activity monitor and doubly labeled water in children. *Med Sci Sports Exerc* **33**(2): 275–281.

Ekelund U, Hildebrand M, Collings PJ (2014). Physical activity, sedentary time and adiposity during the first two decades of life. *Proc Nutr Soc* **73**(2): 319–329.

Escalante Y, Garcia-Hermoso A, Backx K, Saavedra JM (2014). Playground designs to increase physical activity levels during school recess: a systematic review. *Health Educ Behav* **41**(2): 138–144.

Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG (2008). Calibration of two objective measures of physical activity for children. *J Sports Sci* **26**(14): 1557–1565.

Fink GA (2003). *Mustererkennung mit Markov-Modellen: Theorie - Praxis - Anwendungsgebiete*. 1st edition. Stuttgart: Teubner.

Forney Jr G (1973). The viterbi algorithm. *Proc IEEE* **61**(3): 268–278.

Frasso G, Eilers PHC (2015). L- and V-curves for optimal smoothing. *Stat Model* **15**(1): 91–111.

Freedson PS, Melanson E, Sirard J (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc* **30**(5): 777–781.

Freedson PS, Pober DM, Janz KF (2005). Calibration of accelerometer output for children. *Med Sci Sports Exerc* **37**(Suppl 11): S523–530.

Guinhouya BC, Lemdani M, Vilhelm C, Durocher A, Hubert H (2009a). Actigraph-defined moderate-to-vigorous physical activity cut-off points among children: statistical and biobehavioural relevance. *Acta Paediatr* **98**(4): 708–714.

Guinhouya BC, Lemdani M, Vilhelm C, Hubert H, Apete GK, Durocher A (2009b). How school time physical activity is the “big one” for daily activity among schoolchildren: a semi-experimental approach. *J Phys Act Health* **6**(4): 510–519.

Hansen PC (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev* **34**(4): 561–580.

Henauw S de, Verbestel V, Marild S, Barba G, Bammann K, Eiben G, Hebestreit A, Iacoviello L, Gallois K, Konstabel K, Kovacs E, Lissner L, Maes L, Molnar D, Moreno LA, Reisch L, Siani A, Tornaritis M, Williams G, Ahrens W, Bourdeaudhuij I de, Pigeot I, IDEFICS consortium (2011). The IDEFICS community-oriented intervention programme: a new model for childhood obesity prevention in Europe? *Int J Obes* **35**(Suppl 1): S16–23.

Hislop JF, Bulley C, Mercer TH, Reilly JJ (2012). Comparison of epoch and uniaxial versus triaxial accelerometers in the measurement of physical activity in preschool children: a validation study. *Pediatr Exerc Sci* **24**(3): 450–460.

Janz KF, Letuchy EM, Eichenberger Gilmore JM, Burns TL, Torner JC, Willing MC, Levy SM (2010). Early physical activity provides sustained bone health benefits later in childhood. *Med Sci Sports Exerc* **42**(6): 1072–1078.

Jekauc D, Wagner MO, Kahlert D, Woll A (2013). Reliabilität und Validität des MoMo-Aktivitätsfragebogens für Jugendliche (MoMo-AFB). *Diagnostica* **59**(2): 100–111.

Jimenez-Pavon D, Kelly J, Reilly JJ (2010). Associations between objectively measured habitual physical activity and adiposity in children and adolescents: Systematic review. *Int J Pediatr Obes* **5**(1): 3–18.

Jimenez-Pavon D, Fernandez-Vazquez A, Alexy U, Pedrero R, Cuenca-Garcia M, Polito A, Vanhelst J, Manios Y, Kafatos A, Molnar D, Sjostrom M, Moreno LA, Helena Study Group (2013). Association of objectively measured physical activity with body components in European adolescents. *BMC Public Health* **13**: 667.

Joe H, Zhu R (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biom J* **47**(2): 219–229.

Jones MC (1994). Expectiles and M-quantiles are quantiles. *Stat Probabil Lett* **20**(2): 149–153.

Kimm SYS, Glynn NW, Obarzanek E, Kriska AM, Daniels SR, Barton BA, Liu K (2005). Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. *Lancet* **366**(9482): 301–307.

Kneib T (2013). Beyond mean regression. *Stat Model* **13**(4): 275–303.

Koenker R (2005). *Quantile Regression.* New York: Cambridge University Press.

Koenker R, Bassett Jr G (1978). Regression quantiles. *Econometrica* **46**(1): 33–50.

Konstabel K, Veidebaum T, Verbestel V, Moreno LA, Bammann K, Tornaritis M, Eiben G, Molnar D, Siani A, Sprengeler O, Wirsik N, Ahrens W, Pitsiladis Y, IDEFICS consortium (2014). Objectively measured physical activity in European children: the IDEFICS study. *Int J Obesity* **38**(Suppl 2): S135–143.

Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* **380**(9838): 219–229.

McTiernan A (2008). Mechanisms linking physical activity with cancer. *Nat Rev Cancer* **8**(3): 205–211.

