Helge Sveindal Rosfjord

New metrics for analysis and presentation of device-based indices of physical activity

Exploration of a novel way of reducing accelerometer data into analytical and translational indices of physical activity.

Master thesis
Department of Sport Science and Physical Education
Faculty of Health and Sport Sciences
University of Agder, 2021

Summary

Background: Device-based measurements of human behavior are becoming the standard tool to collect precise, valid, and free-living indices of habitual physical activity. However, differences between brands and lack of consensus regarding data reduction strategies, hampers between and within study comparisons. New metrics such as the analytical (average acceleration and intensity gradient) and translational metrics (MX) proposed by Rowlands et al. (2019a) might overcome some of these limitations. Currently the new metrics have not been applied to large representative data sets and comparisons of measurement properties between traditional and new accelerometer metrics are scarce.

Purpose: Provide normative new metrics for population estimates of PA and compare measurement properties of new and traditional count-based measures of physical activity.

Method: A randomly selected cross-sectional sample of 5052 Norwegian adults and older people hip-wore the ActiGraph GT3X+ for seven consecutive days (secondary data analysis using data from the Kan2 survey carried out in 2014-2015) and 3622 were included in the analytical sample. Open-source software (GGIR) was applied to produce the new metrics (analytical and translational metrics). Specific algorithms in RStudio were used to produce visual radar plots and correlation and regression analyses were applied to compare traditional and new metrics and the predictive properties of the two approaches.

Results: New metrics display a negative gradient across age and BMI and a positive gradient across level of education. Comparison between accelerometer signals revealed a modest agreement (r = .460) between new and traditional new and traditional metrics, a 45 % difference in accumulated minutes of MVPA (new metrics showing higher values), and equally good predictions of BMI for intensity gradient and counts per minute (R^2 = 0.062).

Conclusion: We display normative values of new metrics for device-based indices of physical activity in a representative sample of Norwegian adults and older, showing the feasibility and usefulness of this approach. Furthermore, the new approach provides a way of presenting data in a way that is interpretable to a wider audience. Lastly, the appropriateness of the new metrics is further strengthened by the similar predictive properties compared to the traditional metrics. Given the new approach's ability to overcome central limitations of device-based measurements of physical activity, this might be a feasible way forward in order to aid comparability between studies of physical activity and health.

Keywords: accelerometer-assessed physical activity, analytical metrics, translational metrics, average acceleration, intensity gradient (IG), MX metrics, and counts per minute (CPM).

Sammendrag

Bakgrunn: Objektive målinger av fysisk aktivitet har blitt standardverktøyet for å registrere fysisk aktivitet på en valid og reliabel måte. På tross av dette kompliseres sammenlikninger av resultater fra ulike studier som har benyttet slike metoder på grunn av ulike måleapparater samt mangel på konsensus om redusering av rådata fra objektive målemetoder til analyserbare variabler. Nye beregningsmetoder som "analytical metrics" og "translational metrics" er foreslått av Rowlands et al. (2019a) for å øke sammenlikningsevne og minimere utfordringer tilknyttet nåværende praksis. Foreløpig har ikke de nye metodene blitt benyttet i store epidemiologiske studier, og sammenlikninger mellom tradisjonelle og nye beregningsmetoder for reduksjon av data fra objektive målemetoder er begrenset.

Formål: Å benytte nye analysemetoder for framstilling av resultater av objektivt målt fysisk aktivitetsnivå på populasjonsnivå og sammenligne nye og tradisjonelle målemetoders egenskaper til å måle fysisk aktivitet.

Metode: Et tilfeldig utvalg på 5052 norske voksne og eldre hadde på seg ActiGraph GT3X+ i syv påfølgende dager (sekundær datanalyse av Kan2-undersøkelsen utført i 2014-2015) hvor 3622 ble inkludert i de endelige analysene. Statistikkprogrammet R med tilleggspakken GGIR ble benyttet for å produsere datasett med ulike målevariabler (analytical og translational metric), og videre visualisert ved hjelp av radarplott. Korrelasjons- og regresjonsanalyser ble brukt for å sammenlikne metoder og deres evne til å prediktere et helseutfall (kroppsmasseindeks).

Resultater: De nye analysemetodene viser en negativ gradient for alder og kroppsmasseindeks og en positiv gradient på tvers av utdanningsnivå. Sammenlikninger mellom akselerometersignaler redusert på ny og tradisjonell måte avslørte en beskjeden sammenheng (r = .460) mellom måleestimater, en 45 % forskjell i akkumulert moderat til høy fysisk aktivitet (høyere verdier med nye målemetoder), og like gode evner til å predikere BMI for ny metode (intensitetsgradient) sammenliknet med tradisjonell metode (tellinger per minutt) (R^2 = 0.062).

Konklusjon: Vi framstiller populasjonsnivåer av objektivt målt fysisk aktivitet ved hjelp av nye metrics, og viser muligheten og nytten av den nye tilnærmingen til reduksjon av rådata fra objektive målemetoder for fysisk aktivitet. Videre er den nye tilnærmingen letter å tolke for et bredere publikum, og den nye metoden viser tilnærmet lik evne til å forutse et helseutfall som tradisjonell metode. Gitt den nye tilnærmingens evne til å potensielt begrense eksisterende svakheter ved eksisterende metoder, kan nye metrics være en vei fremover mot å øke sammenliknbarheten mellom studier av objektivt registrert fysisk aktivitet og helse.

Nøkkelord: akselerometerbasert estimater av fysisk aktivitet, analytical metrics, translational metrics, gjennomsnitt av akseleresjon (average acceleration), intensitetsgradient (IG), MX metrics og tellinger per minute (CPM)

Acknowledgments

As a wise man once said, “In order to achieve whatever the goals might be, you have to eat one chocolate piece at a time”. I remembered the phrase after the new personal record at 200 m and was on the way home with my father. As a young athlete, I did not reflect much of it other than chocolate in the metaphor. In later years, and especially the last year, I finally found the metaphor meaningful and interpretable after late nights, numerous hours confused, and increase body weight during 2020: be tolerant. Ignore the many hours alone while writing the master thesis with the sun shining through the windows. The times were SPSS acts like the dark side of the force, and you must find the Death Star plans alone. Alternatively, the times were Rstudio fail to rerun the algorithms, and again, and again, and again. So keep calm and carry on!

The last year writes itself in eternity with the help of friends, family, supervisors, and the brilliant mind behind delivering food at home. Primarily, would I like to thank my supervisors: Bjørge Herman Hansen and Monica Klungland Torstveit. In their academic writing contribution, knowledgeable feedback, and encouragement, I have succeeded two academic degrees at the University of Agder. With interest in quantitative research and population-based measuring of PA, I will forever be thankful for having you guys as supervisors.

Secondly, thanks to Alex V. Rowlands (University of Leicester) for comprehending raw accelerometry into readable analytical and translational metrics and guide my way through RStudio. Thirdly, thanks to the primary investigator of the Kan2 survey, Sigmund Alfred Andersen, for the use of Kan2 data. Finally, I would like to thank my fiancé, Randi Victoria, and my parents, Christine and Helge Alvin, for everything during 2020-2021.

Sincerely,

Helge Sveindal Rosfjord
Kristiansand 28.05.2021

Abbreviations

Analytical metrics	Interaction between average acceleration and intensity gradient
AC	Activity counts
ActiGraph GT3X+	Accelerometer brand and model
Av accel	Average acceleration, most active continuous 15 h of a day
Average intensity	Average acceleration
Bouts	Consecutive timeframe of activity with same the same activity
BMI	Body mass index (weight/(height*height)
BMI, Normal	Below 25.0 kg/m^2
BMI, Overweight	Within 25.0-29.9 kg/m^2
BMI, Obese	Everything above 30.0 kg/m^2
β	Beta coefficient
EE	Energy expenditure
E.g.	Exempli gratia, or, for example
Epochs	Time intervals used to assemble acceleration.
CI	Confidence interval
CPM	Counts per minutes
Cut-points	Intensity thresholds determining accelerometer-assessed PA
IG	Intensity Gradient
I.e.	Id est, or that is
Kan2	Norwegian survey from 2014-2015
Mg	Milli-gravitational units
Min	Minutes
ML	Machine learning approaches
MVPA	Moderate to vigorous physical activity
MX metrics	Acceleration above a person's most active accumulated minutes
Non-wear time	Time where the accelerometer is considered not in use
PA	Physical activity
SD	Standard deviation
SEM	Standard Error
SB	Sedentary behavior
Translation metrics	MX metrics
MX metrics	Acceleration above which a person's most active non-consecutive X minutes are spent over the day
VPA	Vigorous physical activity
Wear-time	Time an accelerometer is considered in the use

List of tables

Table 1. Norwegian guidelines for physical activity and sedentary behavior for Norwegian adults.

Table 2. Definitions of new (analytical and translational) and traditional (CPM) metrics and their determination of MVPA and PA recommendations.

Table 3. Comparison between two validation variables (10 and 12 hours per day) and mean values for valid days and valid wear-time (n = 3622).

Table 4. Anthropometric and sociodemographic sample characteristics (n=3622). Data are presented in average (SD) and percentage (%).

Table 5. Accelerometer-assessed PA expressed as mean (SD) values of average acceleration, intensities, and IG (n = 3622).

Table 6. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for age groups (n = 3622).

Table 7. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for BMI categories (n = 3622).

Table 8. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for educational attainment (n = 3622).

Table 9. Comparison of MVPAnew and MVPAold across sex and intensity-variables: expressed as mean values (SD) (n = 3589).

Table 10. Comparison of activity variables, age, sex, educational attainment, and BMI prediction (n = 3583).

List of figures

Figure 1. Dose-response curve for overall physical activity and health benefits. Based on terms described in Haskell et al. (2007).

Feil! Fant ikke referansekilden.

Figure 3. Exemplified illustration of an output-file from ActiLife using the cut-points from Troiano et al. (2008) and further used by the Kan2 survey

Figure 4. Figure A illustrates a low IG with a steeper drop in both times accumulated (y-axis) and activities at high intensities (x-axis). Figure B represents a high IG with more time gathered (y-axis) across the intensity range (x-axis).

Feil! Fant ikke referansekilden.

Figure 6. Selection of participants derived from Kan2 (2014-2015). A total of 3622 participants were approved for this thesis results.

Figure 7. Overview of processing raw accelerometer data to analyzable CPM files.

Figure 8. Illustrated overview over processing raw acceleration through R package GGIR.

Figure 9. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different age groups unequal values for average acceleration and IG.

Figure 10. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different body mass index groups with unequal values for average acceleration and IG.

Figure 11. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different educational attainments with unequal values for average acceleration and IG.

Figure 12. Scatterplot compares CPM and average intensity over the most active continuous 15 hours of all days (n = 3589).

Figure 13. Scatterplot comparison between MVPAold and MVPAnew (n = 3589).

Table of content

Summary 2
Sammendrag 3
Acknowledgments 4
Abbreviations 5
List of tables 6
List of figures 7
1. Introduction 9
2. Theory 12
2.1 Definitions 12
2.1 PA recommendations 13
2.2 Monitoring of population-based PA 17
2.2.1 Criteria methods 17
2.2.2 Self-reported data 19
2.2.3 Device-based measurement of movements 21
2.3.3 ActiGraph 24
2.3.4 Counts per minute (CPM) 25
2.3.5 Analytical and translational metrics 27
3. Materials and methods 32
3.1 Design 32
3.2 Participants 32
3.2 Measurement methods 33
3.2.1 Valid wear-time 33
3.2.2 Counts per minute (CPM) 34
3.2.3 The new analytical and translational metrics 36
3.3 Statistical analysis 38
3.4 Ethics 39
4. Results 40
4.1 Descriptive 40
4.2 Average acceleration and intensity gradient (IG) 41
4.3 Age, BMI, educational attainment, and MX-variables 43
4.3.2 Body mass index (BMI) 44
4.3.3 Educational attainment 46
4.4 Comparison between the various device-measured indices of PA 48
5. Discussion 53
5.1 Results discussion 53
5.2 Method discussion 57
6. Conclusion 64
7. Future research 65
Sources 66
Attachments 71

1. Introduction

Regular *physical activity* (PA) has been shown to prevent and treat numerous diseases (Guthold, Stevens, Riley, & Bull, 2018). In particular, PA effectively reduces non-communicable diseases (NCDs) such as diabetes mellitus type II, cardiovascular disease, and several types of cancers (Guthold et al., 2018). *Insufficient PA*, or physical inactivity, is defined as not meeting current guidelines for PA (Ministers, 2014). Physical inactivity is declared by the World Health Organization (WHO) as one of the leading risk factors for NCDs' development (WHO, 2010). The need for actions was presented through the *Global Action Plan on Physical Activity 2018-2030*: decrease the measured amount of physical inactivity by 15 % within 2030 (WHO, 2018). To achieve targets against physical inactivity, surveillance was considered one of the critical factors in monitoring and evaluating PA and physical inactivity (WHO, 2018).

