METHODOLOGY TO ASSESS STATIC AND DYNAMIC COMPLEXITY OF PHYSICAL ACTIVITY PATTERN

OLEKSII MANDRYCHENKO

A thesis submitted in partial fulfilment of the

requirements of Glasgow Caledonian University

for the degree of Doctor of Philosophy

June 2019

(a) METHODOLOGY TO ASSESS

STATIC AND DYNAMIC COMPLEXITY

OF PHYSICAL ACTIVITY PATTERN

(c) OLEKSII MANDRYCHENKO

(d)(e) A thesis submitted in partial fulfilment of

the requirements of Glasgow Caledonian University

for the degree of Doctor of Philosophy

(g) June 2019

Abstract

Physical activity is an essential determinant of health. Being active helps to reduce risks of numerous non-communicable diseases, all-cause mortality and improves mental health. Despite the positive impact of physical activity, its global levels are in decline, and no clear reasons are known as to why this is happening.

Research in the field concentrated on understanding the root causes of physical activity decline. Use of technology and objective measures allowed researchers to explore physical activity with high accuracy, duration, and resolution. Researchers developed various measures of physical activity behaviour, and this thesis added two new measures of complexity of physical activity pattern.

Our capacity to collect physical activity data appeared to outweigh our capacity to analyse this data and to derive meaningful information beyond simple measures of the volume of activity. Numerous, large-scale, and global multi-national research databases exist for independent research of physical activity data, for example, Biobank and NHANES. However, analysis of the data is challenging - some studies disagreed on the outcomes, some observed different relationships and some studies agreed on the core results. Why research observed different trends from the same data was unknown.

To quantify physical activity, considerable attention has been paid to the measures of the volume and only recently to the pattern of activity. The amount of accumulated physical activity was the current gold standard measure of physical activity. However, how activity was performed, in what order, combination, frequency and variation was also important because the pattern of accumulation could be seen as a direct reflection of people’s behaviour.

Arguably, static and dynamic qualities of human physical activity behaviour did not receive enough attention compared, for example, to physiological studies. To date, there were few studies that looked at complexity of physical activity. Those studies that did investigate complexity of physical activity have found relationships between complexity of physical activity and factors associated with improved health. Clearly, there was a need to investigate if methods largely applied in physiology could be used in studies of physical activity behaviour.

It was possible to analyse the dynamic behaviour of physical activity using simple and effective measures, e.g. by adopting the complexity of physical activity pattern. The static and dynamic complexity of physical activity pattern were the two new measures suggested by this study. This study demonstrated that the complexity of physical activity pattern was related to correlates of physical activity reflecting a person's health. The measures captured similar information to the measure of the volume of physical activity and revealed new relationships, for example, between BMI, education level, gender and alcohol consumption. This study also examined the relationship between these measures and the volume of physical activity.

This was the first study to reveal a degradation of complexity of physical activity behaviour with age. This research added to the pool of evidence suggesting there was a negative correlation between BMI, the volume of PA and complexity of PA pattern. There were some differences between the volume of PA and complexity of PA pattern depending on the ethnicity. For example, Mexican American, on average, demonstrated a more complex PA pattern and higher volume of PA, than non-Hispanic white.

Table of contents

List of figures

List of tables

List of accompanying materials

- Source code on USB
- Source code available at https://github.com/o-mdr/pa-pattern-complexity

List of abbreviations

CDC – Centres for Disease Control and Prevention

EE – energy expenditure

MVPA – Moderate to Vigorous Physical Activity

NHANES – National Health and Nutrition Examination Survey

PA – Physical Activity

List of publications

1. Mandrychenko, Chastin, Granat, Finding optimum outcome measures for physical activity monitoring. BASES Annual Conference. Challenging the Dogma, University of Glasgow, 2010.
2. Mandrychenko, Chastin, Granat. Physical Activity Pattern Complexity Reveals Associations between Pattern, Health and Socio-Demographic Determinants. ACSM's 59th Annual Meeting and 3rd World Congress on Exercise is Medicine, 2012.

Acknowledgement

I am forever grateful to my family – Ievgeniia, Sebastian and Solomia for their support and understanding.

I would like to thank Prof Sebastien Chastin and Prof Malcolm Granat for the continued assistance, enormous support and critical feedback on this study. I appreciate the feedback on the study provided by the research team at Glasgow Caledonian University and also their help with configuring and analysis of physical activity equipment, in particular, Dr Ben Stansfield, Dr Philippa Dall, Dr Janet Finlayson. I am thankful to PhD research students for encouragement and support, in particular, Kym Hennessy, Ellaine Coulter, Paul McCrorie, Lisa Newcombe, Ceri Sellers, Raid AlQaisi, and Stephanie Smith. Matt Little and Dr Jamie Graves provided me with valuable encouragement and support.

I am grateful to Kate McMillan and Dr Miles Gould who proofread this thesis and suggested many helpful corrections that improved the readability of the manuscript.

Finally, this work was funded under Glasgow Research Partnership in Engineering, Joint Research Institute of Mechanics of Materials (2009-2012), Structures and Bioengineering and Scottish Overseas Research Student Awards Scheme (2009-2012). This work would not exist without this financial support.

Author's declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Chapter 1. Background

Physical activity and health

The relationship between physical activity (PA) and health was well established. Low level of PA was associated with poor health, disablement in later life and harmful environmental impact (Haskell *et al.*, 2007; Sallis *et al.*, 2009). There was strong evidence to support the existence of an inverse relationship between the amount of PA, population morbidity and all-cause mortality (Paffenbarger *et al.*, 1993; Young *et al.*, 1993; World Health Organization, 2010; Ekelund *et al.*, 2015). Lack of PA was a modifiable risk factor for several non-communicable diseases including type 2 diabetes, cardiovascular diseases, thromboembolic stroke, colon cancer, breast cancer, depression, anxiety disorders and others (Kesaniemi *et al.*, 2001; Lee *et al.*, 2012). An international report by the World Health Organisation considered inactivity to be the fourth leading cause of mortality around the world (WHO, 2010). WHO issues global progress monitor updates every several years (WHO, 2017).

Reports on the levels of PA worldwide are worrying (Butcher *et al.*, 2008; Kohl *et al.*, 2012). Globally, 31.1% (95% CI 30.9 - 31.2) of adults were physically inactive (Hallal *et al.*, 2012). Evidence suggested there was a sustained trend toward declining levels of PA (Church *et al.*, 2011; Dumith *et al.*, 2011) in industrialised countries. This trend, combined with an ageing population, created a worrying situation with the consistently increasing burden of chronic and non-communicable disease (Knuth and Hallal, 2009). The same situation was rapidly emerging in developing countries where levels of PA were declining while the impact of chronic and non-communicable disease was increasing (Boutayeb, 2006).

The situation was such that some researchers argued that physical inactivity became "pandemic". This kind of terminology was typically applied to communicable diseases that ravage humankind, such as the plague, AIDS, flu, and Ebola. It was, therefore, a testimony to the gravity of the situation and an awakening call to the importance of promoting PA as a modifiable risk factor for health. The public health costs and socioeconomic burden of inactivity were so high that there was a fast emerging drive towards precise infrastructural needs for governmental and educational policies, planning, and training targeting global increase of PA (Kohl *et al.*, 2012; Lee *et al.*, 2012).

To encourage the population to be more active the WHO and governments of most industrialised and some developing countries created PA guidelines (Bull and the Expert Working Groups, 2010). These guidelines had simple messages, for example, that some PA was better than none (Twisk, 2001; Haskell *et al.*, 2007; Tremblay *et al.*, 2011). More specific recommendations were available for different age groups and selected population strata. For example, recommendations for children and adolescents (5 – 18 y.o.) included minimum 60 minutes to several hours of MVPA every day, and vigorous activities should be performed at least 3 days a week, and the time spent sedentary should be minimal (Department of Health, Physical Activity, 2011). For adults (18 – 65 y.o.) an ordinary target level was 30 minutes of MVPA 5 days a week. Some guidelines further clarify that population should accumulate 30 minutes of MVPA in bouts of more than 10 minutes. For the elderly population (65+ y.o.), UK recommendations advised breaking extended sedentary (sitting) periods and accumulated 150 minutes per week of moderate intensity PA in bouts of 10 minutes or more (Sparling *et al.*, 2015).

What made population active or inactive was not well understood. Methodological and technological advances enabled us to obtain scientifically accurate measures and estimates of the amount of PA. Research groups measured the daily average of PA volume in all population inclusive and specific target groups (Schutz, Weinsier and Hunter, 2001; Denkinger *et al.*, 2012). With the realisation that the levels of PA were low and were decreasing, the focus of PA research was shifted towards understanding the actual reasons why populations became less active than was required to stay healthy (Bauman *et al.*, 2002; Owen *et al.*, 2010). Therefore, the goal of understanding the cause and determinants of inactivity was to formulate effective interventions and policies to encourage healthier lifestyles.

There were several theories and speculations as to why some of us were more active than others. Pratt et al. suggested that technological advances changed the way we work, travel and spend our leisure time (Pratt *et al.*, 2012). Such passive behaviour contributed to decreasing levels of PA. Watching TV, playing video games and spending time on a mobile phone contributed to lowering the level of PA in the United States in particular (Fakhouri *et al.*, 2013; Kenney and Gortmaker, 2017; Zhu *et al.*, 2019). Evidence in the UK suggested that pervasive technology allows for a “get-it-easy” cultural attitude. Due to technological advances, there were fewer manual labour jobs on the market. Developed traffic networks and cars resulted in less active commutes even though travel distances and travel time have increased. Housework activities became more straightforward and less demanding (Department of Health and Prevention, 2004).

It appeared there was no single factor that causes a decrease of PA level alone. What seemed to be happening was that a complex interplay of numerous factors acting at environmental, social and individual levels that influence PA behaviour (Sallis JF, Owen N, 2008).

Measures of physical activity

To understand the determinants of physical activity and inactivity, it was essential to be able to measure the activity itself. Academic research of PA aimed to build fundamental models and frameworks to improve mathematical reasoning on cause and effects (Bauman *et al.*, 2002; Barabási, 2005; Sallis JF, Owen N, 2008). Validation research studies looked at numerous ways to measure, standardise and compare PA (Grant *et al.*, 2006; Schneller *et al.*, 2015). Many longitudinal and prospective research studies helped to discover seasonal variations, long term PA trends and population habits (Yasunaga *et al.*, 2008).

Frameworks and methods to measure physical activity were developed and applied to large scale population studies (Tudor-Locke and Myers, 2001; Hawkins *et al.*, 2009; Yang and Hsu, 2010). These studies included data from global data sources that contain and maintain records for 100 000s of participants or even more, such as NHANES and UK Biobank (Doherty *et al.*, 2017). Such significant data sources allowed the analysis of national and international trends (Riddoch *et al.*, 2009; Church *et al.*, 2011; Hallal *et al.*, 2012).

Perhaps one common denominator that linked all the research studies in the field was that measuring PA was not simple or straightforward. In fact, PA was rather difficult to quantify (Schutz, Weinsier and Hunter, 2001; Pearson *et al.*, 2004; Chastin and Granat, 2010).

There were two fundamental ways to measure PA: subjective and objective. The subjective measurement relied on a priori imprecise data collection method of self-reports, questionnaires and recall diaries. This method was cheap, simple and allowed for large scale data collection, but it was also more error-prone. Objective measurements of PA were primarily based on technological observations performed by measuring tools – such as body-worn devices, motion cameras, global positioning tracking systems and others. Despite being less precise, subjective measurements could capture context in which PA was performed, feelings and individual's emotions – this was something that none of the objective measures could infer from the data gathered.

Modern understanding of PA mainly came from the activity monitoring technology (Lee and Shiroma, 2014). This technology helped to facilitate the process of collection, aggregation, storage and analysis of the data. Quantitative studies utilising activity monitoring could have considerable benefits compared to the studies based on questionnaires, recalls, or other subjective measurement techniques (Celis-Morales *et al.*, 2012).

Most measures had a standard unit of measurements, for example, those described by the International System of Units SI. However, there was no standard measure for physical activity. The most commonly used measures of PA based on activity monitoring were step count, accelerometer count, mean times spent in some levels of physical activity or energy expenditure. It was important to note, all these measures used the volume of PA as a proxy measure of PA, but did not quantify the pattern.

Academic research did not have proofs yet to confidently state whether PA was virtually just volume or whether it was more complicated than that. From the recent evidence, PA appeared to be more complicated. For example, PA could be observed in several different dimensions, such as Frequency Intensity Time Type – also known as FITT. Actual physical activity recommendations did not only express the volume of PA, but also specified pattern: how often activity should be performed, how, when, and what kind. Arguably, the complexity of PA pattern was a more direct reflection of behaviour than the volume of PA.

Some measures of PA described a pattern of activity, e.g. GINI index (Chastin and Granat, 2010), a number of changes (Lee and Skerrett, 2001; Zhou *et al.*, 2008), “barcodes” (Paraschiv-Ionescu *et al.*, 2012). This research aimed to expand the understanding of physical activity pattern and its complexity.

Gaps in the research field

Research of physical activity became a new and emerging branch of the epidemiology of public health aimed to understand the causes of physical activity, inactivity and their impact on health.

A range of non-standardised measures, methods, protocols and analysis techniques made studies difficult to compare or even rendered them incomparable. Such diversity prompted the international research community to call for standardisation of PA quantification methods (Wijndaele *et al.*, 2015).

Epidemiological evidence on the determinants of physical activity and inactivity was weak. Most epidemiological models only explained a minimal amount in the variance observed in PA level – 10% to 30% (Bauman *et al.*, 2012). It was interesting to investigate where the models have failed and why research explained so little variance. Bauman et al. asked an important question of why some people were active whereas the others not (Bauman *et al.*, 2012). The answer to this question is still unknown.

The current models only explained less than a third in the variance of activity, suggesting that statistical models didn't account for a significant part of physical activity behaviour. An extract of physical activity variance explained by the studies is presented below (Bauman *et al.*, 2012).

Source	Number of Correlates	Variance (R^2), %
(Rhodes and Smith, 2006)	1	0.2
(Lochbaum *et al.*, 2010)	1	0.6
	1	1
	1	1
	1	1
(Kaewthummanukul and Brown, 2006)	2	30
	3	7
	7	38
	7	25
(Rhodes and Smith, 2006)	13	30
(Trost *et al.*, 2002)	41	21

Table 1. Studies of the correlates of PA

The maximum variance explained $R^2 \approx 0.38$ (Kaewthummanukul and Brown, 2006) with 7 correlates and determinants of PA. The lowest variance explained was $R^2 \approx 0.002$ (Rhodes and Smith, 2006) with 1 correlate.

It was not clear if more correlates of PA could provide a better statistical fit or if some correlates exhibited better statistical correlation than others. It was possible that the currently known measures of PA did not capture enough information to describe the physical activity. Given the variety of measures, it was hard to rationalise a choice of one vs the other. The protocols, tools and methods were not standardised either, which made it hard to compare results of studies (Wijndaele *et al.*, 2015). Some frequently used objective measures of PA, such as step count, or bouts of inactivity, were artificially synthesised from the raw accelerometry data. Software that translated the raw signal into a measure was often proprietary and versioned, making it more error-prone to compare. As an outcome, several different tools could have produced similar measures (e.g. step count), but not the same measurements (values from the different tools could vary within the same experiment). The discrepancies between measurement results was a target of ongoing research (Schneller *et al.*, 2015). It was often challenging to estimate compliance with wear, reliability and device calibration. These technical problems could have introduced a degree of noise that could be difficult to spot and remove from the analysis.

If PA were a fully deterministic process, then more measures of better quality would yield a better statistical fit and hence higher variance explained by the models. In practice, however, it was not what has been observed. More measures and better quality measures did not bring statistical models of better fit (Bauman *et al.*, 2012). It was then reasonable to assume that PA was not a fully deterministic process. There was some degree of free will and arbitrary individual choice that impacted physical activity and inactivity. Some researchers suggested PA could be a derivative of a stochastic random process. Barabasi, for example, suggested a model with PA being a random walk process (Barabási, 2005). Cavanaugh thought of physical activity as a process guided by fractal dynamics (Cavanaugh, Kochi and Stergiou, 2010). However, it was ultimately unknown whether physical activity was indeed random or not, and if it was not, could it contain random input, such as free will?

Other gaps in PA research related to geolocation and funding opportunities. Research was comparatively more advanced in countries with developed economies, such as the UK, USA, Canada, Australia and countries of the EU.

Some correlates did not receive much research attention, such as genetics, some societal, macroeconomic, cultural and geographic factors.

A common observation from research studies was that the high variance of PA was not modelled accurately. An intra-individual variance of PA was 30-45% of the total PA variance, and average inter-individual variance was 55-60% of the total PA variance (Matthews *et al.*, 2002). Even seasonal changes in micro-climate could affect variance in PA. For example, a year-long Nakanojo study showed that habitual PA peaked in spring and autumn and was lowest during winter (Yasunaga *et al.*, 2008). It was not surprising that linear epidemiological models based on volume measures struggled to explain a large percentage of variance in PA.

The epidemiological approach assumed that human PA behaviour was linear, which contrasted with some recent findings (Dumuid *et al.*, 2018; Chinapaw *et al.*, 2019). Non-linear models emerging from statistical physics coudl account for more significant variance in activity by modelling the dynamics of human behaviour. The work of Barabasi showed the decision-making process could model the pattern of human activity. However, these models were a somewhat theoretical invention that could not inform us about determinants. Nonetheless, the models demonstrated that the pattern of PA was less variable than the total observed amount of PA. Given that the behaviour was what researchers and policymakers were trying to influence, it might have been of more benefit to model the pattern of PA rather than the total amount of PA.

Relationship between complexity of PA pattern and sedentary behaviour has not been studied yet, however, it appeared plausible that a more complex pattern could correspond to a healthier lifestyle. Sedentary behaviour is one of the most common behaviours associated with numerous non-communicable diseases and poor health factors even when adjusting for physical activity and exercise (Gardner *et al.*, 2016). Researchers have widely reported the importance of interrupting sedentary behaviour and developed mitigation strategies. For example, in 2015 Chastin et.al. performed a meta-analysis of the relationship between breaks in sedentary behaviour and cardiometabolic health, concluding that the existing evidence suggested that interrupting SB with light-intensity physical activity could help controlling adiposity and postprandial glycemia (Chastin *et al.*, 2015) . Cooper et.al. investigated the relationship between breaks of SB and metabolic variables in people with type 2 diabetes, concluding that higher SB time was linked to a poorer metabolic profile in people newly diagnosed with type 2 diabetes (Cooper *et al.*, 2012). As physical activity contains sedentary behaviour, any interruptions could split the prolonged period of inactivity and could potentially result in the increase of the complexity of PA pattern. The reverse process might also work - an increase in the complexity of behaviour could be related to the reduced sedentary time. However, little research evidence was available to confirm this. This thesis attempted to investigate potential health benefits by having more complex physical activity patterns.

Research outline

This research suggested a novel method to quantify PA pattern complexity and determines whether there was a relationship between PA pattern complexity, PA volume and health determinants. Among other types of research, quantitative correlational research with model testing has been selected for this dissertation.

This study presented a generic method to calculate measures of the static and dynamic complexity of PA pattern. The method was data-agnostic – it could be applied to a variety of input data sources. However, in the scope of this research, only accelerometry data from physical activity monitoring databases were used. The complexity of PA pattern was correlated with health determinants and the results were compared to the correlation of traditionally used measure of the volume of PA.

Quantitative correlational research could uncover a relationship between measures. For example, when one measure increases in value and the other measure increases in value – it is a case of positive correlation. Similarly, if one measure increases in value and the other decreases – it is a negative correlation. However, such correlational research could not state if an increase in value could *cause* a change in another measure. This type of research could only highlight the mere presence of a relationship, but not the reason why such relationship exists.

This research used public objectively-collected physical activity monitoring data from North America collected in 2003 – 2004 and 2005 – 2006 years. The NHANES data was a representative sample of the American population and it was publicly available for research purposes (https://www.cdc.gov/nchs/nhanes, accessed on 20-June-2019). The data set was cleaned and filtered according to the standard and published protocols. Due to this filtering process the weighting coefficients were lost and have not been re-calculated or re-applied to keep this study exploratory rather than to account for the details in a specific population. Several measures of PA and PA pattern were generated, this included one commonly used measure – step count – and two new measures of complexity – static and dynamic complexity of PA pattern. Statistical models were created, separately for each variable: volume and complexity of PA pattern and the results of the statistical modelling were then compared.

Aims and objectives

Complexity of physical activity pattern have not been examined in depth yet. Previous research of physiological complexity defined two broad areas: static complexity and dynamic complexity. Static complexity was attributed to the composition and structure of the behaviour or signal. The dynamic complexity examined the changes in the behaviour or signal over time. Using the largest dataset available at the time of analysis, this research aimed to investigate a set of possible complexity measures and select candidate measures to study static and dynamic complexity of PA pattern as measured via objective physical activity monitoring.

To investigate static complexity, it was hypothesised that a number of fixed-length blocks of physical activity behaviour could represent a static complexity of PA pattern, forming a dictionary of activities. A person with diverse range of activities should have a larger physical activity vocabulary and hence more complex static complexity.

The hypothesis for dynamic complexity was use compression ratio to investigate amount of change in PA pattern over time. The less complex behaviour should be easier to predict and hence compress.

The specific objectives of the research were these:

1. Investigate if a link could be established between the complexity of PA pattern and known determinants and correlates of PA reflecting a person’s health
2. Investigate if new correlates could be established with the complexity of PA pattern, but not with the volume of PA
3. Understand the relationship between the volume of PA and complexity of PA pattern

Chapter 2. Literature review

This research combined developments of new ideas in physical activity research with complexity theory to aid a synthesis of new ideas. The following review was based on themes and introduced key concepts and ideas relevant to both research fields.

Evolution of PA assessment: ecological models

Physical activity research field became active when it was shown, on a global scale, that lack of activity has dramatic contribution to all cause of mortality (Paffenbarger *et al.*, 1993; Young *et al.*, 1993; Kesaniemi *et al.*, 2001; Haskell *et al.*, 2007; Sallis *et al.*, 2009; World Health Organization, 2010). When the risks were understood research started to build up theoretical models of how PA interacted with or was influenced by internal and external factors. Sallis and Owen suggested an ecological model (Sallis JF, Owen N, 2008). Bauman et al. improve the model by adding a timeline to the ecological model and simplifying some of the domains of PA (Bauman *et al.*, 2012). A distinct model was put forward by Barabási, who described PA by a stochastic process – a set of activity states that changed each other with particular power law distributed probabilities (Barabási, 2005). These were not the only models which were used to describe PA, but these models influenced the development of this research.

This research utilised the findings of the three models. It assumed that it was possible to segregate the notion of PA domains from the Sallis and Owen model, added factors based on time proposed Bauman et al. and combined these with the idea of assigning probabilities to the states in a system from the Barabási model. It was thus possible to put PA domains and PA states on a timeline. PA states could have some probabilities that defined the likelihood of occurrence and change. Therefore both PA domain and time influenced a change from one activity to another one. Probability of occurrence of activity could be calculated and related to the domains of PA via correlates in statistical models. A mathematical model capturing these ideas was presented in Chapter 3. Methodology of this thesis.

Any model of physical activity, if applied practically, required measurement: what was there to measure, how to measure it, how long and how sensitive the measurements needed be so that the level of noise did not impact the result. In the PA and health research field PA was initially captured by observations or self-assessment questionnaires. Such subjective data collection was indeed better than nothing. However, it was not shown to be as reliable as objective measurements (Sequeira *et al.*, 1995; Shephard, 2003; Yasunaga *et al.*, 2008). Laboratory and instrument measurements were now more frequently used than before, thanks to the decreased price of the measuring devices and scientifically proven track record of their validity. Where laboratory measurements tended to be more precise, these were often more expensive or required the participating individual to be observed in a laboratory which makes large scale epidemiological studies problematic. Wearable instruments, however, were inexpensive, practical, reliable and valid for most research purposes. Wearable instruments were now precise enough and cheap enough to manufacture to allow large scale data collection efforts, such as NHANSE. The first attempts to use wearable sensors were made in the gait analysis studies in the 1950s. Since then, there was a surge in the number of studies using this technology in free-living PA research (Yang and Hsu, 2010). Laboratory examinations and free-living PA monitoring instruments dramatically reduced chances of human error, which were commonly associated with subjective measurements, like PA questionnaires or diaries (Janz, 2006). Objective measures provided more precise estimates of how much activity people actually do. This allowed for more accurate scientific assessment of the levels of PA (Prince *et al.*, 2008).

Large-scale epidemiological studies emerged and contributed to the national and international PA guidelines (Owen, 1996; Trost *et al.*, 2002; Tudor-Locke *et al.*, 2010; Bauman *et al.*, 2012). For example, Haskell et al. published recommendations for adults in the US regarding levels of PA where the authors suggested that participants maintained "moderate-intensity aerobic (endurance) physical activity for a minimum of 30 min on five days each week or vigorous-intensity aerobic physical activity for a minimum of 20 min on three days each week" (Haskell *et al.*, 2007). Similar guidelines became available in most developed countries, for example, in the UK (Bull and the Expert Working Groups, 2010) and Canada (Tremblay *et al.*, 2011). However, even those who currently did not meet the guideline standard, but engaged in some exercise, benefited from reduced risk of mortality (Wen *et al.*, 2011). A large pooled cohort analysis demonstrated that "being active (7.5+ MET-h/wk) and normal weight (BMI 18.5–24.9) was associated with a gain of 7.2 (95% CI: 6.5–7.9) y of life compared to being inactive (0 MET-h/wk) and obese (BMI 35.0+)" (Moore *et al.*, 2012).

Wearable technology allowed to get time-stamped measurement data. This data allowed researchers to conduct time series and PA pattern analysis. The temporal aspect of time-stamped data provided information beyond “how much” people did and enabled us to investigate the “how” and “when” people were active.

PA was known to be variable and complex, perhaps due to the numerous factors that influence it, for example, mental or physical tiredness, illness or impairment, chronic pain and so forth. These all affected PA to some degree. However, averaged non-timestamped measurements could only partly provide assessment and might not allowed for a distinction between normal and pathological behaviour (Paraschiv-Ionescu *et al.*, 2008). Analysis of PA pattern appeared to be vital as it was shown that the same amount of activity could have a different impact on health (Chastin *et al.*, 2010).

A large pool of diverse activities represented the PA pattern. It included planned behaviour, such as commuting, working, regular exercise and leisure time. This type of behaviour was easy to assess, analyse and predict mathematically due to the deterministic and known-in-advance nature. Additionally, PA included sporadic events or noise – unpredictable changes to the behaviour that introduce a degree of randomness to the behaviour. This addition presented a challenge to quantify and analyse. Its sporadic nature was attributed to the biological internal feedback system. Such feedback allowed humans to stay in equilibrium with the natural environment and, for example, control body weight (Rowland, 1998; Thorburn and Proietto, 2000).

Several trends could be observed in recent years in physical activity and public health: more and better-quality data was collected, and novel methods of analysis were used, primarily based on non-linear mathematics and dynamics. These developments often led to the creation of positive public health guidelines that encouraged the population to be more active and decreased the risk of non-communicable diseases.

Several large-scale studies were conducted looking at the effects of PA or lack of it. For example, studies utilising data from NHANSE (Tudor-Locke *et al.*, 2010), Swedish National March (Bellocco *et al.*, 2010), National Cancer Institute Cohort Consortium (Moore *et al.*, 2012), European Prospective Investigation into Cancer and Nutrition (Ekelund *et al.*, 2015). These studies examined tens or hundreds of thousands of participants and were vastly important for forming governmental policies (Haskell *et al.*, 2007; Bull and the Expert Working Groups, 2010; Tremblay *et al.*, 2011) and expanding public awareness (see, for instance, BBC article http://www.bbc.co.uk/news/health-30812439 last accessed 11-Apr-2017).

Large data sources sometimes also referred to as “big data”, presented new challenges. It was no longer possible to control every individual and ensure strict adherence to protocols and guidelines. Measures that required specific or expensive monitoring tools were prohibitively costly on this massive scale. Therefore more PA in these studies used simple PA measures, such as percentage of waking time or time spent in different activities (sedentary, low, light, moderate etc.) or average number of steps per day (Tudor-Locke *et al.*, 2010); daily average of self-reported time of activity translated into METs an hour per day (Bellocco *et al.*, 2010), self-reported PA levels expressed via metabolic equivalent METs hours per week (Moore *et al.*, 2012), self-reported level of PA expressed via Cambridge Index (Ekelund *et al.*, 2015).

These simple metrics, however, often measured the volume of PA and might not be the most robust representation of PA. Chastin at el. showed that the same volume of PA was observed for Parkinson’s disease patients and healthy controls while their PA pattern was clearly different (Chastin *et al.*, 2010).

There has been a debate as to whether active lifestyle contributed to decreasing BMI, obesity and all-cause of mortality. The causal effects no longer appeared to be clear. Views have recently started changing to highlight that the answer might not be as simple as one would have thought. Trost et al. summarised that the widely reported absence of the relationship between general levels of PA and overweight/obesity status has changed to a negative correlation: the higher levels of PA corresponded to lower levels of obesity (Trost *et al.*, 2002). Correlation by itself did not imply causality - it was not possible to say whether a higher level of PA implied a lower level of obesity or vice versa or the direction works in both ways. In 2008 Bauman et al. suggested that PA could act as obesity preventing factor (Bauman *et al.*, 2008). However, later in 2012, their views have changed. In 2012’s review Bauman et al. described a novel idea, where obesity was thought to be a driver of low levels of PA event that this direction was bi-directional (Bauman *et al.*, 2012). This issue highlighted the embedded complexity of PA and difficulty in understanding the determinants of PA.

One of the significant challenges that the PA research field faces was how to measure physical activity. Research has started to build up common frameworks, models and measures. Their efforts attempted to capture the multidimensional and complex nature of PA behaviour. An overview of the current models is presented next.

Sallis and Owen ecological model

In 2008 Sallis and Owen published an ecological model of PA. An ecological in this context meant that a population interacted with each other and with the environment they live in. The model was based on domains, such as interpersonal, perceived environment, behaviour and government policies. Sallis and Owen suggested that determinants of PA had roots in behavioural ecology developed by Stokols (Stokols, 1996). In the model, PA determinants were the interactions between the environment and population. A change in the environment could result in an adaptive reaction in behaviour. This model captured a comprehensive range of factors and allowed the building of multidimensional domain models acting on different layers. The Sallis and Owen model provided solid theoretical background for understanding the significant domains which had an impact on PA.

The downside of this model, however, was in its practical sophistication. The ecological model could become so complex that it was hard to establish the core drivers of the behavioural change (Sallis JF, Owen N, 2008; Gubbels *et al.*, 2014). This model is presented in Figure 1 below.

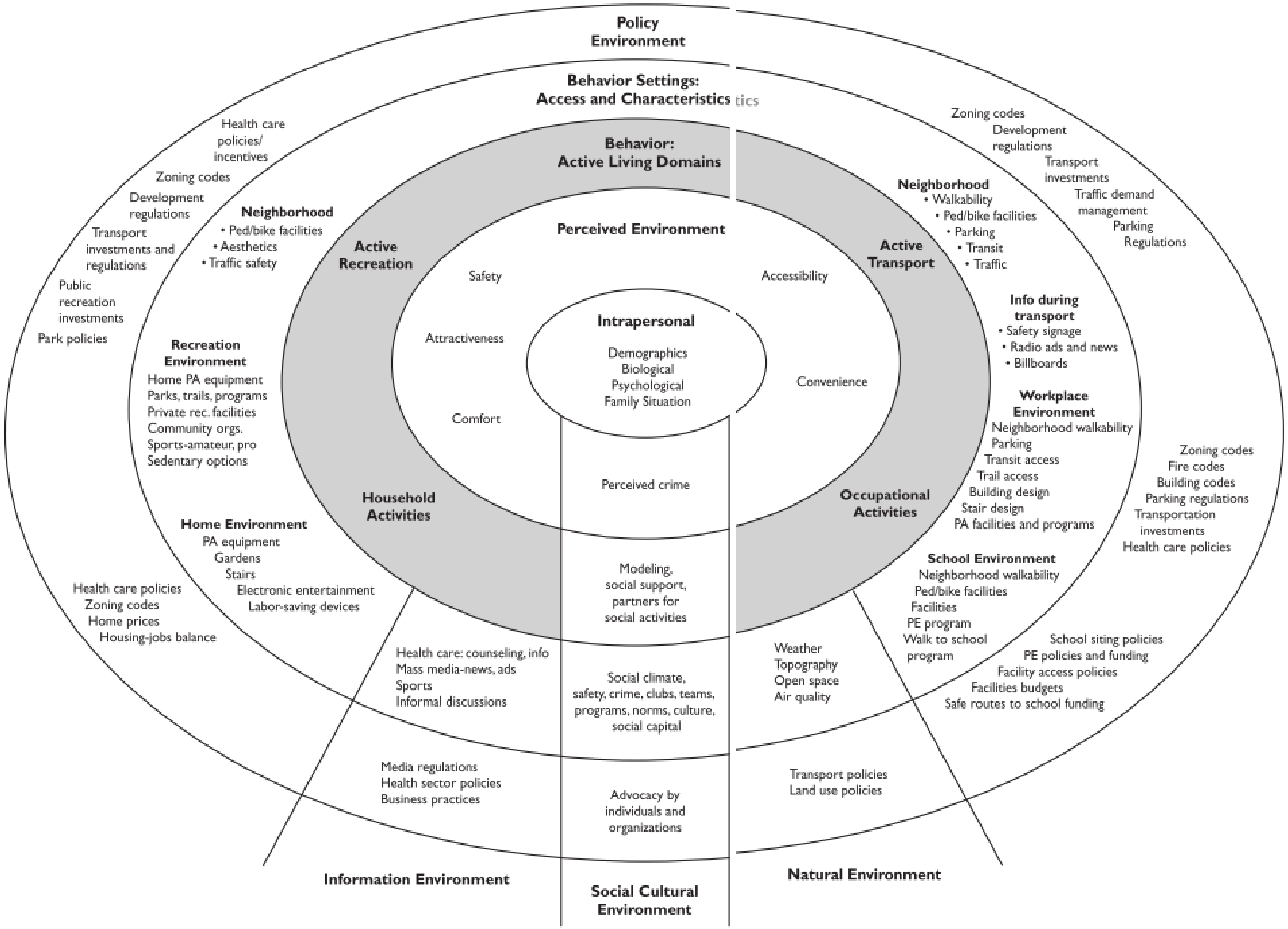

Figure 1. The ecological model of active living developed by Sallis and Owen, 2008

Bauman ecological model

Bauman et al. simplified Sallis and Owen's ecological model (Bauman *et al.*, 2012) and added a life course timeline to highlight the most appropriate age groups and aid in the design of targeted interventions. His group started from the published research on determinants and correlates of PA reflecting a person's health, made a summary of existing knowledge and suggested a simplified model. The ecological model has transformed to include time and domains. Bauman's ecological model is presented in Figure 2 below:

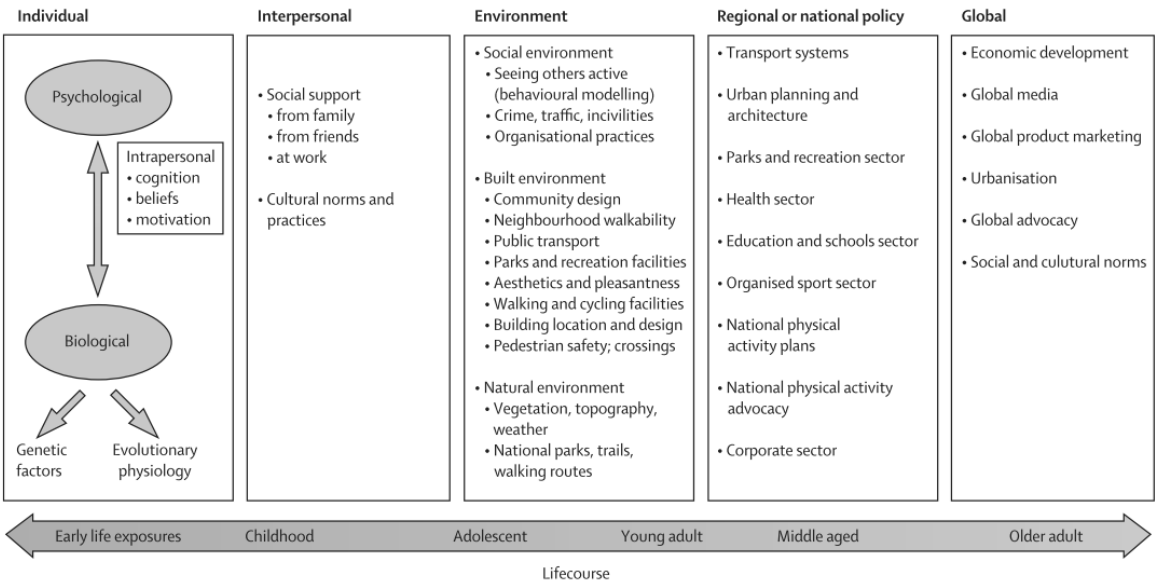

Figure 2. The ecological model of the determinants of PA, Bauman et al., 2012

The ecological model suggested by Sallis and Owen, and the evolved model of Bauman et al. had several common attributes, such as division into domains and the presence of a scale on which determinants acted on proximal determinants (personal, interpersonal), medial (social, environment, context) and distal (such as global economic drivers, urbanisation). The Bauman et al. model simplified domains and added a life course development. An age category related to the specific domain areas where positive changes were the most effective. The models also discussed substantial geographically distributed population, including developing countries and countries with low income.

Other models described similar domains and scales, for example, the model suggested by Glass and McAtee discussed a nested hierarchy of behavioural determinants acting on a time scale through the life course (Glass and McAtee, 2006).

Arguably, the limitations of ecological models were threefold. Firstly, the models were theoretical and did not directly quantify PA, PA pattern or the relationship between determinants of PA and health. A leading role in understanding the impact of the determinants was given to mediators. A mediator was usually a separate correlate acting upon which was necessary to complete cause-effect and observe the modification of behaviour. Secondly, the suggested theoretical models were difficult to quantify from experimental data, as research needed to agree on many practicalities, e.g. measurement scale, quantification metric, standard protocols, acceptable signal-noise ratio and so forth. The most critical limitation was perhaps that the models could only explain some variance of physical activity. For example, a literature review made by Bauman et al. (Bauman *et al.*, 2012) discussed evidence of correlates and determinants of PA with the strength of the evidence, as described by the percentage in the variance of PA explained by an outcome measure, being rather small <1% and sometimes 20-30%. An increase in the number of correlates and determinants did not necessarily meant an increase in the percentage of explained variance of PA (see Figure 3 below).

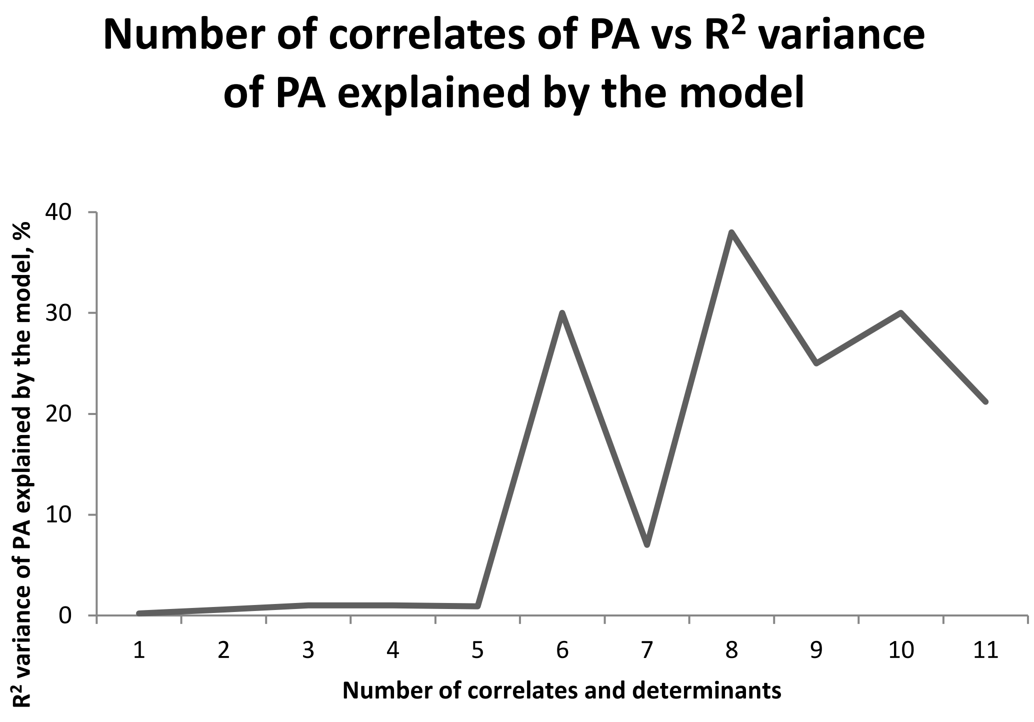

Figure 3. Number of correlates of PA reflecting a person's health vs variance of PA explained by models

Probability of activities: Barabási model

In socio-ecological models, physical activity was viewed as a linear, deterministic process with no random noise. Similarly, human activity has been studied in various domains based on the expected average behaviour and change, for example, insurance calculations, economic output, travel and routing algorithms. Interestingly, in these domains, human activity was conceptualised as a stochastic process – human activity could be described by a set of states and corresponding probabilities associated with these states. People were free to choose what they wish to do. Hence there was a level of uncertainty as to when one state was going to be changed by another (Barabási, 2005).

These models of human activity usually assumed that PA pattern was determined by some exponentially distributed set of events, also known as a random Poisson process. A random Poisson process was a process with an underlying exponential distribution of events or transitions between them. Some of the events occurred with particularly high frequencies, whereas other occurrences were exponentially rare.

Using such simplistic approach, it was possible to calculate probability distribution function for the four activities (sitting/lying, walking, standing, other) for the NHANES 2003-2004 and NHANES 2005-2006. The probability distribution graph for this population was presented on a Figure 4 below. For data overview, filtering procedures, exclusions and inclusion criteria please see Overview of Chapter 3 Methodology.

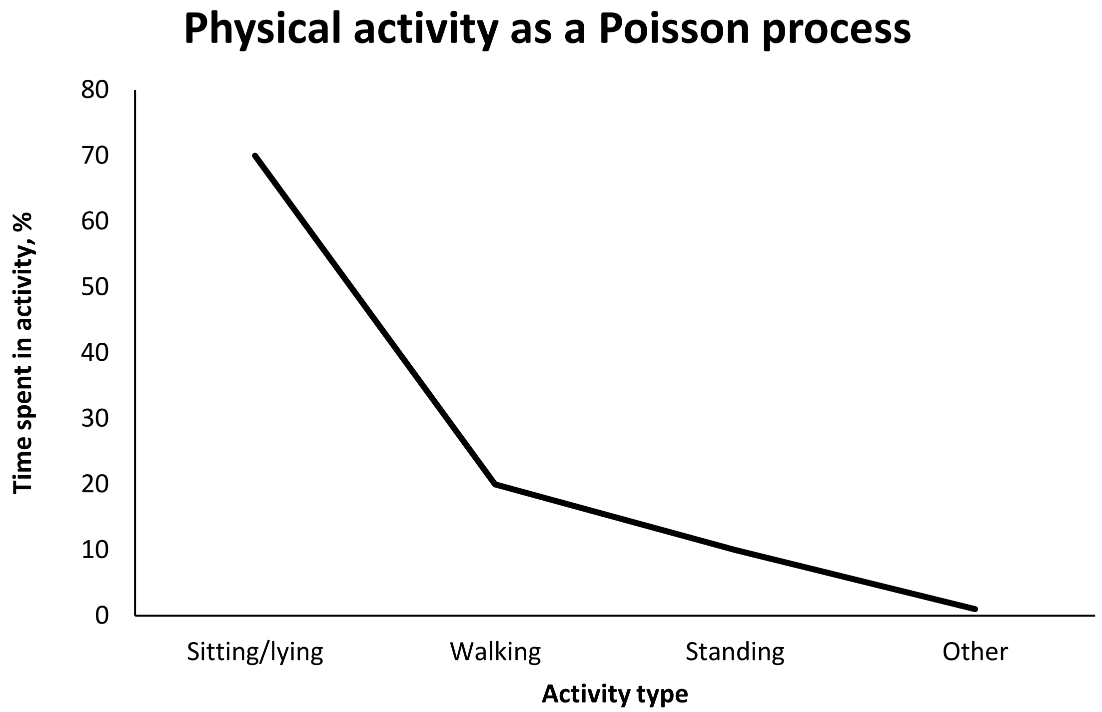

Figure 4. Simple Poisson process of physical activity

Figure 4 showed a probabilistic representation of PA pattern with no memory. Probability of occurrence of the next event did not depend on any previous events, but only on the process-specific global statistics. Many processes could be modelled using Poison distribution, for example, first-come-first-served or last-come-first-served scheduling in computer systems, network traffic, graph problems, or access to services in service-oriented environments.

If PA was indeed a Poisson process, it would have been difficult to imagine that each event in human life was utterly independent of the previous one. And indeed, empirical observations of PA suggested that PA follows dynamics described by some power law, however not precisely a simple Poisson law. Non-Poison dynamics of activity and inactivity was observed in a wide variety of human activity such as email correspondence (Eckmann, Moses and Sergi, 2004), internet activities (Crovella and Bestavros, 1997), disease outbreaks in cities (Eubank *et al.*, 2004), period of inactivity with Parkinson's disease patients (Chastin *et al.*, 2010), healthy people (Chastin and Granat, 2010) older adults (Chastin, Lord and Rochester, 2010). These studies all observed the power law distribution of events. This power law statistics might appear similar to the Poison statistics, but it was more heavy-tailed.

The presence of a tail in the power law statistics tail suggested that there existed a long-range temporal correlation due to memory in a system. This memory in a system described the historical relationship between previous events. Power law statistics allowed PA behaviour with more activity bursts located close to each other and more prolonged bouts of inactivity which fitted behaviour more realistically. There were evidence to suggest that many human activities followed non-Poisson statistics, where activity happened in short bursts, followed by prolonged periods of inactivity (Rybski *et al.*, 2009).

In 2005 Barabási proposed a novel view on modelling human activity (Barabási, 2005). He speculated that human activities were not randomly distributed in time as this had been assumed before. Instead, the dynamics of human activities followed some fundamental law. Barabási hypothesised the driver was based on a decision-making process. He suggested that such dynamics could be explained by a simple queuing process with task selection based on priority. A power law distribution described the waiting time of the other tasks in the queue. The longer tasks were awaiting execution the higher was the probability of their execution. This model could accurately emulate the following bursts of activity found in everyday living (see Figure 5).