Metcalf BS, Hosking J, Jeffery AN, Voss LD, Henley W, Wilkin TJ (2010). Fatness leads to inactivity, but inactivity does not lead to fatness: a longitudinal study in children (EarlyBird 45). *Arch Dis Child* **96**(10): 942–947.

Moraes AC de, Carvalho HB, Siani A, Barba G, Veidebaum T, Tornaritis M, Molnar D, Ahrens W, Wirsik N, Henauw S de, Marild S, Lissner L, Konstabel K, Pitsiladis Y, Moreno LA, IDEFICS consortium (2015). Incidence of high blood pressure in children - effects of physical activity and sedentary behaviors: the IDEFICS study: High blood pressure, lifestyle and children. *Int J Cardiol* **180**: 165–170.

Nagy P, Kovacs E, Moreno LA, Veidebaum T, Tornaritis M, Kourides Y, Siani A, Lauria F, Sioen I, Claessens M, Marild S, Lissner L, Bammann K, Intemann T, Buck C, Pigeot I, Ahrens W, Molnar D, IDEFICS consortium (2014). Percentile reference values for anthropometric body composition indices in European children from the IDEFICS study. *Int J Obes* **38**(Suppl 2): S15–25.

National High Blood Pressure Education Program Working Group on High Blood Pressure in Children and Adolescents (2004). The fourth report on the diagnosis, evaluation, and treatment of high blood pressure in children and adolescents. *Pediatrics* **114**(Suppl 2): 555–576.

Newey WK, Powell JL (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**(4): 819–847.

Ojiambo R, Konstabel K, Veidebaum T, Reilly J, Verbestel V, Huybrechts I, Sioen I, Casajus JA, Moreno LA, Vicente-Rodriguez G, Bammann K, Tubic BM, Marild S, Westerterp K, Pitsiladis YP, IDEFICS consortium (2012). Validity of hip-mounted uniaxial accelerometry with heart-rate monitoring vs. triaxial accelerometry in the assessment of free-living energy expenditure in young children: the IDEFICS Validation Study. *J Appl Physiol* **113**(10): 1530–1536.

Olive L, Byrne D, Cunningham R, Telford R (2012). Effects of physical activity, fitness and fatness on children's body image: the Australian LOOK longitudinal study. *Ment Health Phys Act* **5**(2): 116–124.

Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M (2006). Validation and calibration of an accelerometer in preschool children. *Obesity* **14**(11): 2000–2006.

Pigeot I, Baranowski T, Henauw S de, IDEFICS Intervention Study Group, IDEFICS consortium (2015). The IDEFICS intervention trial to prevent childhood obesity: design and study methods. *Obes Rev* **16**(Suppl 2): 4–15.

Plasqui G, Westerterp KR (2007). Physical activity assessment with accelerometers: an evaluation against doubly labeled water. *Obesity* **15**(10): 2371–2379.

Pratt M, Macera CA, Sallis JF, O'Donnell M, Frank LD (2004). Economic interventions to promote physical activity: application of the SLOTH model. *Am J Prev Med* **27**(Suppl 3): 136–145.

Puyau MR, Adolph AL, Vohra FA, Butte NF (2002). Validation and calibration of physical activity monitors in children. *Obes Res* **10**(3): 150–157.

R Core Team (2015). *R: A Language and environment for statistical computing.* R foundation for statistical computing. Vienna, Austria.

Rauner A, Mess F, Woll A (2013). The relationship between physical activity, physical fitness and overweight in adolescents: a systematic review of studies published in or after 2000. *BMC Pediatr* **13**: 19.

Rethorst CD, Wipfli BM, Landers DM (2009). The antidepressive effects of exercise: a meta-analysis of randomized trials. *Sports Med* **39**(6): 491–511.

Rigby RA, Stasinopoulos DM (2004). Smooth centile curves for skew and kurtotic data modelled using the Box-Cox power exponential distribution. *Stat Med* **23**(19): 3053–3076.

Rigby RA, Stasinopoulos DM (2005). Generalized additive models for location, scale and shape. *J R Stat Soc Ser C Appl Stat* **54**(3): 507–544.

Rippe RC, Meulman JJ, Eilers PHC (2012). Visualization of genomic changes by segmented smoothing using an L0 penalty. *Plos One* **7**(6): e38230.

Sasaki JE, John D, Freedson PS (2011). Validation and comparison of ActiGraph activity monitors. *J Sci Med Sport* **14**(5): 411–416.

Schmid D, Leitzmann MF (2014). Physical Activity Epidemiology. In: Ahrens W, Pigeot I (eds.): *Handbook of epidemiology.* 2nd edition. New York: Springer: pp. 1927–2002.

Schnabel SK, Eilers PHC (2009). Optimal expectile smoothing. *Comput Stat Data An* **53**(12): 4168–4177.

Schulze Waltrup L (2014). *Extensions of semiparametric expectile regression.* PhD thesis. Ludwig-Maximilians-Universität München.

Schulze Waltrup L, Sobotka F, Kneib T, Kauermann G (2015). Expectile and Quantile Regression – David and Goliath? *Stat Model* **15**(5): 433–456.