During the last decade, PA surveillance has evolved considerably nationally and internationally (Hansen et al., 2018a). For example, in Norway, the Kan- and Ungkan-surveys have contributed national representative population levels of PA for a decade. In contrast, the *Global Burden of Disease study* was the first to measure physical inactivity globally (WHO, 2004). This study led to the development of the International Physical Activity Questionnaire (IPAQ) and Global Physical Activity Questionnaire (GPAQ) (WHO, 2004). Both IPAQ and GPAQ are considered desirable as they have been used in several extensive surveys (Guthold et al., 2018). Based on self-reported data from population-based surveys, Guthold et al. (2018) estimated that the targets for decreasing physical inactivity towards 2030 would not be met if current levels of PA continue (Guthold et al., 2018). Such assumption is exemplified due to the prevalence of a quarter of all adults in the world are not attaining enough PA, based on self-reported data (Guthold et al., 2018).

A survey by Hansen et al. (2018a) points out that estimates derived from self-reported data are alone neither precise nor representative enough to estimate the prevalence of meeting PA guidelines. Therefore, the need for objective measurements, like accelerometers, is proposed to comprehend biases connected to questionnaires and be an addition to more accurately measure actual activity level at a population level (Hansen et al., 2018a). However, questionnaires cannot precisely estimate the proportion of populations supervising the PA recommendations successfully or investigate longitudinal associations in PA (LaMunion, Fitzhugh, & Crouter, 2020). As a result, questionnaires might lead to overestimating the reported PA level and the possibility of social desirability bias (LaMunion et al., 2020).

In recent years, accelerometers have been more common to use in surveillance alongside questionnaires (Migueles, Rowlands, Huber, Sabia, & Hees, 2019b). Correspondence between the two is considered essential to get hold of: (1) the populations' activity level at a given time, (2) deviation over time, (3) differences between sex and ages, and (4) reasons for why some people exercise and some do not (Helsedirektoratet, 2015). Accelerometers have been validated for the estimation of physical behavior (Migueles et al., 2019a). Brands like ActiGraph have since the 1990s been developed and frequently used in epidemiological studies worldwide (Migueles et al., 2019b). However, few surveys have made the same collection and processing decisions for the results (Migueles et al., 2019a; Rowlands et al., 2019a). A consensus for standardization of accelerometer methodology is needed to successfully compare results derived from different accelerometer brands, attachment sites, and countries (Rowlands et al., 2019b). Up to 2021 is the *counts per minute* (CPM) approach, accumulated acceleration divided on minutes in use, most commonly used whenever comprehending and analyzing accelerometer-assessed PA (Rowlands et al., 2019a).

In statistical analysis of CPM-approach, specific cut points are most commonly used due to their effectiveness in applying, interpreting, and facilitating PA and health relationships (Rowlands et al., 2019a). Cut points, as 1952 counts/min (moderate to-vigorous-physical activity, MVPA, 3-6 METs, and 5725 counts/min (vigorous activity, ≥ 6 METs), were developed to give accelerometer-assessed PA a biological meaning with estimates spent at a given activity intensity (Freedson, Melanson, & Sirard, 1998). The method is though hampered by (1) specificity and inability to compare to datasets with different cut points, (2) significant differences from obtaining score above or below a cut-point, and (3) consequently scoring zero on vigorous PA due to high cut point (Rowlands et al., 2019a; Rowlands et al., 2019b). Rowlands et al. (2019b) point out the need for an alternative approach to analyze and interpret accelerometer data. A suggestion is analytical and translational metrics combined with the GGIR package derived from R's statistical computing program (Migueles et al., 2019b; Rowlands et al., 2019a). GGIR comprehends and analyzes data extracted from wearable acceleration sensors and enables the use of different accelerometer signals. R, and GGIR, might overcome biases related to accelerometer-assessed PA and improve the population-based research worldwide.

This thesis is limited with theory (e.g., the definition of PA), specific methodological approaches (analytical and translational metrics and CPM), and a selection of health variables (e.g., body mass index), PA variables (e.g., estimates for overall activity and PA intensities), and participants derived from Kan2. Other theory regarding health, estimates of PA (e.g., estimates for weekdays and weekends), and measurement devices (e.g., questionnaires, criteria methods, and other accelerometer brands as Axivity and GENEActiv) were not included. Due to the determined assignments above, CPM-based PA values were considered insignificant to this thesis results, as they are already available in the Kan2 survey. Therefore, the main aim of this thesis was to examine the possibility of analyzing and interpret accelerometer-assessed data and its impact on the results for already existing data. It will include analytical and translational metrics corresponded to the use of the R package GGIR software. Following 1) aim, 2) research question, will be addressed:

1. *Provide normative new metrics for population estimates of physical activity.*
2. *Compare the new metrics to traditional count-based measures of physical activity.*

2. Theory

2.1 Definitions

Physical activity (PA) is commonly defined as any bodily movement produced by skeletal muscles, resulting in an increase in energy expenditure (EE) above the resting metabolic rate (Caspersen, Powell, & Christenson, 1985; Ministers, 2014). The opposite is physical inactivity, which describes insufficient PA as a failure to oversee recommendations (Caspersen et al., 1985). If the bodily movement is a part of an exercise, the activity is considered planned, structured, and repeated to improve or maintain one or more components of physical fitness (Ministers, 2014). Physical fitness is something that people have or strive to achieve, like strength, flexibility, and coordination (Ministers, 2014).

In 1985, researchers categorized PA into intensity terms to classify the activity's energy expenditure (EE) (Caspersen et al.). EE from PA was measured through kilojoules (KJ) and was determined by the amount and interaction of muscles; 4,181 KJ is equal to 1 kilocalorie (kcal) (Caspersen et al., 1985). Today, EE is typically measured through the unit of metabolic equivalent (MET), which estimates PA's oxygen consumption (Ministers, 2014). One MET corresponds to the resting metabolic rate and is approximately 3.5 mL $O_2 \cdot kg^{-1} \cdot min^{-1}$ (Ministers, 2014). Intensity zones such as sedentary, light, moderate, and vigorous are used to identify and describe the activity's content. Light intensity activity, like standing or slow walking (<3,5 km/h), have an EE between 1.5 to 3 METs (Ministers, 2014). Moderate intensity activity as jogging requires 3 to 6 METs, while vigorous-intensity activity (running) requires more than six METs (Ministers, 2014). Lastly, sedentary behavior – often confused with physical inactivity, is defined as any waking activity in a sitting or reclining posture with an EE below the light intensity of ≤ 1.5 MET (Ministers, 2014). Activities typically classified as sedentary behavior consist of screen time (TV, computer, and video games) and reading. For example, an individual engages in SB while sitting or lying down (Hansen, Steene-Johannessen, & Kolle, 2018b). It is stated that SB contribute to development of NCDs worldwide and increases with age (Guthold et al., 2018).

In order to capture contextual information for an individual's physical form, body mass index (BMI, kg/m^2) is often used due to its ease of use and the clear relationship between BMI and prospective health outcomes (Allison et al., 2020). BMI can either be objectively or self-reported (e.g., brief-type self-administered diet history questionnaire, BDHQ) and was first introduced in the mid-1800s (Rothman, 2008; Yoshitake, Okuda, Sasaki, Kunitsugu, &

Hobara, 2012). The primary use of BMI is estimating an individual's risk of weight-related health problems, such as NCDs (Nihiser et al., 2009). Commonly it is used to monitor body fat at a population level as it easily captures and correlates with body fat (Nihiser et al., 2009). In Kan2, BMI was used to classified participants into underweight (i.e., < 18.5 kg/m^2), normal weight (i.e., 18.5-24.9 kg/m^2), overweight (i.e., 25.0-29.9 kg/m^2), and obese (i.e., ≥ 30 kg/m^2). Rothman (2008) classifies BMI as an indirect measurement of body fat, with high usability for public health researchers in capturing population-based data. Golden standard equipment (see subchapter 2.2.1 for more information) as dual-energy x-ray absorptiometry (DEXA) is the preferred choice in monitoring body fat. However, BMI is considered more usable when monitoring at a population level (Rothman, 2008).

2.1 PA recommendations

PA recommendations are public-health suggestions for preventing and treating numerous diseases and emphasized physiological and psychological health benefits for children, adults, the elderly, and pregnant females (Ministers, 2014). Over the years, it is estimated that the number of people who adhere to the recommendations has not increased due to societal developments worldwide (Guthold et al., 2018). These developments might be passive transportation, rising numbers of SB, and escalating growth in the number of deaths caused by NCDs (Guthold et al., 2018). Due to the COVID-19 pandemic ravaging worldwide, it is estimated that the PA level worldwide will decrease if global programs are not implemented in the nearest future (Bentlage et al., 2020). Bentlage et al. (2020) state it is urgent to promote PA under confinement conditions where methodological studies monitor programs over time. Confinement conditions are referred to as activity outdoors or at home (Bentlage et al., 2020). Between 2020 to 2021, it was produced extraordinarily few surveys with population-based measurements capturing PA from wearable device measurements as accelerometers (Bentlage et al., 2020). Thereby, researchers and authorities are relying on PA estimates captured before 2020 without knowing the true prevalence.

Nationally, the Norwegian government has implemented PA targets to involve and improve public health over a decade. Through the governmental document *Folkehelsemeldinga: Gode liv i eit trygt samfunn Meld, St. 19*, the Norwegian health authorities will monitor and establish systematically public health strategies in a long-term perspective (Helse-og-omsorgsdepartmentet, 2018-2019). One of these strategies is *Sammen om aktive liv:*

Handlingsplan for fysisk aktivitet 2020-2029. The strategy is the Norwegian response to the WHOs target for decreasing physical inactivity by 15 % within 2030. Among these, several targets are knowledge development and innovation of objectively measured PA. Every regional and national Norwegian survey from 2020 must prioritize objectively measured PA over self-reported data due to its high reliability (Departementene, 2020). The Norwegian government will then know to facilitate PA targets and the population who oversees the PA recommendations from 2014 (Departementene, 2020).

Table 1. Norwegian guidelines for physical activity and sedentary behavior for Norwegian adults.

1	Adults should attend daily activities per week with a minimum of 150 minutes moderate or 75 minutes vigorous PA. The recommendations can be achieved with a combination of moderate to vigorous intensity
2	The activity can be achieved through several exercises with a minimum of 10 minutes duration.
3	Additional health outcomes can be achieved by increasing the amount of moderate PA up to 300 minutes, vigorous PA up to 150 minutes, or combining the two throughout a week.
4	Physical fitness in the form of strength training is advised two or three times within a week
5	Reduce sedentary behavior.

The international PA recommendations are similar to the 5th edition of the Nordic Nutrition Recommendations (NNR 2012) implemented in 2014 (Ministers; WHO, 2010). The NNR 2012 was established to assess evidence for nutrition and PA into public health-friendly recommendations in the northern European countries (Ministers, 2014). Adequate PA is recommended for the prevention of lifestyle-related diseases as NCDs, alongside a balanced diet. It is evolving evidence that SB increases NCDs' prevalence and not fulfilling the recommendations set in 2014 (Guthold et al., 2018; Helsedirektoratet, 2015; Steene-Johannessen et al., 2020). Therefore, it is recommended to reduce SB: by implementing the following PA recommendations for adults and the elderly (18-85 years old). Individuals with reduced mobility, or coordination, are advised to reduce the risk of falling by performing strength exercises two to three times a week. Adults or the elderly, who cannot achieve the recommendations listed below, are recommended to be physically active to the best of their ability: and what health condition allows (Departementene, 2020).