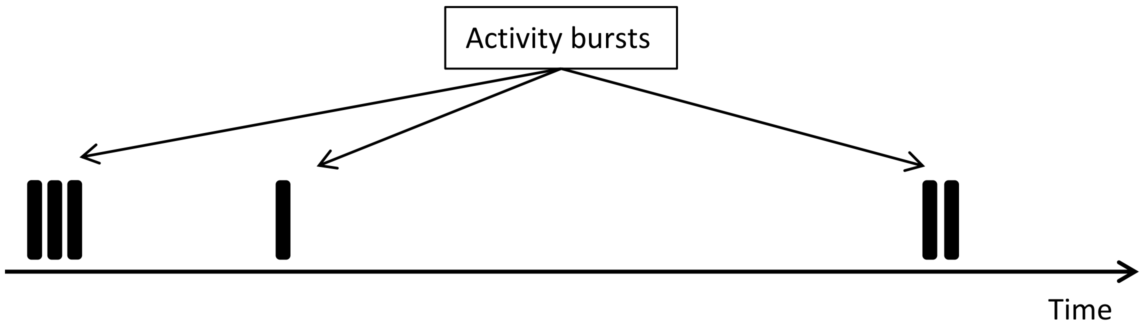

Figure 5. Barabási model explains bursts of physical activity

Barabási model suggested that humans selected and executed tasks with the highest perceived priority. Once finished, the selection process examines the tasks left in the queue and the next one with the highest perceived priority was picked up. A task with lower priority was executed after all the tasks with higher priority were finished. The limitation of this model was that it did not consider the cost or the burden of a task and did not describe the underlying distribution law itself. It was also not clear which task was selected by scheduling if tasks have the same priority.

Measurements of human physical activity

Troiano et al. analysed physical activity of the large, publicly available medical database NHANES to understand the levels of activity in the North American population stratified by age (Troiano *et al.*, 2008a). For the analysis, physical activity was quantified by the number of active minutes a day and compared to self-reported levels of activity. The publication suggested that self-reported levels of physical activity were higher than the levels recorded by accelerometers.

Public guidelines existed for the amount of physical activity required to be healthy (Haskell *et al.*, 2007). The recommended level for healthy adults of 18-65 y.o. was at least 30 minutes of moderate intensity endurance physical activity on 5 days a week or at least 20 minutes of vigorous intensity endurance physical activity on 3 days a week. It was not clear if physical activity could be accumulated in bouts of shorter period.

In a study of chronic pain patients, Paraschiv-Ionescu et al. showed that human physical activity could be bar-coded and complexity measures could be calculated on the PA pattern (Paraschiv-Ionescu *et al.*, 2012).

An example of the barcoding of physical activity could be seen in Figure 6 from the 2013 Michael J. Fox Foundation public competition using Kaggle platform (https://kaggle.com). Kaggle provided a platform for open data and machine learning competition. The 2013 Michael J. Fox Foundation competition was aimed to investigate data collected by mobile phones and provide a method to differentiate Parkinson's disease patients from the matching controls. The figure below presented bar-coded activity: active (black) and inactive (white) pattern collected from the anonymous mobile phone accelerometer study (Chastin and Mandrychenko, 2013).

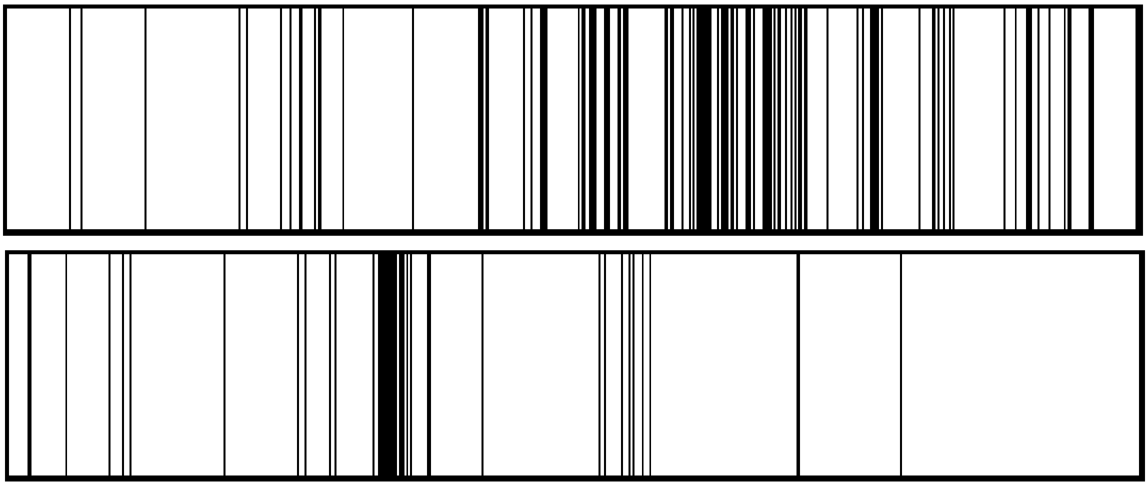

Figure 6. The pattern of active and inactive periods recorded by accelerometer

It was demonstrated that visualisation of PA pattern could help understanding of temporal aspects of PA pattern and that complexity of PA could be considered an excellent discriminative feature. Interestingly, the complexity of PA decreased with the increase of pain intensity.

Developments in complexity theory

Complexity was a common word appearing in scientific publications, sometimes used as a rigorously defined mathematical concept and sometimes as a referral to many details that have to be considered for a particular experiment. Oxford Dictionaries defines complexity as "the state or quality of being intricate or complicated" (http://oxforddictionaries.com/definition/complexity last accessed on 12-Apr-2017). Figure 7 shows growing usage for the word "complexity" retrieved from PUBMED database with an example search query: *(complexity) AND ("2016/01/01"[Date - Publication]: "2016/12/31"[Date - Publication].*

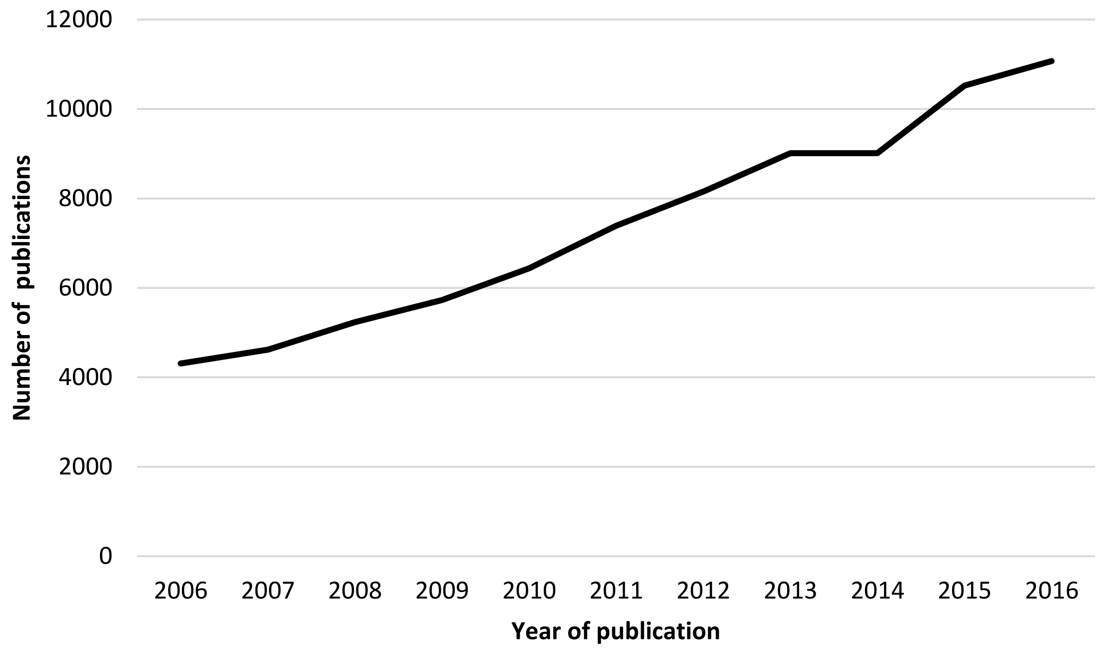

Figure 7. "Complexity" word usage from PUBMED 2006-2016

Clearly, there was a growing interest to understand the complexity and the following section attempted to review how exactly this term was used in research fields.

1. **The complexity of a system was a number of interconnections between the parts**. A simple system was different from complex because it has much fewer parts that were connected and interact with each other. (Sussman, 1999; Moses, 2010)
2. **Complexity was reflected in the fact that the relationship between some applied stimulus and the correspondent effect might not be predictable**. A view typically used in chaotic systems when even a tiny change may result in unpredictable changes (Senge, 1997)

3. **Complexity appeared from a set of interwoven subsystems. A single optimum for the complex system was seldom or practically never exist**. It was challenging to control a complex system, and its behaviour was subject to unexpected change (Rechtin and Maier, 2000)
4. **Complex processes were hard to observe or measure. Attempting to do so might result in high output noisiness, considerable variability or complete system disintegration**. As with observing light, for example, a mere fact of observation, changed the behaviour of a system. It was challenging to monitor a complex system without interfering with its properties or behaviour (Klir, 1988)

There was no clear agreement of what complexity was exactly. However, there was a common theme to all these definitions. The idea that complexity described the pattern either emerging from the structure or the dynamics of a system.

Definition of PA pattern

Paraschiv-Ionescu et al. in their publication of "Nonlinear analysis of human physical activity patterns in health and disease" (Paraschiv-Ionescu *et al.*, 2008) provided several definitions for physical activity pattern. Cavanaugh et al. suggested a definition for ambulatory activity pattern (Cavanaugh, Kochi and Stergiou, 2010). Chastin et al. defined the pattern of sedentary behaviour (Chastin *et al.*, 2010). Riddoch et al. investigated PA pattern in children (Riddoch *et al.*, 2007). For convenience, the definitions are presented below:

1. **Physical activity pattern is a sequence of posture allocation**. This definition used postures as levels of energy spent in the activity, such as lying, sitting, standing, walking, running and so forth, and discards information about the duration of the event. PA pattern was a queue of activities. The length of the pattern corresponded to the number of transitions between postures (Paraschiv-Ionescu *et al.*, 2008)
2. **Physical activity pattern is the duration of walking periods**. As nearly every action in free-living PA required walking, the duration spent walking could act as a temporal representation of PA pattern (Paraschiv-Ionescu *et al.*, 2008)
3. **Physical activity pattern is the time of activity-rest transitions as event series**. Activities were broadly divided into a binary group – a group that only had two states: active (running, walking, standing) or rest (sitting, lying). Transitions between these two states defined primary time series (Paraschiv-Ionescu *et al.*, 2008)
4. **Physical activity pattern is the symbolic sequence from the comparison of successive rest-activity-rest periods**. This approach required encoding a succession of postures with symbols and further combining symbols into words (Paraschiv-Ionescu *et al.*, 2008)
5. **Physical activity pattern is fluctuating step count time series.** Walking was captured by a user-worn pedometer, and the pattern was represented by the fluctuations in the 1-minute summary of step count (Cavanaugh, Kochi and Stergiou, 2010)
6. **Physical activity pattern is a distribution of sedentary bouts with regards to time.** The time spent sedentary on a time continuum presented a sedentary bout. Duration of this bout and the number of the bouts with the same duration could be used to define a pattern (Chastin *et al.*, 2010)

7. **Physical activity pattern is a series of the intensity of PA averaged by time, gender, BMI, or the age of participants**. This representation of PA could assess the differences of accelerometer counts per minute during a particular season of the year for male and female, aged 6-24 (Riddoch *et al.*, 2007).

The definitions of PA pattern discussed above highlight that there was no agreement as to what PA pattern was. A diverse representation of effectively the same term made it a challenge to discuss and compare PA pattern measurements. However, PA pattern measures had a small number of dimensions they operated compared to PA itself. Most commonly, PA pattern was described by one dimension, for example, GINI index statistics (Chastin *et al.*, 2010) or two dimensions, for example, step count per minute bouts (Paraschiv-Ionescu *et al.*, 2008; Cavanaugh, Kochi and Stergiou, 2010).

PA pattern could also appear to be a simplification of the diversity of PA behaviour that allowed to categorise population into groups of behaviour, such as active/non-active, for example, more-active males vs less-active females under the age of 60 (Matthews *et al.*, 2008), compliance with PA guidelines vs non-compliance, multiple sclerosis patients vs healthy controls (Pearson *et al.*, 2004).

It was interesting to note that all PA pattern measures quantified either the static or dynamic qualities of PA pattern. Static quantification referred to the measurement that did not change with time, for example, the total number of steps, mean sedentary time. Dynamic quantification in contrast varied with time, for instance, mean accelerometer count per interval of time (day or minute).

The complexity of physical activity pattern

The traditional way of quantifying PA pattern by using objective measures was to detect the amount of body movement using the accelerometer (Adamo *et al.*, 2009). Since internal and external drivers influenced PA pattern, quantitative measures of volume could only allow for partial assessment and thus did not offer clear indications between normal and abnormal PA pattern (Paraschiv-Ionescu *et al.*, 2008). The introduction of the static and dynamic complexity of PA allowed for an investigation of such dynamics.

Given the current state of research in complexity and physical activity pattern, it was interesting to understand if it was possible to apply complexity theory to physical activity pattern. It was not known if PA pattern was simple or complex, and, if it was complex, it was essential to understand how to measure complexity level. PA has some of the properties generally attributed to complex systems, for example: non-decomposition, unexpected behaviour, and unpredictability.

Complexity theory traditionally examined processes from the outside by providing input and observing the output. Complexity provided an assessment of the amount of change. It allowed for understanding of how much predictability, uncertainty and variability there was in a system.

Complexity was studied in physiology. For example, the complexity of physiological signals changed with age and disease (Goldberger, Peng and Lipsitz, 2002). Some signal simplification and synchronisation were observed in participants with respiratory sinus arrhythmia heart condition (Billman, 2011). Increased gait variability has been known to be a fall predictor in elderly adults (Hausdorff, 2005).

In the recent review by Bravi et al., the variability and complexity analysis methodologies were clustered into two categories: transformations and features. As a second dimension, these categories were further stratified into five domains of variability: statistical, geometric, energetic, informational, and invariant (Bravi, Longtin and Seely, 2011). In total, more than 70 methodologies were discussed, but only a handful of those were applied to the clinical research of PA.

One of the features of a complex system was *non-decomposition*. There was no generic model of PA pattern that could fully replicate the original behaviour of a population. One could create a simple model that could reconstruct PA pattern of an individual. However, it has not been shown that such a model could be transferred to another individual or population while preserving its precision. PA pattern was hard to control and change on a scale of population. One of the many reasons as to why it was difficult to control sometimes attributed to free will. It was possible to forecast PA pattern up to a certain point, after which precision of prediction drastically decreases.

One of the standard measures of complexity was known as Kolmogorov complexity, developed by Kolmogorov, Chaitin and Solomonoff (Fortnow, 2001). They defined complexity as the size of the smallest program that printed the desired end state of a system. For example, consider two PA patterns. Both patterns were recorded using 0 – for being inactive (sitting or lying), and 1 – for being active (standing, walking or running). The first pattern only demonstrated sedentary behaviour and the second pattern was a random sequence of active and inactive. An example Kolmogorov complexity estimate could be calculated as the length of the program that prints the pattern, see Table 2

Input	**Program**	**Complexity**
00000000000000000000	print(“0”,20)	13
00101000011110001001	print(“00101000011110001001”,1)	31

Table 2. Example Kolmogorov complexity estimate

Kolmogorov complexity was a solid theoretical measure. However, it was regarded as generally incomputable for any given input (Shannon, 1948). There were other measures of complexity that have been used to study dynamics, for example, entropy, compression, information content and others, however, there was no universal complexity measure that could be applied to different systems, processes or data (Feldman and Crutchfield, 1998).

Remaining challenges

PA pattern was hard to quantify due to inherent complexity associated with physical activity behaviour. This was why, perhaps, some more straightforward measures of PA and PA pattern were used today. Subjective measures of PA were prevalent in national and international PA guidelines (Haskell *et al.*, 2007; WHO, 2010). Objective measures of the volume of PA were now widely used in large scale longitudinal studies. However, high variability and multi-dimensionality of PA pattern presented statistical challenges for linear models which struggled to fit more than a third of PA variance (Bauman *et al.*, 2012). Models of PA pattern appeared to have a better statistical fit when using non-linear methods and included a measure of dynamics of PA pattern. Some researches started to make a shift in analytical methodology (Paraschiv-Ionescu *et al.*, 2008; Cavanaugh, Kochi and Stergiou, 2010; Chastin, Lord and Rochester, 2010), but more research was needed, particularly in the analysis of PA pattern (Chastin *et al.*, 2010).

How to measure physical activity volume, for example, as step count, or time spent in activities was now well established. Duration of endurance activity was also measured and understood (Haskell *et al.*, 2007; Bull and the Expert Working Groups, 2010). Further guidelines were available describing how to process collected data and which statistical models to use to draw conclusions (Tudor-Locke, Camhi and Troiano, 2012). The situation with the PA pattern was not so clear. Another challenge emerging today was that there was no agreed definition or analytical guidelines for PA pattern. Comparison of studies of PA pattern was difficult, as it was not always possible to directly compare measures of PA pattern.

Chapter 3. Methodology

Overview

This chapter describes a methodology to quantify the complexity of PA pattern. The data used in this research was acquired from the NHANES database which was one of the largest collections of objectively measured PA at the time of the analysis. In this database, PA was measured continuously for up to 7 days using an Actigraph (Actigraph AM-7164, Actigraph, LLC, Fort Walton Beach, Florida previously known as CSA/MTI AM-7164 for both NHANES 2003-2004 and NHANES 2005-2006). The Center for Disease Control and Prevention (CDC, USA) has an example picture of the device used during data collection. The original image was available from https://www.cdc.gov/nchs/tutorials/PhysicalActivity/SurveyOrientation/ DataOverview/Info2c.htm (or this shorten link https://bit.ly/2T1RPD2, accessed 28-Feb-2019). The device manufacturer has stopped providing product information or images for the model AM-7164 as checked on 28-Feb-2019 using https://www.actigraphcorp.com. To maintain an archival record and under a fair use policy Figure 8 illustrates a relative scale of the device as retrieved from the CDC tutorial link above.

Figure 8. Actigraph AM-7164 from NHANES 2003-2004/2005-2006 (source CDC)

This research uses filtering to ensure that only valid data is entered in the analysis. The activity counts, a common measure of the volume of PA, is extracted from the data. Separately, measures of the static and dynamic complexity of PA pattern are calculated. The hypothesis of the study is tested using three statistical models to determine whether the complexity of PA pattern can provide additional information compared to the volume of PA.

An overview of the methodology is illustrated in Figure 9.

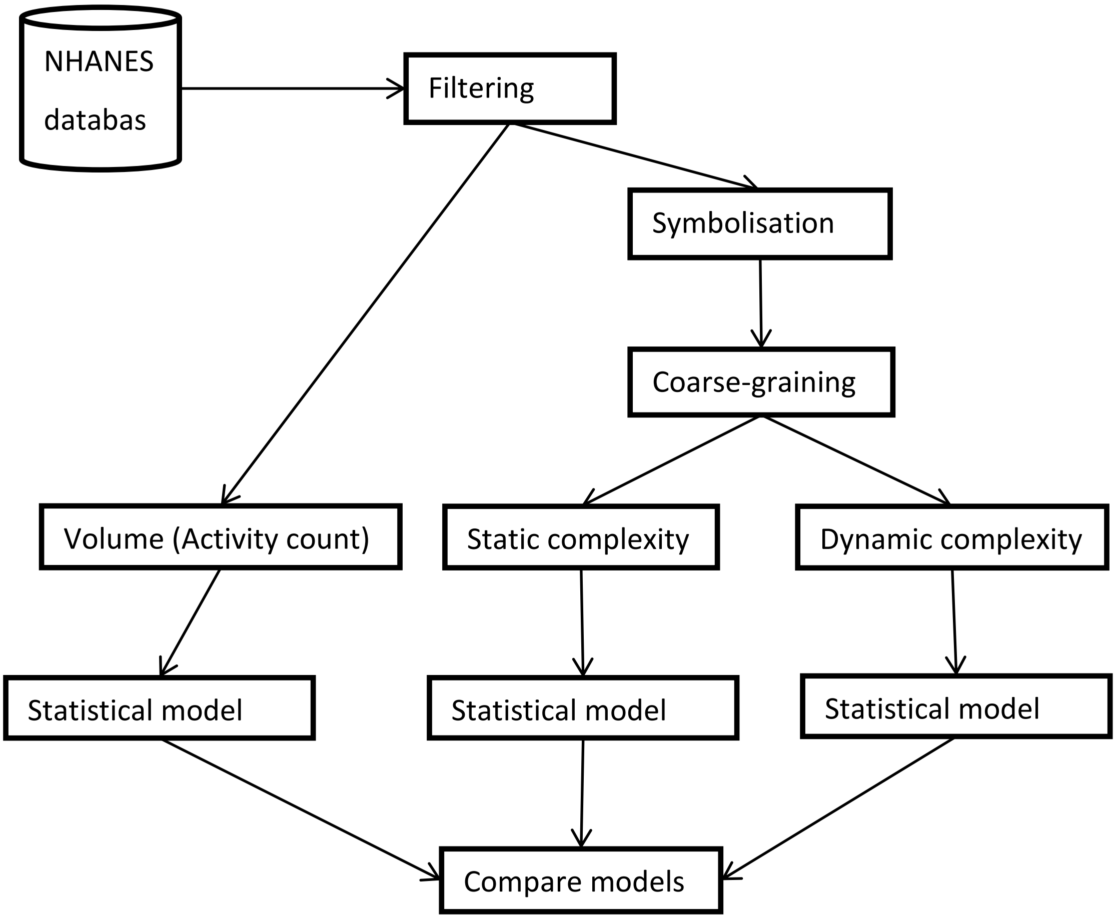

Figure 9. An overview of the methodology

Caspersen et. al. defined *physical activity* as “any bodily movement produced by skeletal muscles that results in energy expenditure” (Caspersen, Powell and Christenson, 1985). However, it is not clear how to extract temporal dynamics of PA for research as a methodology to study PA pattern is in its infancy.

This research assumes that a PA pattern is something that serves as a foundation for a personal decision-making process. Reed and Langford in a book “Data to Life” suggested the existence of “atoms” of PA, which were small, discrete, repetitive building blocks of human lives (Reed and Langford, 2013). Such building blocks can be extracted from the accelerometery data of daily PA. The same building blocks can be used to construct different PA patterns, and it is important to understand how one PA pattern can be different from another.

Measures of physical activity

This research uses data collected by a large nation-wide study called NHANES. NHANES is a recurring data collection effort in North America with a two-year cycle. Data from the two such cycles is analysed: NHANES 2003-2004 (also known as study “C”) and NHANES 2005-2006 (also known as study “D”). These were the first two studies to include accelerometry samples and no newer studies with such data were available at the time of research. Accelerometry, demographics and other data from NHANES are in the public domain and are available for download (http://www.cdc.gov/nchs/nhanes.htm, accessed 18-Nov-2018). The NHANES dataset includes the detailed reports of the data collection protocols, which were described in-depth by Matthews and Tudor-Locke (Matthews *et al.*, 2008; Tudor-Locke *et al.*, 2010). The NHANES dataset was the most comprehensive and standardised collection of records for health, environmental, social, and objectively measured PA data for the North American population.

Newer and larger databases of PA records started to appear, but unfortunately were not available during the research and hence were not considered. For example, see the UK Biobank database (https://www.ukbiobank.ac.uk, accessed 18-Nov-2018).

NHANES studies received ethical approval from The National Center for Health Statistics Research Ethics Review Board. Each participant was provided with information about the study, and an informed written consent was acquired. NHANES 2003-2004 was covered by the protocol number 98-12, and NHANES 2005-2006 was covered by the protocol number 2005-06 (http://www.cdc.gov/nchs/nhanes/irba98.htmc, accessed 18-Nov-2018).

In total, 14629 people participated in accelerometry data collection (7175 participants from NAHNES 2003-2004 and 7454 participants from NHANES 2005-2006). This data set was a representative sample of the US civilian and non-institutionalised population. Weighting coefficients were provided for each section of the research so that a representative sample can be maintained during the analysis of a section. Interviews were conducted at home, followed by a health examination performed at mobile examination centres. Male and female participants were included equally. Likewise, children, adults and older adults were equally sampled. Participants aged six or more with no walking impairments were asked to wear an accelerometer device for seven consecutive days.

All accelerometers were of the same, approved for research purposes model. This device measured acceleration against gravitational force produced by the bodily movements, sampled at a frequency of 10Hz. This instrument was shown to be valid and reliable (Tryon and Williams, 1996). The accelerometer has been widely used in the physical activity research field (Kristensen, P. L. and Møller, N. C. and Korsholm, L. and Wedderkopp, N. and Andersen, L. B. and Froberg, 2008; Troiano *et al.*, 2008b; Riddoch *et al.*, 2009).

Participants received standard instructions on accelerometer wear. The instructions included wearing an accelerometer for the full waking period for seven consecutive days and only removing it for sleep, swimming, bathing or other water activities. After data collection, accelerometers were returned by a pre-paid mail service and compensation was given (40 USD in case of NHANES 2005-2006). Once accelerometers were received, they were checked for calibration and reliability. Then the data recorded by the device was post-processed and saved into exportable file formats. The raw acceleration forces were summed up over a 1-minute epoch into an activity counts using proprietary and validated Actigraph firmware on the device itself.

Before using the dataset, the CDC recommended that the data be filtered. Tudor-Locke et al. published filtering protocols (Tudor-Locke, Camhi and Troiano, 2012). The same filtering protocols were used in this thesis with minor modifications to locate data with at least 10 hours of valid recording time. Changes to the filtering algorithm were described below, and SAS code used was published in this thesis as Appendix 2, Accelerometry data filtering.

The National Cancer Institute provided sample SAS (http://www.sas.com, accessed 18-Nov-2018) code that implemented these filters (available from http://riskfactor.cancer.gov/tools/nhanes_pam, accessed 18-Nov-2018). Since this study dealt with issues of dynamics, it was essential to preserve continuous recording periods that were uninterrupted by non-wear time, and additional filtering stages were added to the recommended CDC filters.

All filters are implemented in the C# language version 4 using Microsoft Visual Studio 2010. Description of the steps and algorithms are included in Appendix 2, Accelerometry data filtering. The filtering process is applied in stages to the whole datasets. The output was independently checked by a second and a third researcher.

The filtering stage ensures that data suitable for this thesis has at least five valid days with at least ten continuous hours of wear-time in each day.

A brief outline of the filtering steps is as follows:

1. Accelerometer CDC reliability flag indicates data is reliable (set either by the device itself or by the firmware during the data extraction stage)
2. Accelerometer CDC calibration flag indicates monitor is calibrated (same as above)
3. The recording containes at least some motion
4. Total recording time is over five days
5. There are five or more days with ten hours of continuous wear time

The following table provides a summary of the retained records:

N	Description	NHANES		
		2003-2004	**2005-2006**	**Total**
0	Provided from source	7175	7454	14629
1	Reliability flag	7134	7175	14309
2	Calibration flag	6804	6862	13666
3	Some motion was present	6800	6857	13657
4	Total recording time was over 5 days	6792	6851	13643
5	5 days with 10h of continuous wear	3287	4133	7420

Table 3. Number of retained records for NHANES data clean up

The required length of individual's physical activity monitoring data set is between 5 days and 7 days (inclusive on both sides). On the lower end of the scale, participants with fewer than 5 days of valid recording time are excluded from the analysis. On the upper scale, participants are instructed to wear the device for 7 consecutive days, so data after the 7th day was also excluded from the analysis.

NHANES study required participants to only remove device for sleep and bathing activities (device was not waterproof). Sleeping time is not accounted for and is excluded from the analyses. Regarding non-wear time, this study only includes days that had more than 10 hours of valid recording time. Non-wear time is excluded using CDC filtering algorithms. Therefore, individual recording time varied between participants. To eliminate the impact of the data length variation the following measures are put in place:

For the measure of volume of PA, a daily average is calculated based on the number of minutes included in the analysis. For example, if the analysed day containes 670 valid wear minutes, an average value is taken for each using 670 minutes. Next, an averaged value is taken across the number of the valid recording days across the recording week.

For the static complexity of PA pattern, there is no daily adjustments made, apart from the last word can be dropped from the analysis if it does not contain enough data for the full word, e.g. when the number of recorded minutes in a day is not divisible by 4. An average is taken across the number of valid recording days.

For the dynamic complexity, an individual value is calculated for each recorded day and average is calculated as a final value.

After the filtering stage, 7420 participants (or 58.87% of the original sample size) are preserved for further analysis and used to extract measures of total activity counts, static and dynamic complexities of PA pattern.

Total activity counts are calculated directly from the cleaned data. Measures of static and dynamic complexities of PA pattern have two further pre-processing steps: data symbolisation and coarse-graining. These steps will be discussed next.

Symbolisation of physical activity pattern

Signal symbolisation usually means grouping some parts of the signal of same or similar values into a new symbol. Other studies may use similar terms to refer to the same process: such as signal categorisation or analogue-to-digital conversion. This research uses symbolisation to convert activity counts into PA intensity symbols.

It was shown that symbolic analysis could provide valuable insight into the internal structure and dynamics of the underlying processes (Daw and Finney, 2002). Paraschiv-Ionescu et al. demonstrated that non-linear symbolic analysis enabled understanding of the relationship between PA pattern and health (Paraschiv-Ionescu *et al.*, 2008). One of the benefits of applying symbolisation to a discrete sequence was that the original complex sequence could be remodelled by a smaller subset of symbols. The output signal had fewer possible states while still maintaining a meaningful amount of information. Hence it provided a simpler representation of the signal, aiding the analysis.

Data for the research was collected by a research grade device. This accelerometer used standard XYZ 3-dimensional space to measure the volume of physical activity at 10 Hz. Each event recorded by the accelerometer had a corresponding timestamp saved next to the XYZ acceleration (John and Freedson, 2012). The 3-dimensional XYZ space was translated into a 1-dimensional space by calculating the magnitude of the total acceleration vector using hardware processing and analogue-to-digital conversions. The signal was then processed by a proprietary algorithm into *activity counts per epoch* using ActiLife software (https://www.actigraphcorp.com/actilife accessed on 28-Feb-2019). For the NHANES data used in this thesis, an epoch span equated to 1 minute and, therefore, activity signal is expressed in *activity counts per minute*. The actual activity count values vary from 0 for no activity to 32767 for the maximum amount of activity (Tudor-Locke *et al.*, 2010). Therefore, activity counts are represented by an activity space ($a \in A$) where acceptable values are between 0 and 32767:

$$A \in [0, 1, 2, \ldots, 32767] \tag{1}$$

An example day extracted from the NHANES raw physical activity data file was plotted in Figure 10. It is possible to visually find a period of inactivity between approximately 1 AM and 9 AM in this case. However, locating such periods of inactivity programmatically is non-trivial. An algorithm to find non-wear time is provided by Tudor-Locke (Tudor-Locke *et al.*, 2010). This algorithm has been modified to locate days with at least 10 hours of valid wear-time. The modified algorithm is included in Appendix 4, Data filtering algorithm with the ability to detect days with N valid hours of accelerometer recording time.

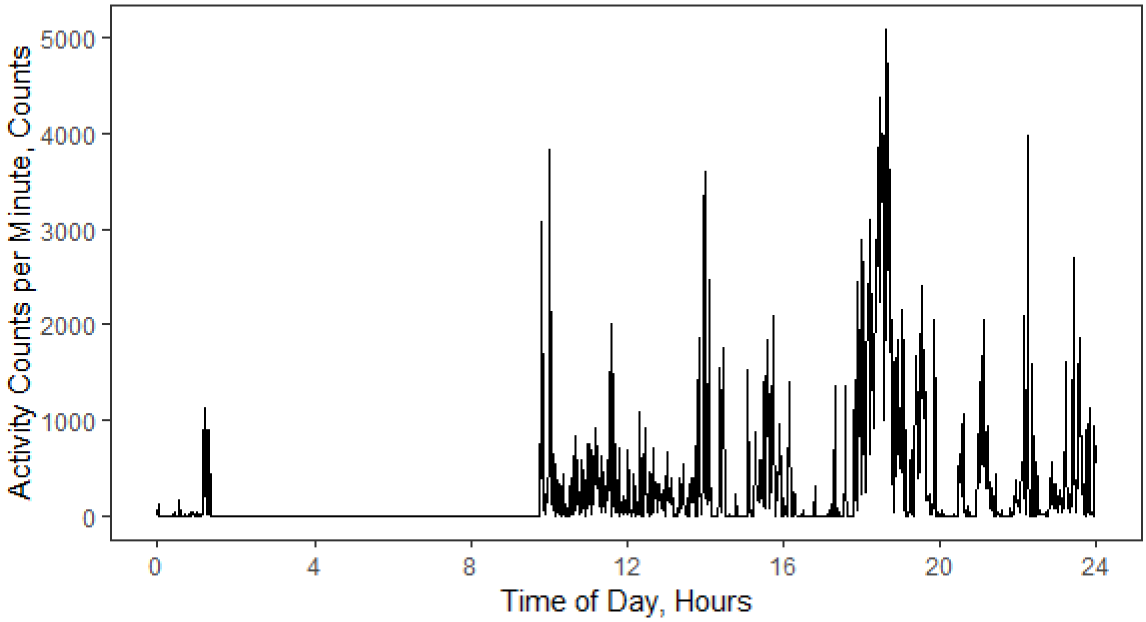

Figure 10. An example line plot of the 24-hour period for activity counts per minute

According to the signal processing protocol suggested by Tudor-Locke et. al., accelerometer activity counts can be transformed into activity intensity, such as Sedentary, Low, Light, Moderate, Vigorous, and Extra Vigorous (Tudor-Locke, Camhi and Troiano, 2012). Low levels of activity count directly correspond to low levels of PA intensity. Physical activity intensity can thus be represented by an intensity space (*i* ∈ *I)*:

$$I = \{Sedentary, Low, Light, Moderate, Vigorous, Extra\ Vigorous\} \tag{2}$$

The goal of symbolisation is a translation from the activity space *A* into the intensity space *I* and can be represented by the symbolisation function σ:

$$\boldsymbol{\sigma}: \boldsymbol{A} \rightarrow \boldsymbol{I} \tag{3}$$

Tudor-Locke et al. recommended that the symbolisation function accounted for the age of participants, as the same value of activity counts required different effort for children, adults and older adults (Tudor-Locke, Camhi and Troiano, 2012). Physical activity records that are available in NHANES studies had participants age ($\tau \in T$) ranging from 6 to 85 y.o.:

$$\mathrm{T} \in [6,7,\dots,85] \tag{4}$$

The symbolisation function that accounts for the age is presented below:

$$\sigma(\tau): A \rightarrow I \tag{5}$$

An example translation for adults of 18 years or older is presented in Table 4 below. For the full mapping of the ages included in the study the reader is referred to Appendix 1. Symbolisation translation.

Intensity	**From activity counts**	**To activity counts**
Sedentary	0	100
Low	101	500
Light	501	2020
Moderate	2021	5999
Vigorous	6000	10000
Extra vigorous	10001	32767

Table 4. Intensity to activity conversion table

The algorithm that is used for symbolisation in this research is outlined below:

1. Build a data structure storing age, activity code, range of accelerometer counts
2. Build a dictionary of participant ID and corresponding age

3. Scan accelerometer data for each participant
    a. Parse each record in the file
    b. Retrieve current value for accelerometer
    c. Lookup a corresponding activity level using [1] and [2]
    d. Append symbol corresponding to activity level to an in-memory string
4. Save data from [3] to a file

After symbolisation, each PA pattern represents a temporal sequence of activities that reflects physical activity behaviour. An individual element of the symbolised sequence is referred to as a *word*. An example of 24-hour symbolised activity is presented on a Figure 11 below

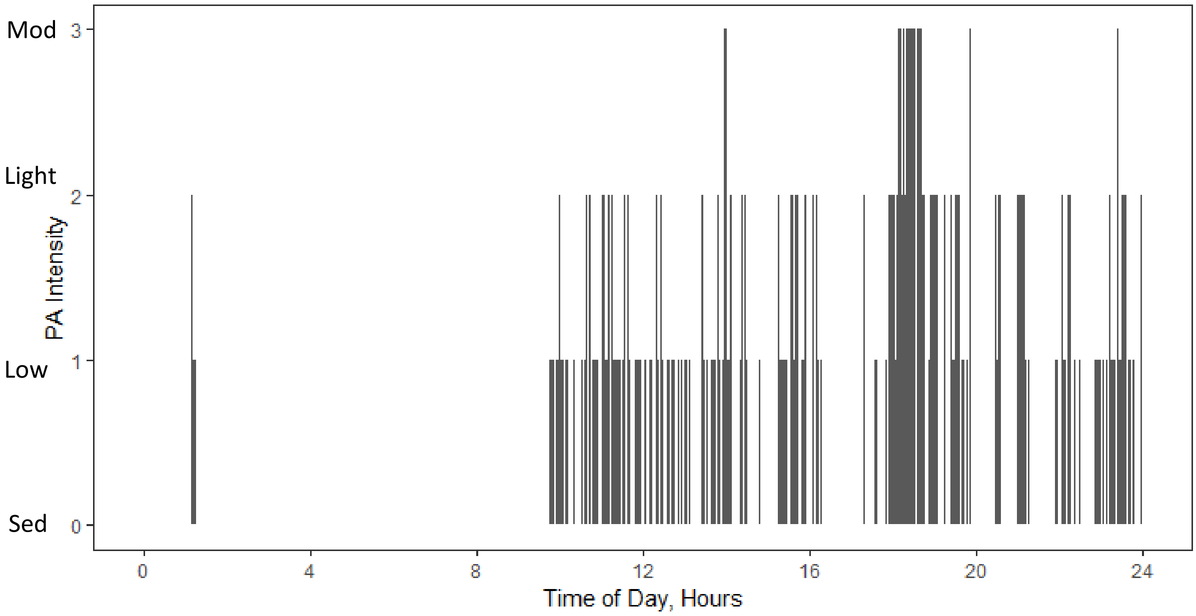

Figure 11. Example symbolisation plot

Figure 12 below represents a 20-minute “zoomed-in” PA pattern for this individual around 14:00 hours. Assuming that symbolisation resolution is 1 minute, it is possible to represent this pattern as a time series of 20 words where 0 is Sedentary, 1 is Low, 2 is Light and so on:

$$PA\ pattern\ (1\ \text{min}) = 22110111232321111321 \qquad (6)$$

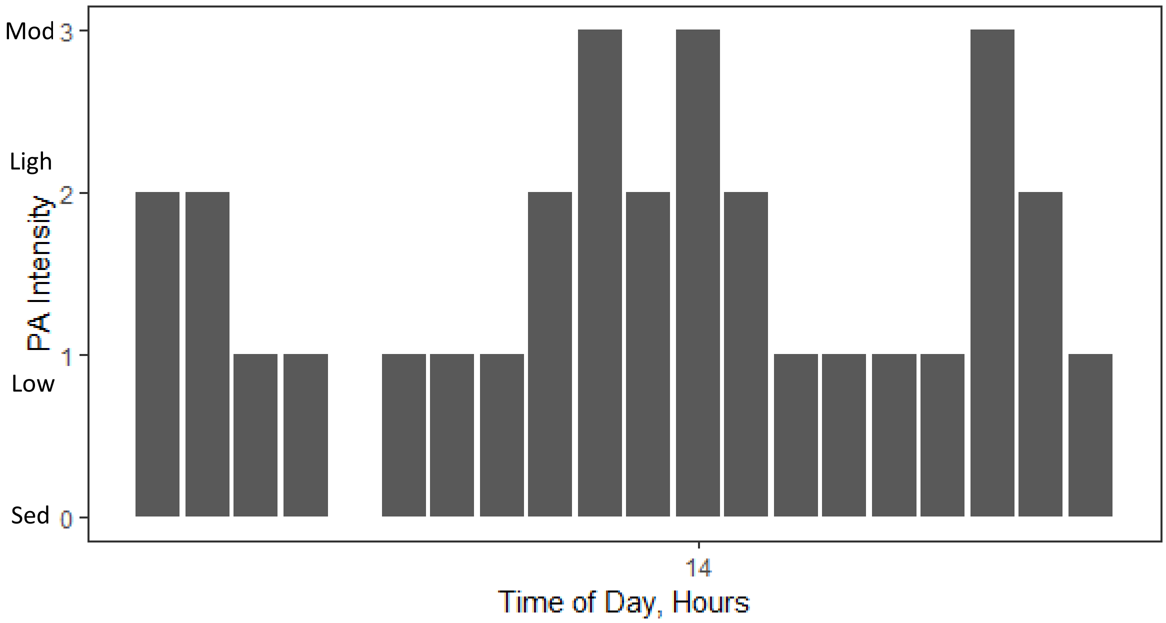

Figure 12. An example 20-minute PA pattern around 14:00

Coarse-graining of physical activity pattern

A process of coarse-graining is applied to each PA pattern to calculate its static and dynamic complexity. This process allows to remove redundant information from PA pattern by combining several symbols of PA intensities into a single word of PA. All unique words that are observed in a pattern thus create a *vocabulary* of PA pattern.

Coarse-graining combines several PA symbols from the sequence into new symbols. Coarse-graining is performed using fixed-length intervals ($\delta \in \Delta$), ranging from 2 to 12 minutes. These values do not bear significant meaning from the point of physical activity research, rather those are bounded by the mathematical constraints. An integer number of less than 2 has no effect. Values larger than 12 minutes create long words that practically do not repeat and make the analysis of PA pattern difficult as most of the words become unique (perhaps apart from total inactivity, e.g. sleep or non-wear time). Thus, intervals of coarse-graining are constrained by the following set:

$$\Delta \in [2,3, \dots , 12] \tag{7}$$

New symbolisation is applied with coarse-graining. It is represented by a translation function that depends on the age of the participants and a coarse-graining interval. The function that translates activity counts *A* into a new coarse-grained sequence of intensities I is presented below:

$$\sigma(\tau, \delta): A \rightarrow I \tag{8}$$

An example of coarse-graining on a 2-minute interval is depicted in Figure 13.

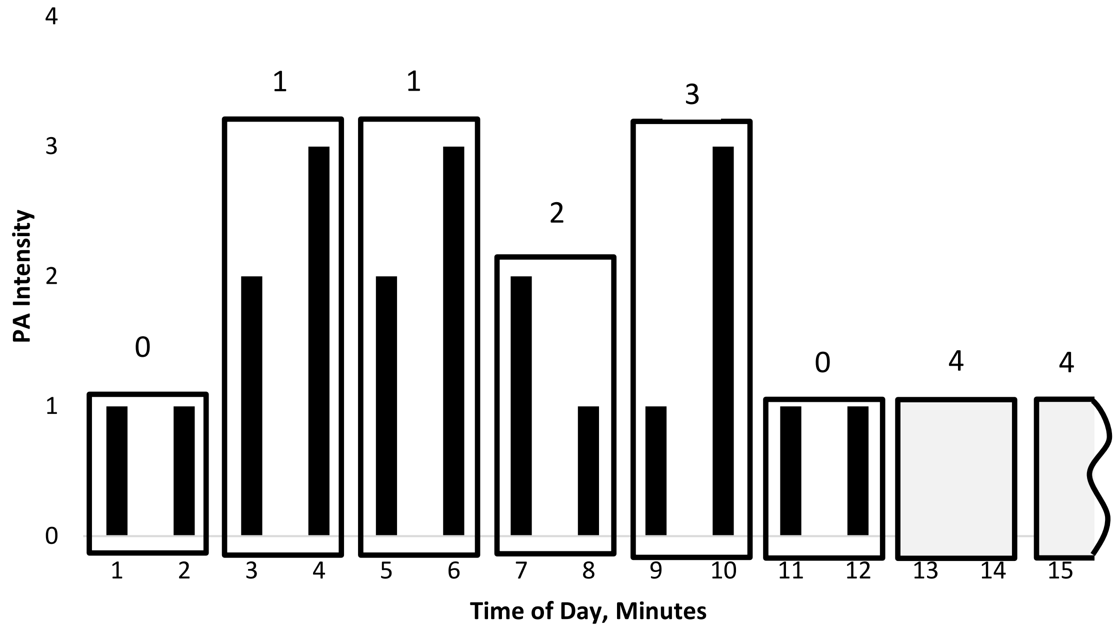

Figure 13. Coarse-graining of symbolic sequence 2 minutes interval

This example shows the PA pattern of a 15-minute interval with PA intensity symbolised on a 2-minute scale. In total, this coarse-grained pattern contains 4 unique words in a sequence of 8 words. Each word is repeated exactly twice. New words of the coarse-grained activity counts are encoded as new numbers: 0 = *Low -> Low*, 1 = *Light -> Moderate*, 2 = *Light -> Low*, etc. Note, the order of activities is important. The intensities in a PA word are referred to as *letters* of the word. For example, the PA word *Low->Low* is created by repeating two letters: *Low* followed by *Low.*

Words *Low -> Light* and *Light -> Low* are different words and are encoded with different numbers during the coarse-graining step. This is because the dynamic complexity of the PA pattern is sensitive to the order of the activities.

An example of a new coarse-grained PA pattern is represented as following:

$$PA\ pattern(2\ \mathrm{min}) = \ 01121133 \tag{10}$$

A coarse-graining vocabulary is a set of all words that can occur. In this example, the PA vocabulary is presented in Table 5:

2-letter word	New coarse-grained value
Low -> Low	0
Light -> Moderate	1
Light -> Low	2
Light -> Moderate	3
Sedentary -> Sedentary	4

Table 5. An example coarse-graining transition map

Using definitions from combinatorics, the size of a PA vocabulary can be described by a permutation (when the order of elements is important, and elements can repeat). If δ is a coarse-graining interval and the number of all possible PA intensities can be explained by the size of I:

$$|I| = sizeof(I) \tag{11}$$

The size of PA vocabulary *V* is a permutation of coarse-grained PA intensity given a fixed-length interval δ and a size of PA word. *V* can be calculated by:

$$|V(I, \delta)| = |I|^{\delta} \tag{12}$$

For example, given 6 possible PA intensities and a coarse-graining interval of 2 minutes there are 36 words. The number of possibilities grows exponentially with the size of the intervals:

$$|V(6, 2)| = 6^2 = 36 \tag{13}$$

$$|V(6, 3)| = 6^3 = 216 \tag{14}$$

$$|V(6, 4)| = 6^4 = 1296 \tag{15}$$

$$\dots \tag{16}$$

$$|V(6,12)| = 6^{12} = 2176782336 \tag{17}$$

The exponential growth of PA vocabulary indicates that there exists a large pool of words that an individual can choose. If all possible words are used with the same likelihood of occurrence, then PA behaviour becomes truly random and has no structure. On the other side of the spectrum, if the behaviour is represented by a single repeated word, then the PA pattern itself is trivial and fully determined. Human behaviour appears to be somewhere in between and ever slightly closer to the trivial case. The size of the PA vocabulary is limited by the number of letters in a word and in practice by the total duration of the recording.