Sobotka F, Kneib T (2012). Geoadditive expectile regression. *Comput Stat Data An* **56**(4): 755–767.

Sobotka F, Schnabel S, Schulze Waltrup L, Eilers PHC, Kneib T, Kauermann G (2014). *Package expectreg.* R package version 0.39.

Stanek 3rd EJ, Well A, Ockene I (1999). Why not routinely use best linear unbiased predictors (BLUPs) as estimates of cholesterol, per cent fat from kcal and physical activity? *Stat Med* **18**(21): 2943–2959.

Stasinopoulos DM, Rigby RA (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *J Stat Softw* **23**(7): 1–46.

Treuth MS, Schmitz K, Catellier DJ, McMurray RG, Murray DM, Almeida MJ, Going S, Norman JE, Pate R (2004). Defining accelerometer thresholds for activity intensities in adolescent girls. *Med Sci Sports Exerc* **36**(7): 1259–1266.

Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M (2008). Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc* **40**(1): 181–188.

Trost SG, Loprinzi PD, Moore R, Pfeiffer KA (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc* **43**(7): 1360–1368.

Trost SG (2007). State of the art reviews: measurement of physical activity in children and adolescents. *Am J Lifestyle Med* **1**(4): 299–314.

Verbestel V, Henauw S de, Bammann K, Barba G, Hadjigeorgiou C, Eiben G, Konstabel K, Kovacs E, Pitsiladis Y, Reisch L, Santaliestra-Pasias AM, Maes L, Bourdeaudhuij I de, IDEFICS consortium (2015). Are context-specific measures of parental-reported physical activity and sedentary behaviour associated with accelerometer data in 2-9-year-old European children? *Public Health Nutr* **18**(5): 860–868.

Viterbi A (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans Inf Theory* **13**(2): 260–269.

Westerterp KR (2009). Assessment of physical activity: a critical appraisal. *Eur J Appl Physiol* **105**(6): 823–828.

Whittaker ET (1923). On a new method of graduation. *P Edinburgh Math Soc* **41**: 63–75.

Witowski V, Foraita R, Pitsiladis Y, Pigeot I, Wirsik N (2014). Using hidden Markov models to improve quantifying physical activity in accelerometer data - a simulation study. *Plos One* **9**(12): e114089.

Witowski V, Foraita R (2014). *HMMpa: Analysing accelerometer data using hidden Markov models.* R package version 1.0.

Yao Q, Tong H (1996). Asymmetric least squares regression estimation: A nonparametric approach. *J Nonparametr Stat* **6**(2-3): 273–292.

Zucchini W, MacDonald IL (2009). *Hidden Markov models for time series: an introduction using R.* Boca Raton: Chapman & Hall / CRC Press.

Curriculum Vitae

Norman Wirsik

Address

Leibniz Institute for Prevention Research and Epidemiology - BIPS
Department Biometry and Data Management
Unit Statistical Modeling of Primary Data
Achterstr. 30, 28359 Bremen, Germany
Phone: +49-421-218 56950
Email: wirsik@bips.uni-bremen.de

Education

2010–present	Ph.D. candidate in Statistics, University of Bremen, Germany
2003–2010	Diploma in mathematics (Dipl.-Math.), University of Bremen, Germany
2003	General qualification for university entrance (Abitur) at Gymnasium Ulricianum Aurich, Germany

---

Working Experience

2010–present	Research associate at BIPS
2008–2010	Student assistant at BIPS
2006–2008	Academic tutor at University of Bremen

Publications

Peer-reviewed publications

Diederich A, Winkelhage J, Wirsik N (2011). Age as a criterion for setting priorities in health care? A survey of the German public view. *Plos One* **6**(8): e23930.

Diederich A, Swait J, Wirsik N (2012). Citizen participation in patient prioritization policy decisions: an empirical and experimental study on patients' characteristics. *Plos One* **7**(5): e36824.

Konstabel K, Veidebaum T, Verbestel V, Moreno LA, Bammann K, Tornaritis M, Eiben G, Molnar D, Siani A, Sprengeler O, Wirsik N, Ahrens W, Pitsiladis Y, IDEFICS consortium (2014). Objectively measured physical activity in European children: the IDEFICS study. *Int J Obesity* **38**(Suppl 2): S135–143.

Moraes AC de, Carvalho HB, Siani A, Barba G, Veidebaum T, Tornaritis M, Molnar D, Ahrens W, Wirsik N, Henauw S de, Marild S, Lissner L, Konstabel K, Pitsiladis Y, Moreno LA, IDEFICS consortium (2015). Incidence of high blood pressure in children - effects of physical activity and sedentary behaviors: the IDEFICS study: High blood pressure, lifestyle and children. *Int J Cardiol* **180**: 165–170.

Witowski V, Foraita R, Pitsiladis Y, Pigeot I, Wirsik N (2014). Using hidden Markov models to improve quantifying physical activity in accelerometer data - a simulation study. *Plos One* **9**(12): e114089.