The Norwegian PA guidelines mentioned in table 1 are based on the *Global recommendations on physical activity for health* published in 2010 (WHO). During the last decades, there has been a knowledgeable evolution of PA and SB health outcomes (Haskell et al., 2007). Notably, this is shown for psychological wellbeing, intellectual health, and PAs impact on different age groups (WHO, 2020). Furthermore, the preventive effects for good wellbeing over time improves with increased total PA (e.g., intensity, duration, and frequency) (Ministers, 2014). In figure 1 below is the dose-response relationship between health and PA described with recommendations from Haskell et al. (2007). The researcher exemplifies moderate PA as brisk walking for at least 30 minutes during a day and vigorous as jogging or an increase in oxygen consumption with a substantial increase in heart rate (Haskell et al., 2007).

The dose-response relationship between health benefits and accumulated PA illustrates the small proportion of activity needed for better wellbeing (Ministers, 2014). For instance, an individual with an insufficient PA during a day has more use for small increases in total PA than individuals with high accumulated PA (Haskell et al., 2007). Additionally, substantial health-promoting effects are found by 30 minutes of brisk walking two times a week in addition to three days with 20 minutes of jogging (Haskell et al., 2007). A sufficient PA is determined by initial health status and interest groups (e.g., elderly, overweight individuals, Taxi-drivers) (Ministers, 2014). Individuals who wish to improve their physical fitness, and reduce the risk of diseases and disabilities, are advised to accumulate a higher amount of PA than 150 minutes mentioned above (Haskell et al., 2007).

It is mention by both Norwegian authorities and the WHO that new national and global guidelines for all age groups will be updated in the next couple of years (Departementene, 2020; 2018). Increasing recognition for sufficient guidelines targeting individuals with disabilities has been prompted during the last couple of years (WHO, 2020). The need for new guidelines for youth, adults, and elderly are considered prioritized through target 4.1 in the *Global action plan on physical activity* (WHO, 2018). The overall objectives are to provide recommendations, based on recent PA data, that lead to health benefits and diminish risk factors (WHO, 2020). SB is estimated to be implemented and replace current guidelines, but it is not revealed when the new recommendations are expected (WHO, 2020).

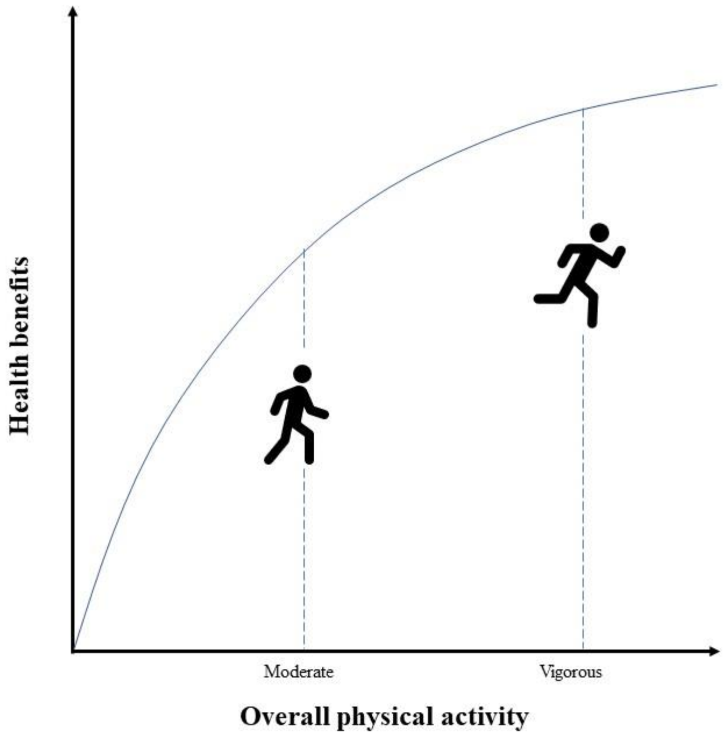

Figure 1. Dose-response curve for overall physical activity and health benefits. Based on terms described in Haskell et al. (2007).

In the past decades, Norway has achieved four national surveys monitoring PA subjectively and objectively among Norwegian adults and the elderly (Hansen et al., 2018a; LaMunion et al., 2020). Researchers from Kan2 (2014-2015) highlighted numerous biases related to self-reported data in the estimation of PA. An over- and under-estimation may occur when monitoring total activity level, the time within activity intensities, and the number of participants who oversee PA recommendations (Hansen et al., 2018a). Additionally, most Kan2-participants have a high socio-economical background, assumed an overestimation of actual activity level (Helsedirektoratet, 2015). Nevertheless, researchers behind Kan2 emphasize that the results represent Norwegian adults aging 20-85 years old (Helsedirektoratet, 2015).

Kan2 participants' PA level, derived from an accelerometer, reports a higher total activity level among participants aging 20-64 than among the elderly (Helsedirektoratet, 2015). A higher prevalence of females (34 %), and participants with body mass index (BMI, kg/m^2) as normal (37 %), achieve recommendations set for PA (Helsedirektoratet, 2015). Results for female PA recommendations are based on higher results achieved within 10 minutes bouts of moderate-to-vigorous activity (MVPA) compared to their male counterparts. Bouts refer to a consecutive activity timeframe with the same intensity over time (Helsedirektoratet, 2015). The possibility of attaining recommendations by dividing movements into 10 minutes bouts is underlined as the main reason for sex differences in adherence (Helsedirektoratet, 2015). Activity levels derived from Kan2 are virtually based on ActiGraph GT3X+ and its possibility to capture PA's acceleration over seven consecutive days. As described in subchapter 2.3 Accelerometer, it is necessary to compare the results validity, and thereby, the representativeness of the results to a “golden standard”. It is further described in 2.2.1 Criteria methods.

2.2 Monitoring of population-based PA

In the last decades, there have been numerous methods to measure PA and its aspect (Migueles et al., 2019b). The development of these methods is crucial for documenting the frequency, distribution, and magnitude of PA within a population and longitudinal data over time (Hansen et al., 2018a). The selection of methodology depends on several factors to identify and implement interventions (WHO, 2018). Factors as content, timeframe, budget, availability of measurements, participants’ requirements, and the researcher's knowledge for comprehending and analyze results are among numerous suggestions (Guthold et al., 2018; LaMunion et al., 2020; Migueles et al., 2019a). Over time there is sufficient evidence that collaboration between questionnaires and objective measurements is needed both nationally and internationally (LaMunion et al., 2020).

2.2.1 Criteria methods

Criteria methods, or criterion validity, can be defined as a method to compare two measurement results to confirm their precision and ability (Steene-Johannessen, Grydeland, & Hansen, 2018). The measurements are often referred to as golden standard equipment (e.g., direct observation, calorimeter, or doubly labeled water) and considered as the most accurate and precise methods for monitoring PA (Steene-Johannessen et al., 2018). Due to specific equipment, costs, and the necessity of knowledge, criterion methods are primarily used with small samples in controlled environments.

Direct observation

Direct observation is a golden standard where an observer oversees and monitors an individual's activity in controlled or everyday activities (Steene-Johannessen et al., 2018). One example of a validated instrument based on direct observation is the System for Observing Play and Recreation in Communities (SOPARC), developed for measuring group behavior and target areas (McKenzie, Cohen, Sehgal, Williamson, & Golinelli, 2006). An observer can use activity codes (i.e., sedentary, moderate, and vigorous) to calculate a summary score for groups and determined where, when, and what kind of activities they are doing (McKenzie et al., 2006). Another method on an individual level is a hand-held personal digital assistant (PDA), combined with software (Noldus Inc., Wageningen, Netherlands), to monitor PA (Lyden, Petruski, Mix, Staudenmayer, & Freedson, 2014). With a menu of 12 categories and a corresponding MET-value, a trained researcher can record everyday activities regardless of area (Lyden et al., 2014). Irrespective of the method, direct observation is hampered by several biases (e.g., researcher's expertise, Hawthorne effect) and usable towards large samples (Steene-Johannessen et al., 2018).

Calorimetry

Calorimetry is a method that captures the human body's heat signature (i.e., directly), or in the surrounding atmosphere (i.e., indirectly calorimetry), and is often used in the comparison of other measurements (Steene-Johannessen et al., 2018). The establishment of calorimetry's validity demonstrates an equivalence between fuel consumed and humans' heat (Kenny, Notley, & Gagnon, 2017). Therefore, changes in temperature in closed chambers (i.e., direct), or arteriovenous oxygen difference (i.e., indirect), can be used as golden standards and as validity methods for EE (Kenny et al., 2017; Steene-Johannessen et al., 2018). Due to functional reasons, indirect calorimetry is the preferred criteria method for measuring whole-body metabolism and EE (Kenny et al., 2017). Concerning accelerometers, indirect calorimetry is typically used to verify the measurement's validity (Steene-Johannessen et al., 2018).

Doubly labeled water

Doubly labeled water is the third criteria method with the possibility to measure total EE and PA by quantifying for EE while resting (Steene-Johannessen et al., 2018). The equipment is carried out by oral intake of a solution containing stable isotopes (i.e., deuterium (^{2}H) and oxygen -18 (^{18}O)) and measuring the time used for body excretion of these. While deuterium and oxygen -18 are excreted through body fluid (e.g., urine), oxygen -18 is also emitted through the air with carbon dioxide (CO_2). By recording the production of CO_2, researchers can compare the excreted isotopes from oxidized fat, carbohydrate, and protein and therefore use this as a measurement for EE (Steene-Johannessen et al., 2018). After an intervention of 1-3 weeks, with pre- (i.e., intake of isotopes) and post-tests, researchers will have accurate results of an individual total EE regardless of age, sex, or everyday life (Steene-Johannessen et al., 2018). Doubly labeled water is though expensive and demands knowledgeable personnel, equipment, and time. Therefore, the method only prioritized small samples of individuals and was not suitable for extensive epidemiological surveys (Steene-Johannessen et al., 2018).

2.2.2 Self-reported data

Self-reported data, or subjective methods, are methods where participants estimate their effort (e.g., PA and sedentary behavior) through premade questionnaires, noted dairies, or interviews (Steene-Johannessen et al., 2018). The methods are cost-effective, can be applied to measure large samples, and provides contextual information of physical or sedentary behavior. The knowledge between PA and its impact on health has, over time, been derived from surveys using self-reported methods (Berntsen & Anderssen, 2018). Today, several standardized methods (e.g., Bouchard's diary, IPAQ, GPAQ) are to be found and used to monitor PA nationally and worldwide. However, each method is vulnerable to biases (i.e., overestimating actual performance) and whether they are examined for validity and reliability (Steene-Johannessen et al., 2018). Large health organization as WHO prioritize self-reported data from questionnaires despite their inaccuracies (WHO, 2018).

Activity dairies

Activity dairies is a self-reported method that continuously records an individual's activity (e.g., intensity, context, duration) throughout an intervention (Steene-Johannessen et al., 2018). Test subjects can record activity every five minutes on a piece of paper, smartphones, or tablets of a different brand. By measuring accumulated time, a researcher can observe an individual's activity level and estimate EE (Steene-Johannessen et al., 2018). An example is a diary developed by Bouchard et al. (1983), where their method records the amount and pattern of daily EE over three consecutive days. While the researchers thought it was suitable to estimate EE, research today has indicated an overestimating actual performance compared to the ActiGraph accelerometer (Bouchard et al., 1983; Martinez-Gomez et al., 2009). It is speculated that this is caused by a behavior modification (i.e., Hawthorne effect) where participants react to be observed by increasing their PA level (Steene-Johannessen et al., 2018). As the method is complex and suffers from several biases, is it unsuited for large samples of children or the elderly (Steene-Johannessen et al., 2018).

Questionnaire

A questionnaire is a handwritten or digital scheme that categorizes individuals by the degree of activity level, monitors populations over time, or is used to detect diseases (Steene-Johannessen et al., 2018). Historically, questionnaires have been the preferred method of monitoring (e.g., PA, SB) from national to international surveys (Berntsen & Anderssen, 2018). It is cost-effective, delivers knowledgeable data, and requires a small amount of effort from participants. Questionnaires are though considered insufficient to estimate the proportion of individuals who achieve PA recommendations and the activity level over time (Hansen et al., 2018a). In addition, PA is regarded as the desired behavior among humans and could overestimate reality (social desirability bias), thus hampering the validity of self-reported data (Hansen et al., 2018b; Steene-Johannessen et al., 2018). Still, questionnaires such as IPAQ and GPAQ are preferred and used by WHO due to the future targets of establishing sufficient surveillance worldwide (Guthold et al., 2018; 2018).