PA vocabulary describes how many possible coarse-grained activities a person can use. PA pattern consists of words put into a time series in the particular position. Practically, the length of PA pattern $|PA\ pattern(\delta)|$ with regards to the coarse-graining interval δ is limited by the total recording duration t and the interval δ. A day consists of 1440 total minutes including sleep time and non-wear time. Therefore, total recording time t is calculated as:

$$t = N_{days} * N_{minutes\ in\ a\ day} \tag{18}$$

The length of PA pattern (*N*) is determined by the recording duration t and the coarse-graining interval δ:

$$|PA\ pattern(\delta)| = \frac{t}{\delta} \tag{19}$$

For example, if recording period *t* is 5 days, each containing 1440 minutes, the length of PA pattern can be calculated as:

$$|PA\, pattern(2\, min)| = \frac{7200}{2} = 3600\, words \quad (20)$$

$$|PA\, pattern(3\, min)| = \frac{7200}{3} = 2400\, words \quad (21)$$

$$\ldots \quad (22)$$

$$|PA\, pattern(12\, min)| = \frac{7200}{12} = 600\, words \quad (23)$$

The relationship between the size of vocabulary $|V(I, \delta)|$ and the length of the coarse-grained PA pattern $|PA\, pattern(\delta)|$ is graphed in Figure 14:

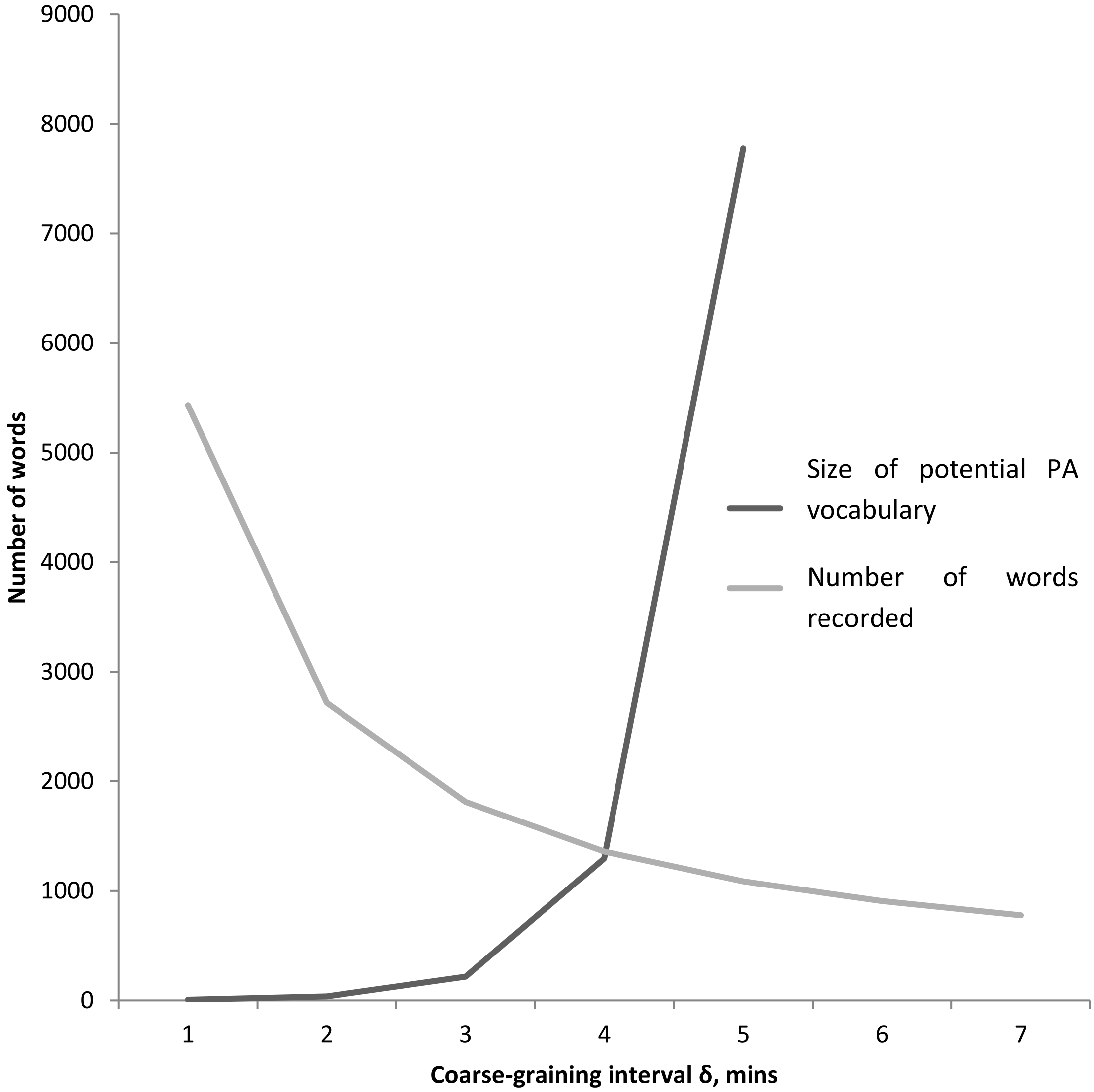

Figure 14. Size of potential PA vocabulary and the number of words recorded

The graph shows the relationship between coarse graining interval (horizontal axis) and number of words (vertical axis). The light grey curve exponentially decreases with the coarse graining interval. This curve is calculated for 5 days of recording and it shows a practical number of words given the fixed-length duration of recording time. For example, for the coarse graining interval of 2 minutes, it was possible to have 3600 words (5 days of recording x 24 hours in each day x 60 minutes in each hour / 2 minutes of coarse graining). For the coarse graining interval of 7 minutes, it is possible to have 1028 words (5 days of recording x 24 hours in each day x 60 minutes in each hour / 7 minutes of coarse graining). The dark grey curve exponentially increases with the coarse graining interval. This curve shows a number of unique words that can be constructed given the coarse graining interval and the 6 possible PA intensities.

As seen in Figure 14, the curves intersect at around $\delta = 4$, creating an optimum point for coarse-graining of PA pattern for the NHANES data set. It makes a balance between the size of PA vocabulary which should be as large as possible for statistical analysis and the number of the recorded words which is defined by the practical length of recording.

This researched uses data for which a coarse-graining interval $\delta = 4$ minutes is an optimum for coarse-graining. It is possible to suggest that there might be some fundamental interval of PA pattern as measured by accelerometry that can be universally applied in PA research.

A new word resulting from the coarse-graining process needs to be symbolised with a new symbol that represents the sequence of 4 fine-grained symbols. The way this is done is important because any new symbolisation can introduce an artificial bias affecting the dynamic complexity of PA pattern. Consider an example where a new symbol needs to be assigned to the following PA word:

$$Word = \{Sedentary, Low, Low, Moderate\} \tag{24}$$

The measure of complexity, in this case Kolmogorov complexity, is sensitive to the actual representation of information (Fortnow, 2001). It is essential to be careful so that no artefacts are introduced purely due to the symbolisation or coarse-graining process. Any non-trivial symbolisation of PA words can introduce some level of noise and artificial bias (Klir, 1988). For example, there are many different ways to represent a word from Equation 24:

$$Possible\ symbol_1 \ = \ SedentaryLowLowModerate \tag{25}$$

$$Possible\ symbol_2 \ = \ \alpha \tag{26}$$

$$Possible\ symbol_3 \ = \ 1 \tag{27}$$

$$Possible\ symbol_4 \ = \ SLoLoM \tag{28}$$

Representations [25] – [28] describe the same word. However, each representation introduces a certain degree of noise to the information which is undesirable for the calculation of dynamic complexity of PA pattern. This problem has been studied in depth in telecommunication where it is essential to transfer information between two points A and B while sending all the information and preserving the size of transmission to the minimum. In 1951 Huffman published an algorithm for creating a minimal possible coding of a signal. This algorithm became known as the Huffman tree prefix code (Huffman, 1951). It is one of the fundamental methods used in information theory for compression. It is practically utilised in many lossless compression algorithms (where all information is preserved in the compressed form).

Huffman encoding is based on the probability of occurrence of a word, and it provides the shortest encoding of the input data. The most probable words have the shortest encoded length.

However, one of the downsides of using this encoding for analysis of dynamic complexity is that the output encoded words had various lengths: more probable words are shorter than less probable ones. This is not desirable for further analysis of the dynamic complexity of PA pattern, as it is harder to compare output words of different lengths. A modification is added to the original Huffman algorithm to preserve the fixed-length of the output words.

The algorithm for coarse-graining of symbolic sequences using modified Huffman encoding is detailed below:

1. Split the input sequence into words of fixed length. For the sake of simplicity, in this example size 2 will be used (the actual word size used in the research is 4).

   Input symbols: {Sedentary, Low, Low, Moderate, Moderate, Low, Sedentary, Low, Low}

   Output words: {[Sedentary Low], [Low Moderate], [Moderate Low], [Sedentary Low], [Low]}

2. Generate probabilities of word occurrence

PA Word	**Probability**
[Sedentary Low]	2/5
[Low Moderate]	1/5
[Moderate Low]	1/5
[Low]	1/5

Table 6. Probability of words in PA pattern

3. Sort the pairs by probability in descending order. Build a prefix tree (Huffman, 1951): start from left-to-right and assign "0" to the left value and "1" to the right value.

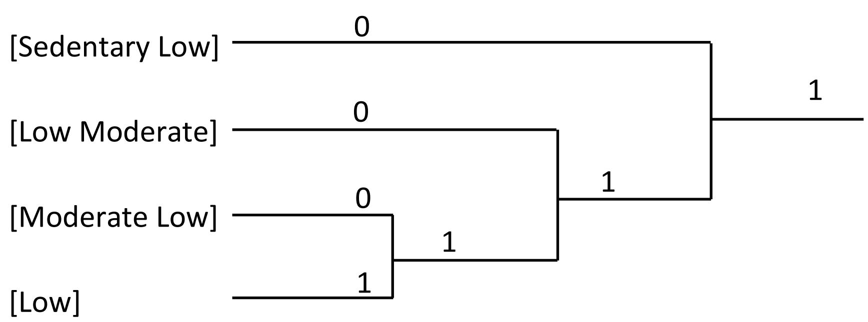

Figure 15. Huffman prefix tree for PA pattern (2-letter words)

4. Extract the encoding of the words following step 3 in the right-to-left direction.

Word	Encoding
[Sedentary Low]	10
[Low Moderate]	110
[Moderate Low]	1110
[Low]	1111

Table 7. New Huffman codes for PA words

5. Ensure all the words have equal length by prepending 0 to the encoded words

Word	Encoding
[Sedentary Low]	0010
[Low Moderate]	0110
[Moderate Low]	1110
[Low]	1111

Table 8. Fixed-length Huffman codes for PA words

Steps 1-4 are the original Huffman coding, and step 5 is the modification that allowed to output words of fixed length. The result tree and encodings are used to calculate measures of static and dynamic complexities of PA pattern.

Correlates and determinants of physical activity

The NHANES 2003-2004 and NHANES 2005-2006 data catalogues were searched for the known or potential correlates of PA reflecting a person's health. The correlates were selected based on the epidemiological models of determinants of PA developed by Bauman (Bauman *et al.*, 2012) and Sallis & Owen (Sallis JF, Owen N, 2008). Following these models, the correlates were arranged into domains, such as Individual, Interpersonal, and Environmental. These correlates were chosen according to two criteria. First, the correlate had to be confirmed by a majority of the reviewing studies (Bauman *et al.*, 2012), and second, the correlate had to be either present in the NHANES database as a separate variable or approximated by a composite score or a proxy measure.

A summary of the extracted correlates is given in Table 9 below

Category	**Correlate**	**Variable**
Individual	Age	Age in years, RIDAGEYR
	Body composition	Body Mass Index, BMXBMI
	Gender	Gender, DMDHRGND
	Self-reported health	Self-reported health, HSD010 Disease burden, ZPD_SCORE
	History of PA	Long-term 10 years, PAQ540 Short-term last year, PAQ500
	Alcohol consumption	Drinks per day, ALQ130
	Smoking habits	Serum cotinine level in the blood, LBXCOT
	Fitness	Shortness of breath on stairs, CDQ010
	Mental health	History of mental health, HSQ480
	Injury	Composite injury score, ZINJURY_SCORE
	Social economic status	Household income, INDHHINC Occupation group, OCC_CODE
	Ethnicity	Ethnicity, RIDRETH1
	Education	Education level, DMDEDUC2+DMDEDUC3
Interpersonal	Marital status	Marital status, DMDMARTL
	Social support	Number of people in the household, DMDHHSIZ Number of close friends, SSQ061

	Origin	Country of birth, DMDHRBRN
	Social desirability	PA level compared to others, PAQ520
Environment	Weather, exposure to sunlight	Vitamin D, LBDVID

Table 9. Summary of the correlates extracted from the NHANES 2003-2006

Age of participants is encoded as an integer number of full years recalled at the time of examination. This varied from 6 to 85 (ages over 85 are capped at 85). The average age is 37.81 ± 23.62 y.o., n = 7420 for the combined dataset from NHANES 2003-2004 and NHANES 2005-2006.

Body mass index (BMI) is calculated as part of the body measurements examination. The population is grouped into Underweight <18.5 kg/m^2 (966 individuals), Normal Weight 18.5 – 25 kg/m^2 (2533 individuals), Overweight 25 – 30 kg/m^2 (2117 individuals), and Obese >30 kg/m^2 (1756 individuals); n = 7372, 48 missing entries.

Gender is encoded as a binary variable. There are 4224 males and 3196 females, n = 7420.

Self-reported general health is described by the following statistics: 753 Excellent, 1818 Very Good, 2141 Good, 856 Fair, 130 Poor, and 1722 missing entries.

Ethnicity is reported as part of the demographics collection. There are 1987 Mexican American, 190 Other Hispanic, 3246 Non-Hispanic White, 1699 Non-Hispanic Black, 298 Other ethnicities; n = 7420.

Education level is categorised as 3675 less than secondary school graduate, 1364 secondary school graduate or equivalent but no post-secondary education, 1312 some college degree or equivalent but not finished, 1064 college graduate or above; n = 7415, 5 missing entries.

Annual household income is self-reported as a range of values in US dollars. Income range and a corresponding number of participants is reported in Table 10 below:

Income range, US dollars	**N**
0 – 4 999	124
5 000 – 9 999	286

10 000 – 14 999	557
15 000 – 19 999	530
20 000 – 24 999	577
25 000 – 34 999	945
35 000 – 44 999	795
45 000 – 54 999	725
55 000 – 64 999	496
65 000 – 74 999	375
75 000 and over	1620
Missing data	390

Table 10. Self-reported annual household income in US dollars

Vitamin D is collected as part of the laboratory examination and measured in ng/ml. It varied from 2 to 86, and the average content is 22.17 ± 8.51 ng/ml, n = 6597, 823 missing entries.

There are several other PA correlates extracted from the questionnaires, such as compound injury score, and compound medical condition score. Specific questions and corresponding statistics for these scores are presented in Appendix 3. Re-coded demographic variables used for statistical analysis.

No information is available about health policies. Different states in the US can have varied regional polices. Quantification of the influence of such policies is a challenge and is beyond the scope of this thesis.

There are several other correlates that are missing too much data to be included in the study. For example, it can have been interesting to see how depression affects PA pattern, but out of 7420 participants, data for 7294 is missing.

Statistical analysis of correlates

General linear regression models are generated separately for the three dependent measures μ_1, μ_2, and μ_3. Three independent models that contain the same correlates have three different dependent variables. Such a statistical experiment makes it possible to understand if measures of volume and complexity of PA pattern can be described by the correlates. The analysis is carried out in SPSS v21 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.). The analysis included merged NHANES 2004-2005 and NHANES 2005-2006 data sets. Statistical analysis does not use weighted samples to account for oversampling in the NHANES dataset. Analysis does not account for the non-random loss of accelerometry data and non-random missing data because this study is exploratory and is not concerned with obtaining results that represented the US population. If any of the covariates are missing for a case, the whole case is excluded from the models.

Collinearity statistics are calculated to identify cases of linear correlation between the input variables. This is done by instructing SPSS to calculate *tolerance* and *variance inflation factor* for the general linear regression models. Tolerance values less than 0.1 indicate that the PA correlates entered in to the model are significantly correlated and are redundant in the model. An exclusion boundary is set for the tolerance level < 0.2 (O'Brien, 2007). The lowest observed tolerance level is 0.69. This indicates that no covariates entered to the model are significantly correlated with each other.

The analysis is divided into the analysis of population stratified by age (children, adults, and older adults) and the analysis of the whole population. Each stage produced 3 results per stage for μ_1, μ_2, and μ_3. The amount of variance explained in the models (R^2 and R^2 adjusted) is compared to ascertain if metrics of the pattern (μ_2 and μ_3) are better explained by the correlates of PA reflecting a person's health than the metrics of the volume of PA (μ_1).

Stage 2 followed a similar design but applied to the whole dataset. The same covariates are entered into the models. The amount of variance explained (R^2 adjusted) is compared to understand if metrics of the complexity of PA pattern (μ_2 and μ_3) are better described by the correlates of PA reflecting a person's health compared to the measure of the volume of PA(μ_1). Their unstandardised regression coefficients are compared to determine the strength of the relationship.

Measures of physical activity

This research derived three measures of PA from NHANES accelerometry data. The first metric is a commonly used measure of the volume of physical activity, and the other two metrics are measures of complexity of PA pattern.

Metric 1: Total physical activity counts

The first measure μ_1 is a *daily average of the total physical activity counts*. This measure is used for comparison analysis:

$$\mu_1 = average\ (sum\ (activity\ count, day), day) \tag{29}$$

Metric 2: Static complexity of PA pattern

It is known that human PA occurs in bursts followed by prolonged periods of inactivity (Barabási, 2005; Zhou et al., 2008). The bursts of such activity appeared to be hard to predict (Barabási, 2005; Chastin and Granat, 2010). Like any dynamic system, PA behaviour has some static qualities associated with the states of the system. These qualities of PA behaviour are associated with the time-independent factors. The static analysis attempts to quantify stationary qualities of PA pattern such as how many different kinds of activities are performed and how often.

Static complexity μ_2 is the size of the vocabulary. It is the number of all unique words that an individual used in their PA pattern. Mathematically, it can be defined as the cardinality of the PA pattern vocabulary. The cardinality of a set is the number of all unique items found in the set. It describes how diverse a PA pattern vocabulary is and how many distinct words of PA are required to reconstruct a PA pattern. Static complexity describes an individual's ability to change PA behaviour in daily activity profile. However, it does not describe the sequence of activities or how those activities are performed:

$$\mu_2 = |\{\text{unique words of length } \delta \text{ occurring in PA pattern}\}| \tag{30}$$

For example:

$$PA\ Pattern = \{Sed, Low, Low, Mod, Mod, Low, Sed, Low, Low\} \tag{31}$$

$$Words_{2\,min} = [Sed\ Low], [Low\ Mod], [Mod\ Low], [Sed\ Low], [Low] \tag{32}$$

$$\mu_2 = 4 \tag{33}$$

Metric 3: Dynamic complexity of PA pattern

The dynamic qualities of PA pattern are more challenging to quantify. There is no standard and general approach that can be applied to PA pattern, and it is an actively developing field of research. It is difficult to quantify the complexity of any given sequence without prior knowledge of the laws driving the behaviour (Lempel and Ziv, 1976). Kolmogorov, Chaitin, and others produced fundamental research on the topic of information theory, entropy and complexity of sequences in the mid-20th century (Kolmogorov, 1965, 1983; Chaitin, 1966). One of the results from their work is a new complexity measure which became known as Kolmogorov complexity. It provides an estimate of how complex, in a general, is any given sequence. In his publication, Kolmogorov noted that this metric is generally incomputable (Kolmogorov, 1965). However, it is often possible to find an approximation of Kolmogorov complexity (Kolmogorov, 1983).

Generic information compression can be used as an approximation of Kolmogorov complexity. Compression algorithms are naturally good at finding repetitions of information in an input sequence. The GZIP, for example, has been used as an estimate for Kolmogorov complexity (Cilibrasi, 2001).

GZIP is originally created by Jean-loup Gailly and Mark Adler for the GNU open source project (https://www.gnu.org/software/gzip/). GZIP is a variation of Lempel-Ziv (Ziv and Lempel, 1977). In GZIP, an input is a sequence of bytes. Compression works in small chunks, where a number of duplicate strings is minimised. Every duplicated string is replaced by a pointer in a form of a pair <distance, length>. By default, distance is limited to 32 KB and the length is limited to 258 bytes. The algorithm scans input from the beginning, locating new chunks. If a chunk is not found in the previous 32 KB, it is treated as a new string and is saved as a new literal in a hash table. Literals are compressed using Huffman tree and the match distances are compressed using another Huffman tree (Huffman, 1951). These two trees are stored in the beginning of a compressed block. A block is not limited in size, as long as it can fit in memory. GZIP selects the best number of blocks as determined by the need to create a new block with new trees. Duplicate strings are found using a hash table. Further details regarding GZIP algorithm are available from the MIT file server ftp://prep.ai.mit.edu/pub/gnu/gzip/. GZIP file format is documented as RFC1952 and it is available from https://www.ietf.org/rfc/rfc1952.txt.

GZIP algorithm is designed to find repetitions in input and therefore provided one of the shortest representations of an input. Given some arbitrary non-random input it is usually possible to compress it, producing a sequence that is shorter than an input. By extracting a ratio between the input and output sequence, it is possible to estimate how much information is shared between the input sequence and the output sequence.

The third measure used in this research, μ_3, is the *dynamic complexity* of PA pattern. It measures inter- and intra-personal temporal variability of PA and helps to quantify temporal aspects of PA pattern. This measure helps to distinguish between the pattern of different dynamical complexity. In the simplest form, dynamic complexity can be calculated as the ratio of the length of the original PA pattern to the length of the compressed PA pattern using the GZIP algorithm:

$$\mu_3(\text{unadjusted}) = \frac{|PA\ pattern|}{|\ \text{GZIP(PA pattern)}|} \tag{34}$$

This form of complexity is easy to obtain, but there are two issues with it.

First, equation 34 provides an inverse relationship between PA pattern and dynamic complexity of PA pattern. The length of the compressed output is shorter for simpler inputs and thus measure μ_3 had greater value. Simpler PA pattern results in an increase in μ_3 and more complex PA pattern results in a decrease in μ_3. This is inconvenient for visualisation purposes, as analysts might intuitively expect higher complexity value to correspond to a greater value of μ_3.

The second issue is that μ_3 is biased with μ_2. It is possible to suggest that larger vocabularies can result in higher dynamic complexities. To account for the two issues mentioned above, a coefficient *k* is introduced:

$$\mu_3 = \mathrm{k} * \frac{|PA\, pattern|}{|\,\mathrm{GZIP(PA\ pattern)}|} \tag{35}$$

Where k is defined as:

$$k = \left(-\frac{1}{\mu_2}\right) \tag{36}$$

The final formula to calculate μ_3 is presented below:

$$\mu_3 = \left(-\frac{1}{\mu_2}\right) * \frac{|\,(PA\, pattern)|}{|\,GZIP(PA\, pattern)|} \tag{37}$$

Chapter 4. Results

Descriptive statistics

Socio-demographic information was retrieved from the NHANES questionnaires. Behavioural and health information was obtained from the laboratory examinations. Age (in years) at the time of the interview was used for the sample stratification: children & adolescents (6 – 18 y.o.), adults (19 – 55 y.o.), older adults (>56 y.o.) and a full-sample. Table 11 below lists the categories of individual, interpersonal and environmental independent variables used in the models:

Characteristics	**Children & adolescents (6 – 18 y.o.)**	**Adults (19 – 55 y.o.)**	**Older adults (>56 y.o.)**	**Full sample**
N	2448	2811	2161	7420
Age (years)	12.2 ± 3.6	37.6 ± 10.7	69.9 ± 8.6	38.6 ± 24.2
BMI category				
Underweight	901	44	21	966
Normal weight	1051	879	603	2533
Overweight	286	980	851	2117
Obese	201	893	662	1756
Missing	9	15	24	48
Gender				
Male	1256	1706	1262	4224
Female	1192	1105	899	3196
Health				
Excellent	238	334	181	753
Very good	448	845	525	1818
Good	493	927	721	2141
Fair	96	333	427	856
Poor	8	31	91	130
Missing	1165	341	216	1722
History of PA (long term)				
Less active	0	1153	1329	2482
Same	0	1362	649	4459

More active	0	296	183	479
Missing	2448	0	0	
History of PA (mid-term)				
Less active	195	647	380	1222
Same	1895	1615	1487	4997
More active	358	549	294	1201
Alcohol consumption (drinks/day)				
Q1 (0 – 1)	0	1005	799	1804
Q2 (1 – 2)	0	258	106	364
Q3 (2 – 4)	0	214	53	267
Q4 (> 4)	0	208	50	258
Missing	2448	1126	1153	4727
Smoking status (serum cotinine)				
Non, <10 ng/dl	1409	1493	1393	4295
Light, 10-<100ng/dl	523	436	282	1241
Moderate, 100-<300ng/dl	141	74	67	282
Heavy, ≥300ng/dl	145	721	349	1215
Missing	230	87	70	387
Shortness of breath on stairs				
Yes	0	366	794	1160
No	0	881	1231	2112
Don’t know	0	0	2	2
Missing	2448	1564	134	4146
Mental health				
Excellent (0 - 5)	1144	2100	1713	4957
Very good (5 - 10)	70	139	69	278
Fair (10 to 20)	43	119	79	241
Poor (> 20)	26	112	83	221
Missing	1165	341	217	1723
Injury				
Low (< 1)	2448	2554	1869	6871

Medium (1 – 2)	0	240	262	502
High (2 – 3)	0	17	28	45
Very high (>3)	0	0	2	2
House income				
Low, \$0 – \$15k	498	429	570	1497
Medium, \$14k– \$35k	761	790	766	2317
High, \$35k– \$65k	519	687	390	1596
Very high, > \$65k	546	773	301	1620
Missing	124	132	134	390
Ethnicity				
Mexican American	905	674	408	1987
Other Hispanic	68	87	35	190
Non-Hispanic White	581	1361	1304	3246
Non-Hispanic Black	784	569	346	1699
Other	110	120	68	298
Education				
Some School	2347	567	761	3675
School Grad	100	726	538	1364
Some College	0	854	458	1312
College Grad	0	663	401	1064
Missing	1	1	3	5
Marital status				
Married	10	1607	1341	2958
Widowed	0	26	448	474
Divorced	0	220	219	439
Separated	2	86	44	132
Never married	957	633	73	1663
Living with a partner	5	238	36	279
Missing	1474	1	0	1475
Number of close friends				
0 – 3	0	571	807	1378
3 – 5	0	299	501	800

5 – 8	0	66	116	182
> 8	0	311	602	913
Missing	2448	1564	135	4147
Origin				
Born in the US	1647	2040	1777	5464
Born in Mexico	553	427	192	1172
Born elsewhere	179	269	148	596
Missing	69	74	43	186
Compare PA with others				
Less active	175	581	257	1013
Same level	1849	1400	908	4157
More active	424	830	996	2250
Vitamin D (ng/ml)				
< 16	625	783	552	1960
16 – 22	603	712	522	1837
22 – 28	565	587	508	1660
> 28	274	480	386	1140
Missing	381	249	193	823
Total valid wear time, mins	5359±1116	5405±1124	5554±1177	5433±1140

Table 11. Demographics and descriptive statistics of the selected sample

Ageing, the volume of PA and complexity of PA pattern

One of the most substantial relationships observed in ANCOVA analysis above was between age and measures of the volume of PA and complexity of PA pattern. It was clear, ageing was related to a decrease of the volume of PA and a decrease of static and dynamic complexity. The following analysis assessed if μ_1 - μ_3 measures were different between 3 age groups: children and adolescents (6 – 18), adults (19 – 55) and older adults (56 – 85).

A probability distribution for μ_1 - μ_3 appeared to be near-normally distributed, as graphed by the bold curve (see Figure 16 – Figure 18). However, at closer look, the figures displayed distribution behaviour closer to a Poisson distribution with skewed statistics either to the left or to the right with a long tail.

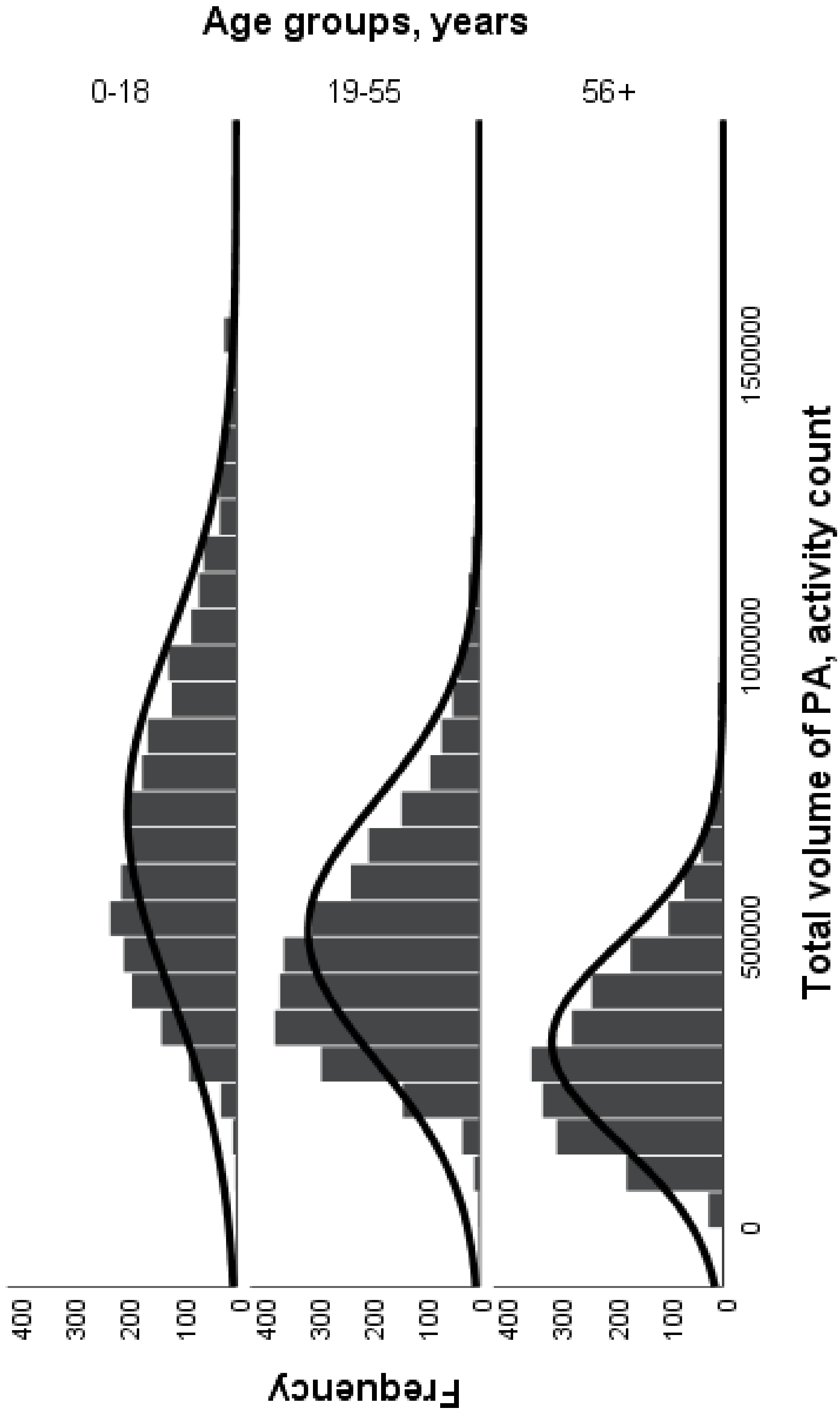

Figure 16. Distribution of volume of PA for age groups

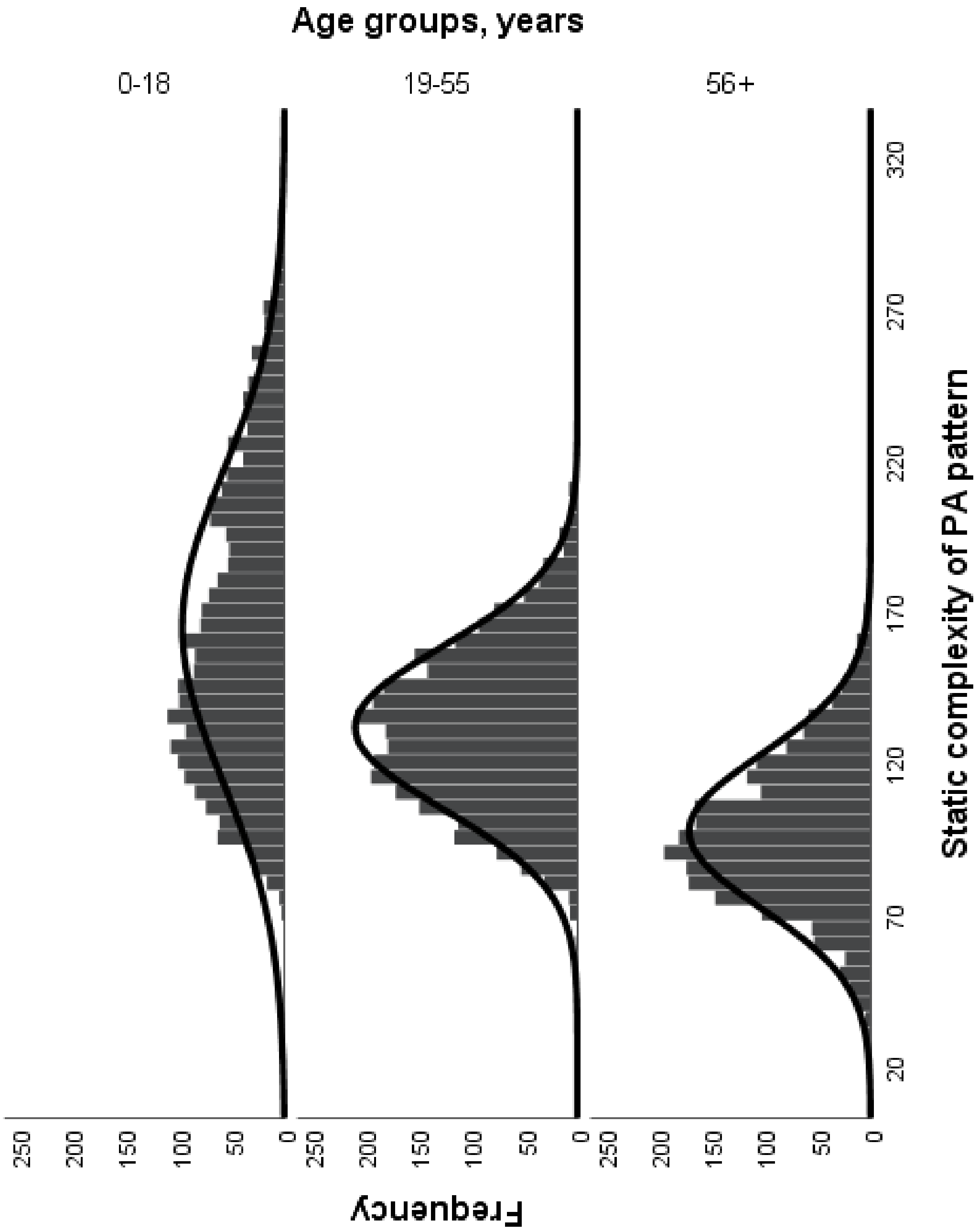

Figure 17. Distribution of static complexity of PA pattern for age groups

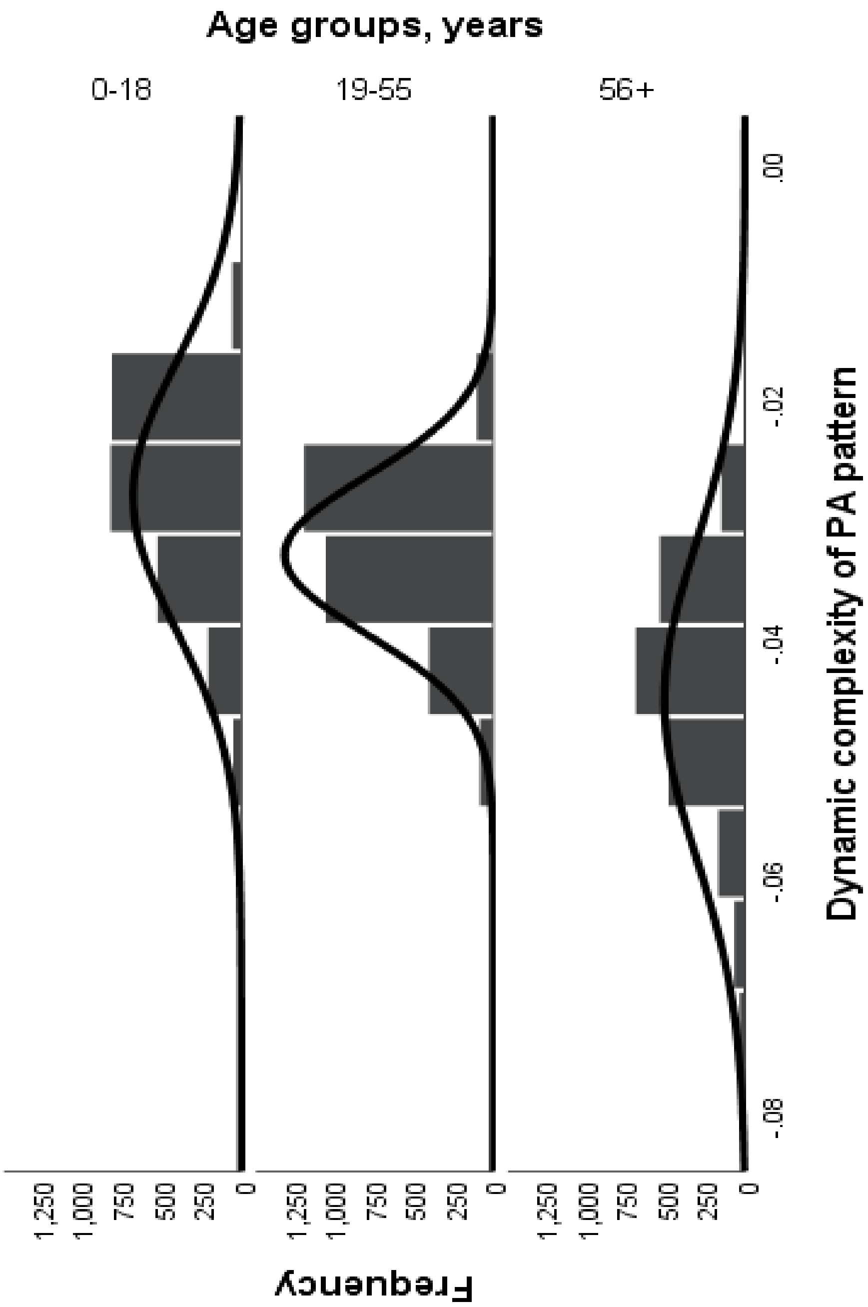

Figure 18. Distribution of dynamic complexity of PA pattern for age groups

Further normality tests demonstrated that μ1 - μ3 were not normally distributed: Kolmogorov-Smirnov test reported significance statistics of p=.000, which indicated the data significantly deviated from the normal distribution.

Descriptive statistics for the age groups is presented below:

Age groups		**Mean**	**Std Dev**
Total volume of PA μ_1, activity counts	6-18	709589.41	306126.415
	19-55	510076.92	222421.688

	56-86	318083.83	171139.337
Static complexity of PA pattern μ_2, number of words	6-18	165.48	50.785
	19-55	132.28	26.832
	56-86	98.00	25.279
Dynamic complexity of PA pattern μ_3, compression units	6-18	-0.0276	0.01118
	19-55	-0.0326	0.00666
	56-86	-0.0448	0.01323

Table 12. Mean and standard deviation for age groups, volume of PA and complexity of PA pattern

Both the volume of PA and the complexity measures of PA pattern appeared to follow three different curves. For children and adolescent, there was a rapid decline of volume and complexity of PA pattern with age. Adults appeared to maintain a slightly declining linear level of volume of PA and complexity of PA pattern. Older adults displayed a slow but constant decline of the volume and the complexity of PA pattern. Based on this, population had been categorised into three age strata: children 6-18 years (n=2448, 12.23±3.61), adults 19 – 55 (n=2811, 37.59±10.74), and older adults 55 – 85 (n=2161, 69.9±8.61).

Scatter plots presented next demonstrate that both the volume of PA and the complexity of PA pattern change with the age:

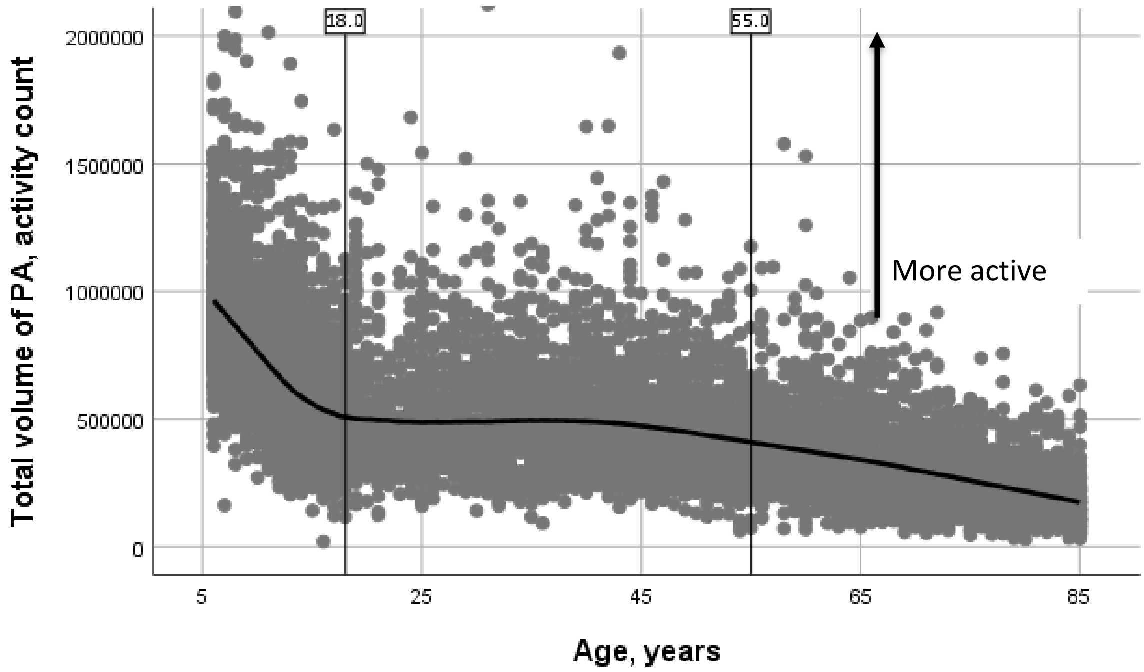

Figure 19. Scatter plot age vs volume of PA

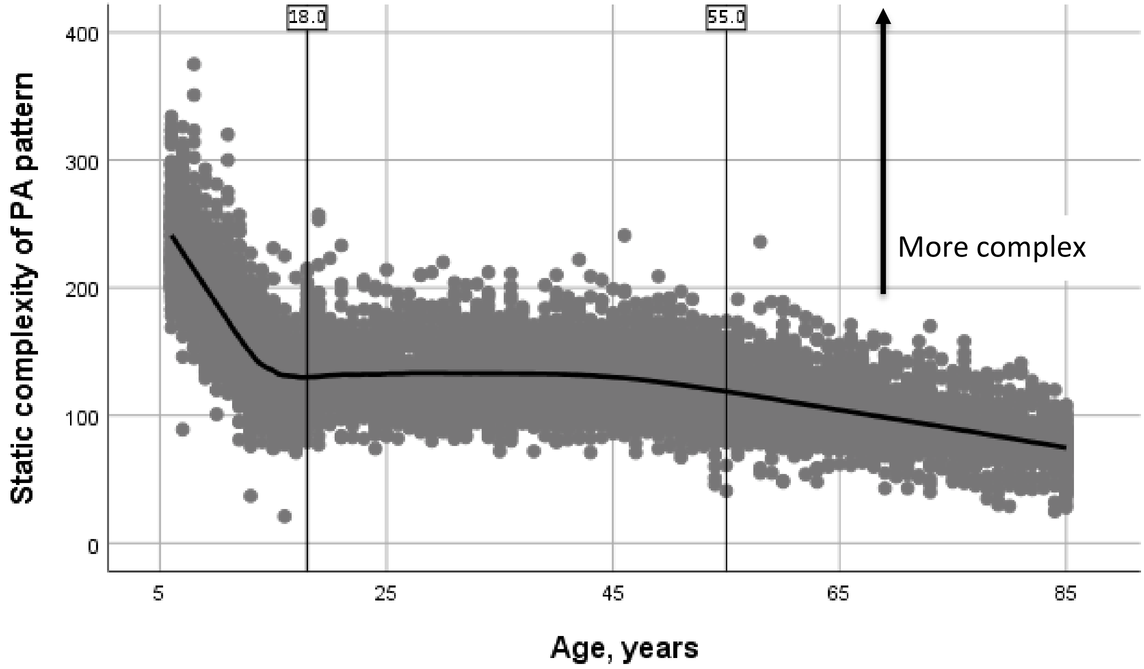

Figure 20. Scatter plot age vs static complexity of PA pattern

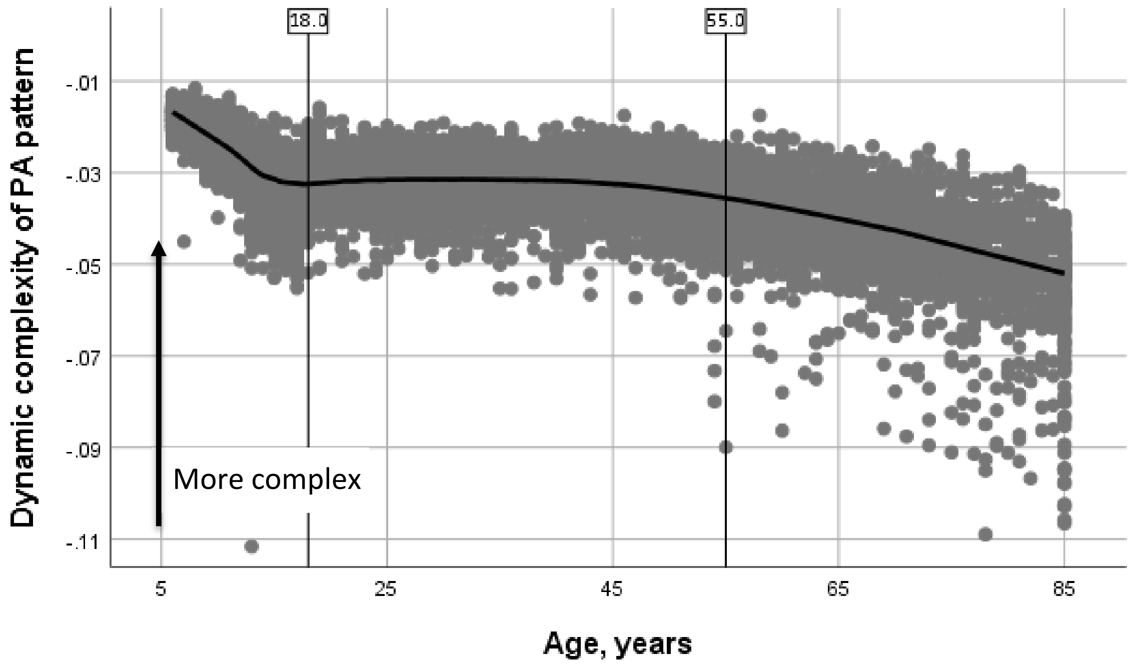

Figure 21. Scatter plot age vs dynamic complexity of PA pattern

The volume of PA, static and dynamic complexity of PA pattern decreased with age. For the participants aged 6 – 18 years this decrease was most rapid.

BMI, the volume of PA and the complexity of PA pattern

Another interesting relationship that was observed in ANCOVA analysis above was between BMI and measures of the volume of PA and complexity of PA pattern. It was demonstrated that an increase in BMI led to a decrease in the volume of PA and a decrease of the static and dynamic complexity. The following analysis assessed if μ_1 - μ_3 measures were different between 4 BMI groups: Underweight <18.5 kg/m2 (966 individuals), Normal Weight 18.5 – 25 kg/m2 (2533 individuals), Overweight 25 – 30 kg/m2 (2117 individuals), and Obese >30 kg/m2 (1756 individuals); n = 7372 included in this analysis and 48 missing entries – excluded from this analysis.