Body mass index (BMI)

In order to capture contextual information of an individual's physical form, body mass index (BMI, kg/m^2) is the most preferred method of choice (Allison et al., 2020). BMI can either be objectively assessed or self-reported (e.g., brief-type self-administered diet history questionnaire, BDHQ) and was first introduced in the mid-1800s (Rothman, 2008; Yoshitake et al., 2012). The primary use of BMI is estimating an individual's risk of weight-related health problems, as NCD (Nihiser et al., 2009). Commonly it is used to monitor body fat at a population level as it easily captures and correlates with body fat (Nihiser et al., 2009). In Kan2 was the measurement used to classified participants into underweight (i.e., < 18.5 kg/m^2), normal weight (i.e., 18.5-24.9 kg/m^2), overweight (i.e., 25.0-29.9 kg/m^2), and obese (i.e., ≥ 30 kg/m^2). Rothman (2008) classifies BMI as an indirect measurement in accessing body fat and high usability for public health researchers in capturing population-based data. Golden standard equipment (see subchapter 2.2.1 for more information) as equipment dual-energy x-ray absorptiometry (DEXA) is the preferred choice in monitoring body fat, but BMI is considered as more usable when monitoring on a population level (Rothman, 2008).

2.2.3 Device-based measurement of movements

Device-based measurements are regarded as methods that do not demand an individual's effort of self-reporting (Steene-Johannessen et al., 2018). These methods (i.e., heart monitors, accelerometers, pedometers) are based on movement where monitors capture the activity precisely by various sensors (accelerometers, gyroscopes). Device-based indices of PA overcome many of the limitations related to self-reported but also have limitations. Motion sensor requires specific knowledge of handling data and equipment to be implemented in surveys. The knowledgeable differences have led to an economic difference in sport science. As these methods are more used in high-income countries (i.e., Norway), there is currently a scarcity of device-based assessment of PA in other parts of the world (Guthold et al., 2018). Nevertheless, there is an agreement that motion sensors (i.e., accelerometer) are more practical to capture PA and should be implemented whenever an opportunity is available (Guthold et al., 2018; Hansen et al., 2018b; Rowlands et al., 2019a).

Pedometer

Walking is the most central component of human behavior – from a child's first steps to older people's walk for leisure and exercise. Measurements as a *pedometer* capture steps either from manufactured monitors (e.g., the Yamax pedometer) or through a health application downloaded on a mobile phone (Steene-Johannessen et al., 2018). The simplicity of using the method has been marked as valuable and suitable equipment for increasing PA in interventions (Harris et al., 2008). A pedometer does not measure intensity, EE, or other activities other than walking and has different algorithms to determine a valid step (Kang, Marshall, Barreira, & Lee, 2009). Compared to other motion sensors under controlled environments, pedometers have been shown to undercount steps at low walking speeds (Harris et al., 2008).

Accelerometer

A motion sensor is a device where the accelerometer is embedded to capture movements and the actual force required for the activity (Steene-Johannessen et al., 2018). It measures data over several days, is user-approachable, and cost-efficient (Steene-Johannessen et al., 2018). Depending on the accelerometer brand, the device is strapped on the thigh, back, arm or hip during everyday activities and exercises. The determination of location impacts the transferability and comparability between epidemiological studies. Besides, standardized methods for calibration, validation, and analyzing are needed to achieve accurate data regardless of accelerometer-brands or models (Basset, Rowlands, & Trost, 2012; Buchan, McLellan, Donnelly, & Arthur, 2019; Rowlands et al., 2019a; Welk, 2005).

Calibration of accelerometers describes converting accelerometer-assessed data to estimates of PA (Ferrahi, Niemlä, Kangas, Korpelainen, & Jämsä, 2019). Traditionally, acceleration has (g; 1 $g = 9.8 \text{m·s}^{-2}$) been sampled, filtered, processed, and referred to as *activity counts* (AC) to illustrate the relationship of EE and activity derived from vertical acceleration (Basset et al., 2012; Ferrahi et al., 2019). According to Chen and Bassett (2005), acceleration is considered desirable for monitoring PA due to its information-rich signals and ability to reflect the energy costs. It is translated into determined periods or epochs (e.g., 1 min), using single- or two-linear regression analysis (Ferrahi et al., 2019). The linear regression is further split into readable cut-points values to determined activity intensities (Ferrahi et al., 2019).

According to Basset et al. (2012), no single regression equation precisely measures EE for all PAs. Therefore, several studies have indicated that pattern recognition with machine-learning (ML) strategy is essential to overcome and improve biases connected to previous methods (Basset et al., 2012; Ferrahi et al., 2019; Rowlands et al., 2018). Pattern recognition is simulated intelligence to classify extracted data from previous experience or statistical information (Basset et al., 2012). It can be divided into ML-based approaches where complex interactions in accelerometer-assessed data can be apprehended to regulate or certify AC or raw acceleration data (Basset et al., 2012; Ferrahi et al., 2019). Regardless of the approach, pattern recognition requires value calibration (validity) of models who utilize the relationship among AC and directly measure EE or PA (Basset et al., 2012). Unit calibration is also used to diminish inter-instrument variability and accurate measuring credibility (Basset et al., 2012).

Determining the use of new or old metrics, it is promising evidence that raw accelerometry can develop precise processing techniques and diminish several biases related to AC (Basset et al., 2012; Ferrahi et al., 2019). The AC approach (i.e., CPM) with cut-point values can capture the overall activity level over 24 hours and is user-friendly for both researcher and user (Rowlands et al., 2018). The calibration is regarded as too specific and an impossible approach to know the prevalence of meeting PA guidelines (Rowlands et al., 2019b). With advanced ML-based models, raw accelerometry is considered the new approach for predicting activity type, intensity, and EE regardless of monitor wear-site (Ferrahi et al., 2019). Research by Rowlands et al. (2019a) is an example of this with the implementation of analytical (e.g., intensity gradient and average acceleration) and translational metrics (e.g., MX).

Over the years, accelerometers have been developed through many brands (e.g., ActiGraph, GENEActiv, Axivity), whereas several researchers have regarded this as a problem (LaMunion et al., 2020; Migueles et al., 2017). The method creates numerous validity (e.g., processing criteria of raw acceleration) and reliability (e.g., differences between accelerometer monitors) issues that differ in researchers' results. Collection protocol, device placement (i.e., hip, wrist, thigh), and sampling frequency (30-100 Hz) differ depending on model and brand (Migueles et al., 2017). It is further problematic when the data processing criteria (e.g., cut-point values vs. analytical and translational metrics) are decided and could affect the credibility of the results. A result of lack of consistency across brands and lack of consensus on data reduction strategies is reduced comparability between surveys, making comparisons impossible unless researchers use the same methods and decisions in every study (Fairclough, Rowlands, & Boddy, 2019a).

Validity can be defined as the degree to which equipment assesses what it is intended to quantify (Basset et al., 2012). Within the quantitative research of PA, criterion-referenced validity (e.g., concurrent or predictive validity) is more aimed towards wearable devices, while the content and construct validity are more used to verify PA questionnaires (Basset et al., 2012). Furthermore, concurrent validity is used to compare or correlate data given at the same time from an ordinary measurement (e.g., accelerometer) towards a golden standard (e.g., indirect calorimetry) (Basset et al., 2012). As validation of wearable monitors is essential for credibility, reliability is considered a premise for validity (Greenfield, Kuhn, & Wojtys, 1998).

Reliability can be defined as the ability to reproduce the same results derived from the exact measurement over a given time (Greenfield et al., 1998). Regarding accelerometers, several methods have been applied to measure intra- and inter-instrument reliability with ActiGraph monitors (ActiGraph, Pensacola, Fl, USA) (Aadland & Ylvisåker, 2015). Intra-instrument reliability is considered the difference amongst a single monitor through numerous trials, while inter-instrument reliability differences among various monitors within a specific test (Mcclain, Sisson, & Tudor-Locke, 2007). The intraclass correlation (ICC) and coefficient of variation (CV) are statistical methods to express this. ICC is a statistical method ranging from

0 to 1.0 where a value of > 0.80 or at least 80 % is considered highly reliable. CV (expressed as a percentage), on the other hand, exemplifies the ratio of the standard deviation (SD) to the mean where low CV values suggest better reliability (Mcclain et al., 2007).

Device measurements intra- and inter-instrument reliability are measured through mechanical and laboratory trials (Mcclain et al., 2007). Automated tests use one or several monitors on a motorized turntable that produce a range of well-regulated and reproducible frequencies as stimulating slow walking or running (Mcclain et al., 2007). In the laboratory, trials are one or several monitors used by individuals during specific exercise protocols (Mcclain et al., 2007). A summarization of tests by Aadland and Ylvisåker (2015) classify the ActiGraph model GT3X+ as a reliable tool for measuring PA among adults under free-living conditions. Additionally, moderate-to-vigorous-physical-activity (MVPA) has been considered the most optimal intensity range, at least for inter-instrument reliability (Mcclain et al., 2007).

2.3.3 ActiGraph

Within PA research, the ActiGraph is categorized as accurate and reproducible to make cross-ethnic assessments and examine the effects of interventions (Grydeland, Hansen, Ried-Larsen, Kolle, & Anderssen, 2014). The brand has been developed with changes within hardware and software over time. From the initial accelerometer model ActiGraph AM7164 to newer generations GT1M and GT3X+, there have been improvements in data storage, signal perception of PA, and researchers' usability (Grydeland et al., 2014). Only the latest models from ActiGraph (wGT3X+, wGT3X-BT, wActiSleep+) are water-resistant and have a wireless interface of collected data. This thesis will concentrate the analysis towards ActiGraph GT3X+ to extract data from the Norwegian survey, Kan2 (Helsedirektoratet, 2015).

The GT3X+ (Pensacola, FL, USA) was released in 2010 with the mission to improve certain aspects from AM7164 and GT1M: see figure 2 below. While the older monitors collected data in user-determined time intervals (epochs), ActiGraph GT3X+ registers acceleration in a range of ± 6 g with a frequency between 30-100 hertz (Hz). The device separates from earlier models by measuring vertical, mediolateral, and anteroposterior axis (Grydeland et al., 2014). By doing so, the values collected from three different angles are transformed into one readable value. It is estimated to correlate well with EE and measure numerous activities better than with the vertical axis alone (Grydeland et al., 2014). Processed data can be recorded by internal memory, be downloaded, and analyzed through computer software (i.e., Actilife, GGIR). The final report from this may differ depending on the cut-point values used and accelerometer brand (e.g., GENEActiv, Axivity, ActiGraph) (Rowlands et al., 2017).

Figure 2. Picture of an accelerometer model, ActiGraph GT3X+, used in Kan2 survey (2014-2015).

Compared to other brands of activity monitors such as the GENEActiv and Axivity, it has been shown that ActiGraph report 11 % higher acceleration compared to GENEActiv alone (Rowlands et al., 2017). Other studies indicate an even higher difference with 13-16 % higher total measured acceleration (Rowlands et al., 2017). Rowlands et al. (2017) phrase that the difference is caused by technical differences among the brands or accelerometer signal processing. Regarding acceleration and MVPA, ActiGraph GT3X+ delivers 9-11 % lower accelerations than the other brands (Rowlands et al., 2017). Regarding sedentary and light-intensity activity (< 40-50 m*g*), the authors determine a high equivalence among the brands despite differentially from earlier results (Rowlands et al., 2017; Rowlands, Yates, Davies, Khunti, & Edwardson, 2016).

2.3.4 Counts per minute (CPM)

In extensive international surveys (e.g., NHANES, UK Biobank, WHS), the *counts per minute* (CPM) approach has been the preferred choice of accelerometer-assessed PA (Migueles et al., 2019a). The approach use cut-points, which can be defined as an analyzing technique to quantify accumulated PA into determined thresholds specific to age, sex, or population groups (Fairclough et al., 2019a). The conventional CPM method is translating the accelerometer output (i.e., counts per unit of time) into thresholds (i.e., expressed as m*g*) and matching the MET expenditure derived from the activity: see figure 3 below (Watson, Carlson, Carroll, & Fulton, 2014). From this, a researcher can determine how many accumulated minutes there are within an intensity zone (Watson et al., 2014). For instance, Rowlands et al. (2019a) explain this by classifying data points above or below a cut-point (e.g., 200 m*g* for MVPA) and translating this into minutes. The simplicity behind this makes this approach a preferable choice among researchers, even though it is hampered with numerous biases (Rowlands et al., 2019a; Watson et al., 2014).