The probability distribution for μ_1 - μ_3 against BMI appeared to be near-normally distributed, as graphed by the bold curve (see Figure 22 – Figure 24). But, at closer look, the figures showed distribution behaviour closer to a Poisson distribution. This distribution has skewed statistics either to the left or to the right with a long tail.

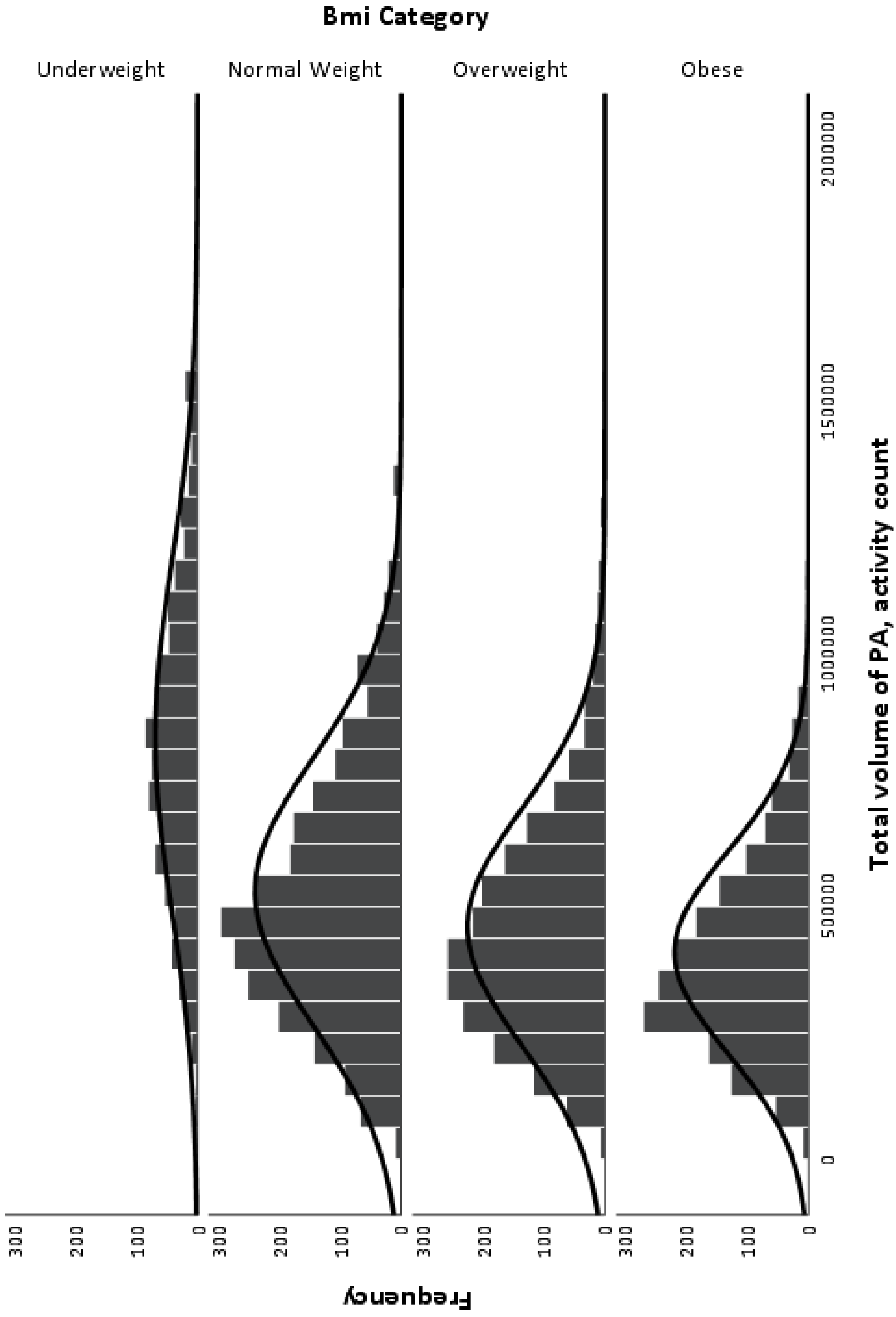

Figure 22. Distribution of volume of PA for BMI groups

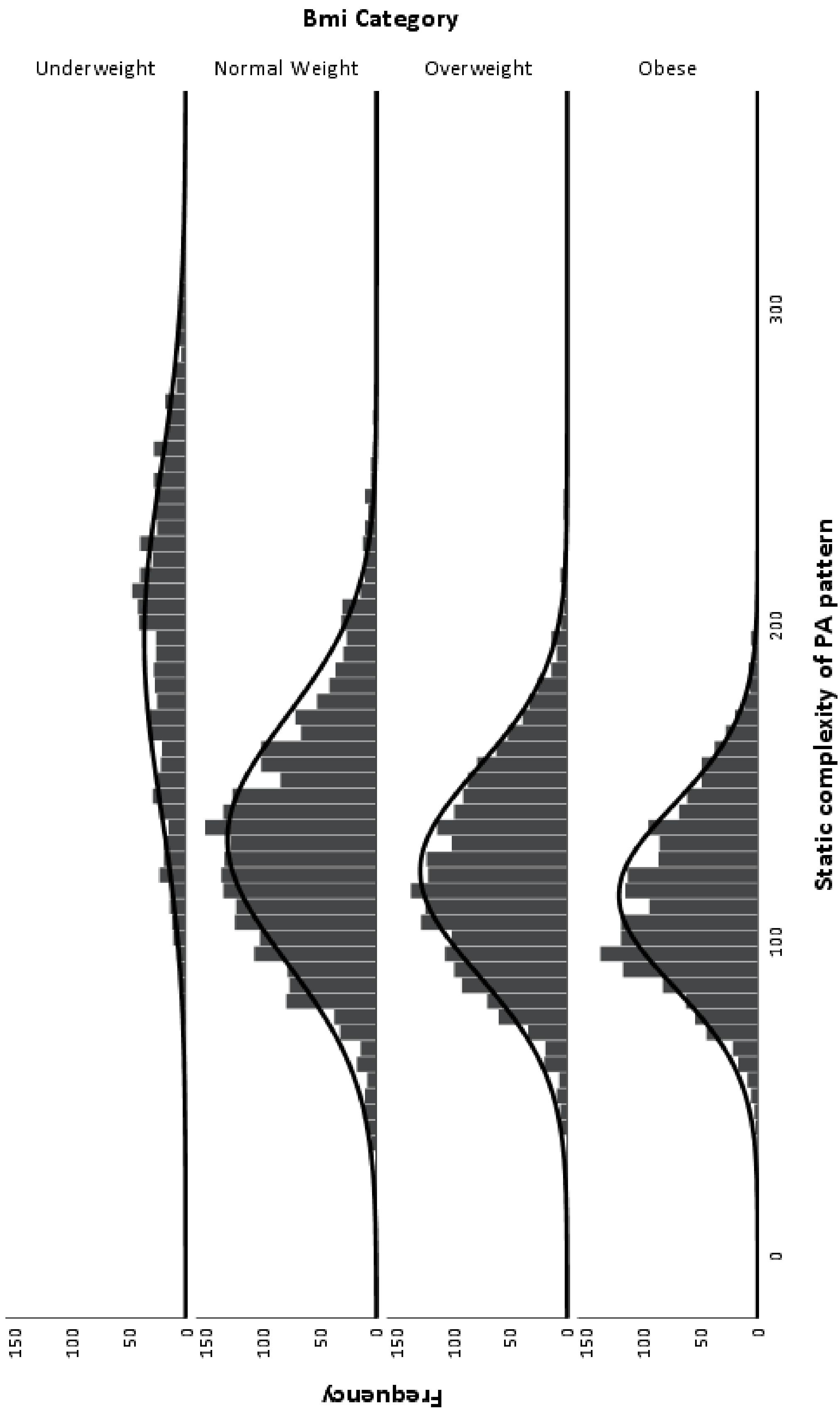

Figure 23. Distribution of static complexity of PA pattern for BMI groups

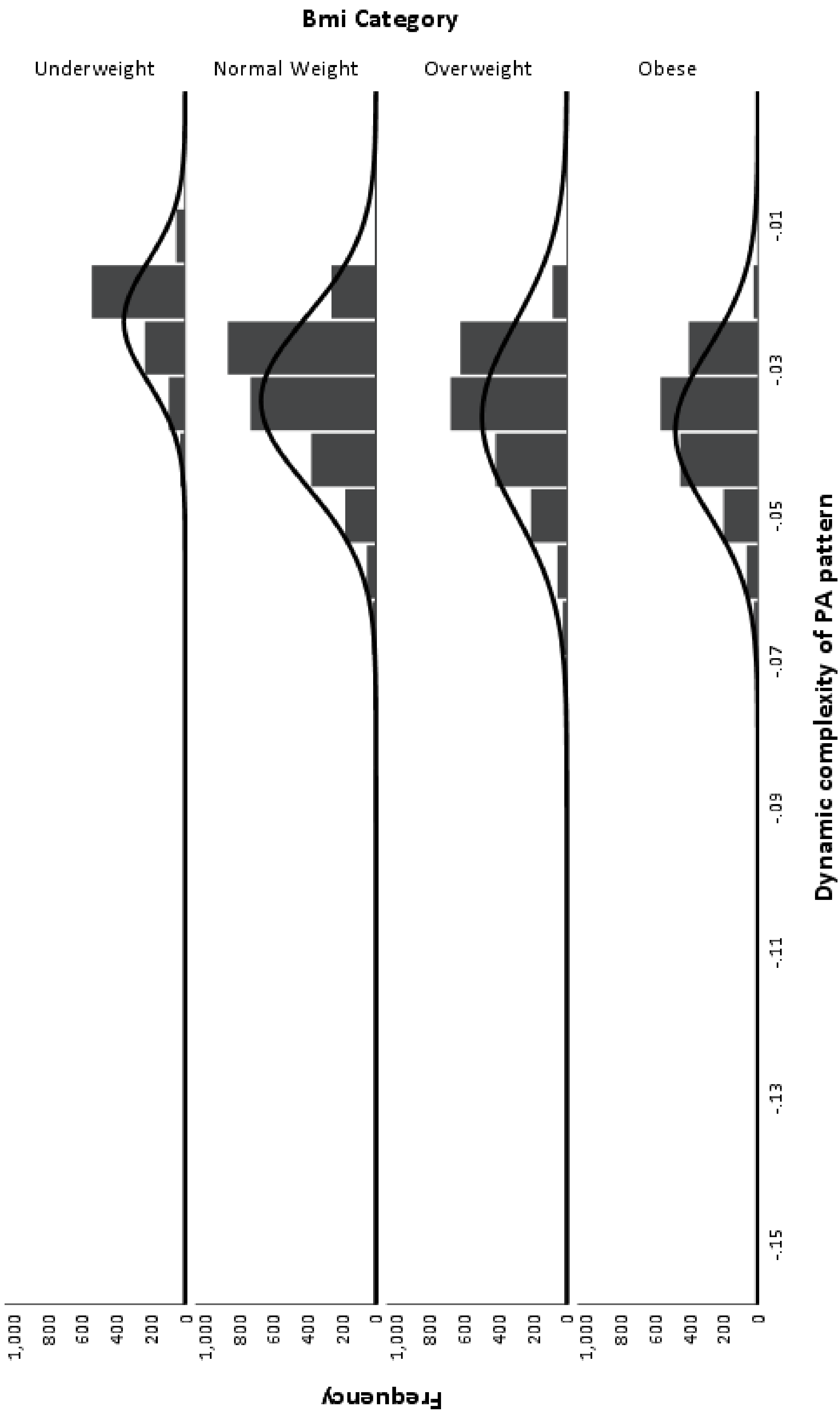

Figure 24. Distribution of dynamic complexity of PA pattern for BMI groups

Further normality tests demonstrated that μ1 - μ3 were not normally distributed: Kolmogorov-Smirnov test reported significance statistics of p=.000, which indicated the data significantly deviated from the normal distribution.

Descriptive statistics for the BMI groups are presented below:

BMI groups		Mean	Std Dev
Total volume of PA μ_1, activity counts	Underweight	837012.69	346273.262
	Normal weight	529428.77	263298.394
	Overweight	459135.78	234057.419
	Obese	409153.67	199144.127
Static complexity of PA pattern μ_2, number of words	Underweight	195.50	52.903
	Normal weight	132.25	38.707
	Overweight	122.23	32.799
	Obese	114.44	28.778
Dynamic complexity of PA pattern μ_3, compression units	Underweight	-0.0233	0.00830
	Normal weight	-0.0342	0.01159
	Overweight	-0.0365	0.01313
	Obese	-0.0383	0.01121

Table 13. Mean and standard deviation for BMI groups, volume of PA and complexity of PA pattern

Both volume of PA and complexity measures of PA pattern appeared to follow three different curves. For children and adolescent there was a rapid decline of volume and complexity of PA pattern with age. Adults appeared to maintain slightly-declining linear level of volume of PA and complexity of PA pattern. Older adults appear to display slow but constant decline of volume and complexity of PA pattern. Based on this, population had been categorised into three age strata: children 6-18 years (n=2448, 12.23±3.61), adults 19 – 55 (n=2811, 37.59±10.74), and older adults 55 – 85 (n=2161, 69.9±8.61).

Scatter plots presented next demonstrate that both volume of PA and complexity of PA pattern change with BMI:

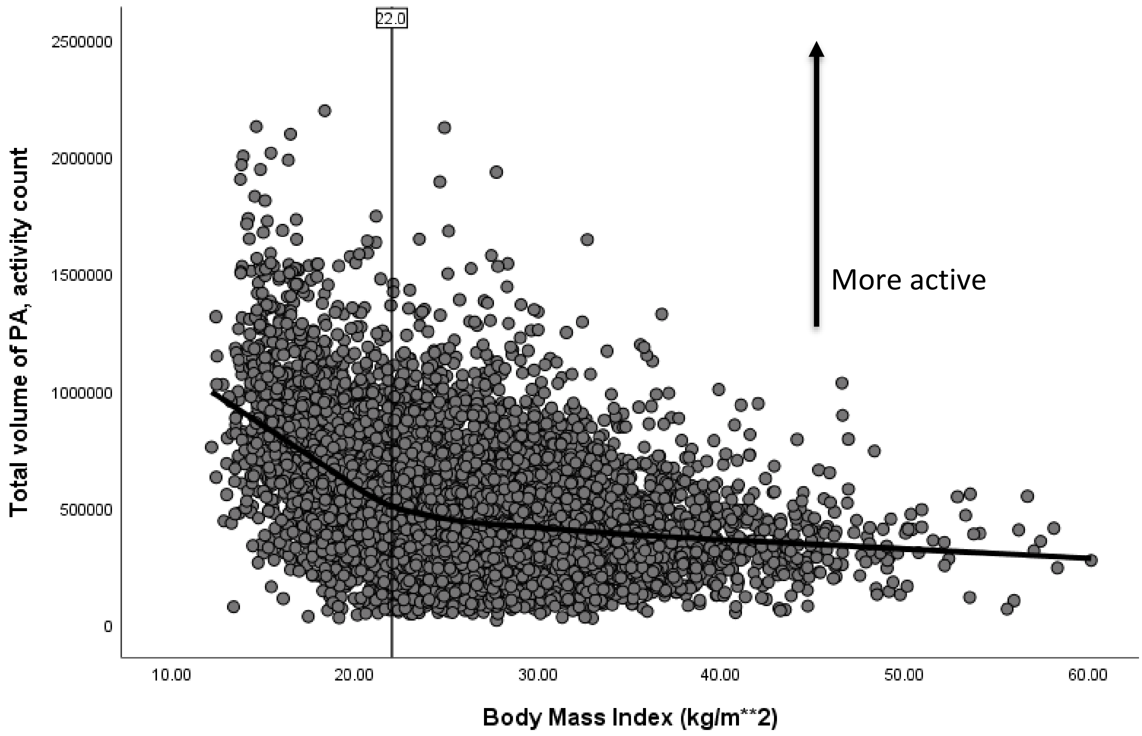

Figure 25. Scatter plot BMI vs volume of PA

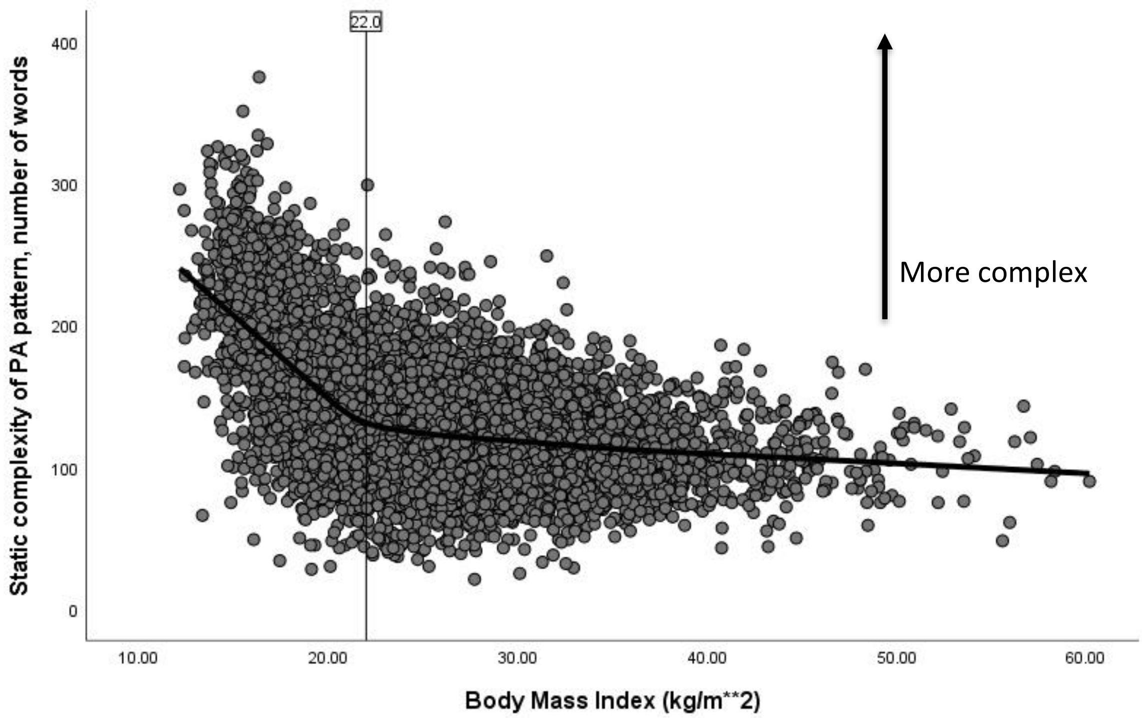

Figure 26. Scatter plot BMI vs static complexity of PA pattern

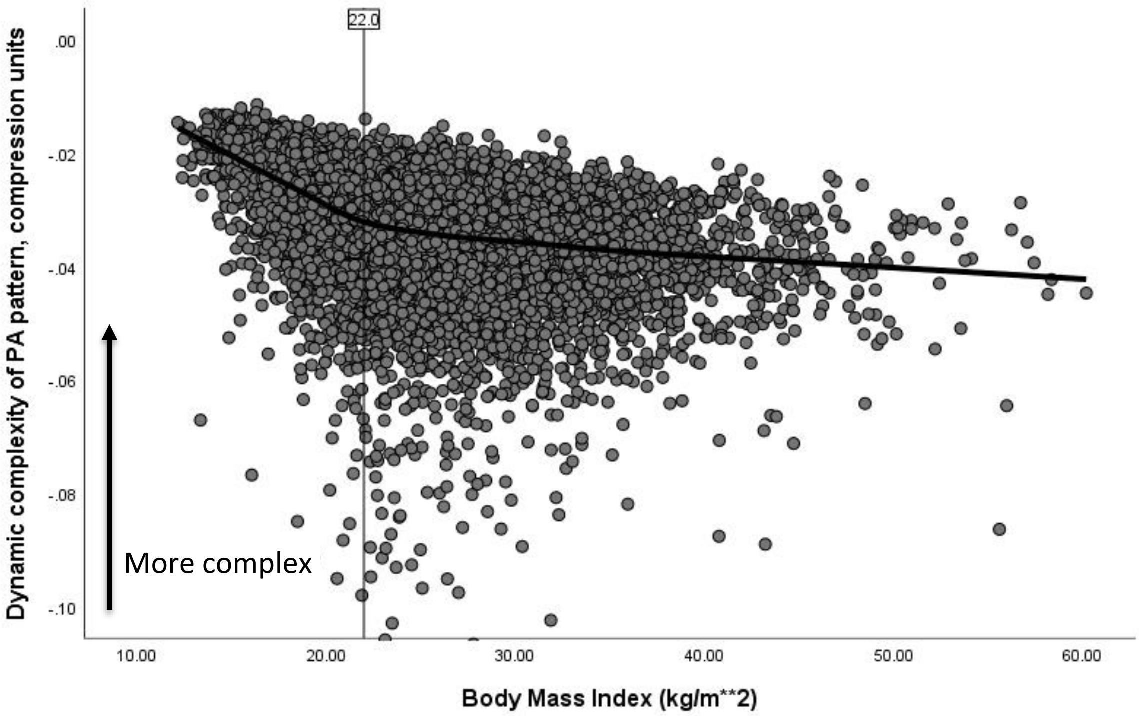

Figure 27. Scatter plot BMI vs dynamic complexity of PA pattern

The total recording time considering only valid wear (in full minutes) was the following: Underweight – 5388±1096 minutes, Normal Weight – 5443±1138 minutes, Overweight – 5503±1144 minutes, and Obese – 5364±1159 minutes.

The volume of PA, static and dynamic complexity of PA pattern decreased with the increase in BMI. For the participants with BMI index 10 – 22 units this decrease was most rapid.

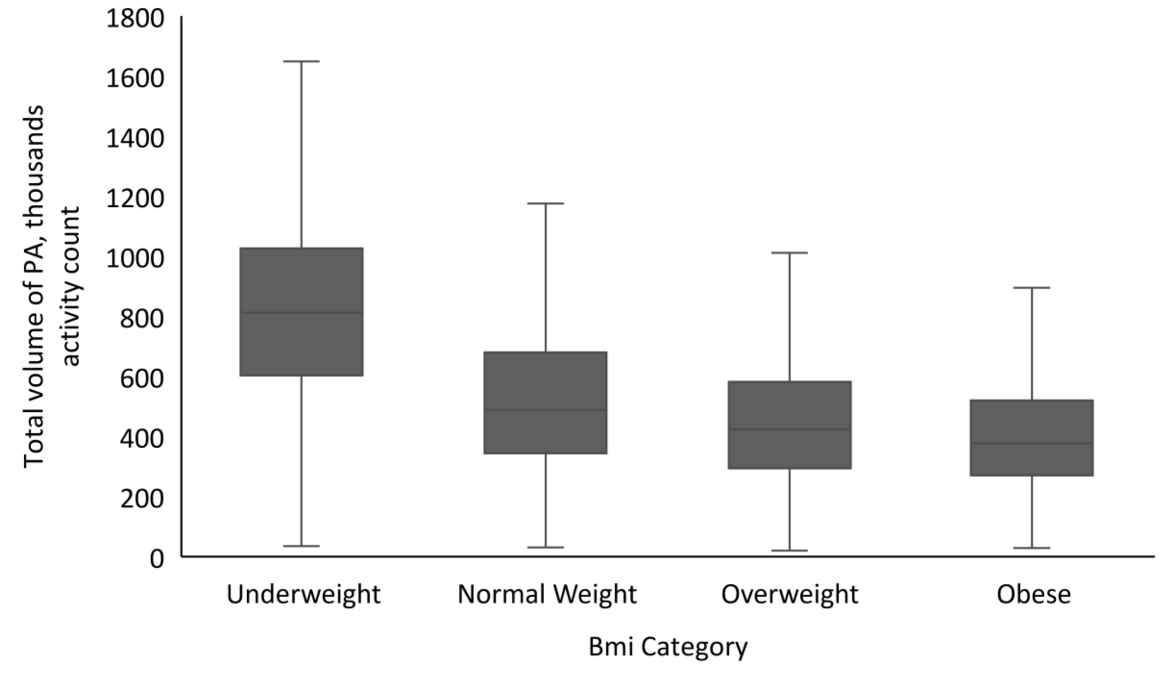

Figure 28. Box plot BMI category vs the volume of PA

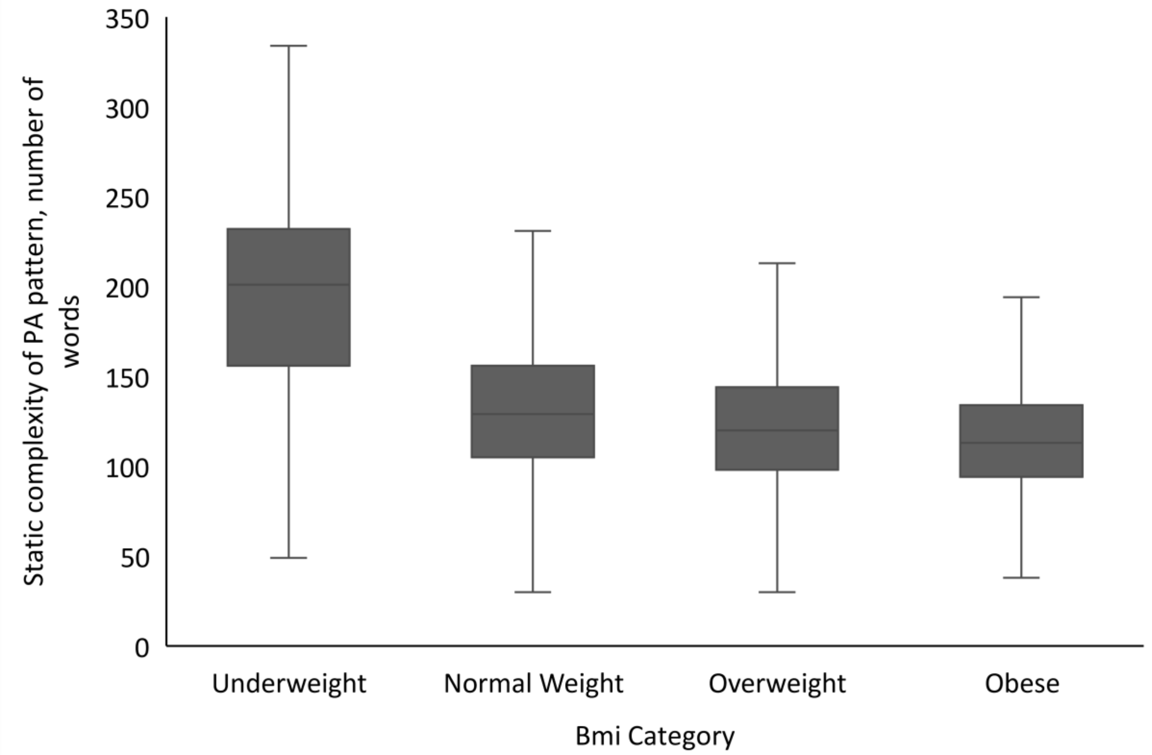

Figure 29. Box plot BMI category vs static complexity of PA pattern

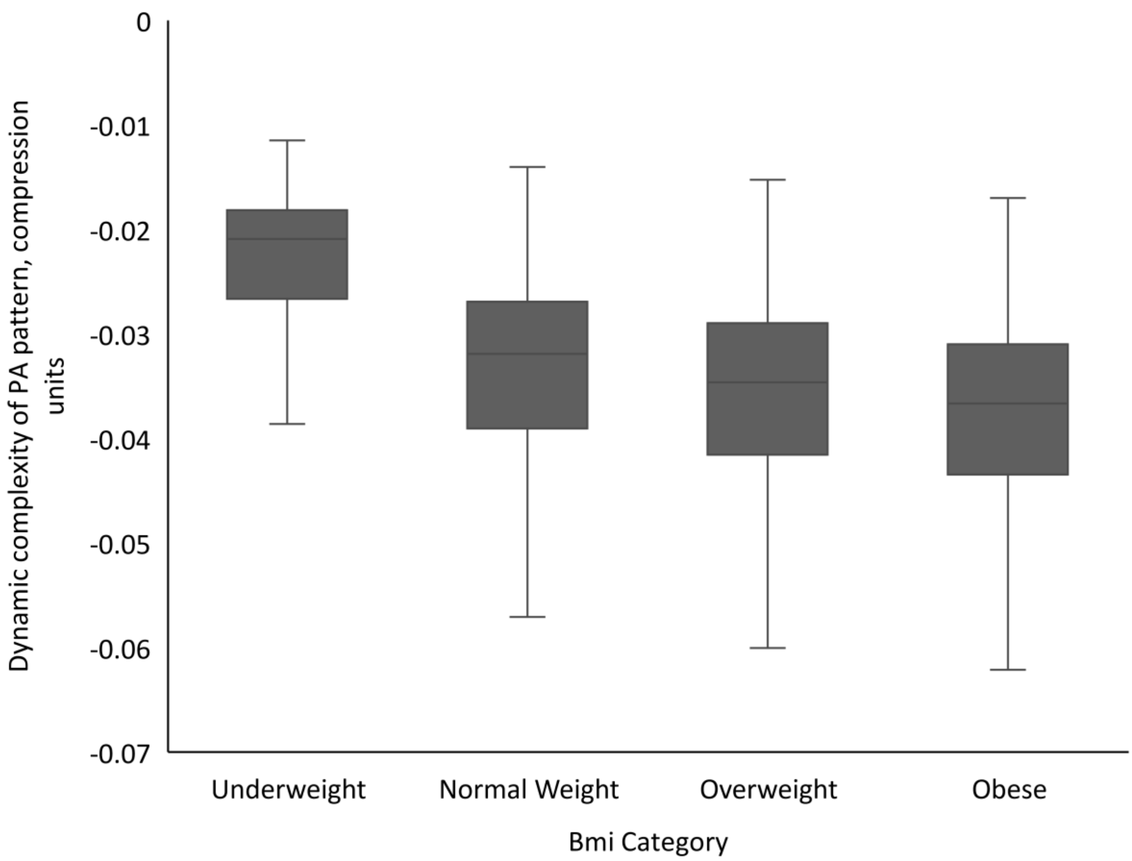

Figure 30. Box plot BMI category vs dynamic complexity of PA pattern

Other correlates of PA, volume and complexity of PA pattern reflecting a person's health

The volume of PA and complexity of PA pattern was statistically different in ethnic groups. as verified using non-parametric Mann-Whitney U test. The test showed that there was a statistically significant difference between Mexican American and non-Hispanic white for volume of PA ($|\mu_1|$=519983, Mann–Whitney U = 2444868, P = 0.000 two-tailed), static complexity of PA pattern ($|\mu_2|$=133.25, Mann–Whitney U = 2436197, P = 0.000 two-tailed) and dynamic complexity of PA pattern ($|\mu_3|$=-0.0345, Mann–Whitney U = 2307361, P = 0.000 two-tailed). Mexican Americans had higher volume of PA and complexity of PA pattern as shown by the rank table below:

Ethnicity		N	Mean Rank
Volume of PA	Mexican American	1987	3009.57
	Non-Hispanic white	3246	2376.69
Static complexity of PA pattern	Mexican American	1987	3013.93
	Non-Hispanic white	3246	2374.02
Dynamic complexity of PA pattern	Mexican American	1987	3078.77
	Non-Hispanic white	3246	2334.33

Table 14. Mann-Whitney U mean rank statistics for Mexican American, non-Hispanic white and measures of the volume of PA, the complexity of PA pattern

Box plots of the relationship between ethnicity and measures of the volume of PA and complexity of PA pattern are presented below:

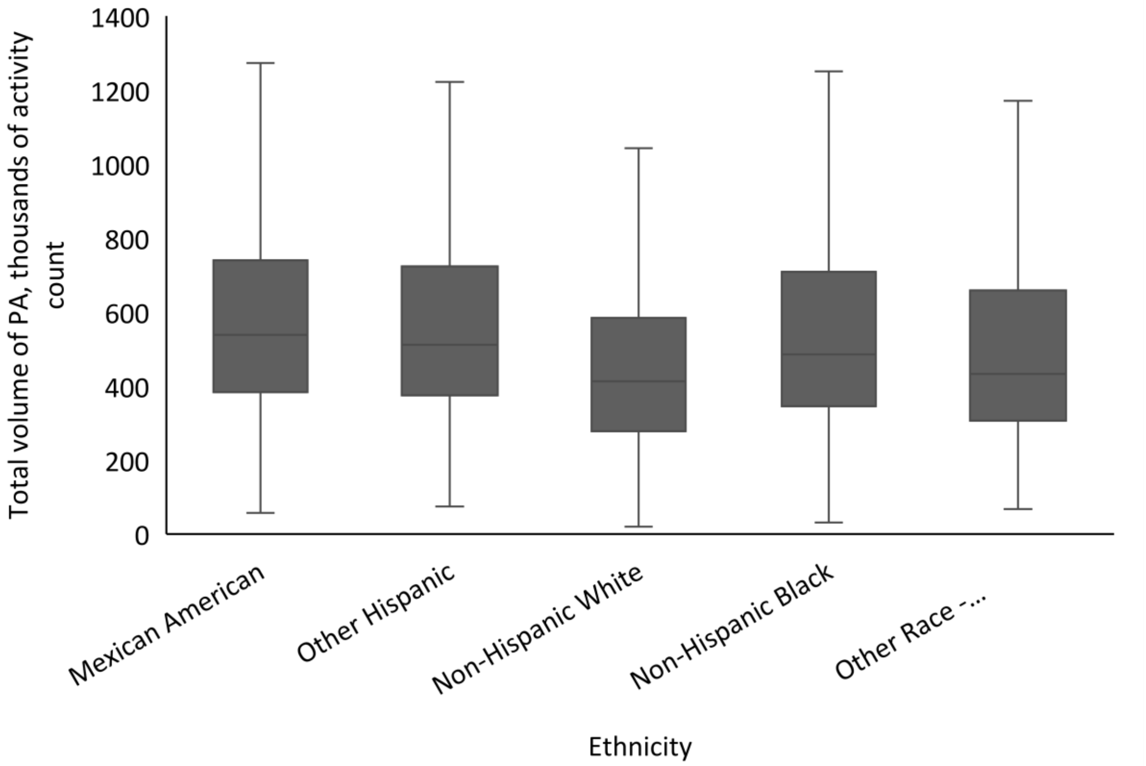

Figure 31. Box plot ethnicity vs volume of PA

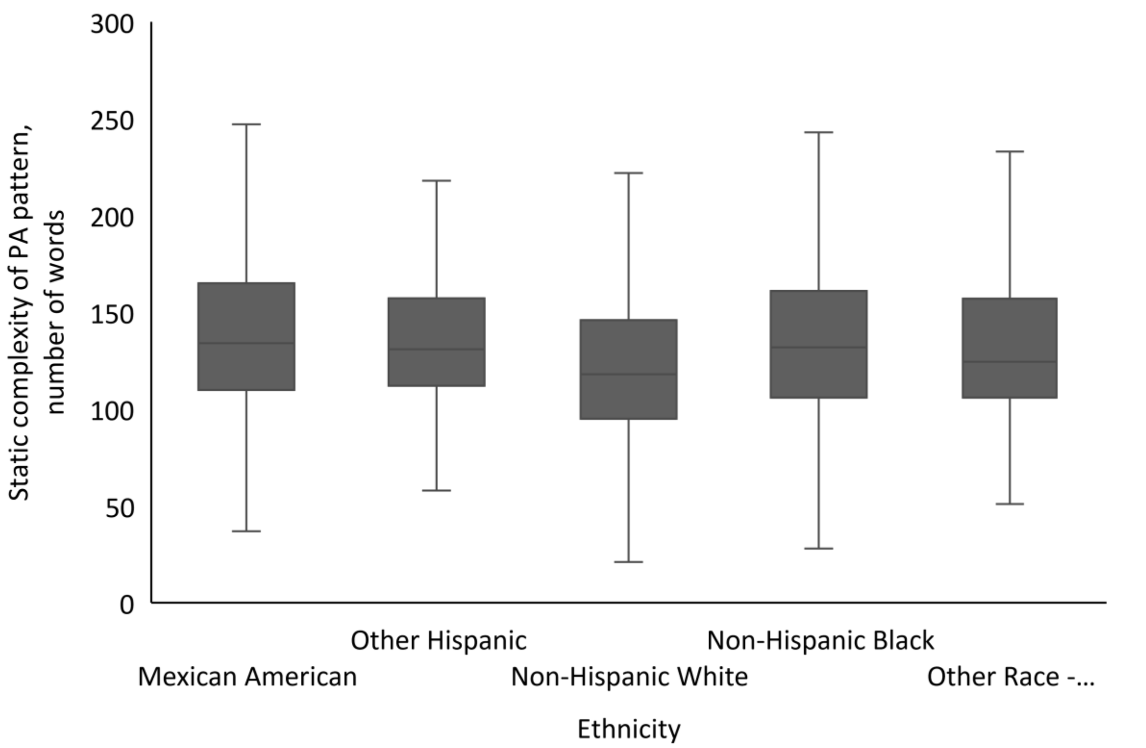

Figure 32. Box plot ethnicity vs static complexity of PA pattern

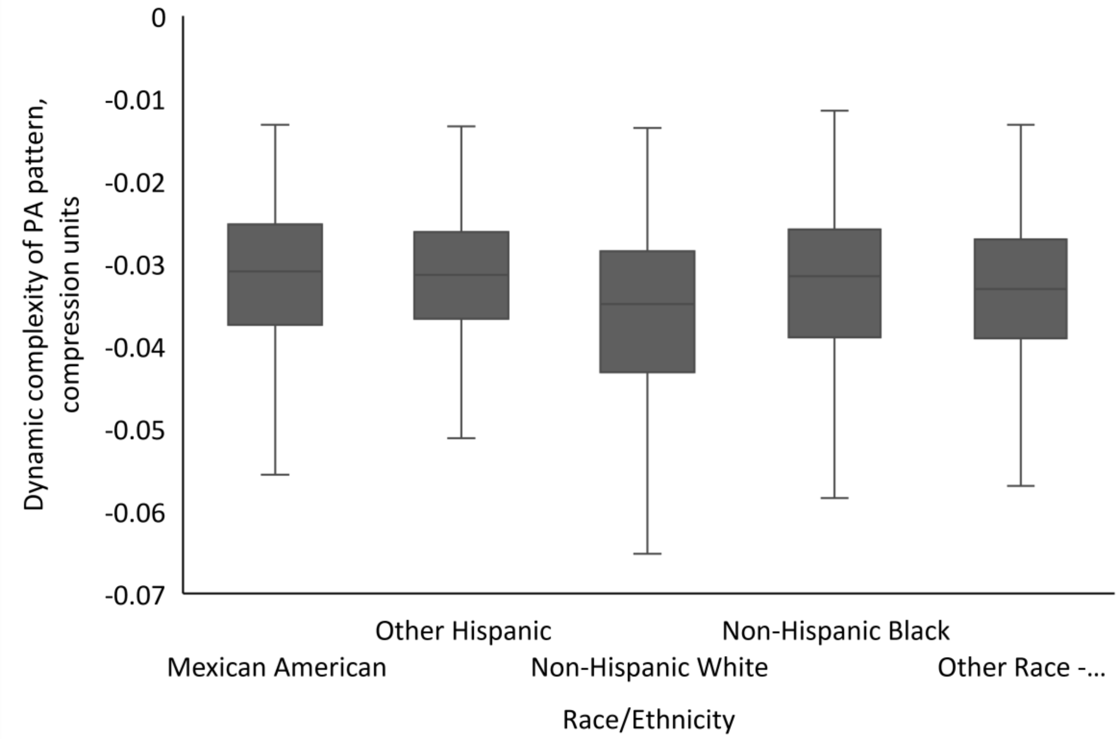

Figure 33. Box plot ethnicity vs dynamic complexity of PA pattern

Self-reported health condition showed a negative correlation with volume of PA (Pearson r = -.233, n = 5698, p = .000), static complexity of PA pattern (Pearson r = -.240, n = 5698, p = .000) and dynamic complexity of PA pattern (Pearson r = -.181, n = 5698, p = .000). A better self-rating of health condition corresponded to higher volume and complexity of PA pattern (in a questionnaire, an increase in the numeric value corresponded to poorer health). The complexity of PA pattern showed a stronger relationship than volume of PA as indicated by the *r* coefficient.

Box plots of the relationship between self-reported health and measures of volume of PA and complexity of PA pattern are presented below:

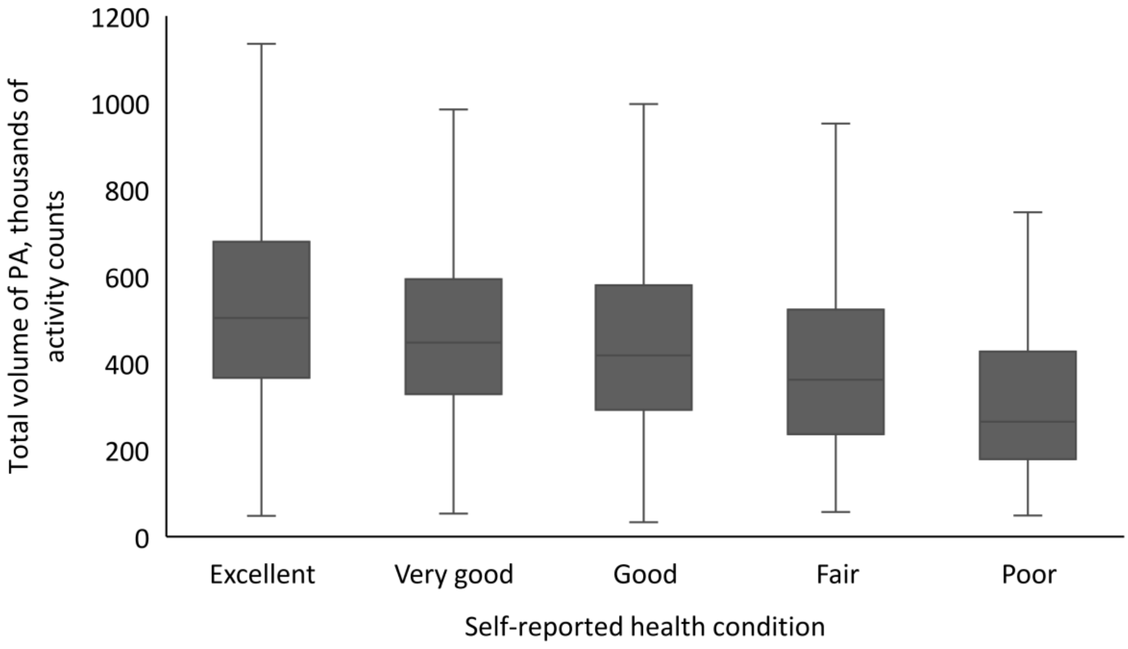

Figure 34. Self-reported health condition vs volume of PA

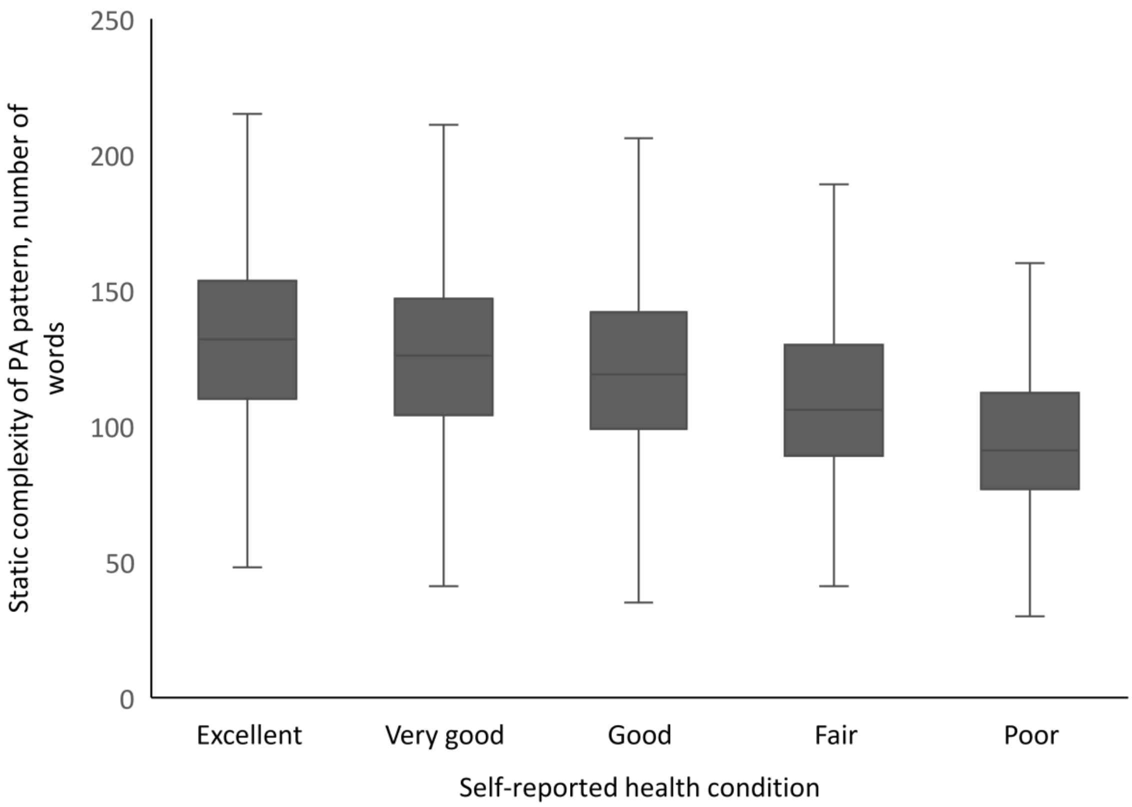

Figure 35. Self-reported health condition vs static complexity of PA pattern

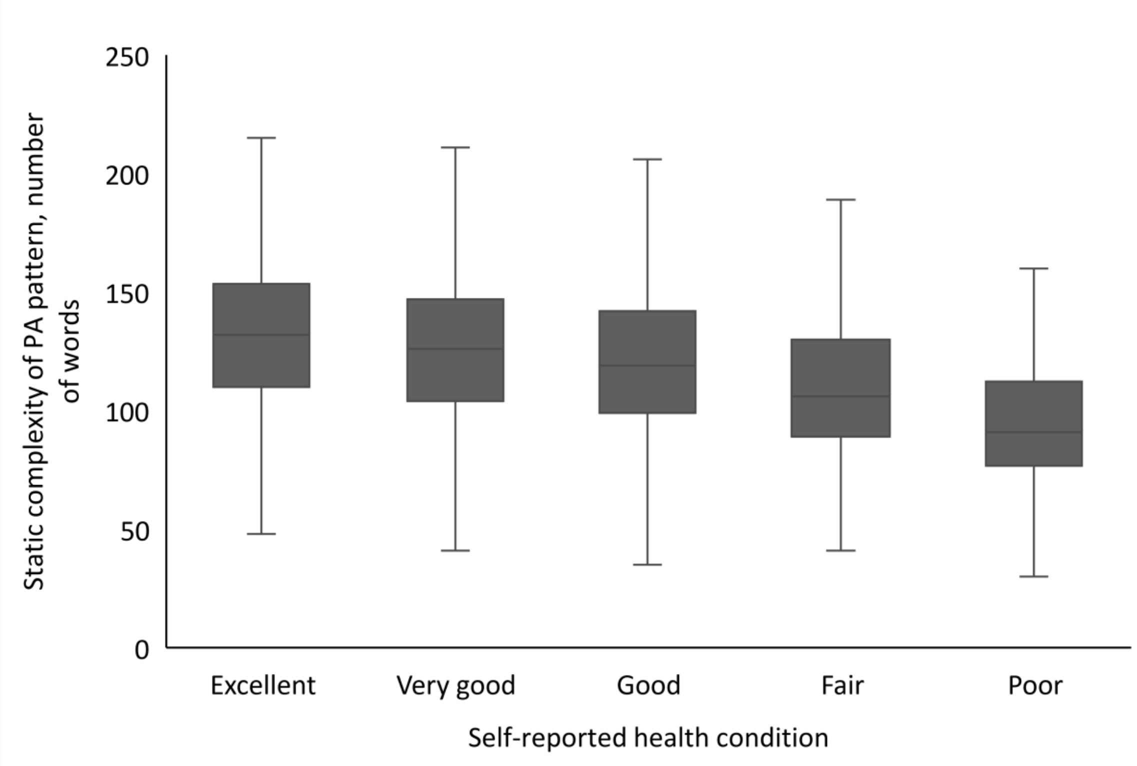

Figure 36. Self-reported health condition vs dynamic complexity of PA pattern

The volume of PA and complexity of PA pattern

Probability distribution of μ_1 - μ_3 for the full population is graphed in Figure 37 – Figure 39.