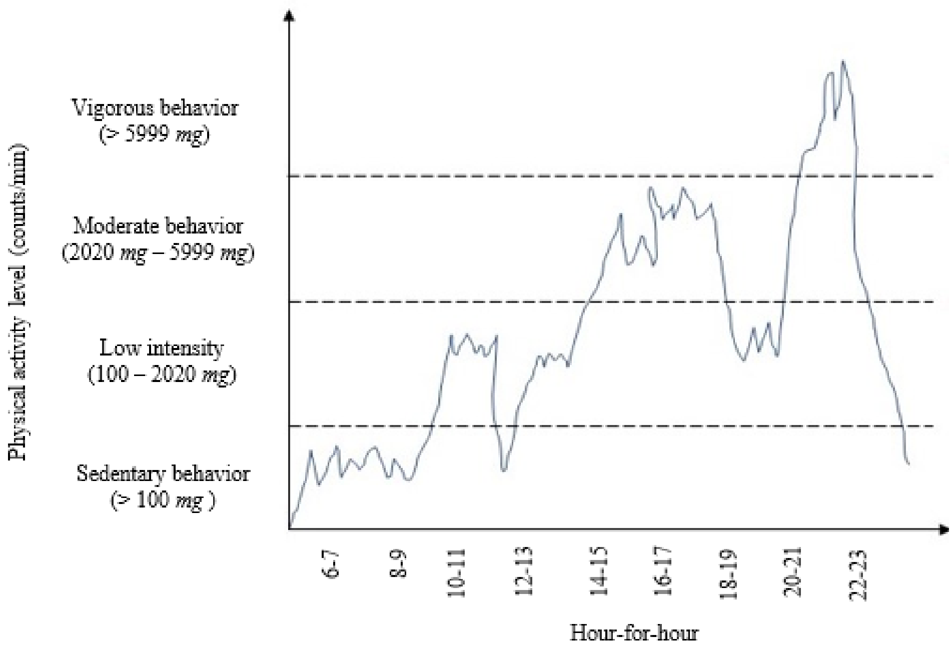

Figure 3. Exemplified illustration of an output-file from ActiLife using the cut-points from Troiano et al. (2008) and further used by the Kan2 survey

One of the biases related to cut-point values can be extracted from the study by Watson et al. (2014). Different accelerometer cut-points were used in the estimation of PA among American adults. Results derived from 11 studies revealed differentially in the determination and use of cut points representing moderate- (range: 191-2743 *mg*) and vigorous-intensity (4945-7526 *mg*) activity (Watson et al., 2014). Consequently, the prevalence among adults who met the PA guidelines was considerably affected, with 4.5-97.6 %, depending on the cut point used (Watson et al., 2014). Watson et al. (2014) exemplified that the use of cut-points may under- or overestimate PA prevalence by varying degrees. The same is observed in children with a variation of 8 % to 96 % who meet the recommended 60 min/d of MVPA based on the choice of cut-points (Migueles et al., 2019a).

Based on results from Migueles et al. (2019a), the CPM approach will not reveal the true prevalence of meeting PA guidelines (Migueles et al., 2019a). It is caused by a lack of consensus of which cut points are most suitable combined with decisions made by the researcher (i.e., wear-location, the decision of cut-points, accelerometer model) (Fairclough, Rowlands, Taylor, & Boddy, 2019b; Migueles et al., 2019a; Rowlands et al., 2019a). In addition, a problem is the substantial amount of data that will be ignored due to the threshold values used in the analysis (Rowlands et al., 2019a). Therefore, several leading authors called for a new approach that addresses the adaptation of standardized accelerometer metrics while using cut points (Fairclough et al., 2019a). It is discussed in the subchapter below.

2.3.5 Analytical and translational metrics

There is an exceptional agreement that a new approach for analyzing and translating accelerometer-assessed PA from accelerometers is needed (Fairclough et al., 2019a; Rowlands et al., 2019a; Steene-Johannessen et al., 2020). Using raw accelerometry with new standardized metrics (i.e., analytical and translational metrics) is considered a potentially beneficial new approach in sports science. Research by Rowlands et al. (2019a) suggests separating metrics used for analyzing and translational for optimal results. At the same time, while analytical metrics (e.g., intensity gradient and average acceleration) capture PA in its whole nature, it provides translational metrics (MX) a visual translation of these analyses (Rowlands et al., 2019a). Thus, it could help facilitate comparability between datasets while preserving meaningful data translation (Rowlands et al., 2019a). Additionally, this substitute will overcome well-known biases related to the traditional use of cut-point values (Rowlands et al., 2019b).

Analytical metrics

It is considerable recognition for capturing intensity and volume distribution of PA: instead of focusing on the MVPA variable alone (Rowlands et al., 2019a). Average acceleration (volume distribution) and intensity gradient (intensity distribution, IG) are two data-driven metrics that capture the 24-h (hours) activity profile derived from accelerometers. Average acceleration can be defined as overall activity level during 24-h or as the most active continuous 15-h during a day (Rowlands et al., 2018). The variable is comparable across studies and populations while being non-specific about the intensity distribution (Rowlands et al., 2019a). IG (Figure 4) is an activity variable expressed negatively and represents the drop in time accumulated as intensity increases (Rowlands et al., 2019a). A steeper drop in IG values reveals little accumulated activity at high intensities, while a more flattened line represents activity well spread across the intensities (Rowlands et al., 2019a). Average acceleration and IG enables the possibility to determine if volume or intensity of PA have alone or cumulative effects on health (Rowlands et al., 2019a)

Furthermore, these metrics are registered and translated through the open-source R package GGIR (Migueles et al., 2019b). The acceleration is either compared to a preferred cut-point (e.g., 200 mg for MVPA) or as a representation of an activity variable matching the accumulated acceleration (e.g., walking) (Rowlands et al., 2019a). While the more commonly used cut-point approach collapses data into categories, the new metrics report the minimum acceleration achieved for a given duration (Rowlands et al., 2019a). For instance, the minimum acceleration equivalent to the most active 60 minutes of the day will be interpreted using cut-points (e.g., 200 m*g*) to distinguish between the intensities (Rowlands et al., 2019a). Thereby, comparability is available between datasets and gives a visual comparison of within- and between-group comparisons.

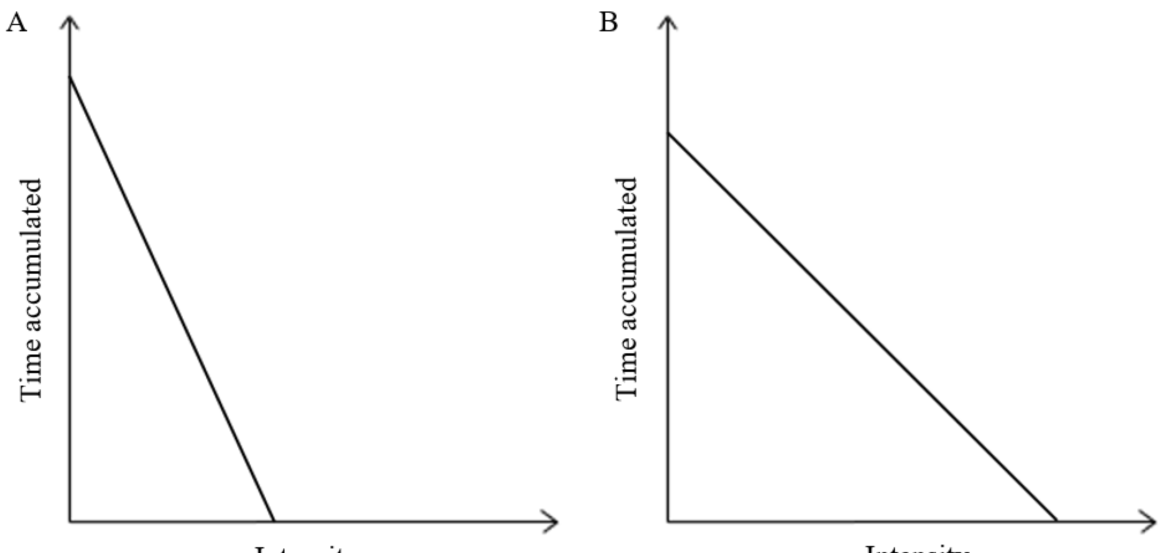

Figure 4. Figure A illustrates a low IG with a steeper drop in both times accumulated (y-axis) and activities at high intensities (x-axis). Figure B represents a high IG with more time gathered (y-axis) across the intensity range (x-axis).

A visual comparison of analytical metrics, thus translational metrics, is well documented by several studies in the last couple of decades (Buchan & McLellan, 2019; Fairclough et al., 2019a; Rowlands et al., 2019a). Up to 2021, the methodology was hampered by absent consensus and knowledge of true PA prevalence among populations (Migueles et al., 2021). Analytical and translational metrics are proposed as alternatives to the CPM approach, but several researchers highlight limitations in the new metrics data management (Buchan et al., 2019). The interpretation of average acceleration values is not relative to specific PA guidelines: whereas IG is (Migueles et al., 2021). However, Migueles et al. (2021) emphasize the need for a standard agreement among researchers that average acceleration reduces confounding effects of PA and may explain aspects of PA-related EE. The GRANDA consensus suggests a combination with IG to achieve prevalence of volume and intensity during a 24-hours profile (Migueles et al., 2021).

Translational metrics

Translational metrics, or MX, have an objective to interpret analytical metrics and illustrate them through suitable illustrations (Rowlands et al., 2019a). The MX refers to the acceleration above which a person's most active non-consecutive X minutes are spent over the day (Migueles et al., 2021). For instance, the M60 metric with a determined MVPA cut-point (e.g., 210 *mg*) will represent 60 minutes of PA accelerations greater than 210 *mg* during a day: see figure 5 below (Rowlands et al., 2019a). However, Rowlands et al. (2019a) exemplifies a situation to use a more stringent cut-point at 250 mg, and the consequences regarded this issue. As a result, high proportions of participants with 210 mg could not succeed PA guidelines set at the 250 mg mark line.

While the CPM approach easily collapses the data into determined MVPA thresholds and monitor everything above, MX metrics can compare themselves to any cut-point of choice (Rowlands et al., 2019a). The use of post-hoc translation can compare accumulated activity towards determined threshold values (e.g., 70 *mg*) and determine the number of participants who meet the PA recommendations (Rowlands et al., 2019a). Translating raw accelerometry can provide a meaningful interpretation of the activity profile, intensity, and volume through selected radar plots (Rowlands et al., 2019a). Radar plots (e.g., Figure 5 below) provides visual translation of analytical and translational analysis (Rowlands et al., 2019a).

Figure 5 describes four opportunities with MX-based radar plots. Firstly, M5 and M2 describe shorter periods but higher intensity during a day. The variables describe intensity better as a vigorous activity than other MX metrics. Secondly, values for PA guidelines are easily determined by higher periods as M60 and M30. The threshold value of 210 *mg* (i.e., brisk walking) illustrates the proportion of healthy, overweight, and obese individuals who successfully accumulated 30 minutes of MVPA during a day. It is possible to use other comparable cut-points (e.g., 200 or 250 *mg*), depending on age, sex, and other contextual aspects. Thirdly, lower MX-variables as M10, M5, and M2 capture higher intensities during a day and can be used to determine vigorous PA guidelines. The fourth opportunity is implementing values of overall activity (average acceleration) and IG (intensity distribution) within the figure. The radar plot (figure 5) is exemplified with descriptions and cut-put values from Rowlands et al. (2019a).