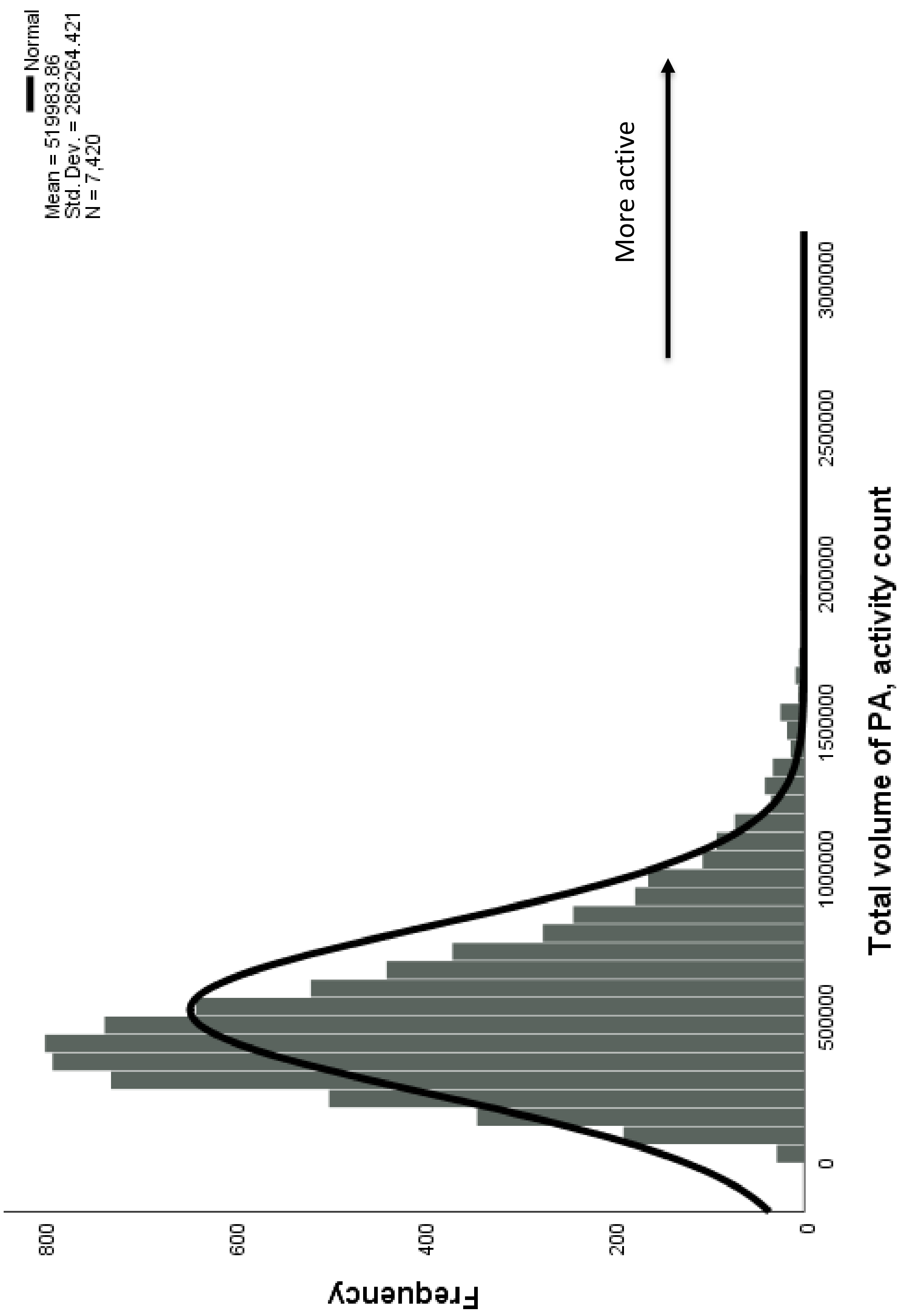

Figure 37. The volume of PA for the full population sample

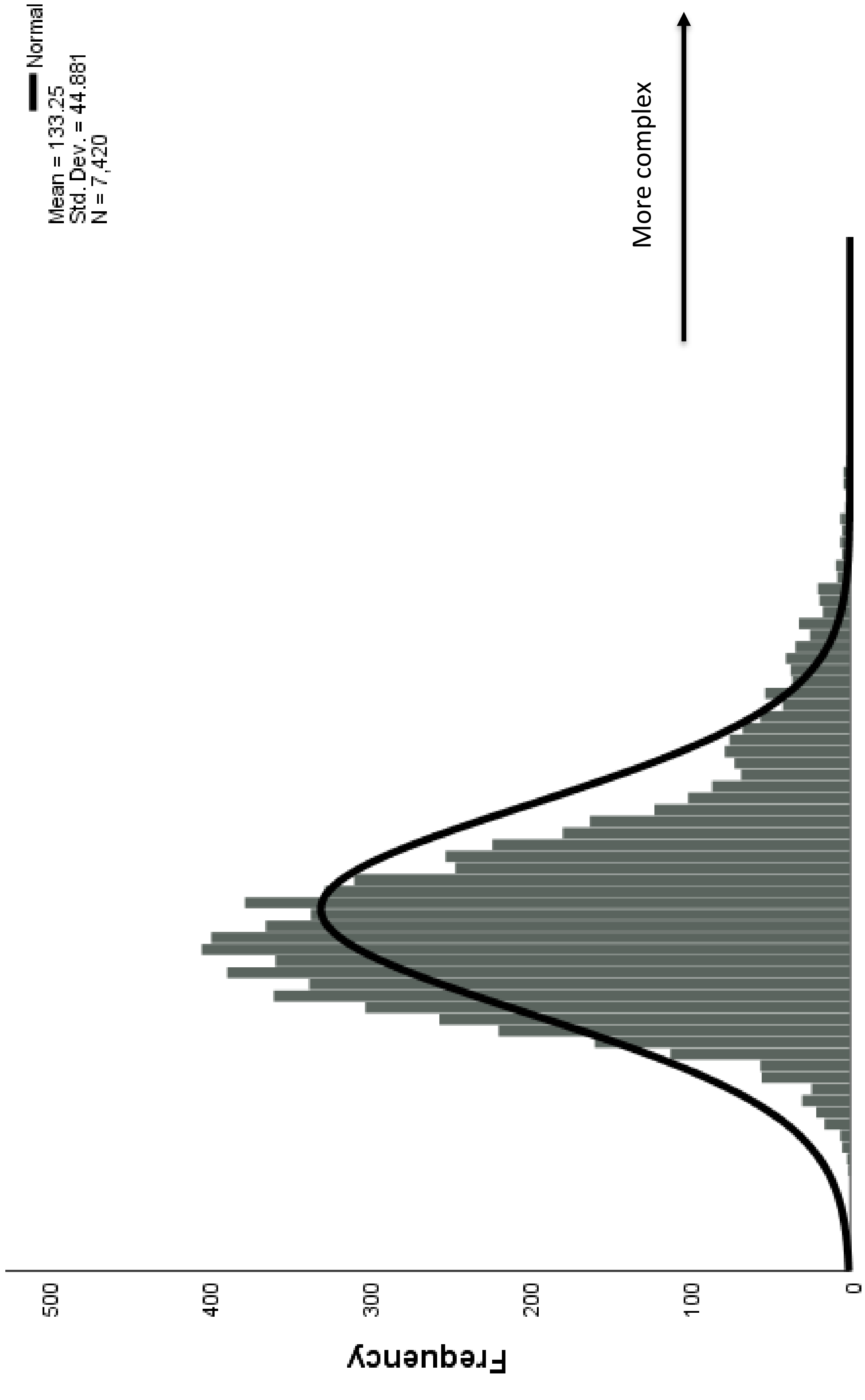

Figure 38. Static complexity of PA pattern for the full population sample

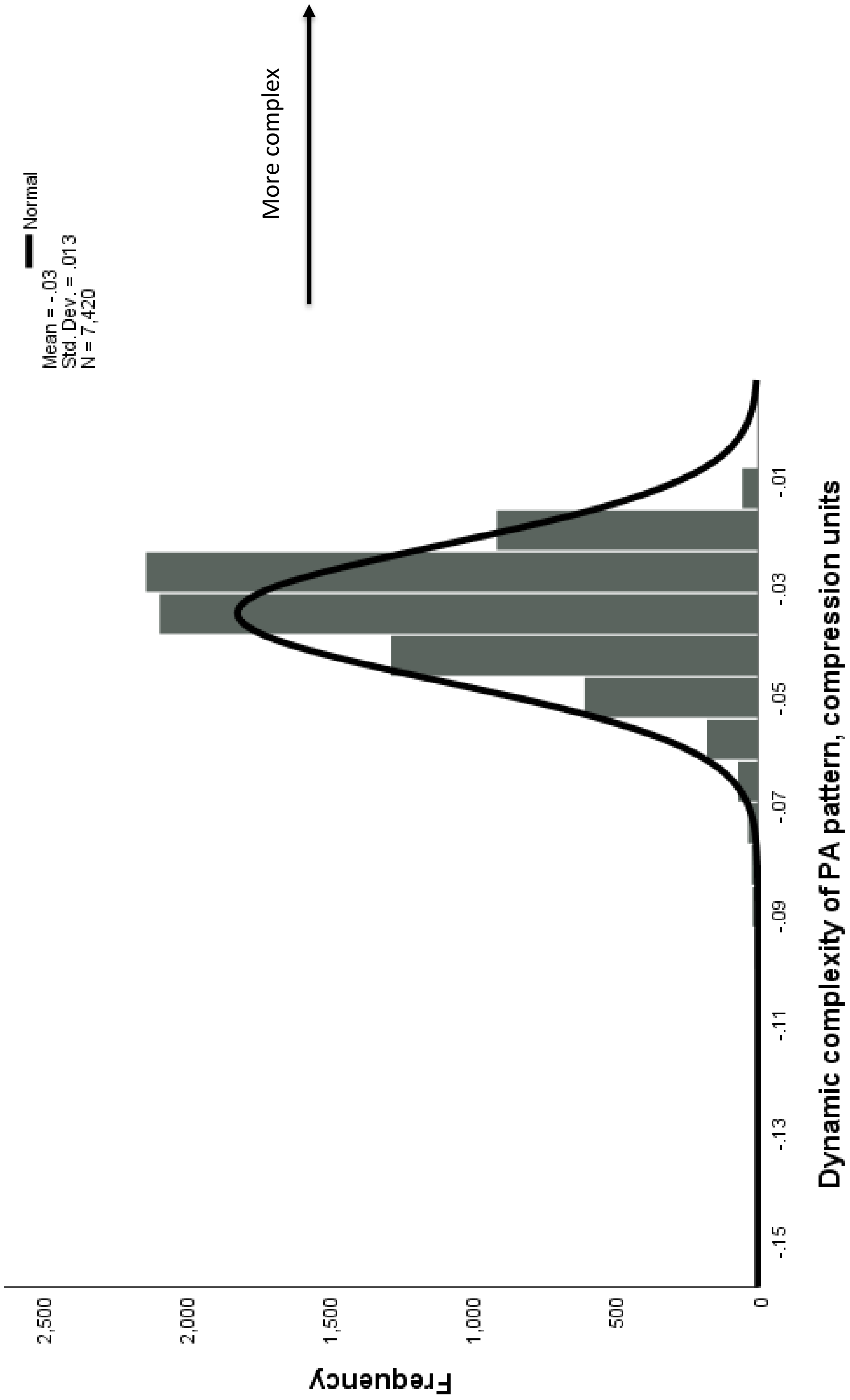

Figure 39. Dynamic complexity of PA pattern for the full population sample

μ_1 - μ_3 appeared to be near-normally distributed. A Kolmogorov-Smirnov test was used to test for normality – μ_1: D(7420) = 0.087, p = .000, μ_2: D(7420) = 0.077, p = .000, μ_3:

D(7420) = 0.078, p = .000. Significance of less than p = .05 highlighted that data was not normally distributed.

Static complexity of PA showed strong positive correlation with volume of PA (Pearson r = .842, n = 7420, p = .000). It appeared to be a good proxy measure of the volume of activity. An increase in the volume of PA corresponds to higher static complexity of PA pattern.

Clearly, there is a quite linear relationship, so static complexity is another measure of volume. It does not appear to carry extra information except at low volume where static complexity climbs up very quickly for small increases in the volume of PA. It probably makes it a more sensitive measure of the behaviour change at a low volume of PA, so in the most inactive population. This probably explains why the models demonstrated a slightly better fit for the complexity of PA pattern, considering this is most of the population.

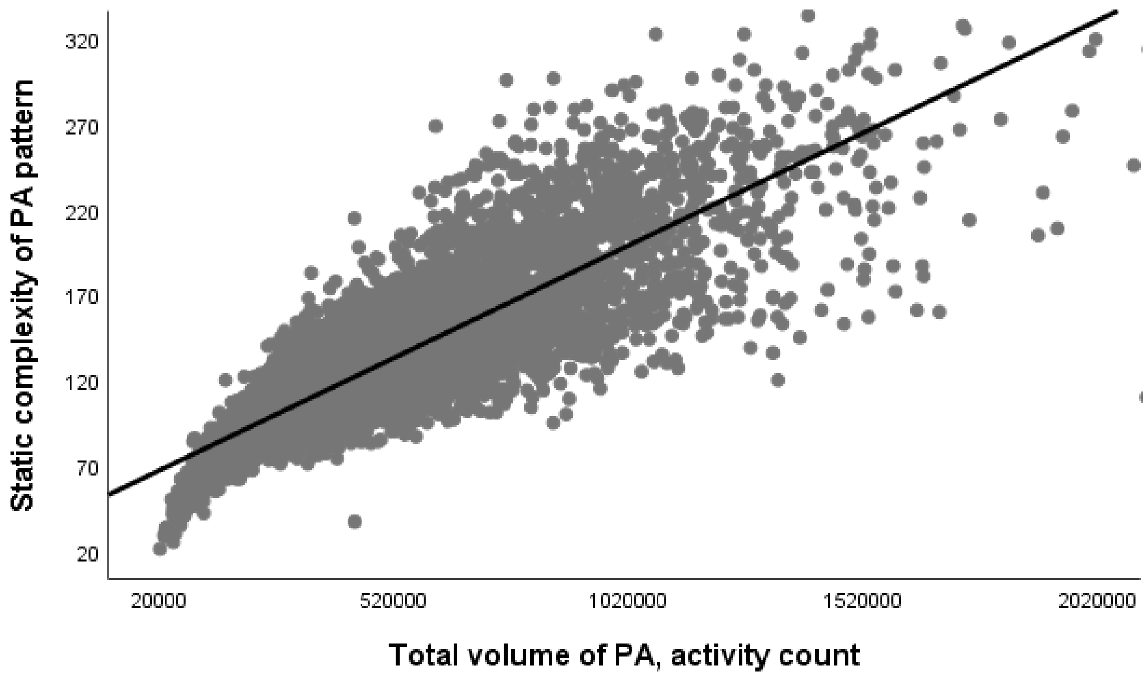

Figure 40. Scatter plot volume of PA vs static complexity of PA pattern

Static complexity of a PA pattern is the size of PA vocabulary. It was interesting to understand what kind of words were used by the participants. On average, the most common word was "0000" which signified that studied population was mostly inactive. The next several words with the highest probability of occurrence were transitions or some form of exercises. An example graph showing the top 10 words is presented in a figure below:

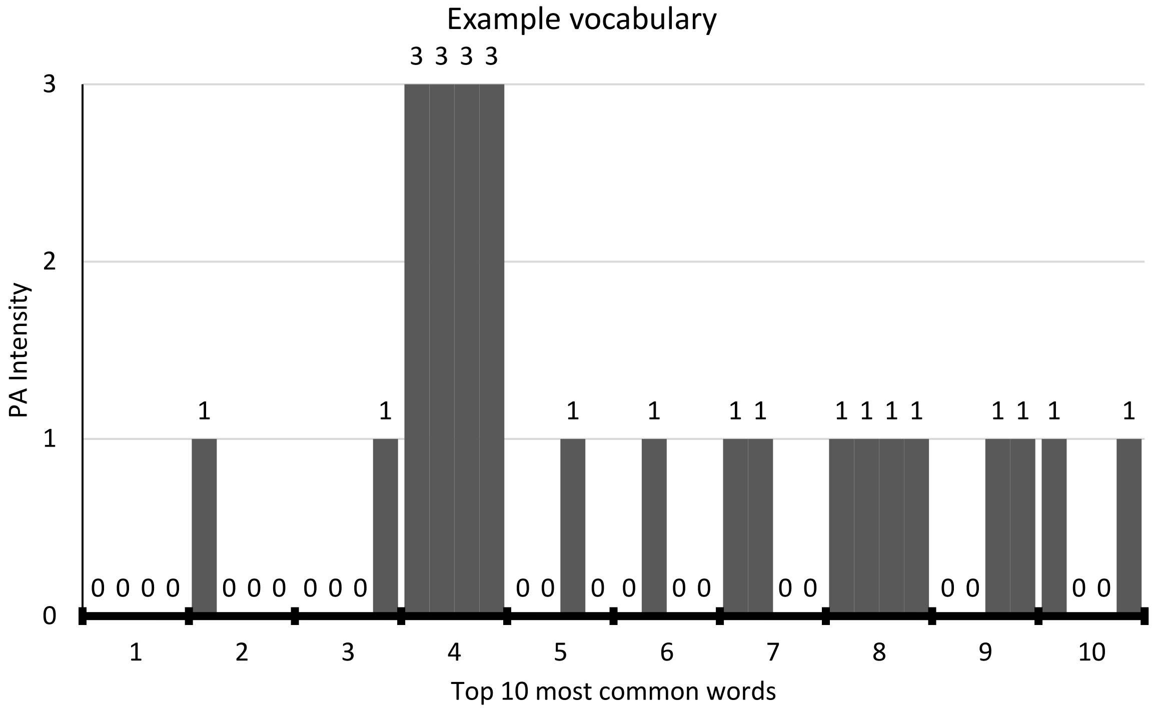

Figure 41. Top 10 words out of 297 in an example vocabulary

The probability distribution of using other words rapidly declined as shown on the next figure

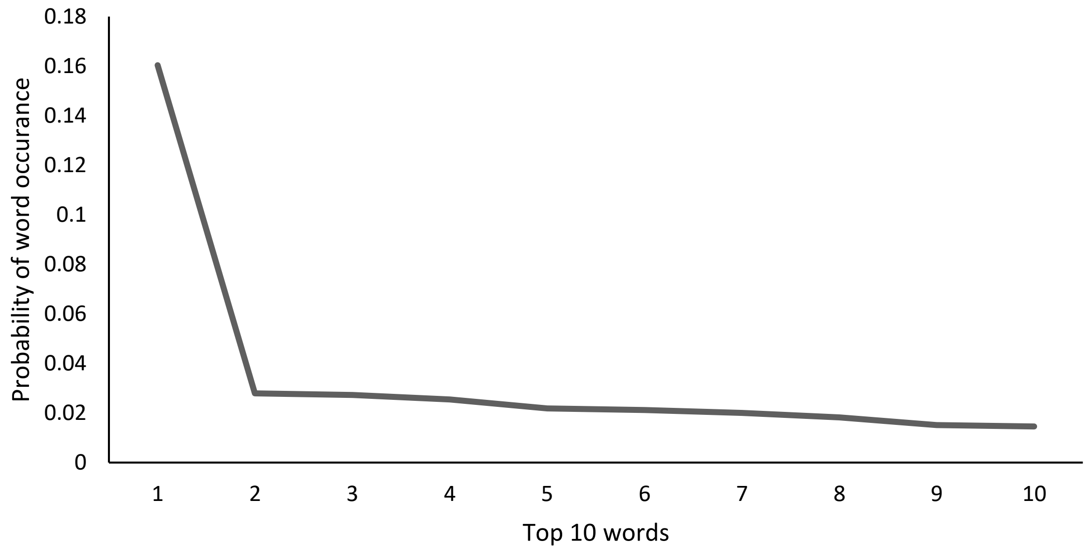

Figure 42. Probability distribution of top 10 words in a vocabulary for subject id 22013 (NHANES 2003-2004)

To see the vocabulary of full population or stratified by age or BMI category please see the Appendices 5 – 12. Physical activity vocabulary was calculated on a 4-minute interval for subjects from NHANES 2003-2004 and 2005-2006.

Dynamic complexity of PA also correlated with PA pattern. The relationship appeared to be logarithmic. Curve estimation regression computed for linear model (R^2 = .492, k = -.05, β = 3.070^{-8}, n = 7420, p = .000) and logarithmic model (R^2 = .704, k = -.269, β = .018, n = 7420, p = .000) demonstrated that the variance was explained better by the logarithmic relationship (see Figure 43).

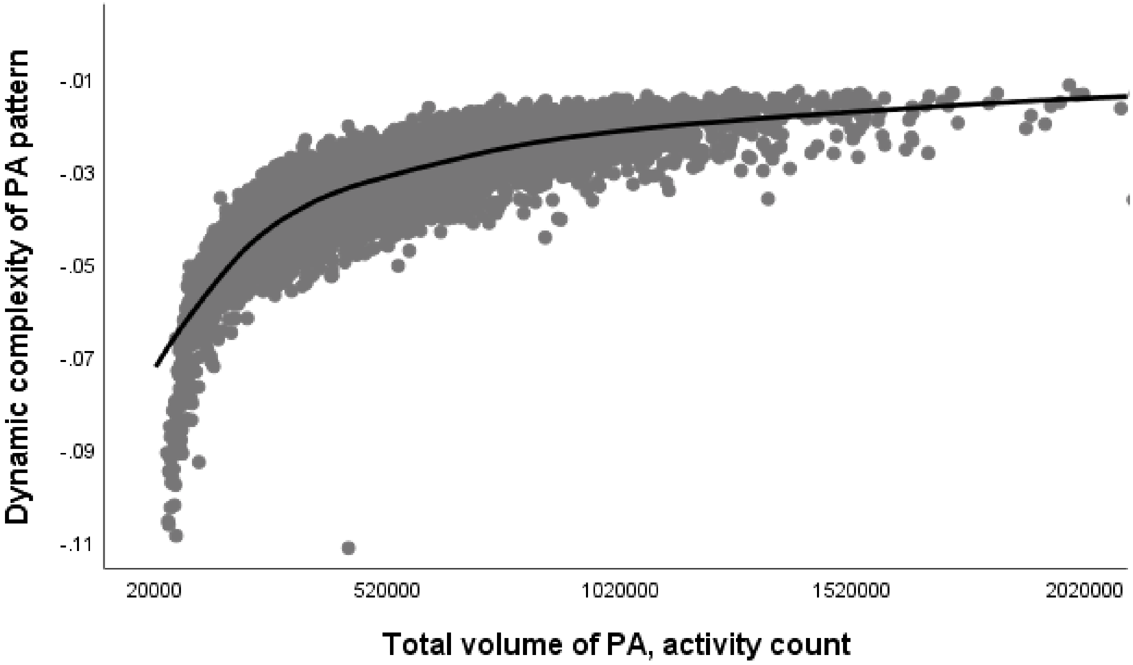

Figure 43. Scatter plot volume of PA vs dynamic complexity of PA pattern

Similarly, dynamic complexity increases rapidly (vertical direction from -1 to 0), when the volume of PA was low, so for the majority of the population who does very little the dynamic complexity is a sensitive measure as the complexity increases, but not the volume.

Between static and dynamic complexity of PA pattern there appeared to be a positive logarithmic relationship. Curve estimation regression computed for the linear model (R^2 = .681, k = -.065, n = 7420, p = .000) and logarithmic model (R^2 = .843, k = -.202, β = .035, n = 7420, p = .000) demonstrated that variance was explained better by the logarithmic relationship (see Figure 44).

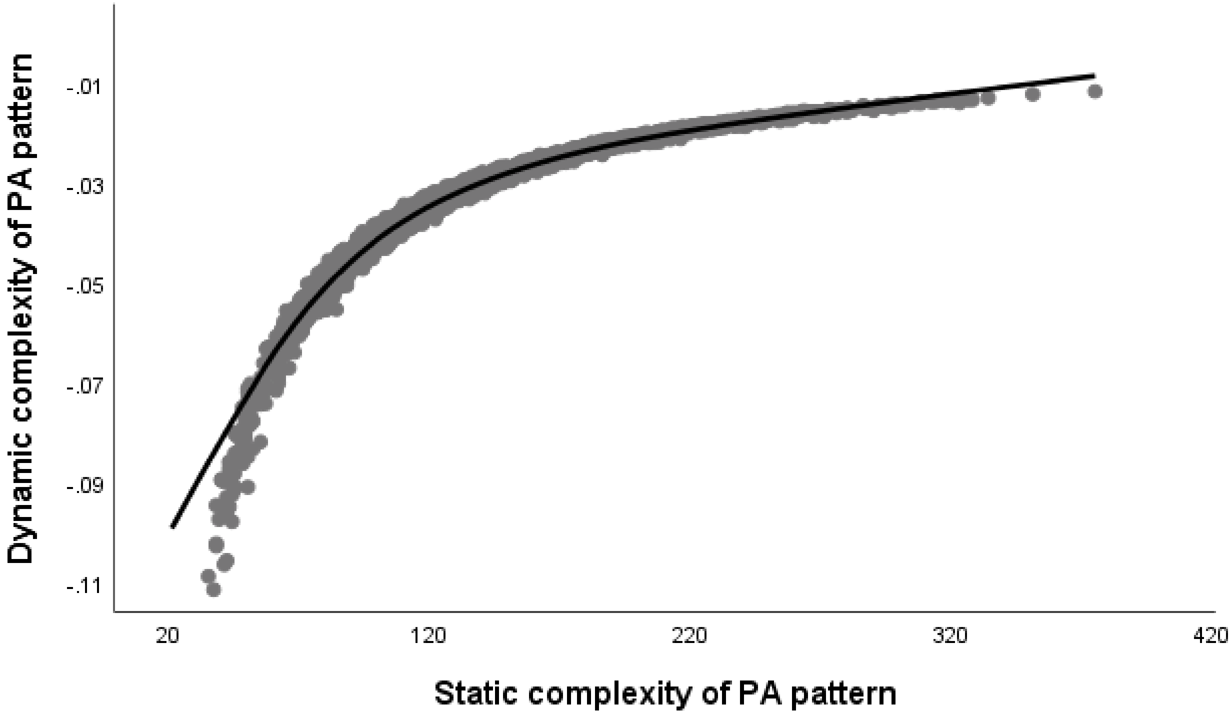

Figure 44. Scatter plot static vs dynamic complexity of PA pattern

Data for the first 100 individuals with the least complex PA pattern showed static complexity in ranges 21-57 (words), whereas population mean value for static complexity was 133.25 (words) and standard deviation was 44.881 (words). The actual words for these participants represented inactivity or low energy expenditure transitions. Therefore, participants with low static complexity exhibited sedentary behaviour.

For a more significant increase of the static complexity of PA pattern, there is small and ever decreasing climb of dynamic complexity. So, the diversity in PA vocabulary, as represented by the static complexity of PA, does not appear to consistently increase the dynamic complexity of PA pattern.

Measures of PA pattern in statistical regression

General Linear Modelling (GLM) was used to establish if and how many independent variables, such as age, BMI, gender, and others could explain dependent variables, such as measures of the volume of PA μ_1, static complexity of PA pattern μ_2, and dynamic complexity of PA pattern μ_3.

Independent variables constituted both quantitative and qualitative variables. There was a single quantitative covariate - age of participants. The other independent variables were not numeric, but rather factors with multiple categories, such as BMI, gender, household income, and so forth. Because of the mix of quantitative, ordinal, and categorical independent variables, ANCOVA GLM tests were used.

It was noted that some of the covariates for the ANCOVA analysis did not necessarily follow a normal distribution. For this reason, it could be difficult to interpret the results, especially when the deviation from the normal distribution was high and the population pool was low. This is why after the ANCOVA analysis was performed, the residuals were examined for signs of non-normality. Some signs of the non-normal distributions were noticed: a small lack of symmetry combined with a heavier-tail in some cases. The normality test was conducted to detect statistically significant, but trivial deviations from normality checking if there could be a real effect on the analysis of covariance. As the number of participants was considerably high (n=7200) it was determined that the ANCOVA results should not be much affected by the small departures from the normal distribution.

Some independent variables were correlated with each other. Therefore, a non-orthogonal statistical ANCOVA design was utilised. A Spearman's rank-order correlation was run to confirm the correlation between self-reported long-term history of PA and short-term history of PA (n = 7240). There was a positive correlation between self-reported long-term and short-term history of PA ($r_s = .143, p = .000$).

Insignificant relationships were present and excluded from the report below. The statistical tolerance threshold was set at the standard $p = .05$ level. Unstandardized β coefficients were reported to investigate the impact of each independent variable on the dependent variable.

Stage 1. Stratified forward-selection ANCOVA

In stage 1, measures $\mu_1 - \mu_3$ were entered as dependent variables to the same ANCOVA models using forward regression. Independent variables used in the model were listed in the table above.

Children and adolescents

Variance explained

1112 children and adolescents aged 6 – 18 y.o. were selected in this analysis. Independent variables explained: 16% of variance of volume of PA μ_1 (R^2 adj = 0.162); 24% of variance of static complexity of PA pattern μ_2 (R^2 adj = 0.243); and 26% of variance of dynamic complexity of PA pattern μ_3 (R^2 adj = 0.257). The variance explained by independent variables is summarised in the table below:

	μ_1	μ_2	μ_3
N	1112	1112	1112
R^2	.172	.253	.267
R^2 adj	.162	.243	.257

Table 15. Variance explained by forward selection ANCOVA in children

Common independent variables

ANCOVA models $\mu_1 - \mu_3$ had four common statistically significant correlates: age, self-reported history of physical activity last month vs last year, comparison of PA level with others and vitamin D. In all models, age was negatively associated with the dependent variable, while self-reported history of physical activity last month vs last year, comparison of PA level with others, and vitamin D were positively associated. Age was linked to lower volume of PA (β = -.23), lower static complexity of PA pattern (β = -.44), and lower dynamic complexity of PA pattern (β = -.46). Self-reported history of PA last month vs last year was positively correlated with μ_1 - μ_3, “more active” responses resulted in higher volume of PA and complexity of PA pattern (μ_1: β=.101, μ_2: β=.1, μ_3: β=.107). Similarly, comparison of the self-reported PA vs others positively correlated with volume of PA μ_1 (β = .165), static complexity of PA pattern μ_2 (β = .129), and dynamic complexity of PA pattern μ_3 (β = .13).

Higher vitamin D level was linked to higher volume of PA μ_1 (β = .162), static complexity of PA pattern μ_2 (β = .069), and dynamic complexity of PA pattern μ_3 (β = .066).

Measure-specific independent variables

Measure-specific unique correlates were reported next. Number of people living in a household had a small positive correlation with volume of PA (β = .084). BMI and education level correlated with static and dynamic complexity of PA pattern, but not with volume of PA. Higher BMI corresponded to lower static complexity of PA pattern (β = -.11) and lower dynamic complexity of PA pattern (β = -.10). Higher education level was linked to higher static complexity (β = .174) and dynamic complexity (β = .201) of PA pattern. Disease burden positively correlated with dynamic complexity of PA pattern (β = .054), an increase in the number of diseases corresponded to an increase in dynamic complexity of PA pattern.

Summary results

Unstandardised ANCOVA coefficients and significance level were presented in the table below:

N=1112		Dependent variable					
	Independent variables	μ_1		μ_2		μ_3	
		β	*p*	β	*p*	β	*p*
Common	Age	-.23	.000	-.44	.000	-.46	.000
	History PA last m-last y	.101	.000	.1	.000	.107	.000
	Compare act with others	.165	.000	.129	.000	.13	.000
	Vitamin D	.162	.000	.069	.014	.066	.018
Unique	N people in household	.084	.004	-	*NS*	-	*NS*
	BMI	-	*NS*	-.11	.000	-.10	.000
	Education	-	*NS*	.174	.000	.201	.000
	Disease burden	-	*NS*	-	*NS*	.054	.041

Table 16. Unstandardised ANCOVA coefficients in forward selection ANCOVA for children (6 – 18 y.o.)

Adults

Variance explained

667 adults aged 19 – 55 y.o. were selected in this analysis. Independent variables explained: 19% of variance of volume of PA μ_1 (R^2 adj = 0.193); 14% of variance of static complexity of PA pattern μ_2 (R^2 adj = 0.143); and 15% of variance of dynamic complexity of PA pattern μ_3 (R^2 adj = 0.151). The variance explained by independent variables were summarised in the table below:

	μ_1	μ_2	μ_3
N	667	667	667
R^2	.220	.171	.179
R^2 adj	.193	.143	.151

Table 17. Variance explained by forward selection ANCOVA in adults

Common independent variables

ANCOVA models μ_1 - μ_3 had four common statistically significant correlates: age, BMI, comparison of levels of self-activity with others, and vitamin D. In all models, age was negatively associated with the dependent variable. Age was linked to lower volume of PA (β = -.14), lower static complexity PA pattern (β = -.11), and to lower dynamic complexity of PA pattern (β = -.10). BMI was negatively corelated with the dependent variables – higher level of BMI corresponded to lower volume of PA and less complex pattern of PA. BMI was linked to lower volume of PA (β = -.12), lower static complexity PA pattern (β = -.14), and to lower dynamic complexity of PA pattern (β = -.14). Comparison of levels of self-activity with others was positively correlated with μ_1 – μ_3, "more active" responses resulted in higher volume of PA and complexity of PA pattern (μ_1: β=.162, μ_2: β=.111, μ_3: β=.138). Higher vitamin D level was linked to higher volume of PA μ_1 (β = .098), static complexity of PA pattern μ_2 (β = .102), and dynamic complexity of PA pattern μ_3 (β = .100).

Measure-specific independent variables

Measure-specific unique correlates are reported next. Disease burden, gender, and alcohol consumption correlated with static and dynamic complexity of PA pattern, but not with volume of PA. There was association between ethnicity and volume, and ethnicity and dynamic complexity of PA. Annual household income, education level and self-reported history of PA last month vs last year correlated with volume of PA.

Higher levels of disease burden corresponded to lower static complexity of PA pattern (β = -.09) and lower dynamic complexity of PA pattern (β = -.09). Male gender was linked to lower static complexity of PA pattern (β = -.10) and lower dynamic complexity of PA pattern (β = -.08). Higher levels of alcohol consumption were linked to higher static complexity of PA pattern (β = .080) and higher dynamic complexity of PA pattern (β = .092). Number of people living in a household had a small positive correlation with volume of PA (β = .084).

Mexican Americans, on average, had higher level of complexity of PA pattern, and higher level of volume of PA than non-Hispanic white. Further bivariate correlation analysis showed that all three metrics were correlated with ethnicity: volume of PA (r_s = -0.159, P = 0.000), static complexity of PA pattern (r_s = -0.07, P = 0.000), and dynamic complexity of PA pattern (r_s = -0.073, P = 0.000). Bar plots showing marginal mean values for $\mu_1 - \mu_3$ are presented below.

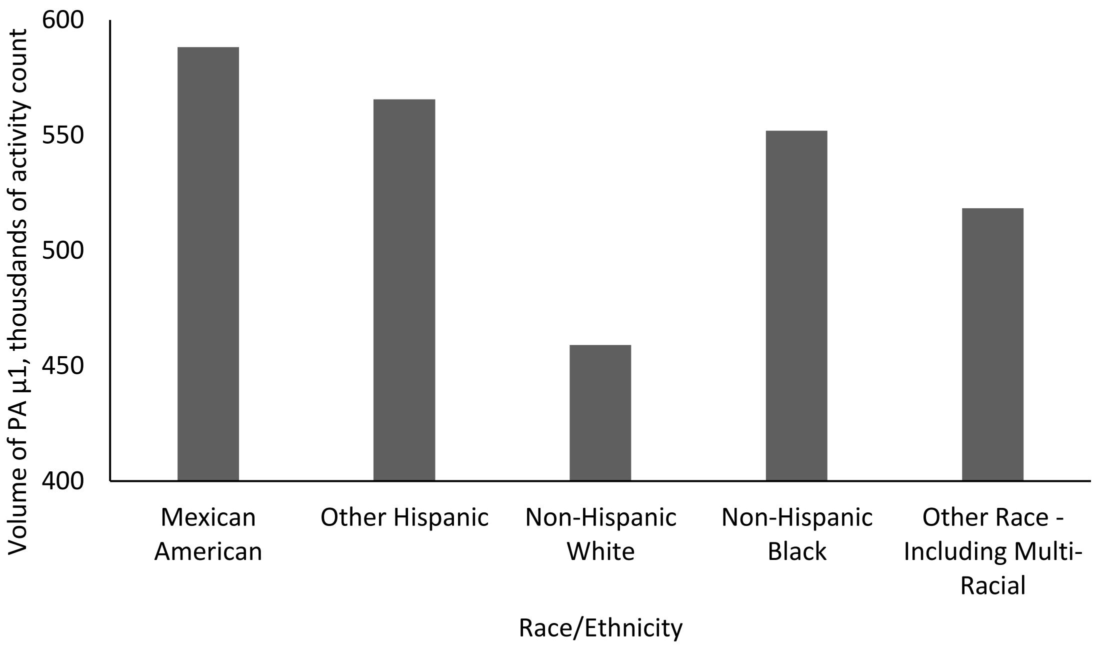

Figure 45. Marginal means volume of PA μ_1 for adults (19 – 55 y.o.)

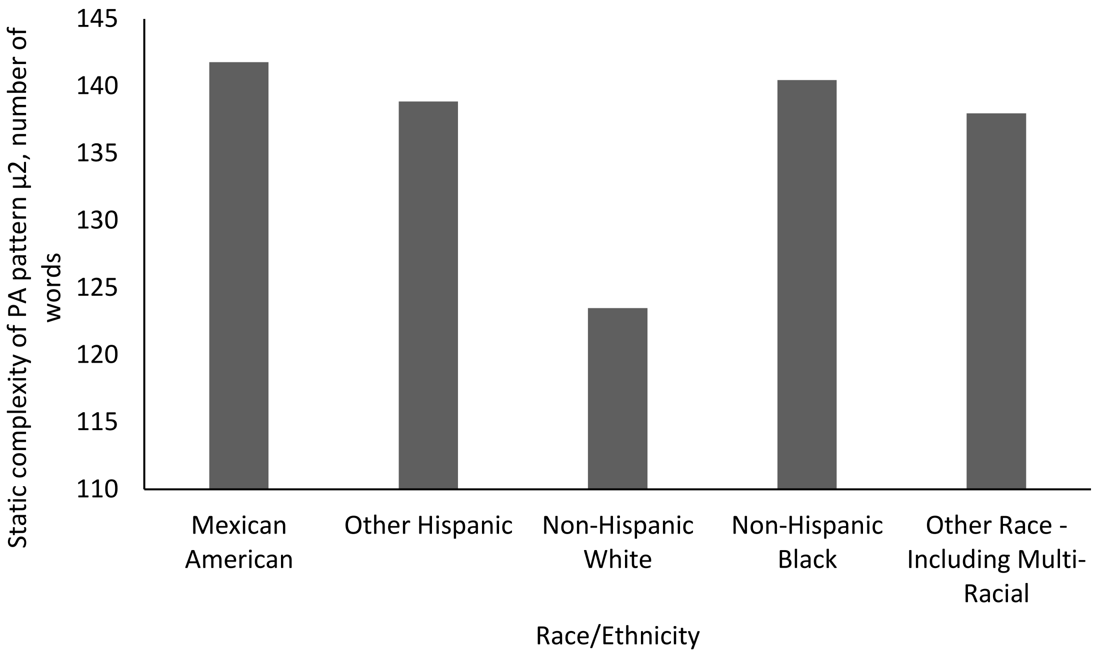

Figure 46. Marginal means static complexity of PA μ_2 for adults (19 – 55 y.o.)

Unlike static complexity of PA pattern μ_2, measure of dynamic complexity of PA pattern μ_3 is reversed (as explained in Metric 3: Dynamic complexity of PA pattern in Chapter 3). The compression units of μ_3 closer to 0 correspond to high dynamic complexity and values going away from 0 to -1 correspond to lower dynamic complexity.

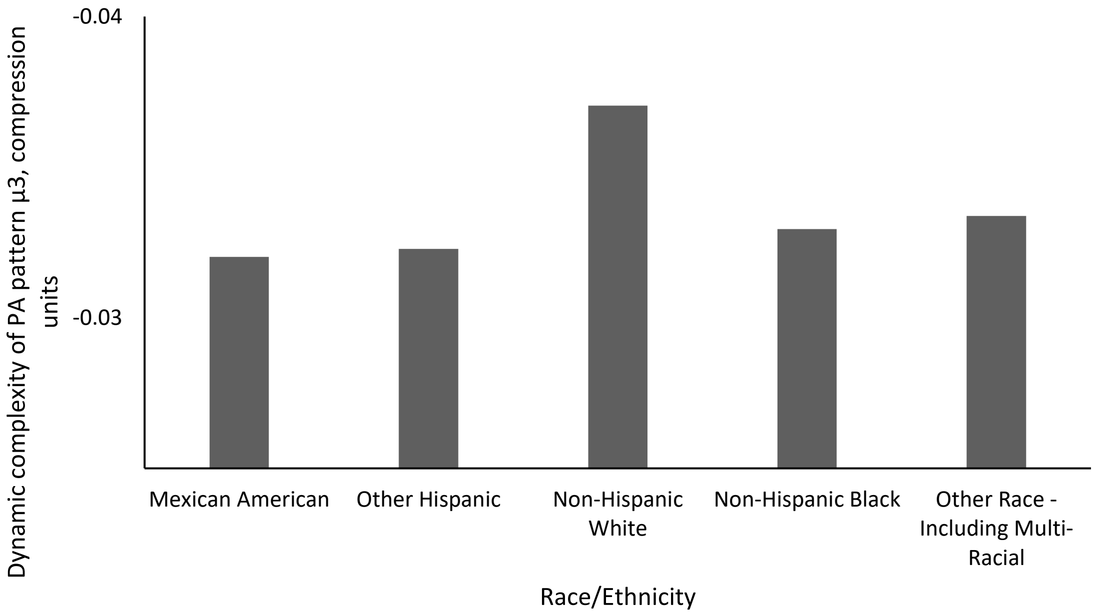

Figure 47. Marginal means dynamic complexity of PA μ_3 for adults (19 – 55 y.o.)

Household income, education level and mid-term history of PA were related to volume of PA. Greater income level was associated with lower level of volume of PA (β = -.109). Higher levels of education were found to be detrimentally linked to the volume of PA (β = -.099). Perceived increase in mid-term self-reported history of PA was positively correlated with volume of PA (β = .075).

Summary results

Unstandardised ANCOVA coefficients and significance level are presented in the table below:

N=667		Dependent variable					
	Independent variables	μ_1		μ_2		μ_3	
		β	*p*	β	*p*	β	*p*
Common	Age	-.14	.000	-.11	.005	-.10	.011
	BMI	-.12	.002	-.14	.000	-.14	.000
	Compare act with others	.162	.000	.111	.006	.138	.001
	Vitamin D	.098	.009	.102	.008	.100	.009
Unique	Disease burden	-	*NS*	-.09	.030	-.09	.019
	Gender	-	*NS*	-.10	.010	-.08	.040
	Alcohol consumption	-	*NS*	.080	.046	.092	.022
	Ethnicity/race	-.11	.005	-	*NS*	.082	.045
	Annual household income	-.12	.009	-	*NS*	-	*NS*
	Education	-.1	.021	-	*NS*	-	*NS*
	History of PA last month – last year	.075	.037	-	*NS*	-	*NS*

Table 18. Unstandardised ANCOVA coefficients in forward selection ANCOVA for adults (19 - 55 y.o.)

Older adults

Variance explained

1725 older adults aged 56 – 85 y.o. were selected in this analysis. Independent variables explained: 30% of variance of volume of PA μ_1 (R^2 adj = 0.302); 36% of variance of static complexity of PA pattern μ_2 (R^2 adj = 0.363); and 35% of variance of dynamic complexity of PA pattern μ_3 (R^2 adj = 0.346). The variance explained by independent variables are summarised in the table below:

	μ_1	μ_2	μ_3
N	1725	1725	1725
R^2	.31	.37	.354
R^2 adj	.302	.363	.346

Table 19. Variance explained by forward selection ANCOVA in older adults

Common independent variables

ANCOVA models $\mu_1 – \mu_3$ had four common statistically significant correlates: age, self-reported health, comparison of PA level with others and vitamin D. In all models, age and self-reported health were negatively associated with the dependent variable. Comparison of PA level with others and vitamin D were positively associated. Age was linked to lower volume of PA (β = -.46), lower static complexity of PA pattern (β = -.52), and lower dynamic complexity of PA pattern (β = -.52). Self-reported health was negatively correlated with μ_1 - μ_3, poor health was related to lower level of volume of PA and complexity of PA pattern (μ_1: β=-.09, μ_2: β=-.13, μ_3: β=-.13). Comparison of the self-reported PA vs others positively correlated with volume of PA μ_1 (β = .148), static complexity of PA pattern μ_2 (β = .161), and dynamic complexity of PA pattern μ_3 (β = .181). Higher vitamin D level was linked to higher volume of PA μ_1 (β = .07), static complexity of PA pattern μ_2 (β = .049), and dynamic complexity of PA pattern μ_3 (β = .053).

Measure-specific independent variables

Measure-specific unique correlates are reported next. Disease burden, gender, and annual household income correlated with static and dynamic complexity of PA pattern, but not with volume of PA. Higher levels of disease burden corresponded to lower static complexity of PA pattern (β = -.07) and lower dynamic complexity of PA pattern (β = -.06). Male gender was linked to lower static complexity of PA pattern (β = -.06) and lower dynamic complexity of PA pattern (β = -.05). Higher levels of annual household income correlated with higher static complexity of PA pattern (β = .066) and higher dynamic complexity of PA pattern (β = -.06).

Mexican Americans, on average, had higher level of volume of PA and static complexity of PA pattern than non-Hispanic white. Further bivariate correlation analysis showed that all three metrics were correlated with ethnicity: volume of PA (r_s = -0.159, P = 0.000), static complexity of PA pattern (r_s = -0.07, P = 0.000), and dynamic complexity of PA pattern (r_s = -0.073, P = 0.000). Bar plots showing marginal mean values for $\mu_1 - \mu_3$ are presented below.