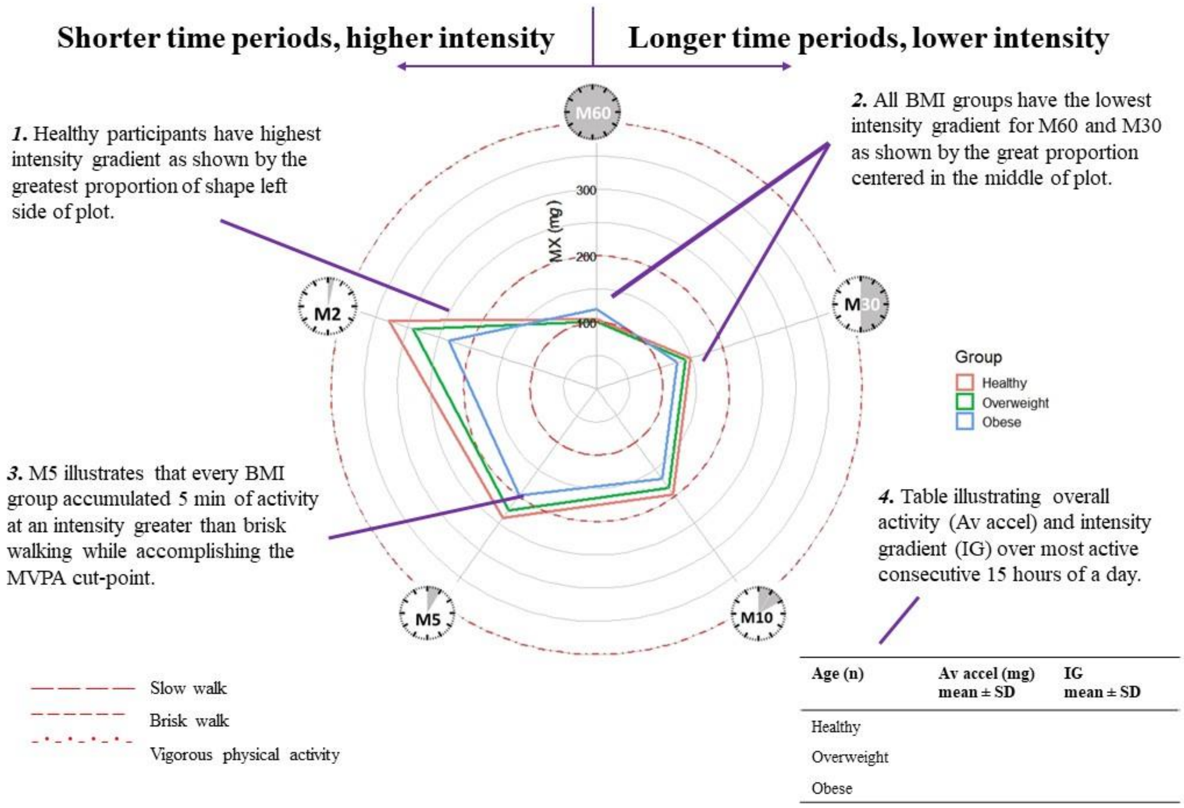

Figure 5. Exemplified radar plot based on RStudio algorithms (see Attachments) and descriptions described in Fig. 1 in Rowlands et al. (2019).

Concerning exemplified activities (i.e., slow walking, brisk walking, and running), threshold values will help the researcher interpret mean values for overall activity and intensity during an average day. Fewer or more MX metrics could be implemented in between- and within-group comparisons while using IG measuring PA guidelines (Rowlands et al., 2019a). MX metrics include a wide range of time variables to estimate overall activity during a day (Migueles et al., 2021). There is no consensus of how many or which MX metrics are preferred, but a number between six to eight variables are proposed by Rowlands et al. (2019a). MX metrics as M5, M15, M30, M60, M120, and $M^{1}/_{3DAY}$ are prompted by Rowlands et al. (2019a), where $M^{1}/_{3DAY}$ represents activity over eight hours of the day. The $M^{1}/_{3DAY}$ is valuable for distinguishing between sleep and active periods of the day (Rowlands et al., 2019a).

Migueles et al. (2021) highlight MX metrics as codependent time-use descriptors, raising the multicollinearity risk. The risk, or collinearity, indicates a condition where two or more variables in a regression model are highly related (Migueles et al., 2021). As MX metrics are at risk for multicollinearity, these time-descriptors will be equal to 1 or -1 in a regression model analysis. However, these calculations only affect individual predictors validity if used alone, but overall carry partial and relative information on activity patterns (Migueles et al., 2021).

3. Materials and methods

3.1 Design

The accelerometer-assessed data in this thesis are retrieved from the Norwegian survey *Kan 2* performed in 2014-2015. On the Norwegian Directorate of Health assignment, Kan 2 was carried out by the Norwegian School of Sport Science to extract representative levels of PA and SB among Norwegian adults. The design is cross-sectional, making it possible to observe PA levels at a given time.

3.2 Participants

Participants, aging from 18-85 years old, were randomly selected through the National Population Register on the criteria of age, ethnic background, population density, and sex. Out of 11 147 potential participants, consented 3180 adults and 409 adults with immigrational backgrounds (Helsedirektoratet, 2015). Participants were recruited by mail with an information letter, written informed consent (Attachments 1), questionnaire (Attachments 2), and an accelerometer (ActiGraph GT3X+) with specific instructions (Attachments 3). Participants were instructed to (1) always wear the monitor at the right hip while awake for seven consecutive days; (2) self-reported minutes of transportation habits (e.g., active or passive transportation) each of the seven days; (3) self-report minutes of activity poorly captured by the monitor (e.g., swimming and strength training).

After the measurement period, both measurements (i.e., accelerometer and questionnaire) were collected by mail from 5102 participants, and accelerometer files were derived. 5052 participants and included in the present study (figure 6). In figure 6 were participants excluded and selected based on (1) post-calibration error greater than 0.01 g (10 mg), (2) fewer than 3 days of valid wear (12 $h \cdot d^{-1}$), (3) wear data was not present for each 15 min period of the 24 h cycle (<0.625 g), and (4) anthropometric data. In total, 5052 were extracted from the Kan2, and 3622 were considered valid with sufficient wear time and included in the study. The exclusion of 1430 individuals was deemed necessary based on exclusion criteria from previous studies (Rowlands et al., 2019b; Vincent et al., 2014).

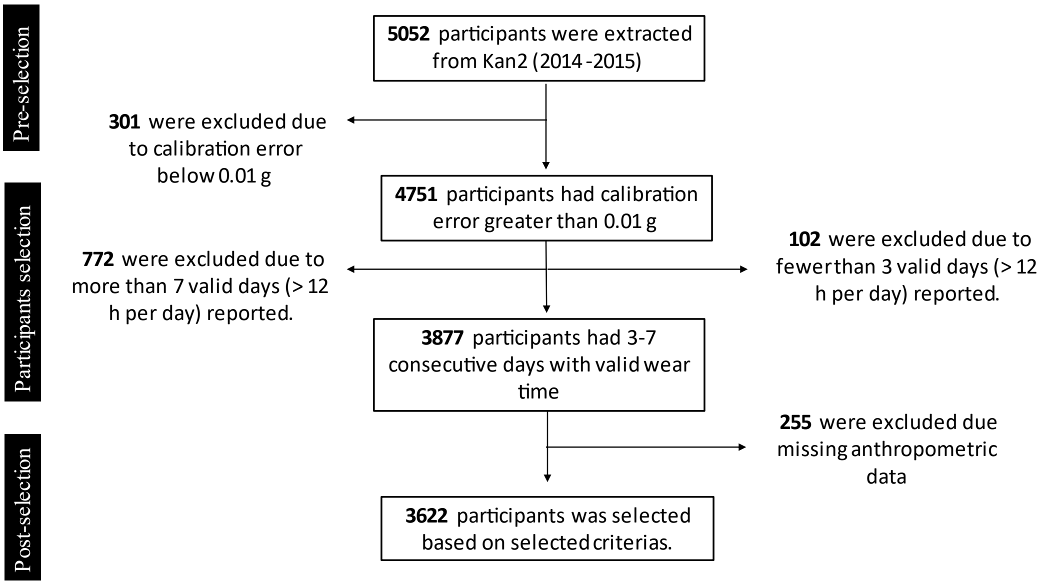

Figure 6. Selection of participants derived from Kan2 (2014-2015). A total of 3622 participants were approved for this thesis results.

3.2 Measurement methods

This thesis used accelerometer-model ActiGraph GT3X+ (ActiGraph, Pensacola, Florida, USA) from the Kan2 survey to monitor PA level among Norwegian adults (Figure 2). The ActiGraph GT3X+ is a triaxial accelerometer that captures acceleration in the range of ±6 g and sampling frequency between 30-100 Hz (Hertz) to the vertical, mediolateral, and anteroposterior axis (Grydeland et al., 2014). Hertz, in this case, refers to the accelerometer's ability to capture 30-100 registrations per second and timeframe concerning duration, frequency, and intensity. In the following thesis, only the vertical axis and 30 Hz were used in the first phase of processing accelerometer-assessed PA. Further, it was used in two different interpretations of raw acceleration signals achieved from the ActiGraph GT3X+: (1) counts per minute (CPM) and (2) analytical and translational metrics.

3.2.1 Valid wear-time

Table 3 illustrate the coherence between the number of valid days compared to CPM values. Valid wear time, concerning both datasets used in this thesis, is the pre-determined criteria for minimum minutes or hours needed to be included in further analysis. To be classified as a valid day in Kan2, a minimum level of 600 minutes (10 $h \cdot d^{-1}$) was the cut-off for all participants. When detecting actual wear and non-wear time, consecutive periods of zero reported registrations lasting at least 60 minutes or more were excluded for the analysis (Helsedirektoratet, 2015). The number of valid days needed to be included in the analyses was set between 2 to 7 consecutive days to achieve the most representative levels of PA. Participants with less than 3 days of valid wear days were omitted.

In dataset 2, involving analytical and translational metrics, validation criteria for minimum wear time were set to 720 minutes (12 $h \cdot d^{-1}$). Additionally, less than 3 valid days were removed on the assumption for an unrepresentative reality of a participant's activity level (Rowlands et al., 2019a). As for the traditional validation criteria, reported days were capped at 7 and further excluded: described in figure 6 above. In total, 874 participants had insufficient reported values and were excluded before further analysis. The validation criteria were set upfront for autumn 2020 by Alex V. Rowlands.

3.2.2 Counts per minute (CPM)

The accelerometer data processing, from raw acceleration signal into readable data, is illustrated in figure 7 below. Following the measurement period, raw acceleration (mg) from ActiGraph GT3X+ were downloaded using the ActiLife and KineSoft software. After five phases (see below), the outcome file (.csv) was downloaded and analyzed using the IBM SPSS Statistics 21. The variable CPM summarizes all acceleration captured by the accelerometer. CPM derives from the researchers' interpretation of acceleration from an accelerometer into interpretable activity counts (i.e., user-specified time intervals, *epochs*) ranging from 5 s to 1 min (Mcclain et al., 2007). In Kan2, extracted data from ActiGraph GT3X+ were integrated into 1 min epochs and interpreted into different cut-points, representing sedentary, light, moderate, and vigorous intensities (Helsedirektoratet, 2015). CPM representing (1) sedentary behavior was determined as total summarized minutes below the cut-point value of 100, (2) light PA between 100 and 2020, (3) moderate PA above 2020, and (4) vigorous PA above 5999 (Helsedirektoratet, 2015). The PA belonging to different intensities can determine the proportion of individuals who oversee national recommendations for PA.

(1) Raw acceleration files extracted from GT3X+ contains:

- Acceleration (mg) over three axes.
- Sampling frequency: 30 Hz.
- All data from start-up time to downloading time.

Software: ActiLife *File format: .gt3x*

(2) Extracted data from raw acceleration files:

- Acceleration (comprehended into activity counts) for vertical axis.
- Acceleration signal converted into steps.
- Data extracted two rounds: 10- and 60-seconds bouts (epoch-period).
- Data from start-up time to downloading time.

Software: ActiLife *File format: .agd*

(3) Preparing of converted datafile before data reduction:

- Every .agd-file with 60-seconds epoch-period was manually inspected for validation and to determine the first day with valid registrations.
- A loading file (.rif) was used to determine the first valid day and 6 consecutive days for futher analysis.
- Sampling frequency: 30 Hz

Software: KineSoft *File format: .agd, .rif*

(4) Data reduction of .agd-files within KineSoft was determined after:

- Valid daytime ≥ 600 min. (10-h)
- Consecutive 0-registrations (two minutes with activity counts > 0 was allowed for accounting for random movements) from 60 minutes or more was defined as time where participant did not wear the accelerometer. Such timeframes were excluded from the analysis.
- Cut-point values and bouts were determined (se chapter 3.2.1)

Software: KineSoft

(5) Establishment of SPSS-database for final analysis:

- Recused datafiles derived from KineSoft was within Microsoft Excel matched with correct ID-number after each reduction of data and opploaded to IBM SPSS Statistics 21 for final analysis. Data was quality checked for outliners and finally matched with data from questionnaire in the database.