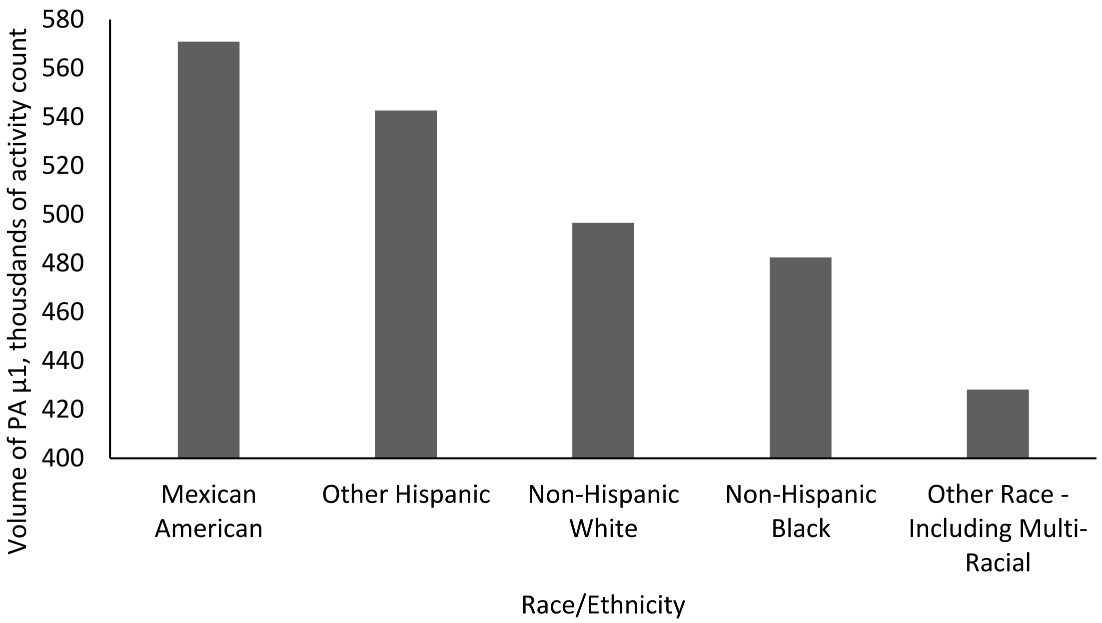

Figure 48. Marginal means volume of PA μ_1 for older adults

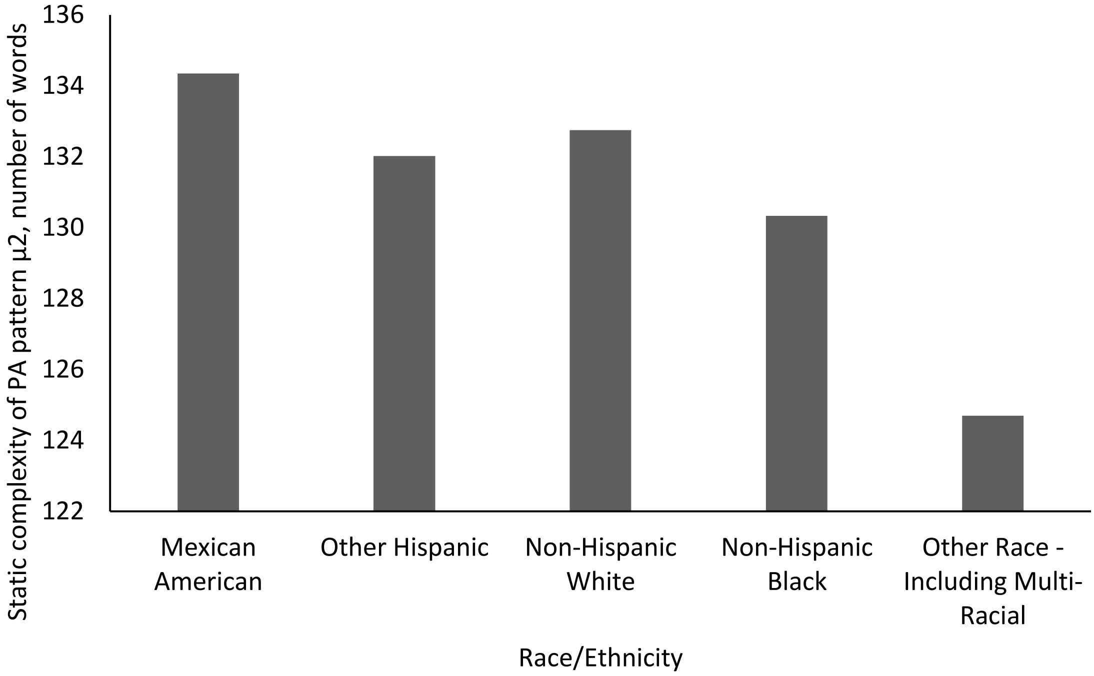

Figure 49. Marginal means static complexity of PA pattern μ_2 for older adults

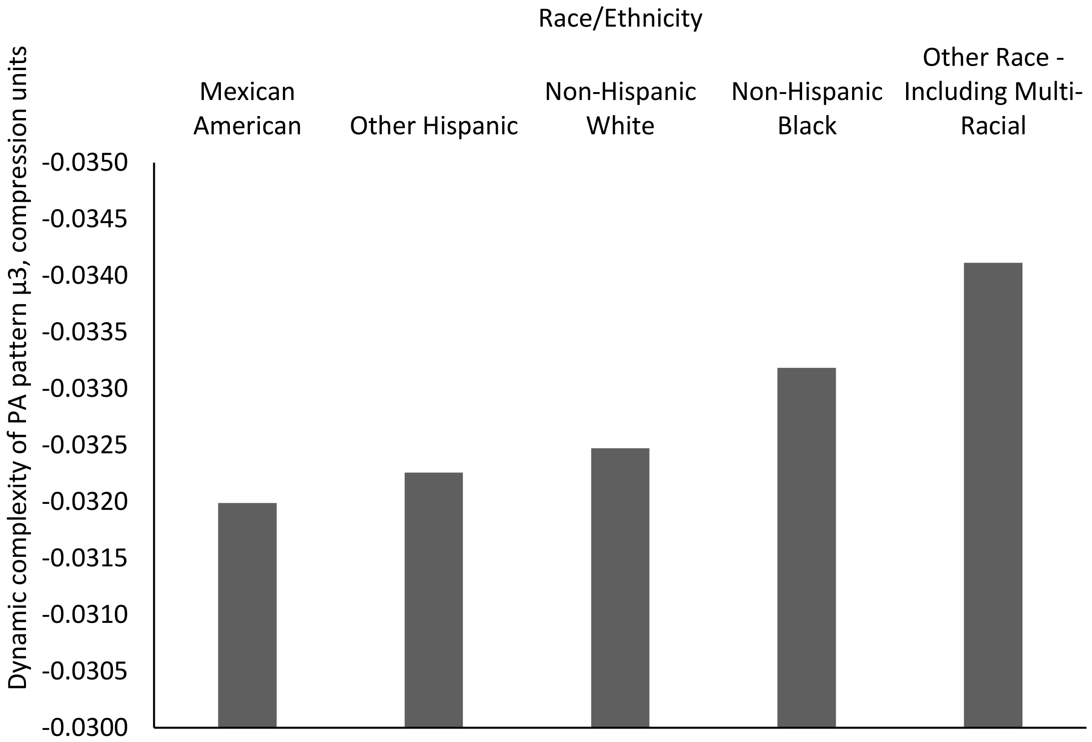

Figure 50. Marginal means dynamic complexity of PA pattern μ_3 for older adults

BMI correlated with volume of PA and static complexity of PA pattern, but not with dynamic complexity of PA pattern. Higher BMI corresponded to lower volume of PA (β = -.12) and lower static complexity of PA pattern (β = -.08). Smoking, education level and shortness of breath only correlated with volume of PA, but not with complexity of PA pattern. Smokers with high value of cotinine (ng/mL) was linked to lower volume PA (β = -.04). Higher levels of education were found to be detrimentally linked to volume of PA (β = -.08). Lastly, self-reported shortness of breath was linked to positive increase in volume of PA (β = .057).

Summary results

Unstandardised ANCOVA coefficients and significance level are presented in the table below:

N=1725		Dependent variable					
	Independent variables	μ_1		μ_2		μ_3	
		β	*p*	β	*p*	β	*p*
Common	Age	-.23	.000	-.44	.000	-.46	.000
	Self-reported health	.101	.000	.1	.000	.107	.000
	Compare act with others	.165	.000	.129	.000	.13	.000
	Vitamin D	.162	.000	.069	.014	.066	.018
Unique	Disease burden	-	*NS*	-.07	.002	-.06	.006
	Gender	-	*NS*	-.06	.00	-.05	.024
	Annual household income	-	*NS*	.066	.006	.06	.013
	Ethnicity	-.09	.000	-.04	.044	-	*NS*
	BMI	-.12	.000	-.08	.000	-	*NS*
	Smoking	-.04	.041	-	*NS*	-	*NS*
	Education	-.08	.001	-	*NS*	-	*NS*
	Shortness of breath	.057	.009	-	*NS*	-	*NS*

Table 20. Unstandardised ANCOVA coefficients in forward selection ANCOVA for older adults (56 – 85 y.o.)

Stage 2. Full sample the backward-selection ANCOVA

In stage 2, measures μ_1 - μ_3 were entered as dependent variables to the same ANCOVA models using backward regression. This time ANCOVA was applied to the full population sample. Independent variables used in the model were listed in Table 11 above.

Variance explained

2805 children, adults and older adults aged 6 – 85 y.o. were selected in this analysis. Independent variables explained: 26% of variance of volume of PA μ_1 (R^2 adj = 0.259); 27% of variance of static complexity of PA pattern μ_2 (R^2 adj = 0.27); and 29% of variance of dynamic complexity of PA pattern μ_3 (R^2 adj = 0.293). The variance explained by independent variables are summarised in the table below:

	μ_1	μ_2	μ_3
N	2805	2805	2805
R^2	.274	.282	.308
R^2 adj	.259	.270	.293

Table 21. Variance explained by the backward selection ANCOVA in the full population sample

Common independent variables

ANCOVA models μ_1 - μ_3 had five common statistically significant correlates: age, BMI, ethnicity, self-reported history of activity last month vs last year, and comparison of levels of self-activity with others. In all models, age was negatively associated with the dependent variable. Age was linked to lower volume of PA (β = -.35), lower static complexity PA pattern (β = -.41), and to lower dynamic complexity of PA pattern (β = -.43). BMI was negatively correlated with the dependent variables – a higher level of BMI corresponded to a lower volume of PA and less complex pattern of PA. BMI was linked to a lower volume of PA (β = -.14), lower static complexity of PA pattern (β = -.15), and to a lower dynamic complexity of PA pattern (β = -.14). Mexican Americans, on average, had a higher volume of PA (β = -.12), a higher level of static complexity of PA pattern (β = -.08) and a higher level of dynamic complexity of PA pattern (β = -.08) than non-Hispanic white. "More active" self-reported history of PA last month vs last year was positively correlated with the volume of PA (β = .066), static complexity of PA pattern (β = .058) and dynamic complexity of PA pattern (β = .056). Comparison of levels of self-activity with others was positively correlated with μ_1 -

μ_3, “more active” responses resulted in a higher volume of PA and complexity of PA pattern (μ_1: β=.142, μ_2: β=.114, μ_3: β=.123).

Measure-specific independent variables

Measure-specific unique correlates are reported next. Higher levels of alcohol consumption were linked to higher static complexity of PA pattern (β = .063) and higher dynamic complexity of PA pattern (β = .065). Male gender was linked to lower static complexity of PA pattern (β = -.09) and lower dynamic complexity of PA pattern (β = -.06). More significant household income was associated with a lower level of volume of PA (β = -.07) and higher dynamic complexity of PA (β = .091). Higher levels of vitamin D were linked to a higher volume of PA μ_1 (β = .083), and higher static complexity of PA pattern μ_2 (β = .064). Higher levels of education were found to be detrimentally linked to the volume of PA (β = -.10). A “no” answer to the question “have you had shortness of breath either when hurrying on the level or walking up a slight hill?” was positively correlated with the volume of PA (β = .07).

Summary results

Unstandardised ANCOVA coefficients and significance level are presented in the table below:

N=667		Dependent variable					
	Independent variables	μ_1		μ_2		μ_3	
		β	*p*	β	*p*	β	*p*
Common	Age	-.35	.000	-.41	.000	-.43	.000
	BMI	-.14	.000	-.15	.000	-.14	.000
	Ethnicity	-.12	.000	-.08	.008	-.08	.013
	Activity last m-last y	.066	.019	.058	.036	.056	.042
	Compare activity with others	.142	.000	.114	.000	.123	.000
Unique	Alcohol consumption	-	*NS*	.063	.031	.065	.03
	Gender	-	*NS*	-.09	.003	-.06	.031
	Annual house income	-.07	.026	-	*NS*	.091	.004

	Vitamin D	.083	.005	.064	.028	-	*NS*
	Education	-.10	.003	-	*NS*	-	*NS*
	Shortness of breath	.07	.017	-	*NS*	-	*NS*

Table 22. Unstandardised ANCOVA coefficients in the backward selection ANCOVA for full population sample (6 - 85 y.o.)

Chapter 5. Discussion

This research explored the complexity of PA pattern and its relationship to the volume of PA and health correlates. Known factors of PA behaviour were analysed against metrics of the volume of PA and complexity of PA pattern to identify a potential explanation of variability and compare outcome results with regards to the effect size in populations (children and adolescents, adults, older adults, and full sample). The analysis stage was broken down into two subcomponents. The first stage used ANCOVA regression on the age-stratified population with a forward selection. The second stage investigated a full population sample to understand which correlates universally remained for the entire population irrespectively of age.

The exact determinants of PA behaviour have not been found yet. The research has focused on the volume of activity to understand how much people did and what was the impact of exercise and activity. The statistical models presented typically accounted for only about 30% of the data variance, so there was a lot of room to improve the understanding (Bauman *et al.*, 2012).

Efforts have been made to identify some of the universal PA determinants and correlates. For example, Bauman et al. published a review suggesting the following main determinants of PA (Bauman *et al.*, 2012): age, gender, and health status or perceived fitness (among others). A pan-European multi-disciplinary study proposed a clustered approach to cover PA determinants across the full life course (Condello *et al.*, 2016). The output cluster from their modelling included several hundred determinants of PA which were grouped into two dimensions: first related to a person and second related to the surrounding society.

Variance explained

The ANCOVA analysis in this study contained 19 independent variables, separately analysed against the volume of PA μ_1, static complexity of PA pattern μ_2, and dynamic complexity of PA pattern μ_3. Surprisingly, nearly in all statistical models, more variance was explained by the static and dynamic complexity of PA pattern compared to the volume of PA. Consider the following table:

ANCOVA analysis		**R^2 adj**		
		μ_1	**μ_2**	**μ_3**
Stage 1	Children and adolescents	.162	.243	**.257**
	Adults	**.193**	.143	.151
	Older Adults	.302	**.363**	.346
Stage 2	Full sample	.259	.270	**.293**

Figure 51. Summary of variance explained in ANCOVA models

The dynamic complexity of PA pattern explained the most variance for children and adolescents. Telford et al. observed that children aged 8 – 12 years old from the 29 Australian elementary schools exhibited clear PA pattern connected to the school activity (Telford *et al.*, 2013). This study reported a low percentage (11% - 50% depending on the day of a week) for boys and girls who met international guidance of 13000 steps/day and 11000 steps/day respectively. A trend was also reported towards lower time spent in MVPA and Light PA with the corresponding increase in sedentary time.

This thesis reported a rapid decline of the volume of PA and a reduced complexity of PA pattern (see Chapter 4 Aging, volume of PA and complexity of PA pattern). Clearly, in this age group, it was easier to predict the complexity of PA pattern rather than the volume. Therefore, other studies of PA behaviour in children will benefit from exploring the complexity of PA pattern, as this study has shown that the independent variables could explain more variance in complexity of PA pattern.

In contrast to children, more variance of the volume of PA was explained in the adult age category. In adults, a slight decline with age of the volume of PA and the complexity of PA pattern was observed (see Chapter 4 Stage 1 Adults). This study demonstrated that adults accumulated less volume of PA and had less complex PA pattern compared to children and adolescents. In literature, significant efforts were made to characterise PA determinants, for example, via ecological model (Bauman *et al.*, 2012), EU-PAD effort (Condello *et al.*, 2016), or study of sedentary behaviour (Chastin *et al.*, 2018). Owen et al. derived an ecological model of sedentary behaviours looking at TV viewing and other screen-focused behaviours, prolonged sitting in the workplace, and time spent sitting in commuting (Owen *et al.*, 2011).

For older adults, the variance of static complexity of PA pattern had the highest R^2 score – μ_2: R^2 adj = .365. As complexity of PA pattern and the volume of PA pattern were found to be related, this result suggested that an increase of static complexity of PA pattern was linked with better health. Dogra et.al looked at sedentary behaviour of older adults suggesting that "breaking up prolonged bouts of sedentary time has on geriatric-relevant [that is related to medical problems and disease of old age] health outcomes." (Dogra *et al.*, 2017).

The last stage of this research included the regression analysis of the full population. Regression models were run with two regression options: enter and backward elimination. As there were no significant differences found, only the backward elimination regression was reported. The independent variables captured 26% of variability of volume of PA (R^2 adj = .259), variability of static complexity was explained by 27% (R^2 adj = .270) and variability for dynamic complexity of PA was predicted by 29% (R^2 adj = .293). Dynamic and static complexity of PA pattern had a slightly better statistical fit, suggesting that the research focused on overall human PA should consider using metrics of the complexity of PA pattern.

In the literature, there was evidence to suggest that physiological complexity was subject to change, for example, heart rate variability (Billman, 2011), gate analysis in elderly fallers (Hausdorff, 2005), physiological complexity affected by age and disease (Goldberger, Peng and Lipsitz, 2002), and physical activity and chronic pain (Paraschiv-Ionescu *et al.*, 2012). A mathematical model of chaos could describe some physiological processes: a system changes its behaviour in a complex way, migrating from one state to another and thus continually adjusting to the influence of multiple factors (Shaffer and Ginsberg, 2017). It was also shown that changes in a chaotic system could not be predicted beyond a short time interval (Lorenz, 1963). Although PA has not been shown to be a chaotic system, some aspects of PA behaviour demonstrated some qualities of these complex, chaotic systems, for example, poor predictive modelling, attraction to sedentary behaviour, complex spikes of activity states.

Common predictors of the volume of PA and the complexity of PA pattern

Predictors of the volume of PA have been discussed in depth before. This research highlighted that several correlates of the volume of PA also correlate with the complexity of PA pattern.

Age

The strongest common predictor for volume and complexity of PA pattern was age with the effect size being two to four times larger than any other independent variable. Age was negatively associated with dependent variables – as the population got older it continually lost accumulated volume and complexity of PA pattern.

The volume of PA and complexity of PA pattern depleted fastest in children/adolescents and older adults age category, with a slight loss during adulthood. The full population sample lost complexity of PA faster compared to the volume of PA. In children, for example, complexity of PA pattern decreased faster with age (μ_2: $\beta = -.44$ and μ_3: $\beta = -.46$) compared to volume of PA (μ_1: $\beta = -.23$). This indicated that growing-up contributed to the noticeable changes in structural and temporal PA behaviour. These changes were also observed in the volume of PA, but with lower effect.

Ageing contributes to the change in the complexity of many physiological processes. The change in the complexity depended on a large pool of factors. Most commonly (but not universally), the complexity of physiological processes decreased with age, for example, heart rate, blood pressure, hormonal responses, and posture (Lipsitz, 2004). Although complexity may also increase as well as decrease, for example, due to a short-term demand and the current state of the system (Vaillancourt and Newell, 2002). As an example of this phenomenon, brain activity, as measured via power spectra and correlation dimension of EEG, could increase with age for participants aged 5 – 60 y.o. (Anokhin *et al.*, 1996). Potentially, the change (in most cases – the decrease) in complexity of physiological processes could be related to the decrease in the complexity of PA behaviour.

Research widely supported this finding in adults and older adults, for example, similar results with regards to the volume of PA and health correlates were reported by Sallis, Prochaska, Taylor, Troiano, Bauman and others (Sallis, 2000; Sallis, Prochaska and Taylor, 2002; Haskell *et al.*, 2007; Keadle *et al.*, 2016). As participants grew older, they accumulated less volume of activity and varied their behaviour more erratically. It has been recommended to achieve a certain level of volume of PA in order to stay healthy, for example, “all healthy adults aged 18 to 65 years need moderate-intensity aerobic (endurance) physical activity for a minimum of 30 min on five days each week or vigorous-intensity aerobic physical activity for a minimum of 20 min on three days each week.” (Haskell *et al.*, 2007). But even smaller amount of activity, for example, “15 min a day or 90 min a week of moderate-intensity exercise”, provided a positive effect on health and decreased the risk of all-cause mortality (Wen *et al.*, 2011). Research of sedentary behaviour confirmed a positive impact of the breaks during inactive behaviour (Owen *et al.*, 2010). The Australian Diabetes, Obesity and Lifestyle study examined the behaviour of 11000 adults from all Australian states and the Northern Territory. This study argued that prolonged sitting increased the risk of cardio-metabolic illnesses. Breaking sedentary behaviour could be related to the increase of static and dynamic complexity of PA pattern, as there would be more non-sedentary words and the activity could become harder to predict.

Unlike some of the clear evidence of the beneficial PA levels in adults and older adults, the relationship between PA levels and health in children and adolescents was not very clear (McKinley, 2003). This was partly because of the little amount of accurate data that was available and a lack of a measuring instrument that could capture a wide range of PA behaviour in children (Riddoch *et al.*, 2004). Given a stronger effect size of complexity of PA pattern compared to the volume of PA, it would be beneficial to use complexity metrics in studies trying to establish connections between PA and health. The use of complexity metrics of PA pattern could help to identify the relationships easier and with better accuracy compared to the volume of PA.

Self-reported metrics of health

Self-reported metrics of health predictably correlated with volume and complexity of PA pattern. History of PA last month vs last year, comparison of self-reported PA level with others, general health condition and shortness of breath on stair incline were found to have associations. Better (healthier) self-reported values showed a positive correlation with volume and complexity of pattern PA. Self-perceived increase of PA correlated with a higher volume of PA and more complex PA pattern.

Historical participation in PA is tracked through life (McAuley and Blissmer, 2000; Ashford, Edmunds and French, 2010). Not-surprisingly some correlations were observed between self-reported measures and complexity and volume of PA pattern.

Shortness of breath on stair incline was positively associated with the volume of PA (β=.058), but not the complexity of PA pattern. This correlate was only present in the older population in which respiratory and heart rate factors were more dominant due to age. It seemed that participant who did more physical activity also indicated shortness of breath. One possible explanation for this could be that they actually engaged in prolonged or higher intensity PA, such as hillwalking or using stairs instead of elevators, and due to ageing, naturally experienced shortness of breath during the ascend.

BMI

Body composition, measured via body-mass index, was found to be a correlate with measures of volume and complexity of PA pattern. It was one of the strongest correlates after age.

It was shown that the volume and complexity of PA pattern decreased with the increase in BMI. The complexity of PA pattern had similar associations with BMI as the volume of PA. In the ANCOVA analysis of the full population, for example, BMI negatively correlated with the static complexity with β=-.146, dynamic complexity β=-.143, and volume of PA β=-.141. Participants in the underweight or normal-weight BMI category had a more complex pattern and performed more volume of PA.

In literature, however, the relationship between the volume of PA and BMI was not clear. Koenman et al. suggested there was moderate evidence of negative correlation of BMI and volume of PA (Koeneman *et al.*, 2011). Studies reported either negative correlation (King *et al.*, 1997; Nitz and Choy, 2007) or an absence of such correlation (Garcia and King, 1991).

Ethnicity

Ethnicity was a common predictor for volume and complexity of PA pattern. In the analysis of the full population, Mexican Americans and Other Hispanic ethnicities had more complex PA pattern (static complexity β=-.08, dynamic complexity β=-.08) and accumulated more volume of PA (β=-.12) than Non-Hispanic Black and Non-Hispanic White ethnicities. These findings have confirmed the results reported in the literature (CDC, 2005; Palacios-Ceña *et al.*, 2011).

Participants of different ethnic backgrounds showed that they accumulated volume and complexity of PA pattern differently. There could be several explanations for this. In the analysis of US youths using the same NHANES data, Belcher et al. suggested there could be cultural or biological reasons for this. The research paper noted: “non-Hispanic White (NHW) males spent 3–4 fewer min/day in vigorous PA than Mexican American (MA) (p=.004) and non-Hispanic Black (NHB) (p<.001) males” (Belcher *et al.*, 2010). Saffer et al. presented significant race/ethnicity/gender differences of PA levels and concluded that 30–65% of these were due to education, socioeconomic status, time constraints, and locational attributes. (Saffer *et al.*, 2013).

Vitamin D

Vitamin D was another common correlate for the three models. Due to the nature of its production in the human body, it can be viewed as a proxy measure to exposure to the sunlight (Luxwolda *et al.*, 2013). It can be used as an objective indicator of how much time participants spent outside. For children, this study reported that the volume of PA had a positive relationship with the Vitamin D (β=.163); likewise, static and dynamic complexity (β=.064, and β=.066 respectively) were positively associated. Quite possibly, this was because outside bouts of activity were longer and therefore the time spend outside had a higher impact on achieving the goals of PA volume and had less impact on the static and dynamic complexity of PA pattern. For adults, higher vitamin D corresponded to the greater volume of PA (β=.1) and a higher level of complexity of PA (static complexity β=.103, dynamic complexity β=.104). Vitamin D influenced both how much PA participants did and how diverse and complex PA pattern was. Participants were less regimented if they could go outside or had access to outdoors.

Wanner et al. investigated the relationship between PA and Vitamin D serum levels and suggested that higher levels of PA increased the level of Vitamin D serum. The research group additionally suggested that in a more active population, there could be other factors contributing to Vitamin D serum than exposure to sunlight (Wanner *et al.*, 2015).

Unique predictors of volume of PA and complexity of PA pattern

Gender

Male gender was associated with a higher level of complexity of PA pattern and more volume of PA. The association for the volume of PA, has been confirmed by some previous studies reviewed by Bauman et al. (Bauman *et al.*, 2012). However, the complexity of PA pattern has not been reported yet. This study demonstrated that North American male adults had higher complexity of PA pattern (static complexity β=-.099, dynamic complexity β=-.08) and higher volume of PA (β=-.063) than females. A similar relationship was observed in the older adult population, but not in children. In the analysis of full population, males demonstrated higher complexity of PA pattern, but not the volume of PA (β=-.085 for static complexity and β=-.06 for dynamic complexity).

Results of this study confirmed similar findings in wider research, where males were found to be more physically active than females (Troiano *et al.*, 2008a; Belcher *et al.*, 2010; Saffer *et al.*, 2013).

Number of people living in a household

A total number of people living in a household was positively correlated with the volume of PA in children. More people living in the same household corresponded to a higher volume of PA (β=.081). No such relationship was observed with the complexity of PA pattern or in other age groups.

This could be one of the factors that relate to family and social environment. Perhaps having siblings or living with an extended family encouraged younger participants to be more active. However, there was no confirmation in the wider research community to this observation.

Smoking

Correlations with smoking were observed in youths and older adults. Young smokers (6 – 18 years old) with a higher level of cotinine (ng/mL) also accumulated higher volume of PA (β=.095). There was no such relationship for adults. However, for older adults (>55 years) this relationship was flipped – higher level of cotinine was associated with a lower volume of PA (β=-.05). Relationship between smoking and complexity of PA pattern was not observed, suggesting that smoking did not relate to the complexity of PA pattern.

Relationship between smoking and PA is not simple. While smoking is known for causing profound adverse effects on health (Fielding, 1985) there was no substantial evidence to suggest its correlation with PA (Bauman *et al.*, 2012). Perhaps some smokers attempt to compensate bad habit with physical activity. Some studies reported smokers who had a higher volume of PA, while others have reported the opposite. A systematic review of 2008 suggested that smoking and physical activity are more or less independent behaviours, citing an interplay of complex factors, such as “physiological (e.g., lung function) to psychological (e.g., depression) to socio-demographic (e.g., education)” (Kaczynski *et al.*, 2008).

Education level

Relationship with education level was positive for static (β=.175) and dynamic (β=.201) complexity of PA pattern in children and adolescents suggesting that education led to more stimulation and more complex pattern of PA and less routine life. This relationship was not found for the volume of PA. More years completed in education were associated with more complex PA pattern, but not volume. In a systematic review published by Sallis et al. in 2002, researches did not find publications that mentioned the number of years of completed education as a correlate with PA (Sallis, Prochaska and Taylor, 2002). A possible explanation for this could be somewhat longer time spent studying, watching TV or playing video games or in other sedentary behaviours, but also greater access and participation in more leisure PA, for example, sports clubs, gym, team sports after school. The number of completed years in education was likely related to the age of children. Care should be taken during further analysis to remove the statistical bias between age and education level.

For adults and older adults, higher education level was associated with a lower volume of PA (β=-.076). This is perhaps due to the higher education being commonly associated with an office or otherwise sedentary job, whereas a lower level of education could be linked to manual labour that could help accumulating higher volume of PA during work hours. Wider research did not appear to have a consensus on this topic. For example, King et al. and Finkelstein et al. found the education level was inversely associated with the volume of PA (King *et al.*, 1997; Finkelstein *et al.*, 2008). Shaw and Spokane examined the association of education and PA in early older adults (54-72 y.o.) and found that education was positively correlated with education (Shaw and Spokane, 2008). Several papers did not find a significant relationship between education and PA (Garcia and King, 1991; Burton, Shapiro and German, 1999; McAuley *et al.*, 2007).

Disease burden

Disease burden was calculated as a total number of illnesses/diseases that have been diagnosed previously. Disease burden was inversely associated with the complexity of PA pattern in children (dynamic complexity of PA pattern only), adults, and older adults, but not in full population. It was not associated with the volume of activity.

This score appeared to impact the way of how the activity was accumulated and what activity was performed. Such association could be reflective of a compensative behaviour where, with an increased number of reported illnesses and diseases, the PA pattern became less complex. Participants used fewer words in daily activity, and their PA behaviour was easier to predict.

Studies that examine the effect of the disease on PA, could gain benefits by using measures of complexity of PA pattern (Paraschiv-Ionescu *et al.*, 2008; Chastin *et al.*, 2010).

Country of birth

Country of birth is known to be a proxy measure of parental migration (Besharat Pour *et al.*, 2014). This correlate was only observed for children. Being born in one of the 50 states of the US was associated with higher dynamic complexity score (β=.053). Static complexity of PA pattern or volume of PA did not have a significant relationship with this correlate, nor was it present for adults, older adults or full population. Country of birth could result in a cultural and environmental impact for children that related to the dynamic complexity of PA.

In wider research, country of birth appeared as a correlate, but the relationship was not straightforward. For example, in a study of young Australian women, those born in Asia had significantly higher sitting time (Uijtdewilligen *et al.*, 2014). Wilkinson et al. investigated gender and country of birth as a correlate of PA among 1154 Mexican-Heritage adolescents living in Houston, Texas. They suggested that using country of birth as a correlate might not be as effective as using gender-specific psychosocial correlates (Wilkinson *et al.*, 2017).

Alcohol consumption

Amount of alcohol consumed was positively associated with static and dynamic complexity in full population and dynamic complexity in adults. It was not found to be related to the volume of PA. One of the possible explanations for this could be that drinking culture is related to the potential diversity of behaviour, for example, via peer socialising, but has no impact on the volume of PA.

Clearly, alcohol consumption has a deep negative impact on health – as much as 4% of global disease burden is caused by alcohol (Room, Babor and Rehm, 2005). However, its relationship with the volume of PA was not widely established. A 13-year longitudinal study of 903 young Finn adolescents/adults observed health behaviour change with smoking, alcohol use and physical activity and concluded that there was no significant change of leisure time PA (Paavola, Vartiainen and Haukkala, 2004). However, a different 4-year longitudinal study of 6,981 older adults (65+ y.o.) highlighted a significant risk of losing mobility associated with small-to-moderate amounts of alcohol consumption compared to no alcohol consumption (LaCroix *et al.*, 1993). It is possible to suggest that social drinking may slightly increase complexity of PA pattern, however, heavy drinking may reduce it due to the numerous negative impacts of alcohol consumption on internal organs, physical and mental health.

Annual household income

This research reported a negative association of volume of PA with household income in adults and full population and a positive association with the complexity of PA pattern. Participants living in households with more significant income were associated with higher complexity of PA pattern and lower volume, suggesting a better access to facilities with more diverse behaviour, and perhaps lower willingness to participate in a higher-intensity volume of PA.

A higher total annual household income was shown to reflect a higher socioeconomic status (Grusky, 1990). Publications in walkability research highlighted associations of PA, environment and other complex socioeconomic factors (Sallis *et al.*, 2009). Research evidence in PA suggested the presence of a direct positive correlation between both education (Trost *et al.*, 2002) and household income (Plonczynski, 2003) with inactivity. King et al. showed, that middle-aged and older women with higher income level performed more physical activity compared to those with lesser income (King *et al.*, 2000).

Volume of PA

Activity measuring technology is commonly used to record the volume of PA performed daily. Volume of PA is now an essential measure of physical activity and only a handful number of studies looked at the exact process of accumulation of this volume – a pattern of PA. The volume of PA can be estimated by a single metric, for instance, accelerometry count, step count or time spent in a particular type of activity. A single measure of the complexity of PA pattern is more elusive, and a new robust metric could potentially plug this gap. The focus of current research in PA is on identifying what determinants increase or decrease the volume of PA (Bauman *et al.*, 2012; Condello *et al.*, 2016). However, there is no clear agreement about the exact determinants of PA. Some research papers even expressed disappointment with the inability of volumetric measures and methodologies to fully explain the pervasive physical inactivity (Baranowski, Anderson and Carmack, 1998). Intervention studies appeared to have failed to create long-lasting change in the population's PA (Orleans, 2000).

Fundamentally, human physical activity could be viewed as a reflection of a decision-making process (Barabási, 2005). It could be useful to understand how transformations in PA pattern relate to adjustment in the volume of PA and vice versa. Not only how much activity was performed, but also how this activity was accumulated that was also shown to be important (Chastin and Granat, 2010; McVeigh *et al.*, 2016; Bellettiere *et al.*, 2017). It was clear that research of PA pattern could add more information about the benefits of activity and downsides of inactivity.

This thesis explored PA pattern via static and dynamic complexity associated with it. Interventions to promote PA should also aim to change the complexity of PA pattern. A metric, sensitive to complexity variation of PA pattern could be used as a powerful tool to detect behavioural transformation and help to evaluate interventions.

PA pattern

There are research papers that have touched on the complexity of PA pattern. For example, Paraschiv-Ionescu et al. demonstrated that a PA pattern was affected by the disease (Paraschiv-Ionescu *et al.*, 2008). The research of sedentary behaviour showed that the pattern of energy accumulation could influence health (Chastin and Granat, 2010; Healy *et al.*, 2011). However, few other studies have attempted to investigate the relationship of PA pattern, its complexity and health correlates.

Nonetheless, application of variability and complexity analysis methods led to many important discoveries in physiology, for instance, heart rate variability in 1965 (Bayevsky and Ivanov, 2002). Only a few complexity analysis approaches have been used to study human PA behaviour, among those were: fractal analyses (Paraschiv-Ionescu *et al.*, 2008), information entropy, sample entropy, Lempel-Ziv complexity (Paraschiv-Ionescu *et al.*, 2012), detrended fluctuation analysis, entropy rate, approximate entropy (Cavanaugh, Kochi and Stergiou, 2010), Lorenz curve, GINI index statistics (Chastin *et al.*, 2010). These methods were applied in small-population studies, but even then, they have shown great potential.

It appeared that PA pattern became more complex with higher volume of PA. A paper published by Cavanaugh et.al. demonstrated that participants with higher level of activity exhibited higher level of complexity of PA pattern (Cavanaugh, Kochi and Stergiou, 2010).

In this thesis, volume of PA was calculated as a sum of accelerometry counts over the eligible days of PA monitoring. Static complexity was calculated as a total number of unique words in the eligible days. Dynamic complexity of PA pattern was calculated as a compression ratio of the eligible days. Dynamic complexity was further adjusted for static complexity via the coefficient *k* (see Equation 36) to remove possible dependency between the two complexity measures. As an explorative research, static complexity and dynamic complexity were not adjusted for volume of activity, however, further research is necessary to investigate a deeper relationship between the measures.

This thesis contributed to a small pool of evidence that complexity of PA pattern was an independent measure of PA behaviour. It put forward two measures of complexity: static complexity that dealt with the number of PA words, and dynamic complexity that described how easy it was to predict PA pattern. These metrics were based on static and dynamic qualities of PA behaviour.

The results demonstrated that the complexity of PA pattern could be explained, by known PA correlates. In most cases, static and dynamic complexity of PA pattern showed the same relationships with age, BMI, ethnicity, self-reported measures of health and vitamin D serum. Clearly, the complexity of PA pattern was at least as good of a measure of PA behaviour as the volume of PA. Additionally, new potential correlates became significant with the complexity of PA pattern that were not significant with the volume of PA, for example, alcohol consumption and gender.

This study found that that volume of PA and complexity of PA pattern were non-normally distributed. Most of the population were less active and had less complex PA pattern (see Chapter 4 Volume of PA and complexity of PA pattern) indicating that most population was inactive.

Symbolisation of PA pattern

PA pattern was symbolised on a 4-minute interval during a coarse-graining process. Analysis in this study suggested that a 4-minute word was the most suited for the analysis of the complexity of PA pattern, as it allowed for the most substantial diversity of unique words, and at the same time, it was constrained by the total recording duration. Other studies can choose another fixed-length duration to be able to observe longer trends. However, the recording duration ought to be much longer than seven days. For example, studies concerned with seasonal variation may choose a 1-day word or 1-week word given there is more than a year's worth of data recording per individual.

Somewhere between a too-short and too-long word, there may be a fundamental constant of PA pattern measured by accelerometry. Further research is needed to ascertain if a 4-minute interval is actually the most suited for PA research as measured by standard 1-minute epoch accelerometers.

It is possible to think of a 4-minute-word as a relatively short monitoring bout. Frequently, our activities take longer time, for example, sedentary behaviour, commuting, walking or exercising. Having a short monitoring period could allow to investigate smaller trends inside the behaviour. However, relative to the currently widely used 1-minute monitoring bout, a 4-minute bout is clearly 4 times as long. A longer monitoring bout allows for softening the boundaries of the behaviour.

This study did not attempt to perform a semantic analysis of the 4-minute words. But it was noticed that there were some words synonyms to each other. These were the words with the same volume, but different order of activities. More often, transitions between activities were recorded as synonyms, such as: "sit-sit-sit-stand", "stand-sit-sit-sit" and similar. Future studies may consider investigating a potential relationship between words-synonyms and health-related measures.

Measures of the static complexity in this thesis did not adjust for synonyms, as the main focus was to investigate the unique and distinct words, however other studies may want to have a softer approach that could allow to categorise and cluster behaviour into symbolic groups - mathematical components of the behaviour.

Static complexity

Static complexity described how many unique words were available in PA pattern. This measure was useful to mitigate the problem of a high probability of appearance of some words, for example, sedentary behaviour. Sedentary behaviour was known to account for up to 70% of PA. A sedentary lifestyle was also linked with a handful number of low activity words, such as sit to stand transitions.

It would be interesting to find some universal and common words in populations to study PA pattern. Some evidence suggested that there was a deterministic non-linear relationship between volume and pattern of PA (Barabási, 2005) with the same amount of PA can be accumulated via various PA pattern. It was unclear what that pattern may look like and how to locate similarities between different pattern. Static complexity of PA pattern provided a method to achieve that.

This study found that patterns of PA observed in NHANES data set contained a mixture of long periods of inactivity, predictable activity (commuting, active work or sports and exercises) and some sporadic bursts of activity. These occasional spontaneous bursts of activity presented an interesting research problem of how to quantify it, why it occurred, and what was its role? Some temporal and intermittent behaviour was also reported and analysed by other researches (Cavill, Biddle and Sallis, 2001; Barabási, 2005).

Dynamic complexity

The measure of dynamic complexity used in this research was based on a compression algorithm. Compression metric looked at the number of repetitions in sequence and made predictions about subsequent words. If PA pattern was easy to predict, it was easy to compress, resulting in smaller output and lesser dynamic complexity of PA pattern. If PA pattern was more complex and harder to predict, it became more difficult to compress it, and the output was longer.

Some PA words repeated daily, like sleep. Other words could be seen at arbitrary intervals along the day, presenting themselves as short bursts of activity for a few minutes. There were some accidental words that never repeated or were only seen a handful number of times. This suggested that PA behaviour is mostly predictable and periodic with some short intervals of hard to predict spikes of activity.

In the wider research community, the dynamics of PA pattern was described by fractals. Fractal dynamics could quantify repetitive nature and found similarities observed at different scales (Paraschiv-Ionescu *et al.*, 2008). This study evaluated the fractal dynamics technique and concluded that there was a common control system to regulate PA behaviour, which could be described in part by fractals.

Practical use and implications

Recent technological advances, for example, Internet of Things, Big Data, artificial intelligence, pervasive virtual and physical activity trackers have created an information-rich, interconnected, and hybrid environment.

It is reasonable to assume that more and more objectively measured PA data will become available for research. More devices are likely to appear, measuring with higher accuracy, frequency and precision. These devices shall become cheaper to manufacture paving the way to be used in developing countries where physical activity research has not been conducted on a large scale due to the associated costs. Data-rich multimillion databases started to appear already containing records of human behaviour on a truly global scale. The size of these databases will likely grow. There will be more data, globally accessible and ethically collected making it suitable for research. There will be better quality data, fewer device failures, with non-wear time being clearly marked. So, even though globally PA is in decline, researchers should get *more information* about the underlying reasons for such decline.

This work contributed to the pool of evidence that studying the complexity of PA pattern was important. Static and dynamic complexity measures of PA pattern correlated with the volume of PA. So, the complexity of PA pattern could be used as a proxy measure for the volume of PA. Clearly, there was a relationship between how much we do and how we do it, especially when the volume of activity gets above the population's mean where correlation became linear (see Chapter 4 Volume of PA and complexity of PA pattern).

With that in mind, the complexity of PA pattern was more sensitive to change when the volume of activity was low. This result may be of interest to studies of sedentary behaviour where a small change in behaviour may result in a significant positive impact on an individual's physical and mental health. The proposed complexity measures are relatively easy to calculate in practice and have demonstrated their usefulness. Including complexity of PA pattern in other research may add some benefits, for example, this study has shown that it was possible to capture some relationships that were not observed with the volume of PA, such as alcohol consumption and gender.

Limitations

This research touched on several interesting topics, however, due to time limit some of these topics were left without additional investigations. Further studies are encouraged to explore these directions.

Due to the nature of research, this study can only report a *correlation*, but not the *causation* of PA determinants.

This study's main focus was on using the known determinants of the volume of PA and not health outcomes correlated with the complexity of PA pattern. Longitudinal studies targeting the complexity of PA pattern are required to investigate if known correlates of PA reflecting a person's health *can cause* complexity to increase or decrease.

The data set used in this research was obtained from the two NHANES databases: 2003-2004 (also known as study "C") and 2005-2006 (also known as study "D"). The

NHANES database contained representation and weighting coefficients for the non-industrialised North American population. This thesis only explored some PA aspects of the data set and utilised several filtering stages. After filtering, participants with missing data were excluded, thus making the population sample non-representative. The analysis did not account for the non-random loss of accelerometry data and non-random missing data because this study was exploratory and was not concerned with obtaining the results that represented the US population.

This study has analysed a handful of known physical activity correlates, but many correlates were either not analysed or excluded at first stages of the analysis. Several physical activity correlates were not analysed because there was no clear scientific consensus regarding their relationship with physical activity. These were: [demographic and biological correlates] marital status of parents; [psychosocial variables] perceived competence, self-efficacy, attitude, perceived behavioural control, value of health status, barriers to physical activity; [social and cultural variables] perceived parental role models, parental activity, support for physical activity, support for physical activity from parents and family; [transport activity outcome] neighbourhood design, transport environment, social environment, aesthetics, etc.

Some of the important correlates had insufficient or missing data. For example, depression score was interesting to investigate, however, the metric had too many missing records, and the results were not statistically significant to report.

Newer and larger data sets became available after the analysis stage but were not included. It would be beneficial to apply the methods discussed in this study to a different data source to validate and confirm the results.

In the literature review, this study has identified more than 70 measures of complexity. It would have been infeasible to explore every measure within the scope of a single study. Nonetheless, it would be interesting to investigate if other measures could be successfully applied to PA pattern research.

Analysis of large volume of data brings some downsides, for example data length variability has been recorded between participants. Participants were advised to remove device before sleep and bathing activities.

Chapter 6. Conclusions

In the physical activity research community, more scientific effort was directed towards the investigation of the volume of PA. Specifically, more and more research papers have focused on the determinants of the volume of PA and its causal relationship with health.

It is understood that the pattern of PA, or the specific order of activities, played some role, but the research was sparse. When a measure of the volume of PA was used in research, it automatically removed any additional information contained in the temporal PA pattern, therefore reducing the amount of information that can be used for research.

No other study has attempted to quantify the static and dynamic complexity of PA pattern or investigate its relationship with the volume of PA and its determinants. This research attempted to address this gap.

The complexity of PA pattern was quantified using two metrics: static complexity and dynamic complexity:

1. Static complexity of PA pattern represented a vocabulary of unique words observed in PA pattern on a fixed-length interval. It estimated the diversity of behaviour.
2. Dynamic complexity of PA pattern represented the unpredictability in the temporal pattern. It quantified the amounts of spontaneity in PA behaviour.

The volume of PA and complexity of PA pattern were related. In general, more volume of PA corresponded to more complex PA pattern, and hence more complex behaviour. However, such a relationship was not found to be linear. More change in the complexity of PA pattern was observed at the lower spectrum of the volume of PA, suggesting that an *inactive* population accumulated complexity of PA pattern *faster* than active population. Clearly, studies targeting sedentary behaviour or other forms of low energy expenditure behaviour could benefit from using the complexity of PA pattern.

This study analysed physical activity monitoring data of 7200 participants using NHANES 2003-2004 and 2005-2006 dataset. It confirmed a robust negative relationship between age, the volume of PA and complexity of PA pattern.

The decline of the levels of physical activity with age was widely reported in the literature. However, this was the first study to reveal such a relationship between age and complexity of PA pattern. The relationship between BMI and the volume of PA has not been made clear in the literature. This research added to the pool of evidence suggesting there was a negative correlation between BMI, the volume of PA and complexity of PA pattern. There were some differences between the volume of PA and complexity of PA pattern depending on the ethnicity. For example, Mexican American, on average, demonstrated a more complex PA pattern and accumulated more volume of PA, than non-Hispanic white. There was a correlation between self-reported measures of health and volume of PA and complexity of PA pattern. Better self-reported health corresponded to higher volume and complexity of PA.

There were some unique correlates only appearing for the complexity of PA pattern, such as disease burden, annual household income, alcohol consumption, gender and Vitamin D. These unique correlates were found in some age groups and were absent in others.

Review of the aims

This research applied complexity theory to analyse static and dynamic complexity of PA pattern. A novel idea was put forward to analyse PA pattern via the complexity of the underlying PA behaviour.

Complexity theory as a tool provided a large pool of measures that could be used to investigate PA behaviour. Some complexity measures were particularly well suited for some problems and perform poorly if applied to others. Understanding which measures could be used to study PA pattern complexity was a non-trivial task. This research assessed PA pattern via three measures:

- The volume of PA – a common measure of PA behaviour for comparison purposes;
- Static complexity of PA pattern – to estimate the number of unique words in PA vocabulary using fixed-time interval;
- Dynamic complexity of PA pattern – to quantify the amount of change in the time series in PA pattern.