Software: IBM SPSS Statistics 21

Figure 7. Overview of processing raw accelerometer data to analyzable CPM files.

PA recommendations are in Kan2 assessed by accumulated moderate to vigorous PA divided into 10 minutes *bouts*. Bouts, in this case, illustrates a consecutive timeframe of activity with the same intensity values that represent the activity level over a more extended period (Helsedirektoratet, 2015). The 10-minute bout allows two drops in intensity, for instance, stopping for red lights or break within a workout, as a response to maintain as many bouts as possible. As mentioned in the GRANDA consensus, recent guidelines have removed the 10-minute bout requirement for MVPA in adults (Migueles et al., 2021). In table 2, the Kan2 used the following approach to determine the proportion of participants who met the recommendations for PA.

3.2.3 The new analytical and translational metrics

Data management for analytical and translational metrics from raw acceleration files to establishing a final database is described in figure 8 below. In advance of analysis, raw acceleration data were uploaded to ActiLife (ActiGraph, Pensacola, USA) and saved as GT3X+ (.csv) file format (phase 1). The .csv-format was corrected for gravitational acceleration (raw accelerations) and used to generate 5 s epoch files expressed as milli-gravitational units (mg). Furthermore, the .csv files were processed and analyzed in R (http://cran.r-project.org) using the GGIR package (version 2.3.0). The raw accelerometer signal was processed in GGIR for (1) autocalibration using local gravity as reference, (2) detection of outliners, non-wear times, inactivity, and sleep periods, and (3) finally expressed as time use per day (.csv).

The reason for splitting into five different parts (figure 8) is, by default, avoiding re-doing of all analyses when implementing slight differences. Especially *part 5* is interesting, as this part calculates analytical and translational metrics used in this thesis. Average acceleration and IG are attained and further expressed through MX metrics: $M_{1/3DAY}$, M120, M60, M30, M15, and M5. Average acceleration was determined as overall activity based on the most active continuous 15-h of a valid day. Alex V. Rowlands recommended the variable as 15 h is a proxy for waking day and wear time covered across the week. Alternative variables as M16, M14, M13, M12, M11, M10, was proposed but not used in this thesis. The M16 refers to the most active continuous 16 hours during a day. In addition, was IG expressed as intensity distribution over the most active continuous 15 h of the day. Articulated as a negative value, a low negative result for the IG variable will be similar to high-intensity distribution during a day.

The determination of MX metrics was based on research by Rowlands et al. (2019a). Suggested variables as M6h, M4h, M45, M20, M10, and M2, was proposed but not used. The MX metrics' primary purpose is to express the average acceleration towards determined intensity variables (> 70 *mg*: slow walking; > 160 *mg*: brisk walking; 260 *mg*: vigorous activity; > 450 *mg*: running). An intensity threshold of 70 *mg* was used for estimating $MVPA_{TOTAL}$ and thereby PA guidelines. The MVPA threshold was determined by

Hildebrand, Hees, Hansen, and Ekelund (2014). In the use of ActiGraph GT3X+ among a selection of adults (n = 30), 69.1 *mg* was concluded as representing 3 METs: or moderate PA (Hildebrand et al., 2014; Ministers, 2014). Similar was conducted for the VPA_{TOTAL}, where the threshold value of 260 *mg* is equal to vigorous activity. An additional threshold of 450 *mg* was proposed apprehend, implicating acceleration values as running, but not implemented. Based on MX metrics, similar criteria were set from Kan2 (see table 2 for definitions) to achieve the PA guidelines.

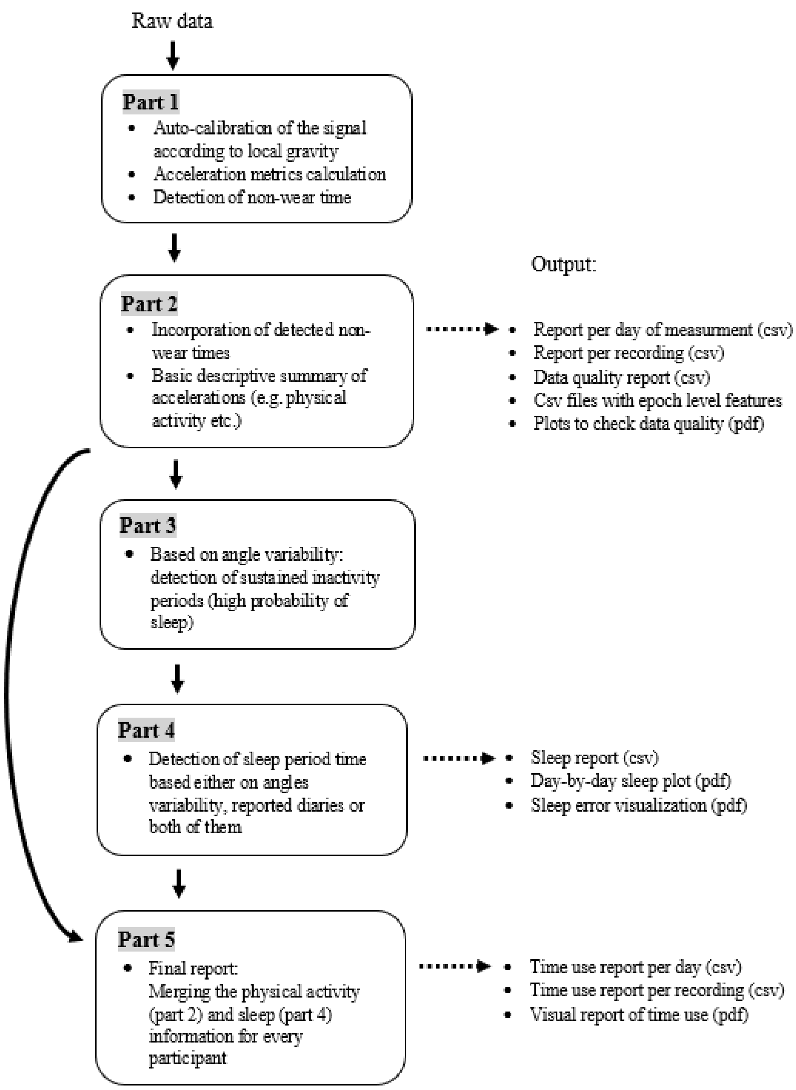

Figure 8. Illustrated overview over processing raw acceleration through R package GGIR.

Table 2. Definitions of new (analytical and translational) and traditional (CPM) metrics and their determination of MVPA and PA recommendations.

	Activity variable	Definition
New metrics	Average acceleration (average intensity)	*Average acceleration can be defined as overall activity level during 24-h or as the most active continuous 15-h during a day.*
	Intensity gradient (IG)	*IG is an activity variable expressed negatively and represents the drop in time accumulated as intensity increases.*
	MX metrics	*The MX refers to the acceleration above which a person's most active non-consecutive X minutes are spent over the day.*
	Moderate-to-Vigorous-Physical-Activity$_{new}$	*MVPA threshold (70 mg) is used to compare average acceleration values above which a person's most active non-consecutive X minutes are spent during a day. Derives from the work by Hildebrand et al. (2014).*
	PA recommendations $_{new}$	*1) Minimum 21.4 minutes MVPA each day to overcome 150 min/week with moderate PA: achieved by intensity threshold of 70 mg.* *2) A combination of moderate to vigorous intensity oversees 150 min/week.*
Traditional metrics	Counts per minute (CPM)	*CPM is based on researchers' interpretation of acceleration from an accelerometer into interpretable activity counts (user-specified time intervals, epochs) ranging from 5 s to 1 minute.*
	PA recommendations $_{old}$	*1) Minimum 21.4 minutes MVPA each day to overcome 150 min/week with moderate PA.* *2) An average of minimum 10.7 minutes of vigorous PA per day must be performed within the 10-minute bouts to overcome 75 min/week with vigorous PA, or* *3) A combination of moderate to a vigorous intensity above 150 min/week.*
	Moderate-to-Vigorous-Physical-Activity$_{old}$	*It is defined as a threshold value equal to or above 2020 counts per minute.*

3.3 Statistical analysis

Statistical analysis, and the creation of specific scatterplots and tables, was performed in IBM SPSS 25 (IBM Corp., Armonk, NY). The level for significance was at 0.05 level of margin. The age of participants was reduced to three groups (e.g., 18-34, 35-64, 64, and older). Due to few underweighted participants, these were classified as BMI normal, and participants with university degrees intermingled into one university variable: regardless of the level of a university degree. Descriptive data are presented as mean values, standard deviation (SD) or standard error (± SEM), and 95 % confidence interval (CI): unless others are stated. Radar plots were generated with algorithms (see Attachments) in RStudio Desktop (version 1.4.1106) and used MX metrics: $M_{1/3DAY}$, M120, M60, M30, M15, and M5 to illustrate acceleration above which a person's most active non-consecutive X minutes are spent over the day. A Chi-square sample test of independence and an independent-samples t-test (sex as group variable) was applied to assess differences between categorical variables and the continuous data.

Person correlation and univariate linear regression models were used to assess agreement between new and traditional metrics. The statistical test compared the predictive properties of the two approaches using BMI used as an outcome variable, with age, sex, educational background, wear time (10 $h \cdot d^{-1}$) included as covariates. Post hoc testing was carried out using the Bonferroni correction.

3.4 Ethics

Methods and results have followed the Helsinki declaration and been clarified by the Norwegian Centre for Research Data (NSD) and the Faculty of Ethical Committee (FEK) belonging to the University of Agder. Anonymization of participants was maintained without the possibility to connect any of the data to an individual.

4. Results

4.1 Descriptive

Table 3 illustrates the mean values of new (i.e., analytical and translational metrics) and traditional (i.e., CPM)accelerometer-derived metrics of PA capturing the number of valid days and wear time per day. A paired-sample t-test was conducted to compare different variables determining a valid day and wear time. There was a significant difference in days between traditional (M=6.5, SD=1.0) and new variable (M=6.6, SD=0.8) measuring valid days; $t(3613)= -9.99$, $p < 0.01$. Similar findings were found for traditional (M=876.6, SD=71.3) and new variable (M=964.2, SD=64.5) towards valid wear time per day; $t(3606)= -138.45$, $p < 0.01$. The new variables for validity (i.e., 12-h^{-1}) have higher mean values and percentages of participants with 7 measured consecutive days. Specifically, the results indicate that the number of minutes and days is considered valid whenever the new validity variables are used.

Table 3. Comparison between two validation variables (10 and 12 hours per day) and mean values for valid days and valid wear-time (n = 3622).

	New (12 h/d)	**Traditional (10 h/d)**
No. of valid days*	6.6 (0.8)	6.5 (1.0)
No. of valid wear minutes*	964.2 (64.5)	876.6 (71.3)
3	1.2 %	1.4 %
4	2.4 %	3.1 %
5	6.0 %	7.3 %
6	15.2 %	18.0 %
7	75.2 %	69.6 %

A total of 25 participants from the traditional variable are not included due to less than 3 valid days (0.6 %).

*P < 0.01

It was a significant difference in traditional validation criteria between participants with 1 and 2 valid days compared to those with 3-7 consecutive days. Considering that no significant differences were found in CPM between individuals with 3-7 valid days, participants with less than 3 valid days (12 $h \cdot d^{-1}$) were excluded from further analysis. The descriptive characteristics of the analyzed sample (i.e., those with a full set of demographics, anthropometric, and accelerometer variables) are presented in Table 4. Among the included sample were 2001 females (age 51.0 ± 15.5 years; height 166.8 ± 6.2 cm; weight 69.0 ± 12.4

kg; BMI 24.8 ± 4.2 kg/m^2) and 1621 males (52.0 ± 15.8 years; height 180.0 ± 6.8 cm; weight 84.4 ± 13.3 kg; BMI 26.1 ± 3.7 kg/m^2) participants.