These three measures were calculated using accelerometer records from NHANES data set. This data set represented the two most extensive studies available for research at the time of analysis. Statistical models (ANCOVA) were used to investigate whether the measures of complexity of PA pattern could capture the same information that was captured by the volume of PA.

Now, referring back to the aims, this research has put forward three:

1. Investigate if determinants and correlates of PA reflecting a person's health can be identified by the complexity of PA pattern
2. Investigate if new correlates can be established with the complexity of PA pattern, but not with the volume of PA
3. Understand the relationship between the volume of PA and complexity of PA pattern

The results demonstrated that the aims of this research have been met.

Concerning the aim #1 - statistically significant relationships were found between independent health correlates and factors with the dependent metrics of the complexity of PA pattern. For example, in ANCOVA analysis of full population (see Chapter 4 > Stage 2. Full sample backward-selection ANCOVA) age was negatively associated with static complexity PA pattern (β = -.41), and dynamic complexity of PA pattern (β = -.43); BMI was negatively correlated with static complexity of PA pattern (β = -.15), and dynamic complexity of PA pattern (β = -.14); Mexican Americans on average had higher level of static complexity of PA pattern (β = -.08) and higher level of dynamic complexity of PA pattern (β = -.08) than non-Hispanic white; "more active" self-reported history of PA last month vs last year was positively correlated static complexity of PA pattern (β = .058) and dynamic complexity of PA pattern (β = .056); comparison of levels of self-activity with others was positively correlated, "more active" responses resulted in higher complexity of PA pattern (μ_2: β=.114; μ_3: β=.123).

Regarding the aim #2 – new statistically significant relationships were found between the metrics of the complexity of PA pattern and health correlates, such as the same relationships were not present for the measure of the volume of PA. For example, a higher level of alcohol consumption was linked to higher static complexity of PA pattern (β = .063) and higher dynamic complexity of PA pattern (β = .065). Male gender was linked to higher static complexity of PA pattern (β = -.09) and higher dynamic complexity of PA pattern (β = -.06). Other relationships were also present, see Chapter 4 Stage 1..

In most cases, the measures of complexity of PA pattern provided a slightly better statistical model fit (higher R^2 adj score). For example, independent variables explained: 26% of variance of volume of PA μ_1 (R^2 adj = 0.259); 27% of variance of static complexity of PA pattern μ_2 (R^2 adj = 0.27); and 29% of variance of dynamic complexity of PA pattern μ3 (R2 adj = 0.293) in ANCOVA analysis of full population. Clearly, the complexity of PA pattern provided similar insights into PA behaviour as the measure of the volume of PA.

Finally, considering the aim #3 – there was a strong positive correlation between the complexity of PA pattern and volume of PA. More complex pattern corresponded to higher levels of volume of PA. Static complexity of PA pattern showed strong linear correlation with volume of PA (Pearson r = .842, n = 7420, p = .000); dynamic complexity of PA pattern demonstrated logarithmic relationship with volume of PA (R2 = .704, k = -.269, β = .018, p = .000). Complexity of PA pattern changed much faster at the lower scale of the volume of PA.

Future work

This study created a pathway for further analysis of PA pattern and complexity of PA pattern. The complexity of PA pattern was shown to be related to health correlates (e.g. age, vitamin D, BMI category and ethnic groups). There were several directions to pursue from here, and some interesting questions remained unanswered.

Was PA pattern simple or complex? It appeared to be simple most of the times, perhaps due to the prevalence of sedentary behaviour. However, bursts of activity presented an analytical challenge and appeared to be of more complex if not even chaotic nature. How could one quantify a simple sequence with such complex spikes? Were there other measures of complexity that could analyse PA pattern better?

Clearly, future research should consider the complexity of PA pattern as a separate factor. This could be particularly useful for the analysis of sedentary behaviour or studies targeting low or light PA.

Was it possible to achieve some gains in health by targeting specifically the complexity of PA pattern rather than volume? It was interesting to see if changing PA pattern alone could be more practical compared to asking participants to accumulate more volume of PA. Large-scale and experimental studies should consider breaking PA behaviour and introducing more variability as one of the possible ways to improve health. A naïve example of this may be a reminder-based study with randomised prompts to behavioural change.

This research limited its analysis to the fixed length 4-minute word. Another plausible direction may be to explore non-fixed length word for complexity analysis.

Appendices

Appendix 1. Symbolisation translation

Age (years)	Category code: 0 - Sedentary, 1 - Low, 2 - Light, 3 - Moderate, 4 - Vigorous, 5 - Extra Vigorous	Min accelerometer count (including)	Max accelerometer count (excluding)
6	0	0	100
6	1	100	500
6	2	500	1400
6	3	1400	3758
6	4	3758	10000
6	5	10000	32767
7	0	0	100
7	1	100	500
7	2	500	1515
7	3	1515	3947
7	4	3947	10000
7	5	10000	32767
8	0	0	100
8	1	100	500
8	2	500	1638

8	3	1638	4147
8	4	4147	10000
8	5	10000	32767
9	0	0	100
9	1	100	500
9	2	500	1770
9	3	1770	4360
9	4	4360	10000
9	5	10000	32767
10	0	0	100
10	1	100	500
10	2	500	1910
10	3	1910	4588
10	4	4588	10000
10	5	10000	32767
11	0	0	100
11	1	100	500
11	2	500	2059
11	3	2059	4832
11	4	4832	10000
11	5	10000	32767
12	0	0	100

12	1	100	500
12	2	500	2220
12	3	2220	5094
12	4	5094	10000
12	5	10000	32767
13	0	0	100
13	1	100	500
13	2	500	2393
13	3	2393	5375
13	4	5375	10000
13	5	10000	32767
14	0	0	100
14	1	100	500
14	2	500	2580
14	3	2580	5679
14	4	5679	10000
14	5	10000	32767
15	0	0	100
15	1	100	500
15	2	500	2781
15	3	2781	6007
15	4	6007	10000

15	5	10000	32767
16	0	0	100
16	1	100	500
16	2	500	3000
16	3	3000	6363
16	4	6363	10000
16	5	10000	32767
17	0	0	100
17	1	100	500
17	2	500	3239
17	3	3239	6751
17	4	6751	10000
17	5	10000	32767
18	0	0	100
18	1	100	500
18	2	500	2020
18	3	2020	5999
18	4	5999	10000
18	5	10000	32767

Table 23. Symbolisation translation

Appendix 2. Accelerometry data filtering

Step 1

This step splits large input files from NHANES 2003-2004 and NHANES 2005-2006 databases into small one- per-person files.

Step 2

This step checks if the data reliability flag (PAXSTAT variable) was set to false and excludes the whole file from further analysis.

Step 3

This step checks if the monitor in calibration flag (PAXCAL variable) was set to false and excludes the whole file from further analysis.

Step 4

This step checks if all device intensity was 0 throughout recording period and excludes the whole file from further analysis.

Step 5

This step checks if the total recording time was more than 5 days, if not it excludes the whole file from further analysis.

Step 6

This step checks wear time to ensure there were at least 10 continuous hours of wear-time, if not it excludes the whole file from further analysis. The algorithm was adapted from http://riskfactor.cancer.gov/tools/nhanes_pam/create.pam_perday.sas with modifications that are clearly marked. The input to the filter is data from one the day of recording. The output is a Boolean value: TRUE if day has at least 10 continuous hours of wear-time, otherwise FALSE.

Routine name: IS_DAY_VALID (meta code)

```
// Step 1. Set up

// minimum length for the non-wear period, must be > 1 minute.
validNonWearPeriodBoundary = 60

// number of non-wear periods
nonWearPeriodsCount = 0

// starting minute for the non-wear period
startIdxNonWear = 0

// ending minute for the non-wear period
stopIdxNonWear = 0

// duration for the non-wear period
durationNonWear = 0

// counter for the number of minutes with intensity between 1 and 100
lowActivityCount = 0

// indicator for starting to count the non-wear period
startNonWearIndicator = false

// indicator for resetting and starting over
resetIndicator = false

// indicator for stopping the non-wear period
stopNonWearIndicator = false

// get the number of records in a recording day
max = Records.Count

// Step 2.  CDC filter and marking of the minutes as wear/non-wear

// loop through all the records
for (int i = 0 i < max i++)

        // set the current record
        record = Records[i]

        // first iteration
        if (i == 0)

                // reset counter for non-wear periods
                nonWearPeriodsCount = 0
        end_if

        // ****** CONTRIBUTION_TO_CDC_START *******
        // valid end of non-wear period detected
        if (stopNonWearIndicator)
```

Step 7

This step removes not valid days and saves only valid days in the output folder

Appendix 3. Re-coded demographic variables used for statistical analysis

Total number of people in the Household

DMDHHSIZ: Total number of people in the Household

Value	Description	Count
1 to 6	Range of Values	8936
7	7 or more people in the Household	820
.	Missing	0

Table 24. Encoding of the DMDHHSIZ question from National Health and Nutrition Examination Survey, NHANES

Race/Hispanic origin

RIDRETH1: Recode of reported race and Hispanic origin information

Value	Description	Count
1	Mexican American	1355
2	Other Hispanic	1076
3	Non-Hispanic White	2973
4	Non-Hispanic Black	2683
5	Other Race - Including Multi-Racial	1669
.	Missing	0

Table 25. Encoding of the RIDRETH1 question from National Health and Nutrition Examination Survey, NHANES

General health condition

HSD010: Next I have some general questions about {your/SP's} health. Would you say {your/SP's} health in general is ...

Value	Description	Count
1	Excellent	606
2	Very good,	1652
3	Good,	2493
4	Fair, or	1212
5	Poor?	203

7	Refused	0
9	Don't know	0
.	Missing	2999

Table 26. Encoding of the HSD010 question from Physical Activity Questionnaire, NHANES

Activity comparison last month – last year

PAQ500: How does the amount of activity that you reported {for SP} for the past 30 days compare with {your/his/her} physical activity for the past 12 months? Over the past 30 days, {were you/was he/she}.

Value	**Description**	**Count**
1	More active	1201
-1	Less active	1222
0	In all other cases	4997

Table 27. Encoding of the PAQ500 question from Physical Activity Questionnaire, NHANES

Compare activity with 10 years ago

PAQ540: Compared with {yourself/himself/herself} 10 years ago, would you say that {you are/SP is}

Value	**Description**	**Count**
1	More active	479
-1	Less active	2482
0	In all other cases	4459

Table 28. Encoding of the PAQ540 question from Physical Activity Questionnaire, NHANES

Average number of alcoholic drinks/day for the past 12 months

ALQ130: In the past 12 months, on those days that you drank alcoholic beverages, on the average, how many drinks did you have? If less than 1, entered 1; if more than 95, entered 95. Question extracted from the alcohol use questionnaire of NHANSE. Minimum age for which data is available was 20 year olds.

Value	Description	Count
1-36	Range of values	2692
99	Don’t know	1
Excluded	Children under age of 20	2498
Missing		2229

Table 29. Encoding of the ALC130 question from Alcohol Use questionnaire, NHANES

Occupation group

OCD240 and OCD241: occupation group code for the current job. The two datasets available contained different occupation codes. A common occupation encoding was adopted to represent major groups of employment. Questions extracted from the occupation survey of NHANSE. Mapping for NHANES 2003-2004 occupation list is provided below:

Code	Input occupation category	Output occupation category
1	1='Executive, administrators, and managers' 2='Management related occupations'	Executive, management, business
2	9='Supervisors and proprietors, sales occupations' 10='Sales representatives, finance, business, & commodities ex. retail' 11='Sales workers, retail and personal services' 12='Secretaries, stenographers, and typists' 13='Information clerks' 14='Records processing occupations' 15='Material recording, scheduling, and distributing clerks' 16='Miscellaneous administrative support occupations' 24='Personal service occupations' 25='Farm operators, managers, and supervisors'	Office and admin service

3	17='Private household occupations' 19='Waiters and waitresses' 20='Cooks' 21='Miscellaneous food preparation and service occupations' 23='Cleaning and building service occupations' 26='Farm and nursery workers' 27='Related agricultural, forestry, and fishing occupations' 28='Vehicle and mobile equipment mechanics and repairers' 29='Other mechanics and repairers' 30='Construction trades' 31='Extractive and precision production occupations' 32='Textile, apparel, and furnishings machine operators' 33='Machine operators, assorted materials' 34='Fabricators, assemblers, inspectors, and samplers' 35='Motor vehicle operators' 36='Other transportation and material moving occupations' 37='Construction laborers' 38='Laborers, except construction' 39='Freight, stock, and material movers, hand' 40='Other helpers, equipment cleaners, hand packagers and laborers'	Manual labourer
4	6='Writers, artists, entertainers, and athletes'	Creative worker
5	3='Engineers, architects and scientists' 5='Teachers' 8='Technicians and related support occupations'	Technical worker, engineering
6	4='Health diagnosing, assessing and treating occupations' 22='Health service occupations'	Care and support
7	18='Protective service occupations' 41='Military occupations'	Arm forces, defence, protection
8	7='Other professional specialty occupations'	Other

	98='Blank but applicable';	

Table 30. Nhanes 2003-2004 occupation list mapping

Mapping for NHANES 2005-2006 occupation list is provided below:

Code	Input occupation category	Output occupation category
1	1 Management Occupations	Executive, management, business
2	2 Business, Financial Operations Occupations 7 Legal Occupations 16 Sales & Related Occupations 17 Office, Administrative Support Occupations	Office and admin service
3	14 Building & Grounds Cleaning, Maintenance Occupations 18 Farming, Fishing, Forestry Occupations 19 Construction, Extraction Occupations 20 Installation, Maintenance, Repair Occupations 21 Production Occupations 22 Transportation, Material Moving Occupations	Manual labourer
4	9 Arts, Design, Entertainment, Sports, Media Occupations	Creative worker
5	3 Computer, Mathematical Occupations 4 Architecture, Engineering Occupations 5 Life, Physical, Social Science Occupations 8 Education, Training, Library Occupations	Technical worker, engineering
6	6 Community, Social Services Occupations 10 Healthcare Practitioner, Technical Occupations 11 Healthcare Support Occupations 13 Food Preparation, Serving Occupations 15 Personal Care, Service Occupations	Care and support

7	12 23	Protective Service Occupations Armed Forces	Arm forces, defence, protection
8	98 99	Text present but uncodable Blank but applicable	Other

Table 31. Nhanes 2005-2006 occupation list mapping

Minimal age of participants was 16 year olds.

Value	**Description**	**Count**
1-36	Range of values	2692
99	Don't know	1
Excluded	Children under age of 16	2089
Missing		2638

Table 32. Encoding of the ALC130 question from Alcohol Use questionnaire, NHANES

Number of days when mental health was not good

HSQ480: Now thinking about your mental health, which includes stress, depression, any problems with emotions, for how many days during the past 30 days was your mental health not good?

Value	**Description**	**Count**
0-30	Range of values	5692
99	Don't know	5
Missing		1723

Table 33. Encoding of the HSQ480 question from Current Health Status questionnaire, NHANES

Disease burden

A compound z-score was computed based on the Medical Condition Questionnaire, Diabetes Questionnaire, Osteoporosis Questionnaire, and Kidney Conditions – Urology Questionnaire the MCQ score is a standardized z-score of the sum of answers (DIQ010, KIQ022, MCQ010, MCQ160A, MCQ160B, MCQ160C, MCQ160D, MCQ160E, MCQ160F, MCQ160G, MCQ160K, MCQ160L, MCQ160M, MCQ220, OSQ060), All variables equally contributed to the final score. Relationship of the MCQ score with general medical health is reverse, i.e. all variables contribute negatively to general medical health condition.

Doctor told you have diabetes (used for Disease burden score)

DIQ010: Other than during pregnancy, have you been ever told by a doctor or health professional that you have diabetes or sugar diabetes?

Value	Description	Count
1	Yes	499
0	In all other cases	6921

Table 34. Encoding of the DIQ010 question from Diabetes Questionnaire, NHANES

Ever told you had weak/failing kidneys (used for Disease burden score)

KIQ022: {Have you/Has SP} ever been told by a doctor or other health professional that {you/s/he} had weak or failing kidneys? Do not include kidney stones, bladder infections, or incontinence.

Value	Description	Count
1	Yes	95
0	In all other cases	7325

Table 35. Encoding of the KIQ022 question from Kidney Conditions – Urology Questionnaire, NHANES

Ever been told you have asthma (used for Disease burden score)

MCQ010: Has a doctor or other health professional ever told {you/SP} that {you have/s/he/SP has} asthma?

Value	Description	Count
1	Yes	924
0	In all other cases	6496

Table 36. Encoding of the MCQ010 question from Medical Conditions Questionnaire, NHANES

Doctor ever said you had arthritis (used for Disease burden score)

MCQ160A: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had arthritis?

Value	Description	Count
1	Yes	1312
0	In all other cases	6108

Table 37. Encoding of the MCQ160A question from Medical Conditions Questionnaire, NHANES

Ever told had congestive heart failure (used for Disease burden score)

MCQ160B: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had congestive heart failure?

Value	Description	Count
1	Yes	154
0	In all other cases	7266

Table 38. Encoding of the MCQ160B question from Medical Conditions Questionnaire, NHANES

Ever told you had coronary heart disease (used for Disease burden score)

MCQ160C: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had coronary heart disease?

Value	Description	Count
1	Yes	209
0	In all other cases	7211

Table 39. Encoding of the MCQ160C question from Medical Conditions Questionnaire, NHANES

Ever told you had angina/angina pectoris (used for Disease burden score)

MCQ160D: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had angina, also called angina pectoris?

Value	Description	Count
1	Yes	171
0	In all other cases	7249

Table 40. Encoding of the MCQ160D question from Medical Conditions Questionnaire, NHANES

Ever told you had heart attack (used for Disease burden score)

MCQ160E: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had a heart attack (also called myocardial infarction)?

Value	Description	Count
1	Yes	220
0	In all other cases	7249

Table 41. Encoding of the MCQ160E question from Medical Conditions Questionnaire, NHANES

Ever told you had a stroke (used for Disease burden score)

MCQ160F: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had a stroke?

Value	Description	Count
1	Yes	145
0	In all other cases	7275

Table 42. Encoding of the MCQ160F question from Medical Conditions Questionnaire, NHANES

Ever told you had emphysema (used for Disease burden score)

MCQ160G: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had emphysema?

Value	Description	Count

1	Yes	93
0	In all other cases	7327

Table 43. Encoding of the MCQ160G question from Medical Conditions Questionnaire, NHANES

Ever told you had chronic bronchitis (used for Disease burden score)

MCQ160K: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had chronic bronchitis?

Value	**Description**	**Count**
1	Yes	264
0	In all other cases	7156

Table 44. Encoding of the MCQ160K question from Medical Conditions Questionnaire, NHANES

Ever told you had any liver condition (used for Disease burden score)

MCQ160L: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had any kind of liver condition?

Value	**Description**	**Count**
1	Yes	144
0	In all other cases	7276

Table 45. Encoding of the MCQ160L question from Medical Conditions Questionnaire, NHANES

Ever told you had a thyroid problem (used for Disease burden score)

MCQ160M: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had a thyroid problem?

Value	**Description**	**Count**
1	Yes	454
0	In all other cases	6966

Table 46. Encoding of the MCQ160M question from Medical Conditions Questionnaire, NHANES

Ever told you had cancer or malignancy (used for Disease burden score)

MCQ220: Has a doctor or other health professional ever told {you/SP} that {you/s/he} . . . had cancer or a malignancy of any kind?

Value	Description	Count
1	Yes	443
0	In all other cases	6977

Table 47. Encoding of the MCQ220 question from Medical Conditions Questionnaire, NHANES

Ever told had osteoporosis/brittle bones (used for Disease burden score)

OSQ060: Has a doctor ever told {you/SP} that {you/s/he} had osteoporosis, sometimes called thin or brittle bones?

Value	Description	Count
1	Yes	307
0	In all other cases	7113

Table 48. Encoding of the OSQ060 question from Osteoporosis Questionnaire, NHANES

Composite Injury score (INJ)

Similarly a composite z-score was calculated based on the Osteoporosis Questionnaire, and the INJ score is a standardized z-score of the sum of the answers (OSQ010A, OSQ010B, OSQ010C). All variables equally contributed to the final score. Relationship of the INJ score with general health is reverse, i.e. all variables contribute negatively to general health condition.

Broken or fractured a hip (Composite Injury score)

OSQ010A: Has a doctor ever told {you/SP} that {you/s/he} had broken or fractured {your/his/her} . . . hip?

Value	Description	Count
1	Yes	65
0	In all other cases	7355

Table 49. Encoding of the OSQ010A question from Osteoporosis Questionnaire, NHANES

Broken or fractured a wrist (Composite Injury score)

OSQ010B: Has a doctor ever told {you/SP} that {you/s/he} had broken or fractured {your/his/her} . . . wrist?

Value	Description	Count
1	Yes	429
0	In all other cases	6991

Table 50. Encoding of the OSQ010B question from Osteoporosis Questionnaire, NHANES

Broken or fractured spine (Composite Injury score)

OSQ010C: Has a doctor ever told {you/SP} that {you/s/he} had broken or fractured {your/his/her} . . . spine?

Value	Description	Count
1	Yes	104
0	In all other cases	7316

Table 51. Encoding of the OSQ010C question from Osteoporosis Questionnaire, NHANES

Marital status

DMDMARTL: Marital Status. Calculated based on the demographic examination of NHANSE; n = 5947, 1475 missing entries.

Value	Description	Count
1	Married	2958
2	Widowed	474
3	Divorced	439
4	Separated	132
5	Never married	1663
6	Living with partner	279
.	Missing	1475

Table 52. Encoding of the DMDMARTL question from Demographic Examination, NHANES

Total number of people in the household

DMDHHSIZ: total number of people in the household, extract from the demographic examination of NHANSE.

Value	Description	Count
1	1 person	695
2	2 people	1789
3	3 people	1155
4	4 people	1244
5	5 people	1077
6	6 people	451
7	7 or more people	517
Missing	No answer recorded	492

Table 53. Encoding of the DMDHHSIZ question from Demographic Examination, NHANES

Country of origin

DMDHRBRN: In what country were you born? Question extracted from the demographic examination of NHANSE.

Value	Description	Count
1	Born in 50 US States or Washington, DC	5098
2	Born in Mexico	1099
3	Born elsewhere	556
7	Refused	1
9	Don’t know	1
Missing		665

Table 54. Encoding of the DMDHRBRN question from Demographic Examination, NHANES

Appendix 4. A modified filtering algorithm to find N valid hours

```
*---------------------------------------------------------------------------------------------------*;
*create.is_pam_valid_day.sas                                                          *;
*                                                                                      *;
*Purpose: to detect if a single day of activity monitoring contains N hours   *;
*     of valid recording time. Invalid time is determined from non-wear       *;
*     periods                                                                    *;
*                                                                                      *;
*Source code is based on                                                           *;
*   http://riskfactor.cancer.gov/tools/nhanes_pam/create.pam_perday.sas *;
*                                                                                      *;
*List of modifications                                                             *;
* - Comments added to explain flow of execution                                    *;
* - Code changes are marked with comments:                                         *;
*     /* CHANGE_STARTED */                                                         *;
*     /* CHANGE_STOPPED */                                                         *;
*---------------------------------------------------------------------------------------------------*;

data is_day_valid

 /******************** SETUP ********************/
 // minimum length for the non-wear period
 validNonWearPeriodBoundary = 60

 // number of non-wear periods
 nonWearPeriodsCount = 0

 // first minute for the non-wear period
 startIdxNonWear = 0

 // last minute for the non-wear period
 stopIdxNonWear = 0

 // duration for the non-wear period
 durationNonWear = 0

 // counter for the number of minutes with intensity between 1 and 100
 lowActivityCount = 0

 // indicator for beginning of the non-wear period
 startNonWearIndicator = false

 // indicator for resetting and starting over
 resetIndicator = false

 // indicator for stopping the non-wear period
 stopNonWearIndicator = false

 // get the number of records in a recording day
 max = Records.Count

 /******************** PROCESSING ********************/
 // loop through all the records
 for (int i = 0 i < max i++)

  // set the current record
  record = Records[i]

  // first iteration
```

```
    if (i == 0)
     // reset counter for non-wear periods
     nonWearPeriodsCount = 0
    end_if

 /* CHANGE_STARTED */
    // valid end of non-wear period detected
    if (stopNonWearIndicator)
     //Mark minutes in the non-wear period as NonWearTime
     for (int j = startIdxNonWear j < stopIdxNonWear j++)
       Records[j].Type = RecordType.NonWearTime
     end_for
    end_if
 /* CHANGE_STOPPED */

    // reset state if
    // a) first run
    // b) valid end of non-wear period detected
    // c) non-valid end of non-wear period detected,
    //    activity in a filter window was low but valid
    if (i == 0 || stopNonWearIndicator || resetIndicator)
     startIdxNonWear = 0
     stopIdxNonWear = 0
     durationNonWear = 0
     lowActivityCount = 0
     resetIndicator = false
     startNonWearIndicator = false
     stopNonWearIndicator = false
    end_if

    // the non-wear period starts with a zero count
    if (record.PaxInten == 0 AND NOT startNonWearIndicator)
     startIdxNonWear = i
     startNonWearIndicator = true
    end_if

    // accumulate the number of the non-wear minutes
    if (startNonWearIndicator AND record.PaxInten == 0)
    //keep track of the ending minute for the non-wear period
    stopIdxNonWear = i
    end_if

    // keep track of the number of minutes with intensity between 1-100
    if (record.PaxInten > 0 AND record.PaxInten <= 100)
    lowActivityCount = lowActivityCount + 1
    end_if

    // before reaching the 3 consecutive minutes of 1-100 intensity,
    // if encounter one minute with zero intensity, reset the counter
    if (NOT startNonWearIndicator AND record.PaxInten == 0)
    lowActivityCount = 0
    end_if

    // duration of non-wear period
    durationNonWear = stopIdxNonWear - startIdxNonWear + 1

    // a non-wear period ends with 3 consecutive minutes of 1-100 intensity,
    // one missing count, or
    // one minute with >100 intensity
    if (lowActivityCount == 3 || record.PaxInten > 100)
```

```
    if (durationNonWear < validNonWearPeriodBoundary)
      // Reset if less than nonWearPeriodsCount minutes of non-wear
      resetIndicator = true
    else
      stopNonWearIndicator = true
    end_if_else
   end_if

   // Last minute of the day
   if (i == max - 1 AND durationNonWear >= validNonWearPeriodBoundary)
    stopNonWearIndicator = true
   end_if

   //Increment non-wear periods counter
   if (stopNonWearIndicator)
    nonWearPeriodsCount = nonWearPeriodsCount + 1
   end_if

  end_for //end global loop

/* CHANGE_STARTED */

 /******************** CHECKS ********************/

 // check if the day has N valid and continuous hours

 // set records to wear time if they not specifically marked as non-wear
 int continiousWearTime = 0

 // scan through every record
 foreach (var record in Records)

   // non-wear time detected
   if (record.Type == RecordType.NonWearTime)
    // Reset counter for wear time
    continiousWearTime = 0
     // wear time detected
   else
    // Increment continuous counter
    continiousWearTime++
   end_if_else

   // check if continuous wear time > than pre-configured value
   // which is by default 10 hours
   if (continiousWearTime >= REQUIRED_MINUTES_FOR_VALID_DAY)
    // Yes, we have 10 or more hours of non-breaking wear time
    // return TRUE
    return true
   end_if

 end_for_each

 // no valid continuous time greater than threshold was found
 // return FALSE
 return false

/* CHANGE_STOPPED */


/* CHANGE_STARTED */
```

```
data has_more_then_n_valid_days(int expectedDays)

  // Calculate number of valid days a person has
  int actual = Days.Count(d => d.IsValid)

  // is it greater or equals to the threshold value
  return actual >= expectedDays

/* CHANGE_STOPPED */
```

Appendix 5. PA vocabulary of a 4-minute interval for full population

#	Word				Probability of occurrence
1.	0	0	0	0	0.33880543621203900
2.	1	0	0	0	0.03652477636339210
3.	0	0	0	1	0.03611893895346640
4.	1	1	1	1	0.02699540669028560
5.	2	2	2	2	0.02560893903234490
6.	0	1	0	0	0.02354205802575740
7.	0	0	1	0	0.02340690031861980
8.	1	1	0	0	0.01837271471870700
9.	0	0	1	1	0.01815331738819770
10.	1	1	1	0	0.01365094942014030
11.	0	1	1	1	0.01344689774977880
12.	0	1	1	0	0.01198805222792920
13.	0	0	0	2	0.01135746761494570
14.	2	0	0	0	0.01097224561201330
15.	1	0	0	1	0.01079074202223740
16.	2	2	2	1	0.01024383939634900
17.	1	2	2	2	0.01023948652785910
18.	1	1	0	1	0.00992281130081733
19.	1	0	1	1	0.00985512493568280
20.	2	1	1	1	0.00908402032292931
21.	1	1	1	2	0.00885301959738969
22.	1	0	1	0	0.00845319699725240
23.	2	2	1	1	0.00832730152201855
24.	0	1	0	1	0.00830619607491351
25.	1	1	2	2	0.00819181530657345
26.	2	1	0	0	0.00797100870542614
27.	0	0	1	2	0.00780301178715507
28.	1	1	2	1	0.00730639245396079
29.	1	2	1	1	0.00728597939759427
30.	2	2	1	2	0.00711578274914001
31.	2	1	2	2	0.00709603990227225
32.	1	2	2	1	0.00659366490608659
33.	0	0	2	2	0.00607456631443259
34.	2	2	0	0	0.00578742532679068
35.	3	3	3	3	0.00542087087477111
36.	2	1	1	2	0.00531078615871532
37.	2	2	1	0	0.00513757238119724
38.	0	1	2	2	0.00509338248170325
39.	0	0	2	1	0.00499864821238745
40.	2	1	1	0	0.00495335181969630
41.	0	1	1	2	0.00476922826134739
42.	1	2	0	0	0.00471649900614030
43.	2	1	2	1	0.00470726654081194
44.	1	2	1	2	0.00464753596918744
45.	1	2	1	0	0.00445930735522999

46.	0	1	2	1	0.00443729469264428
47.	0	2	2	2	0.00439779778853647
48.	2	2	2	0	0.00422976232000461
49.	0	2	0	0	0.00405377432609866
50.	0	0	2	0	0.00400593159974802

Table 55. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for full population NHANES 2003-2004, NHANES 2005-2006

Appendix 6. PA vocabulary of a 4-minute interval for subjects aged 6 – 18 years old

#	Word				Probability of occurrence
1.	0	0	0	0	0.28995002358367600
2.	1	0	0	0	0.03444747672201550
3.	0	0	0	1	0.03421793507979250
4.	2	2	2	2	0.02629370756587770
5.	0	1	0	0	0.02330151189543020
6.	0	0	1	0	0.02323164093941500
7.	1	1	1	1	0.02141157520283150
8.	1	1	0	0	0.01634841945982960
9.	0	0	1	1	0.01622722001147520
10.	0	0	0	2	0.01222632516548430
11.	2	0	0	0	0.01199426201863530
12.	1	1	1	0	0.01172213452700370
13.	0	1	1	1	0.01159077566988830
14.	0	1	1	0	0.01099865372553770
15.	1	2	2	2	0.01079969529955220
16.	2	2	2	1	0.01073969592298260
17.	1	0	0	1	0.01046138438659700
18.	1	1	0	1	0.00905705844708860
19.	1	0	1	1	0.00895252358332533
20.	2	1	1	1	0.00875877611887998
21.	1	1	1	2	0.00857178317679279
22.	2	2	1	1	0.00855623458036547
23.	1	1	2	2	0.00843313349285405
24.	1	0	1	0	0.00826291576427388
25.	2	1	0	0	0.00820136987811598
26.	0	0	1	2	0.00804189116793136
27.	0	1	0	1	0.00802471956494645
28.	2	2	1	2	0.00750486053989908
29.	2	1	2	2	0.00748309149431529
30.	3	3	3	3	0.00734838578399905
31.	1	2	1	1	0.00688974588477615
32.	1	1	2	1	0.00687588863631689
33.	0	0	2	2	0.00664324083409131
34.	1	2	2	1	0.00653847487573215
35.	2	2	0	0	0.00634430947628679
36.	2	1	1	2	0.00555145092278621
37.	2	2	1	0	0.00555082883571090
38.	0	1	2	2	0.00547547261665570
39.	2	1	1	0	0.00513416439166946
40.	0	0	2	1	0.00506315979380540
41.	0	2	2	2	0.00498064703528726
42.	0	1	1	2	0.00488135500573871
43.	2	1	2	1	0.00483704187325977
44.	1	2	0	0	0.00482533308187935

45.	1	2	1	2	0.00477750557251616
46.	2	2	2	0	0.00476242521120179
47.	0	2	0	0	0.00472080325668921
48.	0	0	2	0	0.00466415965400913
49.	0	1	2	1	0.00443829649420586
50.	1	2	1	0	0.00440563003858502

Table 56. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects aged 6 - 18 years old NHANES 2003-2004, NHANES 2005-2006

Appendix 7. PA vocabulary of a 4-minute interval for subjects aged 19 – 55 years old

#	Word				Probability of occurrence
1.	0	0	0	0	0.31867220184284800
2.	1	0	0	0	0.03598483833180800
3.	0	0	0	1	0.03541193091631480
4.	2	2	2	2	0.02873806466536800
5.	1	1	1	1	0.02850999769299990
6.	0	1	0	0	0.02306855013112810
7.	0	0	1	0	0.02292597983914710
8.	1	1	0	0	0.01802880538611290
9.	0	0	1	1	0.01769982131547270
10.	1	1	1	0	0.01365644428839320
11.	0	1	1	1	0.01338101031733250
12.	0	0	0	2	0.01241402932101310
13.	2	0	0	0	0.01193509616725910
14.	0	1	1	0	0.01176494028115920
15.	2	2	2	1	0.01148209019811960
16.	1	2	2	2	0.01138587887057640
17.	1	0	0	1	0.01090052878086960
18.	1	0	1	1	0.01002167894004050
19.	1	1	0	1	0.01001572277666010
20.	2	1	1	1	0.00965260454003092
21.	1	1	1	2	0.00941620983973477
22.	2	2	1	1	0.00907179903486111
23.	1	1	2	2	0.00892487371180337
24.	2	1	0	0	0.00844428216805611
25.	1	0	1	0	0.00840097566284088
26.	0	1	0	1	0.00832663521402469
27.	0	0	1	2	0.00823956930765701
28.	2	1	2	2	0.00807474401156463
29.	2	2	1	2	0.00804775966792205
30.	1	1	2	1	0.00772833854046666
31.	1	2	1	1	0.00770436004009093
32.	1	2	2	1	0.00718837064451928
33.	0	0	2	2	0.00678116032544913
34.	2	2	0	0	0.00648102577630758
35.	2	1	1	2	0.00584790281470805
36.	0	1	2	2	0.00551865975215880
37.	2	2	1	0	0.00551457934336612
38.	0	0	2	1	0.00537519310856314
39.	2	1	2	1	0.00519869231413284
40.	2	1	1	0	0.00518669218102614
41.	3	3	3	3	0.00518226785087691
42.	1	2	1	2	0.00506162295291294
43.	1	2	0	0	0.00499169206678598
44.	0	1	1	2	0.00497762918334558

45.	0	2	2	2	0.00493053548144883
46.	2	2	2	0	0.00472444239204233
47.	1	2	1	0	0.00465797383403811
48.	0	1	2	1	0.00456948453305269
49.	0	2	0	0	0.00434183184726571
50.	0	0	2	0	0.00423571836306288

Table 57. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects aged 19 - 55 years old NHANES 2003-2004, NHANES 2005-2006

Appendix 8. PA vocabulary of a 4-minute interval for subjects over 56 years old

#	Word				Probability of occurrence
1.	0	0	0	0	0.42033832465536400
2.	1	0	0	0	0.03958030404912680
3.	0	0	0	1	0.03919207966387210
4.	1	1	1	1	0.03135066081923420
5.	0	1	0	0	0.02443048357820730
6.	0	0	1	0	0.02423101065091540
7.	1	1	0	0	0.02111320704987510
8.	0	0	1	1	0.02092511925707640
9.	2	2	2	2	0.02076290223247640
10.	1	1	1	0	0.01582878041678060
11.	0	1	1	1	0.01563523482714040
12.	0	1	1	0	0.01339907268893100
13.	1	0	0	1	0.01102103212567680
14.	1	1	0	1	0.01078268581601110
15.	1	0	1	1	0.01066094853786760
16.	0	0	0	2	0.00899885667580922
17.	1	0	1	0	0.00873667808441677
18.	2	1	1	1	0.00871285307500711
19.	0	1	0	1	0.00859846820649975
20.	2	0	0	0	0.00856203317600822
21.	1	1	1	2	0.00843901635185022
22.	1	2	2	2	0.00811366517270750
23.	2	2	2	1	0.00807142857683180
24.	1	1	2	1	0.00724520915753516
25.	1	2	1	1	0.00719061227742796
26.	2	2	1	1	0.00709953167683869
27.	2	1	0	0	0.00709442570949945
28.	1	1	2	2	0.00696489531716288
29.	0	0	1	2	0.00696453889762647
30.	1	2	2	1	0.00588259936401042
31.	2	2	1	2	0.00546272881555626
32.	2	1	2	2	0.00538449915792139
33.	0	0	2	2	0.00451123869338120
34.	2	1	1	0	0.00444499969943325
35.	0	0	2	1	0.00443576433711628
36.	0	1	1	2	0.00437112494713784
37.	2	1	1	2	0.00433948478784945
38.	0	1	2	1	0.00426420906024687
39.	1	2	1	0	0.00426169078846334
40.	2	2	0	0	0.00425435579344572
41.	1	2	0	0	0.00423524333243175
42.	2	2	1	0	0.00417902616587742
43.	0	1	2	2	0.00410735237637509
44.	1	2	1	2	0.00396166641796065

45.	2	1	2	1	0.00392101579456607
46.	3	3	3	3	0.00354773649364042
47.	2	1	0	1	0.00311234676027618
48.	0	2	2	2	0.00304456289227429
49.	1	0	1	2	0.00301914431802125
50.	2	2	2	0	0.00298288381924168

Table 58. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects over 56 years old NHANES 2003-2004, NHANES 2005-2006

Appendix 9. PA vocabulary of a 4-minute interval for subjects in underweight BMI category

#	Word				Probability of occurrence
1.	0	0	0	0	0.25026903037447500
2.	1	0	0	0	0.03222677939754810
3.	0	0	0	1	0.03207533231545330
4.	2	2	2	2	0.02195227104567790
5.	0	1	0	0	0.02165313729965940
6.	1	1	1	1	0.02163145727015440
7.	0	0	1	0	0.02157816616268790
8.	1	1	0	0	0.01608639862207980
9.	0	0	1	1	0.01598565664774800
10.	1	1	1	0	0.01195368130145810
11.	0	1	1	1	0.01171310387320120
12.	2	0	0	0	0.01090418661332210
13.	0	0	0	2	0.01084882020948330
14.	0	1	1	0	0.01066079523684420
15.	1	0	0	1	0.01044565325601500
16.	1	2	2	2	0.01029450363158610
17.	2	2	2	1	0.01015266665121650
18.	3	3	3	3	0.00999290151252809
19.	1	1	0	1	0.00935575814540723
20.	1	0	1	1	0.00920592617296363
21.	2	1	1	1	0.00910757655056349
22.	1	1	1	2	0.00889572509681270
23.	2	2	1	1	0.00848363290256475
24.	1	1	2	2	0.00842736626930443
25.	1	0	1	0	0.00815376955974997
26.	0	1	0	1	0.00791430705456289
27.	2	1	0	0	0.00768700887241181
28.	2	1	2	2	0.00752779158766856
29.	0	0	1	2	0.00752386740622073
30.	2	2	1	2	0.00742883003200082
31.	1	1	2	1	0.00729255788883413
32.	1	2	1	1	0.00723567782726662
33.	1	2	2	1	0.00663465842409275
34.	0	0	2	2	0.00581787615994653
35.	2	1	1	2	0.00578049871268729
36.	2	2	0	0	0.00555545256828357
37.	2	2	2	3	0.00539180575997405
38.	2	1	1	0	0.00528527529994687
39.	3	2	2	2	0.00518585638482730
40.	2	2	1	0	0.00516921823652281
41.	2	1	2	1	0.00509595959215300
42.	0	1	2	2	0.00509292315504320
43.	1	2	1	2	0.00500585529497150
44.	0	1	1	2	0.00492404913126468

45.	2	2	3	2	0.00486396931613350
46.	2	3	2	2	0.00480246572928291
47.	0	0	2	1	0.00480095503911309
48.	1	2	0	0	0.00469705445545154
49.	0	1	2	1	0.00447029068430012
50.	2	3	3	3	0.00443060061776643

Table 59. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects in underweight BMI category NHANES 2003-2004, NHANES 2005-2006

Appendix 10. PA vocabulary of a 4-minute interval for subjects in normal-weight BMI category

#	Word				Probability of occurrence
1.	0	0	0	0	0.33923556128023900
2.	1	0	0	0	0.03692070005378730
3.	0	0	0	1	0.03654703774215240
4.	2	2	2	2	0.02646505972423280
5.	1	1	1	1	0.02568546211771270
6.	0	1	0	0	0.02406226232714660
7.	0	0	1	0	0.02398264653445390
8.	1	1	0	0	0.01835877078338790
9.	0	0	1	1	0.01824251595615580
10.	1	1	1	0	0.01346353025577700
11.	0	1	1	1	0.01321973698079290
12.	0	1	1	0	0.01220202565886470
13.	0	0	0	2	0.01171444841222790
14.	2	0	0	0	0.01123363123115060
15.	1	0	0	1	0.01089638529452730
16.	2	2	2	1	0.01023354575623630
17.	1	2	2	2	0.01023130909467870
18.	1	1	0	1	0.00972547530250057
19.	1	0	1	1	0.00968861045913418
20.	2	1	1	1	0.00870938988777322
21.	1	0	1	0	0.00861953487163420
22.	1	1	1	2	0.00854341863862253
23.	0	1	0	1	0.00838600528960679
24.	2	2	1	1	0.00812489516998540
25.	2	1	0	0	0.00812036793051509
26.	1	1	2	2	0.00793890249466149
27.	0	0	1	2	0.00786897650839978
28.	2	2	1	2	0.00704210247780658
29.	1	2	1	1	0.00703406070502514
30.	1	1	2	1	0.00701877658492106
31.	2	1	2	2	0.00699149786972776
32.	1	2	2	1	0.00643943769056105
33.	0	0	2	2	0.00614804418890006
34.	2	2	0	0	0.00582327907534793
35.	3	3	3	3	0.00571446265193966
36.	2	1	1	2	0.00517777653240047
37.	2	2	1	0	0.00516985819282424
38.	0	1	2	2	0.00515585891583736
39.	0	0	2	1	0.00505477538362323
40.	2	1	1	0	0.00491334125601666
41.	0	1	1	2	0.00479934068536676
42.	1	2	0	0	0.00478410344028759
43.	2	1	2	1	0.00462863116672923
44.	1	2	1	2	0.00454655967825640

45.	1	2	1	0	0.00447736046304823
46.	0	1	2	1	0.00446137672729458
47.	0	2	2	2	0.00444829921968733
48.	2	2	2	0	0.00436025416007180
49.	0	0	2	0	0.00425944486972948
50.	0	2	0	0	0.00425789535002091

Table 60. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects in normal-weight BMI category NHANES 2003-2004, NHANES 2005-2006

Appendix 11. PA vocabulary of a 4-minute interval for subjects in overweight BMI category

#	Word				Probability of occurrence
1.	0	0	0	0	0.35620187943951800
2.	1	0	0	0	0.03698831253590720
3.	0	0	0	1	0.03646723640773740
4.	1	1	1	1	0.02832165730353760
5.	2	2	2	2	0.02691236450920560
6.	0	1	0	0	0.02317843424765120
7.	0	0	1	0	0.02311277017598100
8.	1	1	0	0	0.01880573980701580
9.	0	0	1	1	0.01849656014409790
10.	1	1	1	0	0.01392443820213800
11.	0	1	1	1	0.01378597574125260
12.	0	1	1	0	0.01210458846410620
13.	0	0	0	2	0.01116194548612590
14.	2	0	0	0	0.01086518378806780
15.	1	0	0	1	0.01070266900751200
16.	1	2	2	2	0.01031840908346460
17.	2	2	2	1	0.01028651438694840
18.	1	1	0	1	0.01001540947462640
19.	1	0	1	1	0.00994104998914563
20.	2	1	1	1	0.00916731056185254
21.	1	1	1	2	0.00888835436360316
22.	2	2	1	1	0.00841631995923097
23.	1	1	2	2	0.00831641216662815
24.	1	0	1	0	0.00828908342433045
25.	0	1	0	1	0.00825073285286146
26.	2	1	0	0	0.00798165224346810
27.	0	0	1	2	0.00780533157128304
28.	1	2	1	1	0.00738837958463095
29.	1	1	2	1	0.00736914231449448
30.	2	2	1	2	0.00722361401444645
31.	2	1	2	2	0.00717007606526901
32.	1	2	2	1	0.00660371746870021
33.	0	0	2	2	0.00601754571038598
34.	2	2	0	0	0.00584178592756979
35.	2	1	1	2	0.00529533759225153
36.	0	1	2	2	0.00505177301838307
37.	2	2	1	0	0.00501232949250426
38.	0	0	2	1	0.00499944189390550
39.	3	3	3	3	0.00495382774391374
40.	2	1	1	0	0.00487819056503726
41.	1	2	0	0	0.00471722893934924
42.	0	1	1	2	0.00469287320337394
43.	1	2	1	2	0.00465665354502498
44.	2	1	2	1	0.00464711770368278

45.	1	2	1	0	0.00443087799166656
46.	0	1	2	1	0.00442252665068963
47.	0	2	2	2	0.00438660889711104
48.	2	2	2	0	0.00417463424152588
49.	0	2	0	0	0.00391839214727283
50.	0	0	2	0	0.00382634003750356

Table 61. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects in overweight BMI category NHANES 2003-2004, NHANES 2005-2006

Appendix 12. PA vocabulary of a 4-minute interval for subjects in obese BMI category

#	Word				Probability of occurrence
1.	0	0	0	0	0.36374246027637600
2.	1	0	0	0	0.03773569339190380
3.	0	0	0	1	0.03727956074210320
4.	1	1	1	1	0.03026235063667010
5.	2	2	2	2	0.02493605492942790
6.	0	1	0	0	0.02426831028568420
7.	0	0	1	0	0.02393602795249000
8.	1	1	0	0	0.01912523632861600
9.	0	0	1	1	0.01880456468873870
10.	1	1	1	0	0.01452899359873090
11.	0	1	1	1	0.01432485714101480
12.	0	1	1	0	0.01228795900419870
13.	0	0	0	2	0.01141703179058900
14.	1	0	0	1	0.01095998754438350
15.	2	0	0	0	0.01080018511499350
16.	1	1	0	1	0.01042212599114250
17.	1	0	1	1	0.01037002528427210
18.	2	2	2	1	0.01029771410192870
19.	1	2	2	2	0.01014658475096680
20.	2	1	1	1	0.00955060856560905
21.	1	1	1	2	0.00927364367205677
22.	1	0	1	0	0.00857689007890008
23.	0	1	0	1	0.00847984366024030
24.	2	2	1	1	0.00845147500679643
25.	1	1	2	2	0.00831047212486584
26.	2	1	0	0	0.00793726277658691
27.	0	0	1	2	0.00785989015073875
28.	1	1	2	1	0.00766838090399510
29.	1	2	1	1	0.00758044196788017
30.	2	2	1	2	0.00694899417121577
31.	2	1	2	2	0.00694790835035855
32.	1	2	2	1	0.00681509265468315
33.	0	0	2	2	0.00620874720394168
34.	2	2	0	0	0.00580979425806409
35.	2	1	1	2	0.00527401738965212
36.	2	2	1	0	0.00524017250126812
37.	0	1	2	2	0.00507298617749977
38.	0	0	2	1	0.00503340976573620
39.	2	1	1	0	0.00492710762955739
40.	0	1	1	2	0.00474345859496103
41.	2	1	2	1	0.00468922847540553
42.	1	2	0	0	0.00465024188567611
43.	1	2	1	2	0.00460952300538837
44.	1	2	1	0	0.00449513407432010

45.	0	2	2	2	0.00440518704992655
46.	0	1	2	1	0.00440000426790640
47.	2	2	2	0	0.00414115208267222
48.	0	2	0	0	0.00375387173529866
49.	0	0	2	0	0.00369998239203491
50.	2	1	0	1	0.00365484541214327

Table 62. Top 50 words of physical activity vocabulary calculated on a 4-minute interval for subjects in obese BMI category NHANES 2003-2004, NHANES 2005-2006

Glossary

Health is a human condition with physical, social, and psychological dimensions, each characterized on a continuum with positive and negative poles.

Physical activity (PA) refers to any bodily movement produced by contraction of skeletal muscle that substantially increases energy expenditure.

The *dose (volume or amount) of PA* is characterised by frequency, intensity, duration, type and is equal to the energy expended in physical activity. The dose of PA is sometimes referred to by a FITT principle (Frequency, Intensity, Time, and Type).

Physical activity pattern is a temporal sequence of physical activities.

Bibliography

Adamo, K. B., Prince, S. a, Tricco, A. C., Connor-Gorber, S. and Tremblay, M. (2009) 'A comparison of indirect versus direct measures for assessing physical activity in the pediatric population: a systematic review.', *International journal of pediatric obesity : IJPO : an official journal of the International Association for the Study of Obesity*, 4(1), pp. 2–27. doi: 10.1080/17477160802315010.

Anokhin, A. P., Birbaumer, N., Lutzenberger, W., Nikolaev, A. and Vogel, F. (1996) 'Age increases brain complexity', *Electroencephalography and Clinical Neurophysiology*, 99(1), pp. 63–68. doi: https://doi.org/10.1016/0921-884X(96)95573-3.

Ashford, S., Edmunds, J. and French, D. P. (2010) 'What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis', *British Journal of Health Psychology*. John Wiley & Sons, Ltd (10.1111), 15(2), pp. 265–288. doi: 10.1348/135910709X461752.

Barabási, A. (2005) 'The origin of bursts and heavy tails in human dynamics', *Nature*, 435(May), pp. 207–211. doi: 10.1038/nature03526.1.

Baranowski, T., Anderson, C. and Carmack, C. (1998) 'Mediating variable framework in physical activity interventions: How are we doing? how might we do better?', *American Journal of Preventive Medicine*, 15(4), pp. 266–297. doi: https://doi.org/10.1016/S0749-3797(98)00080-4.

Bauman, A., Allman-Farinelli, M., Huxley, R. and James, W. P. T. (2008) 'Leisure-time physical activity alone may not be a sufficient public health approach to prevent obesity--a focus on China.', *Obesity reviews : an official journal of the International Association for the Study of Obesity*. England, 9 Suppl 1, pp. 119–126. doi: 10.1111/j.1467-789X.2007.00452.x.

Bauman, A. E., Reis, R. S., Sallis, J. F., Wells, J. C., Loos, R. J. F. and Martin, B. W. (2012) 'Correlates of physical activity: why are some people physically active and others not?', *Lancet*. Elsevier Ltd, 380(9838), pp. 258–71. doi: 10.1016/S0140-6736(12)60735-1.

Bauman, A. E., Sallis, J. F., Dzewaltowski, D. a and Owen, N. (2002) 'Toward a better understanding of the influences on physical activity', *American Journal of Preventive*

Medicine, 23(2), pp. 5–14. doi: 10.1016/S0749-3797(02)00469-5.

Bayevsky, R. and Ivanov, G. (2002) 'HRV Analysis under the usage of different electrocardiography systems (methodical recommendations)'. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:HRV+Analysis+under+the+usage+of+different+electrocardiography+systems+(Methodical+recommendations)#0 (Accessed: 20 February 2014).

Belcher, B. R., Berrigan, D., Dodd, K. W., Emken, B. A., Chou, C.-P. and Spruijt-Metz, D. (2010) 'Physical activity in US youth: effect of race/ethnicity, age, gender, and weight status.', *Medicine and science in sports and exercise*, 42(12), pp. 2211–2221. doi: 10.1249/MSS.0b013e3181e1fba9.

Bellettiere, J., Winkler, E. A. H., Chastin, S. F. M., Kerr, J., Owen, N., Dunstan, D. W. and Healy, G. N. (2017) 'Associations of sitting accumulation patterns with cardio-metabolic risk biomarkers in Australian adults.', *PloS one*, 12(6), p. e0180119. doi: 10.1371/journal.pone.0180119.

Bellocco, R., Jia, C., Ye, W. and Lagerros, Y. T. (2010) *Effects of physical activity, body mass index, waist-to-hip ratio and waist circumference on total mortality risk in the Swedish National March Cohort*, *European Journal of Epidemiology*. doi: http://dx.doi.org/10.1007/s10654-010-9497-6.

Besharat Pour, M., Bergström, A., Bottai, M., Kull, I., Wickman, M., Håkansson, N., Wolk, A. and Moradi, T. (2014) 'Effect of Parental Migration Background on Childhood Nutrition, Physical Activity, and Body Mass Index', *Journal of obesity*, 2014. doi: https://doi.org/10.1155/2014/406529.

Billman, G. E. (2011) 'Heart rate variability - a historical perspective.', *Frontiers in physiology*, 2(November), p. 86. doi: 10.3389/fphys.2011.00086.

Boutayeb, A. (2006) 'The double burden of communicable and non-communicable diseases in developing countries.', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(3), pp. 191–9. doi: 10.1016/j.trstmh.2005.07.021.

Bravi, A., Longtin, A. and Seely, A. J. E. (2011) 'Review and classification of variability analysis techniques with clinical applications.', *Biomedical engineering online*. BioMed Central Ltd, 10(1), p. 90. doi: 10.1186/1475-925X-10-90.

Bull, F. and the Expert Working Groups, A. (2010) *Physical Activity Guidelines in the UK: Review and Recommendations*.

Burton, L. C., Shapiro, S. and German, P. S. (1999) 'Determinants of physical activity initiation and maintenance among community-dwelling older persons', *Preventive Medicine*. Academic Press Inc., 29(5), pp. 422–430. doi: 10.1006/pmed.1999.0561.

Butcher, K., Sallis, J. F., Mayer, J. a and Woodruff, S. (2008) 'Correlates of physical activity guideline compliance for adolescents in 100 U.S. Cities.', *The Journal of adolescent health : official publication of the Society for Adolescent Medicine*, 42(4), pp. 360–8. doi: 10.1016/j.jadohealth.2007.09.025.

Caspersen, C. J., Powell, K. E. and Christenson, G. M. (1985) 'Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research.', *Public health reports (Washington, D.C. : 1974)*, 100(2), pp. 126–131.

Cavanaugh, J. T., Kochi, N. and Stergiou, N. (2010) 'Nonlinear analysis of ambulatory activity patterns in community-dwelling older adults.', *The journals of gerontology. Series A, Biological sciences and medical sciences*, 65(2), pp. 197–203. doi: 10.1093/gerona/glp144.

Cavill, N., Biddle, S. and Sallis, J. F. (2001) 'Health Enhancing Physical Activity for Young People : Statement of the United Kingdom Expert Consensus Conference', pp. 12–25.

CDC (2005) 'Trends in leisure time physical inactivity by age, sex and race/ethnicity United States - 1994–2004', *CENTERS FOR DISEASE CONTROL AND PREVENTION*, 54, pp. 991–994.

Celis-Morales, C. A., Perez-Bravo, F., Ibañez, L., Salas, C., Bailey, M. E. S. and Gill, J. M. R. (2012) 'Objective vs. self-reported physical activity and sedentary time: effects of measurement method on relationships with risk biomarkers', *PloS one*. Public Library of Science, 7(5), pp. e36345–e36345. doi: 10.1371/journal.pone.0036345.

Chaitin, G. J. (1966) 'On the Length of Programs for Computing Finite Binary Sequences', *J. ACM*. New York, NY, USA: ACM, 13(4), pp. 547–569. doi: 10.1145/321356.321363.

Chastin, S., Baker, K., Jones, D., Burn, D., Granat, M. H. and Rochester, L. (2010) 'The pattern of habitual sedentary behavior is different in advanced Parkinson's disease.', *Movement disorders : official journal of the Movement Disorder Society*, 25(13), pp. 2114–20. doi: 10.1002/mds.23146.

Chastin, S. F. M., DeCraemer, M., Oppert, J.-M. and Cardon, G. (2018) 'Dynamics of Sedentary Behaviours and Systems-Based Approach: Future Challenges and Opportunities in the Life Course Epidemiology of Sedentary Behaviours', in Leitzmann, M. F., Jochem, C., and Schmid, D. (eds) *Sedentary Behaviour Epidemiology*. Cham: Springer International Publishing, pp. 595–616. doi: 10.1007/978-3-319-61552-3_26.

Chastin, S. F. M., Egerton, T., Leask, C. and Stamatakis, E. (2015) 'Meta-analysis of the relationship between breaks in sedentary behavior and cardiometabolic health', *Obesity*. John Wiley & Sons, Ltd, 23(9), pp. 1800–1810. doi: 10.1002/oby.21180.

Chastin, S. and Granat, M. (2010) 'Methods for objective measure, quantification and analysis of sedentary behaviour and inactivity.', *Gait & posture*, 31(1), pp. 82–6. doi: 10.1016/j.gaitpost.2009.09.002.

Chastin, S., Lord, S. and Rochester, L. (2010) 'Chasing a theoretical basis for the definition and analysis of nonlinear patterns of activity in older adults.', *The journals of gerontology. Series A, Biological sciences and medical sciences*, 65(11), pp. 1254-5-8. doi: 10.1093/gerona/glq085.

Chinapaw, M. J., Wang, X., Andersen, L. B. and Altenburg, T. M. (2019) 'From Total Volume to Sequence Maps: Sophisticated Accelerometer Data Analysis.', *Medicine and science in sports and exercise*. United States, 51(4), pp. 814–820. doi: 10.1249/MSS.0000000000001849.

Church, T. S., Thomas, D. M., Tudor-Locke, C., Katzmarzyk, P. T., Earnest, C. P., Rodarte, R. Q., Martin, C. K., Blair, S. N. and Bouchard, C. (2011) 'Trends over 5 decades in U.S. occupation-related physical activity and their associations with obesity.', *PloS one*, 6(5), p. e19657. doi: 10.1371/journal.pone.0019657.

Cilibrasi, R. (2001) 'Clustering by Compression', pp. 1–28.

Condello, G., Ling, F. C. M., Bianco, A., Chastin, S., Cardon, G., Ciarapica, D., Conte, D., Cortis, C., De Craemer, M., Di Blasio, A., Gjaka, M., Hansen, S., Holdsworth, M., Iacoviello, L., Izzicupo, P., Jaeschke, L., Leone, L., Manoni, L., Menescardi, C., Migliaccio, S., Nazare, J.-A., Perchoux, C., Pesce, C., Pierik, F., Pischon, T., Polito, A., Puggina, A., Sannella, A., Schlicht, W., Schulz, H., Simon, C., Steinbrecher, A., MacDonncha, C., Capranica, L. and on behalf of the DEDIPAC consortium (2016) 'Using concept mapping in the development of the EU-PAD framework (EUropean-Physical Activity Determinants across the life course): a DEDIPAC-study', *BMC Public Health*, 16(1), p. 1145. doi: 10.1186/s12889-016-3800-8.

Cooper, A. R., Sebire, S., Montgomery, A. A., Peters, T. J., Sharp, D. J., Jackson, N., Fitzsimons, K., Dayan, C. M. and Andrews, R. C. (2012) 'Sedentary time, breaks in sedentary time and metabolic variables in people with newly diagnosed type 2 diabetes', *Diabetologia*, 55(3), pp. 589–599. doi: 10.1007/s00125-011-2408-x.

Crovella, M. E. and Bestavros, A. (1997) 'Self-Similarity in World Wide Web Tra c : Evidence and Possible Causes 1 Introduction', 5(6), pp. 835–846.

Daw, C. S. and Finney, C. E. A. (2002) 'A review of symbolic analysis of experimental data'.

Denkinger, M. D., Lukas, A., Herbolsheimer, F., Peter, R. and Nikolaus, T. (2012) 'Physical activity and other health-related factors predict health care utilisation in older adults: the ActiFE Ulm study.', *Zeitschrift für Gerontologie und Geriatrie*, 45(4), pp. 290–7. doi: 10.1007/s00391-012-0335-1.

Department of Health, Physical Activity, H. I. and P. (2011) 'Start Active, Stay Active: A report on physical activity from the four home countries' Chief Medical Officers'.

Department of Health, P. A. and Prevention, H. I. and (2004) *At least five a week. Evidence on the impact of physical activity and its relationship to health*. London.

Dogra, S., Ashe, M. C., Biddle, S. J. H., Brown, W. J., Buman, M. P., Chastin, S., Gardiner, P. A., Inoue, S., Jefferis, B. J., Oka, K., Owen, N., Sardinha, L. B., Skelton, D. A., Sugiyama, T. and Copeland, J. L. (2017) 'Sedentary time in older men and women: an international consensus statement and research priorities', *British Journal of Sports Medicine*. British Association of Sport and Excercise Medicine, 51(21), pp. 1526–1532. doi: 10.1136/bjsports-2016-097209.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., Owen, C. G., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S. and Wareham, N. J. (2017) ‘Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study’, *PLOS ONE*. Public Library of Science, 12(2), pp. 1–14. doi: 10.1371/journal.pone.0169649.

Dumith, S. C., Hallal, P. C., Reis, R. S. and Kohl, H. W. (2011) ‘Worldwide prevalence of physical inactivity and its association with human development index in 76 countries.’, *Preventive medicine*. Elsevier Inc., 53(1–2), pp. 24–8. doi: 10.1016/j.ypmed.2011.02.017.

Dumuid, D., Stanford, T. E., Martin-Fernandez, J.-A., Pedisic, Z., Maher, C. A., Lewis, L. K., Hron, K., Katzmarzyk, P. T., Chaput, J.-P., Fogelholm, M., Hu, G., Lambert, E. V, Maia, J., Sarmiento, O. L., Standage, M., Barreira, T. V, Broyles, S. T., Tudor-Locke, C., Tremblay, M. S. and Olds, T. (2018) ‘Compositional data analysis for physical activity, sedentary time and sleep research.’, *Statistical methods in medical research*. England, 27(12), pp. 3726–3738. doi: 10.1177/0962280217710835.

Eckmann, J.-P., Moses, E. and Sergi, D. (2004) ‘Entropy of dialogues creates coherent structures in e-mail traffic.’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(40), pp. 14333–7. doi: 10.1073/pnas.0405728101.

Ekelund, U., Ward, H. A., Norat, T., Luan, J., May, A. M., Weiderpass, E., Sharp, S. S., Fagherazzi, G., Overvad, K., Østergaard, J. N., Tjønneland, A., Johnsen, N. F., Mesrine, S., Trichopoulou, A., Lagiou, P., Trichopoulos, D., Li, K., Kaaks, R., Ferrari, P., Licaj, I., Jenab, M., Bergmann, M., Boeing, H., Palli, D., Sieri, S., Panico, S., Tumino, R., Agudo, A., Vineis, P., Peeters, P. H., Monnikhof, E., Bueno-de-mesquita, H. B., Ram, J., Huerta, M., Ardanaz, E., Arriola, L., Hedblad, B., Wirf, E., Sund, M., Johansson, M., Key, T. J., Travis, R. C., Khaw, K., Brage, S., Wareham, N. J. and Riboli, E. (2015) ‘Physical activity and all-cause mortality across levels of overall and abdominal adiposity in European men and women : the European Prospective Investigation into Cancer and Nutrition Study (EPIC) 1 – 6 The EPIC cohort’, (C), pp. 1–6. doi: 10.3945/ajcn.114.100065.

Eubank, S., Guclu, H., Kumar, V. S. A. and Marathe, M. V (2004) ‘Modelling disease outbreaks in realistic urban social networks’, 429(May), pp. 180–184. doi: 10.1038/nature02534.1.

Fakhouri, T. H. I., Hughes, J. P., Brody, D. J., Kit, B. K. and Ogden, C. L. (2013) ‘Physical Activity and Screen-Time Viewing Among Elementary School–Aged Children in the United States From 2009 to 2010’, *JAMA Pediatrics*, 167(3), pp. 223–229. doi:

10.1001/2013.jamapediatrics.122.

Feldman, D. P. and Crutchfield, J. P. (1998) ‘Measures of statistical complexity: Why?’, *Physics Letters A*, 238(4–5), pp. 244–252. doi: 10.1016/S0375-9601(97)00855-4.

Fielding, J. (1985) ‘Smoking: Health Effects and Control’, *New England Journal of Medicine*, 313(8), pp. 491–498. doi: 10.1056/NEJM198508223130807.

Finkelstein, E. A., Brown, D. S., Brown, D. R. and Buchner, D. M. (2008) ‘A randomized study of financial incentives to increase physical activity among sedentary older adults’, *Preventive Medicine*. Academic Press Inc., 47(2), pp. 182–187. doi: 10.1016/j.ypmed.2008.05.002.

Fortnow, L. (2001) ‘Kolmogorov Complexity’, in Downey, R. and Hirschfeldt, D. (eds) *Aspects of Complexity, Minicourses in Algorithmics, Complexity, and Computational Algebra, NZMRI Mathematics Summer Meeting, Kaikoura, New Zealand, January 7--15, 2000*. de Gruyter ({de Gruyter} Series in Logic and Its Applications).

Garcia, A. W. and King, A. C. (1991) ‘Predicting Long-term Adherence to Aerobic Exercise: A Comparison of Two Models’, *Journal of Sport and Exercise Psychology*, 13(4), pp. 394–410. doi: 10.1123/jsep.13.4.394.

Gardner, B., Smith, L., Lorencatto, F., Hamer, M. and Biddle, S. J. H. (2016) ‘How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults’, *Health Psychology Review*. Routledge, 10(1), pp. 89–112. doi: 10.1080/17437199.2015.1082146.

Glass, T. a and McAtee, M. J. (2006) ‘Behavioral science at the crossroads in public health: extending horizons, envisioning the future.’, *Social science & medicine (1982)*, 62(7), pp. 1650–71. doi: 10.1016/j.socscimed.2005.08.044.

Goldberger, A. L., Peng, C.-K. and Lipsitz, L. a (2002) ‘What is physiologic complexity and how does it change with aging and disease?’, *Neurobiology of aging*, 23(1), pp. 23–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11755014.

Grant, P. M., Ryan, C. G., Tigbe, W. W. and Granat, M. H. (2006) ‘The validation of a novel activity monitor in the measurement of posture and motion during everyday activities.’, *British journal of sports medicine*, 40(12), pp. 992–7. doi:

10.1136/bjsm.2006.030262.

Grusky, D. (1990) 'The Past , Present , and Future of Social Inequality'.

Gubbels, J. S., Van Kann, D. H., de Vries, N. K., Thijs, C. and Kremers, S. P. (2014) 'The next step in health behavior research: the need for ecological moderation analyses - an application to diet and physical activity at childcare.', *The international journal of behavioral nutrition and physical activity*, 11(1), p. 52. doi: 10.1186/1479-5868-11-52.

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W. and Ekelund, U. (2012) 'Global physical activity levels: surveillance progress, pitfalls, and prospects.', *Lancet*. Elsevier Ltd, 380(9838), pp. 247–57. doi: 10.1016/S0140-6736(12)60646-1.

Haskell, W. L., Lee, I., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., Macera, C. A., Heath, G. W., Thompson, P. D., Bauman, A. and others (2007) 'Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association', *Medicine and science in sports and exercise*. WILLIAMS & WILKINS, 39(8), p. 1423.

Hausdorff, J. M. (2005) 'Gait variability : methods , modeling and meaning Example of Increased Stride Time Variability in Elderly Fallers Quantification of Stride-to-Stride Fluctuations', 9, pp. 1–9. doi: 10.1186/1743-Received.

Hawkins, M. S., Storti, K. L., Richardson, C. R., King, W. C., Strath, S. J., Holleman, R. G. and Kriska, A. M. (2009) 'Objectively measured physical activity of USA adults by sex, age, and racial/ethnic groups: a cross-sectional study.', *The international journal of behavioral nutrition and physical activity*, 6, p. 31. doi: 10.1186/1479-5868-6-31.

Healy, G. N., Matthews, C. E., Dunstan, D. W., Winkler, E. a H. and Owen, N. (2011) 'Sedentary time and cardio-metabolic biomarkers in US adults: NHANES 2003-06.', *European heart journal*, 32(5), pp. 590–7. doi: 10.1093/eurheartj/ehq451.

Howley, E. T. (2001) 'Type of activity: resistance, aerobic and leisure versus occupational physical activity.', *Medicine and science in sports and exercise*, 33(6 Suppl), pp. S364-9-20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11427761.

Huffman, D. A. (1951) 'Minimum-Redundancy Codes *', 27.

Janz, K. F. (2006) 'Physical activity in epidemiology: moving from questionnaire to objective measurement.', *British journal of sports medicine*. England, 40(3), pp. 191–192. doi: 10.1136/bjsm.2005.023036.

John, D. and Freedson, P. (2012) 'ACTIGRAPH AND ACTICAL PHYSICAL ACTIVITY MONITORS: A PEEK UNDER THE HOOD', *Medicine and science in sports and exercise*, pp. S86-9. doi: 10.1249/MSS.0b013e3182399f5e.

Kaczynski, A. T., Manske, S. R., Mannell, R. C. and Grewal, K. (2008) 'Smoking and Physical Activity: A Systematic Review', *American Journal of Health Behavior*, 32(1), pp. 93–110.

Kaewthummanukul, T. and Brown, K. C. (2006) 'Determinants of employee Participation in Physical Activity: Critical Review of the Literature', *AAOHN*, 54(6).

Keadle, S. K., McKinnon, R., Graubard, B. I. and Troiano, R. P. (2016) 'Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys', *Preventive Medicine*, 89, pp. 37–43. doi: https://doi.org/10.1016/j.ypmed.2016.05.009.

Kenney, E. L. and Gortmaker, S. L. (2017) 'United States Adolescents' Television, Computer, Videogame, Smartphone, and Tablet Use: Associations with Sugary Drinks, Sleep, Physical Activity, and Obesity', *The Journal of Pediatrics*, 182, pp. 144–149. doi: https://doi.org/10.1016/j.jpeds.2016.11.015.

Kesaniemi, A., Danforth, E., Jensen, M., Kopelman, P., Lefebvre, P. and Reeder, B. (2001) 'Dose-response issues concerning physical activity and health: an evidence-based symposium', *Medicine & Science in Sports & Exercise*, pp. 351–358.

King, A. C., Castro, C., Wilcox, S., Eyler, A. A., Sallis, J. F. and Brownson, R. C. (2000) 'Personal and Environmental Factors Associated With Physical Inactivity Among Different Racial-Ethnic Groups of U . S . Middle-Aged and Older-Aged Women', 19, pp. 354–364. doi: 10.1037/AB78.

King, A. C., Kiernan, M., Oman, R. F., Kraemer, H. C., Hull, M. and Ahn, D. (1997) 'Can we identify who will adhere to long-term physical activity? Signal detection methodology as a potential aid to clinical decision making.', *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*. United States, 16(4), pp. 380–389.

Klir, G. (1988) 'A review of: "DEALING WITH COMPLEXITY: AN INTRODUCTION TO THE THEORY AND APPLICATION OF SYSTEMS SCIENCE", by Robert L. Flood and Ewart R. Carson. Plenum Press, New York and London, 1988. XV + 288 pp.', *International Journal of General Systems*, 14(3), p. 285. doi: 10.1080/03081078808935014.

Knuth, A. and Hallal, P. (2009) 'Temporal trends in physical activity: a systematic review', *J Phys Act Health.*, 6(5), pp. 548–59.

Koeneman, M. a, Verheijden, M. W., Chinapaw, M. J. M. and Hopman-Rock, M. (2011) 'Determinants of physical activity and exercise in healthy older adults: a systematic review.', *The international journal of behavioral nutrition and physical activity*, 8, p. 142. doi: 10.1186/1479-5868-8-142.

Kohl, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G. and Kahlmeier, S. (2012) 'The pandemic of physical inactivity: global action for public health.', *Lancet*. Elsevier Ltd, 380(9838), pp. 294–305. doi: 10.1016/S0140-6736(12)60898-8.

Kolmogorov, A. N. (1965) 'Three approaches to the definition of the concept "quantity of information"', *Problems Inform. Transmission*, 1(1), pp. 3–11.

Kolmogorov, A. N. (1983) 'Combinatorial foundations of information theory and the calculus of probabilities', *Russ. Math. Surv.*, 38, pp. 29–40.

Kristensen, P. L. and Møller, N. C. and Korsholm, L. and Wedderkopp, N. and Andersen, L. B. and Froberg, K. (2008) 'Tracking of objectively measured physical activity from childhood to adolescence: The European youth heart study', *Scandinavian Journal of Medicine & Science in Sports*, 18(2), pp. 171--178. doi: 10.1111/j.1600-0838.2006.00622.x.

LaCroix, A. Z., Guralnik, J. M., Berkman, L. F., Wallace, R. B. and Satterfield, S. (1993) 'Maintaining Mobility in Late Life. II. Smoking, Alcohol Consumption, Physical Activity, and Body Mass Index', *American Journal of Epidemiology*, 137(8), pp. 858–869. doi: 10.1093/oxfordjournals.aje.a116747.

Lee, I.-M. and Shiroma, E. J. (2014) 'Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges.', *British journal of sports medicine*, 48(3), pp. 197–201. doi: 10.1136/bjsports-2013-093154.

Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N. and Katzmarzyk, P. T. (2012)

‘Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy.’, *Lancet*. Elsevier Ltd, 380(9838), pp. 219–29. doi: 10.1016/S0140-6736(12)61031-9.

Lee, I. M. and Skerrett, P. J. (2001) ‘Physical activity and all-cause mortality: what is the dose-response relation?’, *Medicine and science in sports and exercise*, 33(6 Suppl), pp. S459-71–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11427772.

Lempel, A. and Ziv, J. (1976) ‘On the Complexity of Finite Sequences’, *IEEE Transactions on Information Theory*, 22(1), pp. 75–81. doi: 10.1109/TIT.1976.1055501.

Lipsitz, L. (2004) *Physiological Complexity, Aging, and the Path to Frailty*, *Science of aging knowledge environment : SAGE KE*. doi: 10.1126/sageke.2004.16.pe16.

Lochbaum, M. R., Rhodes, R. E., Stevenson, S. J., Surles, J., Stevens, T. and Wang, C. K. J. (2010) ‘Does gender moderate the exercising personality? An examination of continuous and stage-based exercise.’, *Psychology, health & medicine*, 15(1), pp. 50–60. doi: 10.1080/13548500903443449.

Lorenz, E. N. (1963) ‘Deterministic Nonperiodic Flow’, *Journal of the Atmospheric Sciences*. American Meteorological Society, 20(2), pp. 130–141. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Luxwolda, M. F., Kuipers, R. S., Kema, I. P., van der Veer, E., Dijck-Brouwer, D. A. J. and Frits, M. A. J. (2013) ‘Vitamin D status indicators in indigenous populations in East Africa’, *European Journal of Nutrition*, 52(3), pp. 1115–1125.

Matthews, C. E., Ainsworth, B. E., Thompson, R. W. and Bassett, D. R. (2002) ‘Sources of variance in daily physical activity levels as measured by an accelerometer.’, *Medicine and science in sports and exercise*, 34(8), pp. 1376–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12165695.

Matthews, C. E., Chen, K. Y., Freedson, P. S., Buchowski, M. S., Beech, B. M., Pate, R. R. and Troiano, R. P. (2008) ‘Amount of time spent in sedentary behaviors in the United States, 2003-2004.’, *American journal of epidemiology*, 167(7), pp. 875–81. doi: 10.1093/aje/kwm390.

McAuley, E. and Blissmer, B. (2000) ‘Self-efficacy determinants and consequences of physical activity', *Exerc Sport Sci Rev*, 28(2), pp. 85–88.

McAuley, E., Morris, K. S., Motl, R. W., Hu, L., Konopack, J. F. and Elavsky, S. (2007) 'Long-term follow-up of physical activity behavior in older adults.', *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*. United States, 26(3), pp. 375–380. doi: 10.1037/0278-6133.26.3.375.

McKinley, M. B. E. L. ; P. J. R. ; J. M. W. W. ; M. C. (2003) 'How active are we? Levels of routine physical activity in children and adults', *Proceedings of the Nutrition Society*, 7(62), pp. 681–701.

McVeigh, J. A., Winkler, E. A. H., Howie, E. K., Tremblay, M. S., Smith, A., Abbott, R. A., Eastwood, P. R., Healy, G. N. and Straker, L. M. (2016) 'Objectively measured patterns of sedentary time and physical activity in young adults of the Raine study cohort', *International Journal of Behavioral Nutrition and Physical Activity*, 13(1), p. 41. doi: 10.1186/s12966-016-0363-0.

Moore, S. C., Patel, A. V., Matthews, C. E., Berrington de Gonzalez, A., Park, Y., Katki, H. a., Linet, M. S., Weiderpass, E., Visvanathan, K., Helzlsouer, K. J., Thun, M., Gapstur, S. M., Hartge, P. and Lee, I. M. (2012) 'Leisure Time Physical Activity of Moderate to Vigorous Intensity and Mortality: A Large Pooled Cohort Analysis', *PLoS Medicine*, 9(11), pp. 1–14. doi: 10.1371/journal.pmed.1001335.

Moses, J. (2010) 'Flexibility and Its Relation to Complexity and Architecture', in Aiguier, M., Bretaudeau, F., and Krob, D. (eds) *Complex Systems Design {&} Management: Proceedings of the First International Conference on Complex System Design {&} Management CSDM 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 197–206. doi: 10.1007/978-3-642-15654-0_14.

Nitz, J. C. and Choy, N. L. L. (2007) 'Changes in activity level in women aged 40-80 years.', *Climacteric : the journal of the International Menopause Society*. England, 10(5), pp. 408–415. doi: 10.1080/13697130701624955.

O'brien, R. M. (2007) 'A Caution Regarding Rules of Thumb for Variance Inflation Factors', *Quality & Quantity*, 41(5), pp. 673–690. doi: 10.1007/s11135-006-9018-6.

Orleans, C. (2000) 'Promoting the maintenance of health behavior change: Recommendations for the next generation of research and practice', *Health psychology :*

official journal of the Division of Health Psychology, American Psychological Association, 19, pp. 76–83. doi: 10.1037//0278-6133.19.Suppl1.76.

Owen, N. (1996) ‘Strategic initiatives to promote participation in physical activity’, II(3), pp. 213–218.

Owen, N., Healy, G. N., Matthews, C. E. and Dunstan, D. W. (2010) ‘Too much sitting: the population health science of sedentary behavior.’, *Exercise and sport sciences reviews*, 38(3), pp. 105–13. doi: 10.1097/JES.0b013e3181e373a2.

Owen, N., Sugiyama, T., Eakin, E. E., Gardiner, P. a, Tremblay, M. S. and Sallis, J. F. (2011) ‘Adults’ sedentary behavior determinants and interventions.’, *American journal of preventive medicine*. Elsevier Inc., 41(2), pp. 189–96. doi: 10.1016/j.amepre.2011.05.013.

Paavola, M., Vartiainen, E. and Haukkala, A. (2004) ‘Smoking, alcohol use, and physical activity: A 13-year longitudinal study ranging from adolescence into adulthood’, *Journal of Adolescent Health*, 35(3), pp. 238–244. doi: https://doi.org/10.1016/j.jadohealth.2003.12.004.

Paffenbarger, R. J., Hyde, R., Wing, A., Lee, I., Jung, D. and Kampert, J. (1993) ‘The association of changes in physical-activity level and other lifestyle characteristics with mortality among men’, *N Engl J Med*, 25(328(8)), pp. 538–45.

Palacios-Ceña, D., Alonso-Blanco, C., Jiménez-Garcia, R., Hernández-Barrera, V., Carrasco-Garrido, P., Pileño-Martinez, E. and Fernández-de-las-Peñas, C. (2011) ‘Time trends in leisure time physical activity and physical fitness in elderly people: 20 year follow-up of the Spanish population national health survey (1987-2006)’, *BMC Public Health*, p. 799. doi: 10.1186/1471-2458-11-799.

Paraschiv-Ionescu, A., Perruchoud, C., Buchser, E. and Aminian, K. (2012) ‘Barcoding human physical activity to assess chronic pain conditions.’, *PloS one*, 7(2), p. e32239. doi: 10.1371/journal.pone.0032239.

Paraschiv-Ionescu, a., Buchser, E., Rutschmann, B. and Aminian, K. (2008) ‘Nonlinear analysis of human physical activity patterns in health and disease’, *Physical Review E*, 77(2), p. 21913. doi: 10.1103/PhysRevE.77.021913.

Pearson, O. R., Busse, M. E., Deursen, R. W. M. V. A. N. and Wiles, C. M. (2004) ‘Review

Quantification of walking mobility in neurological disorders', 97(8), pp. 463–475. doi: 10.1093/qjmed/hch084.

Plonczynski, D. J. (2003) 'Physical Activity Determinants Of Older Women: What Influences Activity?', 12(4), pp. 213–222.

Pratt, M., Sarmiento, O. L., Montes, F., Ogilvie, D., Marcus, B. H., Perez, L. G. and Brownson, R. C. (2012) 'The implications of megatrends in information and communication technology and transportation for changes in global physical activity.', *Lancet*. Elsevier Ltd, 380(9838), pp. 282–93. doi: 10.1016/S0140-6736(12)60736-3.

Prince, S. a, Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S. and Tremblay, M. (2008) 'A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review.', *The international journal of behavioral nutrition and physical activity*, 5, p. 56. doi: 10.1186/1479-5868-5-56.

Rechtin, E. and Maier, M. (2000) *The Art of Systems Architecting, Second Edition*, *Art of Systems Architecting*. CRC Press (Systems Engineering). doi: 10.1201/9781420058529.

Reed, M. and Langford, J. (2013) *Data to life*.

Rhodes, R. E. and Smith, N. E. I. (2006) 'Personality correlates of physical activity: a review and meta-analysis.', *British journal of sports medicine*, 40(12), pp. 958–65. doi: 10.1136/bjsm.2006.028860.

Riddoch, C. J., Bo Andersen, L., Wedderkopp, N., Harro, M., Klasson-Heggebø, L., Sardinha, L. B., Cooper, A. R. and Ekelund, U. (2004) 'Physical activity levels and patterns of 9- and 15-yr-old European children.', *Medicine and science in sports and exercise*, 36(1), pp. 86–92. doi: 10.1249/01.MSS.0000106174.43932.92.

Riddoch, C. J., Leary, S. D., Ness, a. R., Blair, S. N., Deere, K., Mattocks, C., Griffiths, a., Davey Smith, G. and Tilling, K. (2009) 'Prospective associations between objective measures of physical activity and fat mass in 12-14 year old children: the Avon Longitudinal Study of Parents and Children (ALSPAC)', *Bmj*, 339(nov26 2), pp. b4544–b4544. doi: 10.1136/bmj.b4544.

Riddoch, C. J., Mattocks, C., Deere, K., Saunders, J., Kirkby, J., Tilling, K., Leary, S. D., Blair, S. N. and Ness, A. R. (2007) 'Objective measurement of levels and patterns of physical activity.', *Archives of disease in childhood*, 92(11), pp. 963–9. doi: 10.1136/adc.2006.112136.

Room, R., Babor, T. and Rehm, J. (2005) 'Alcohol and public health', *The Lancet*, 365(9458), pp. 519–530. doi: https://doi.org/10.1016/S0140-6736(05)17870-2.

Rowland, T. W. (1998) 'The biological basis of physical activity', *Medicine and science in sports and exercise*, 30(3), p. 392—399. doi: 10.1097/00005768-199803000-00009.

Rybski, D., Buldyrev, S. V, Havlin, S., Liljeros, F. and Makse, H. a (2009) 'Scaling laws of human interaction activity.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), pp. 12640–5. doi: 10.1073/pnas.0902667106.

Saffer, H., Dave, D., Grossman, M. and Leung, L. A. (2013) 'Racial, Ethnic, and Gender Differences in Physical Activity.', *Journal of human capital*, 7(4), pp. 378–410. doi: 10.1086/671200.

Sallis, J. F. (2000) 'Age-related decline in physical activity: a synthesis of human and animal studies.', *Medicine and science in sports and exercise*. United States, 32(9), pp. 1598–1600.

Sallis, J. F., Bowles, H. R., Bauman, A., Ainsworth, B. E., Bull, F. C., Craig, C. L., Sjöström, M., De Bourdeaudhuij, I., Lefevre, J., Matsudo, V., Matsudo, S., Macfarlane, D. J., Gomez, L. F., Inoue, S., Murase, N., Volbekiene, V., McLean, G., Carr, H., Heggebo, L. K., Tomten, H. and Bergman, P. (2009) 'Neighborhood environments and physical activity among adults in 11 countries.', *American journal of preventive medicine*. American Journal of Preventive Medicine, 36(6), pp. 484–90. doi: 10.1016/j.amepre.2009.01.031.

Sallis, J. F., Prochaska, J. J. and Taylor, W. C. (2002) 'A review of correlates of physical activity of children and adolescents', *Med Sci Sports Exerc*, 32(963).

Sallis JF, Owen N, F. E. (2008) 'Ecological models of health behavior', *Glanz K, Rimer BK, and Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice*, pp. 465–482.

Schneller, M. B., Pedersen, M. T., Gupta, N., Aadahl, M. and Holtermann, A. (2015) 'Validation of five minimally obstructive methods to estimate physical activity energy expenditure in young adults in semi-standardized settings.', *Sensors (Basel, Switzerland)*.

Multidisciplinary Digital Publishing Institute, 15(3), pp. 6133–51. doi: 10.3390/s150306133.

Schutz, Y., Weinsier, R. L. and Hunter, G. R. (2001) 'Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures.', *Obesity research*, 9(6), pp. 368–79. doi: 10.1038/oby.2001.48.

Senge, P. M. (1997) 'THE FIFTH DISCIPLINE', *Measuring Business Excellence*, 1(3), pp. 46–51. doi: 10.1108/eb025496.

Sequeira, M. M., Rickenbach, M., Wietlisbach, V., Tullen, B. and Schutz, Y. (1995) 'Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey.', *American journal of epidemiology*. UNITED STATES, 142(9), pp. 989–999.

Shaffer, F. and Ginsberg, J. P. (2017) 'An Overview of Heart Rate Variability Metrics and Norms', *Frontiers in Public Health*, 5, p. 258. doi: 10.3389/fpubh.2017.00258.

Shannon (1948) 'Mathematical theory of communication.pdf', *The Bell System Technical Journal*.

Shaw, B. A. and Spokane, L. (2008) 'Examining the Association Between Education Level and Physical Activity Changes During Early Old Age', *J Aging Health.*, 20(7), pp. 767–787. doi: 10.1177/0898264308321081.Examining.

Shephard, R. J. (2003) 'Limits to the measurement of habitual physical activity by questionnaires'.

Sparling, P. B., Howard, B. J., Dunstan, D. W. and Owen, N. (2015) 'Recommendations for physical activity in older adults', *BMJ*. BMJ Publishing Group Ltd, 350. doi: 10.1136/bmj.h100.

Stokols, D. (1996) 'Translating socio economical thoery into guidelines.PDF', *Am J Health Promt*, 10(4), pp. 282–98.

Sussman (1999) 'The New Transportation Faculty: The Evolution to Engineering Systems.', *Transportation Quarterly*.

Telford, R. M., Telford, R. D., Cunningham, R. B., Cochrane, T., Davey, R. and Waddington, G. (2013) 'Longitudinal patterns of physical activity in children aged 8 to 12 years: the LOOK study.', *The international journal of behavioral nutrition and physical activity*, 10, p. 81. doi: 10.1186/1479-5868-10-81.

Thorburn, a W. and Proietto, J. (2000) 'Biological determinants of spontaneous physical activity.', *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 1, pp. 87–94.

Tremblay, M. S., Warburton, D. E. R., Janssen, I., Paterson, D. H., Latimer, A. E., Rhodes, R. E., Kho, M. E., Hicks, A., Leblanc, A. G., Zehr, L., Murumets, K. and Duggan, M. (2011) 'New Canadian physical activity guidelines.', *Applied physiology, nutrition, and metabolism = Physiologie appliquée, nutrition et métabolisme*, 36(1), pp. 36-46-58. doi: 10.1139/H11-009.

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T. and McDowell, M. (2008a) 'Physical activity in the United States measured by accelerometer.', *Medicine and science in sports and exercise*, 40(1), pp. 181–8. doi: 10.1249/mss.0b013e31815a51b3.

Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T. and McDowell, M. (2008b) 'Physical activity in the United States measured by accelerometer.', *Medicine and science in sports and exercise*, 40(1), pp. 181–8. doi: 10.1249/mss.0b013e31815a51b3.

Trost, S. G., Owen, N., Bauman, A. E., Sallis, J. F. and Brown, W. (2002) 'Correlates of adults' participation in physical activity: review and update.', *Medicine and science in sports and exercise*, 34(12), pp. 1996–2001. doi: 10.1249/01.MSS.0000038974.76900.92.

Tryon, W. W. and Williams, R. (1996) 'Fully proportional actigraphy: A new instrument', *Behavior Research Methods, Instruments, & Computers*, 28(3), pp. 392–403. doi: 10.3758/BF03200519.

Tudor-Locke, C., Brashear, M. M., Johnson, W. D. and Katzmarzyk, P. T. (2010) 'Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women.', *The international journal of behavioral nutrition and physical activity*, 7, p. 60. doi: 10.1186/1479-5868-7-60.

Tudor-Locke, C., Camhi, S. M. and Troiano, R. P. (2012) 'A catalog of rules, variables, and definitions applied to accelerometer data in the National Health and Nutrition Examination Survey, 2003-2006.', *Preventing chronic disease*, 9, p. E113. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3457743&tool=pmcentrez&r

endertype=abstract.

Tudor-Locke, C. E. and Myers, a M. (2001) ‘Challenges and opportunities for measuring physical activity in sedentary adults.’, *Sports medicine (Auckland, N.Z.)*, 31(2), pp. 91–100. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11227981.

Twisk, J. W. (2001) ‘Physical activity guidelines for children and adolescents: a critical review.’, *Sports medicine (Auckland, N.Z.)*, 31(8), pp. 617–27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11475323.

Uijtdewilligen, L., Twisk, J. W. R., Singh, A. S., Chinapaw, M. J. M., van Mechelen, W. and Brown, W. J. (2014) ‘Biological, socio-demographic, work and lifestyle determinants of sitting in young adult women: a prospective cohort study.’, *The international journal of behavioral nutrition and physical activity*, 11, p. 7. doi: 10.1186/1479-5868-11-7.

Vaillancourt, D. E. and Newell, K. M. (2002) ‘Changing complexity in human behavior and physiology through aging and disease’, *Neurobiology of Aging*, 23(1), pp. 1–11. doi: https://doi.org/10.1016/S0197-4580(01)00247-0.

Wanner, M., Richard, A., Martin, B., Linseisen, J. and Rohrmann, S. (2015) ‘Associations between objective and self-reported physical activity and vitamin D serum levels in the US population’, *Cancer Causes & Control*, 26(6), pp. 881–891. doi: 10.1007/s10552-015-0563-y.

Wen, C. P., Wai, J. P. M., Tsai, M. K., Yang, Y. C., Cheng, T. Y. D., Lee, M. C., Chan, H. T., Tsao, C. K., Tsai, S. P. and Wu, X. (2011) ‘Minimum amount of physical activity for reduced mortality and extended life expectancy: A prospective cohort study’, *The Lancet*, 378(11), pp. 1244–1253. doi: 10.1016/S0140-6736(11)60749-6.

WHO (2010) *Global recommendations on physical activity for health. World Health Organization*. Geneva, Switzerland.

WHO (2017) *Noncommunicable Diseases Progress Monitor 2017*. doi: 978 92 4 151302 9.

Wijndaele, K., Westgate, K., Stephens, S. K., Blair, S. N., Bull, F. C., Chastin, S. F., Dunstan, D. W., Ekelund, U., Esliger, D. W., Freedson, P. S., Granat, M. H., Matthews, C. E., Owen, N., Rowlands, A. V, Sherar, L. B., Tremblay, M. S., Troiano, R. P., Brage, S. and Healy,

G. N. (2015) ‘Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus.’, *Medicine and science in sports and exercise*. doi: 10.1249/MSS.0000000000000661.

Wilkinson, A. V, Miller, E. E., Koehly, L. M., Daniel-MacDougall, C. R. and Forman, M. R. (2017) ‘Correlates of Physical Activity Differ by Sex and Country of Birth Among Mexican-Heritage Youth’, *Journal of Immigrant and Minority Health*. Springer New York, 19(2), pp. 246–253. doi: 10.1007/s10903-016-0451-x.

World Health Organization (2010) *Global status report on noncommunicable diseases*. doi: 978 92 4 156422 9.

Yang, C.-C. and Hsu, Y.-L. (2010) ‘A review of accelerometry-based wearable motion detectors for physical activity monitoring.’, *Sensors (Basel, Switzerland)*, 10(8), pp. 7772–88. doi: 10.3390/s100807772.

Yasunaga, A., Togo, F., Watanabe, E., Park, H., Park, S., Shephard, R. J. and Aoyagi, Y. (2008) ‘Sex, age, season, and habitual physical activity of older Japanese: The Nakanojo study’, *Journal of Aging and Physical Activity*, 16, pp. 3–13.

Young, D., Haskell, W., Jatulis, D. and Fortmann, S. (1993) ‘Associations between Changes in Physical Activity and Risk Factors for Coronary Heart Disease in a Community-based Sample of Men and Women: The Stanford Five-City Project’, *Am. J. Epidemiol.*, 138(4), pp. 205–216.

Zhou, T., Kiet, H. a. T., Kim, B. J., Wang, B.-H. and Holme, P. (2008) ‘Role of activity in human dynamics’, *EPL (Europhysics Letters)*, 82(2), p. 28002. doi: 10.1209/0295-5075/82/28002.

Zhu, Z., Tang, Y., Zhuang, J., Liu, Y., Wu, X., Cai, Y., Wang, L., Cao, Z.-B. and Chen, P. (2019) ‘Physical activity, screen viewing time, and overweight/obesity among Chinese children and adolescents: an update from the 2017 physical activity and fitness in China—the youth study’, *BMC Public Health*, 19(1), p. 197. doi: 10.1186/s12889-019-6515-9.

Ziv, J. and Lempel, A. (1977) ‘A universal algorithm for sequential data compression’, *IEEE Transactions on Information Theory*, 23(3), pp. 337–343. doi: 10.1109/TIT.1977.1055714.

Index