Approximately 42% and 56 % of females and males were categorized as overweight or obese, respectively. Overall, the prevalence of overweight and obese was higher for male participants with an average BMI of 26.1 kg/m^2 (± 3.7 kg/m^2). Results for females 24.8 kg/m^2 was significant different ($p < 0.01$). In educational attainment, younger participants had a 65 % higher college degree than the elderly ($p < 0.01$). Similar results were found among male participants with a 19 % difference in the number of reported college degrees. According to reported data, 1 out of 5 elderly participants (i.e., > 64 years old) has low educational attainment, compared to only 1-2 % of young adults.

Table 4. Anthropometric and sociodemographic sample characteristics (n=3622). Data are presented in average (SD) and percentage (%).

	18-34 yr		**35-64 yr**		**> 64 yr**		**Total**	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
N (%)	370 (18.5)	257 (15.9)	1198 (59.9)	929 (57.3)	433 (21.6)	435 (26.8)	2001 (55.2)	1621 (44.8)
Age (years)	28.2 (4.1)	28.3 (4.0)	50.1 (8.2)	50.1 (8.6)	71.8 (5.7)	71.7 (5.3)	51 (15.5)	52 (15.8)*
Height (cm)	166.3 (6.4)*	180.5 (7.9)	167.4 (6.1)	180.5 (6.5)	165.6 (5.8)	178.2 (6.6)	166.8 (6.2)	180.0 (6.8)*
Weight (kg)	65.9 (12.8)*	81.1 (13.1)	70.4 (13.1)	85.9 (12.7)	68.6 (11.8)	83.2 (13.1)	69.0 (12.4)	84.4 (13.3)*
Body Mass Index	23.9 (4.7)*	24.9 (4.0)	25.1 (4.3)	26.4 (3.5)	25.0 (3.9)	26.2 (4.0)	24.8 (4.2)*	26.1 (3.7)
Normal (%)	69***	57	57	38	53	41	58	42
Overweight (%)	20***	35	31	47	36	44	30	44
Obese (%)	10***	6	11	14	10	13	11	12
Educational attainment								
Low (10 y)**	1	2	6	6	22	19	8	9
Middle (13 y)**	34	37	36	38	44	39	37	38
High (>13 y)**	65	61	59	56	34	42	55	52

BMI categories: Normal <25; Overweight: 25.0-29.9; Obesity: >29.9; total of 6 participants had insufficient values for educational attainment
* $P < 0.01$: significant difference between sex of total sample
** $P < 0.01$: significant difference between all age groups
*** $P < 0.01$: significant difference between sex in the same age group

4.2 Average acceleration and intensity gradient (IG)

The descriptive accelerometer-assessed average acceleration and IG variables are presented in Table 5. Values were collapsed across sexes due to small differences but included as a covariate in the analyses exploring differences across age. There was a significant effect of age on average acceleration, and the Bonferroni's post hoc indicated that the mean value for participants aging 18-34 years old (M = 24.2, SD = 13.4, $CI_{95\%}$ = [23.4, 25.2]) was significantly different from the age group 35-64 years old (M = 22.5, SD = 11.2) and 65 and older (M =17.7, SD 9.5). Overall, increased age was associated with lower accelerometer-assessed average acceleration.

Table 5. Accelerometer-assessed PA expressed as mean (SD) values of average acceleration, intensities, and IG (n = 3622).

	18-34 yr	35-64 yr	> 64 yr	Total
N (%)	627 (17)	2127 (59)	868 (24)	3622 (100)
Average acceleration$_{mg}$	24.2 (13.4)*	22.5 (11.2)	17.7 (9.5)	21.6 (11.5)
Intensity gradient	-2.42 (0.3)*	-2.49 (0.3)	-2.75 (0.4)	-2.53 (0.4)
MVPA$_{TOTAL}$	84.3 (85.6)*	74.2 (64.2)	51.6 (49.7)	70.6 (66.3)
VPA$_{TOTAL}$	5.9 (10.6)*	4.3 (9.4)	1.4 (6.2)	3.9 (9.1)

18-34: participants are aging 18-34 years old; 35-64: participants are aging 35-64 years old; < 64: participants are aging 64 years old and older; M15 h Accel: overall daily activity based on most active continuous 15 hours of an average day expressed as mg-values; M15 h IG: intensity distribution over the most active straight 15 hours of an average day; MVPA$_{TOTAL}$: threshold 70 *mg*, min/day; VPA$_{TOTAL}$: threshold 260 *mg*, min/day.

** P < 0.05: significant difference between all age groups

* P < 0.01: significant difference between all age groups

A similar association with age was prevalent for the IG-variable. There was a significant difference among the age groups ($p < 0.01$). The youngest age group (M = -2.42, SD = 0.3, $CI_{95\%}$ = [-2.45, -2.40]), middle-age group (M = -2.49, SD = 0.3, $CI_{95\%}$ = [-2.50, 2.47]), and the oldest age group (M = -2.25, SD = 0.4, $CI_{95\%}$ = [-2.77, 2.73]) all differed significantly from each other. Table 5 further display lower levels of MVPA$_{TOTAL}$ and VPA$_{TOTAL}$ with increasing age. Young adults aging 18-34 years old (M = 84.3, SD 85.6, $CI_{95\%}$ = [79.2, 89.5]) had the highest total MVPA volume compared to participants 34-64 years old (M = 74.2, SD = 64.2, $CI_{95\%}$ = [71.5, 77.0]) and elderly (M = 51.6, SD = 49.7, $CI_{95\%}$ = [47.2, 55.9]). Results suggest that an increase in age decrease values for total MVPA with a mean difference of 33 minutes $CI_{95\%}$ = [24.5, 41.0]) between the youngest and oldest part of the sample. Similar significant level was found towards total vigorous PA of all days ($p < 0.01$). Participants within 18-34 years old (M = 5.9, SD = 10.6, $CI_{95\%}$ = [5.3, 6.7]) stood out with the highest total volume achieved on the highest intensity level. Adults within 34-64 age group (M = 4.3, SD = 9.4, $CI_{95\%}$ = [3.8, 4.7]) and elderly (M = 1.4, SD = 6.2, $CI_{95\%}$ = [0.8, 2.0]) were significant independently of each other: as too the youngest age group.

4.3 Age, BMI, educational attainment, and MX-variables

The MX variables express the acceleration above which a person's most active non-consecutive X minutes (MX) are accumulated, enabling researchers to focus on a person's most active periods of the day, which is registered as the minimum acceleration for a given duration of time. Due to minor differences between sexes and aid readability, all data are collapsed across sex, but sex is included as a covariate. Table 6 shows the MX variables across age. Significance is illustrated using an asterisk. In summary, results indicate lower MX-values with increasing age, indicating that increasing age is associated with lower levels of mg thresholds for the most active non-consecutive minutes for all durations.

Table 6. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for age groups (n = 3622).

	18-34 yr	**35-64 yr**	**> 64 yr**	**Total**
N (%)	627 (17)	2127 (59)	868 (24)	3622 (100)
$M_{1/3}h_{mg}$	13.5 (13.2)*	12.3 (10.7)	10.1 (8.8)	12.0 (10.9)
$M120_{mg}$	47.5 (25.2)	45.9 (23.5)	37.2 (21.5)	44.1 (23.7)
$M60_{mg}$	86.7 (38.8)*	82.2 (37.3)	63.6 (34.9)	78.5 (38.0)
$M30_{mg}$	137.1 (61.7)*	125.6 (54.1)	91.0 (43.9)	119.3 (55.7)
$M15_{mg}$	181.0 (75.7)*	161.8 (64.6)	113.3 (50.5)	153.5 (67.8)
$M5_{mg}$	233,2 (89.2)*	202.8 (74.5)	141.0 (57.5)	193.3 (80.1)

18-34: participants are aging 18-34 years old; 35-64: participants are aging 35-64 years old; < 64: participants are aging 64 years old and older; $M_{1/3}$h (M8h), M120, M60, M30, M15, M5: minimum acceleration for most active accumulated 8 hours, 120 min, 60 min, 30 min, 15 min, and 5 min of all days; M15h accel: overall daily activity based on most active continuous 15 hours of all days (mg); M15h IG: intensity distribution over the most active straight 15 hours of all day.

* P < 0.01: significant difference between age groups

A selection of the MX variables is further illustrated in figure 9. The figure illustrates a lower area within the curve for the oldest age group than the younger age groups, illustrating similar findings as Table 6 but slightly more illustrative. In this radar plot, the red dotted line indicates threshold values (mg) for the listed activities. For example, M15 – the acceleration above which a person's most active non-consecutive 15 minutes (M15) are accumulated, the thesis observes that the youngest age group has an M15 value above the threshold for doing 15 minutes of brisk walking for a given day. In contrast, the oldest age group has an M15-value of below the same threshold. Similarly, it was observed that the mean M30-value for the age groups is well below the threshold for doing 30 minutes of brisk walking, indicating that the average individual does not meet guidelines if they are operationalized *as doing 30 minutes of brisk walking per day*.

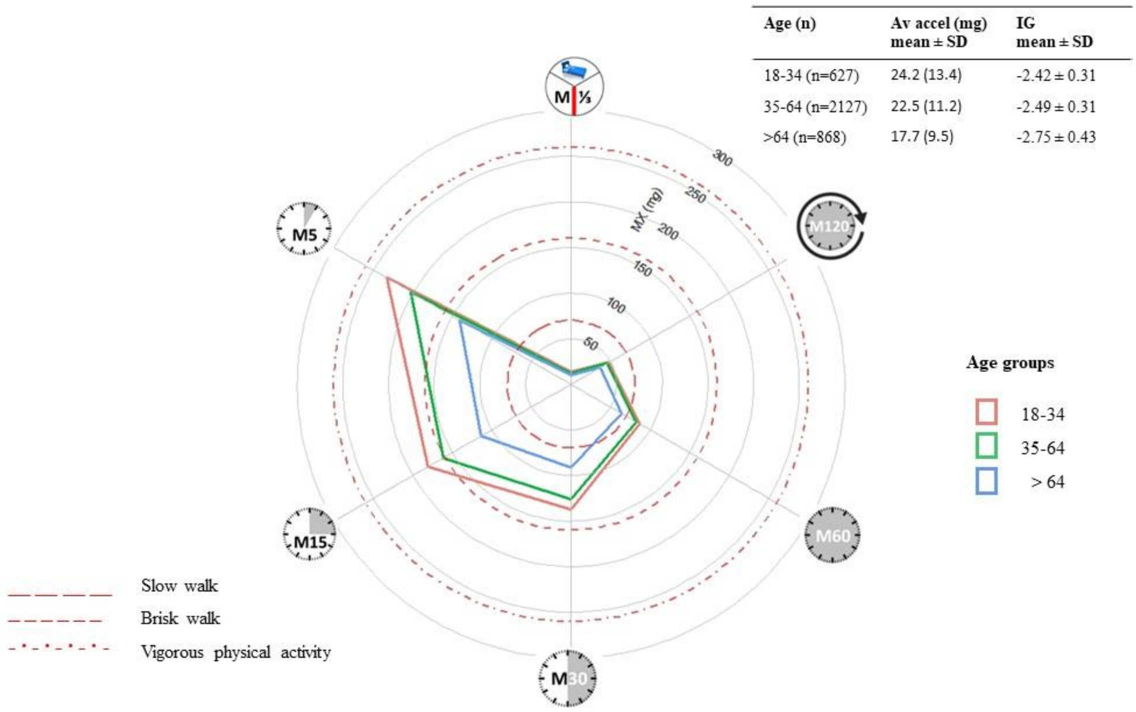

Figure 9. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different age groups unequal values for average acceleration and IG.

Exploring results for M60, the acceleration above which a person's most active non-consecutive 60 minutes are accumulated indicates that young adults (red line) and adults (green line) have an acceleration level that represents intensities greater than slow walking for 60 minutes of a day. Further, the highest measured values for all age groups are found for the M5, where young adults and adults achieved acceleration values well above the threshold representing brisk walking. The intensity limit of 260 *mg*, representing vigorous activity, was not met by either age groups or within selected MX-variables. An indication that values representing vigorous activity are equal or lower than five minutes during a day.

4.3.2 Body mass index (BMI)

Implementation of a similar methodological approach, as for age groups, was performed for BMI categories as to age groups below. Section of MX-variables is described and illustrated through table 7 and figure 10. Results for normal, overweight, and obese participants are illustrated with a red, green, and blue line. Indication of IG and average intensity reveals decreased overall activity and increased intensity distribution with increased BMI. Results imply significant disparities between BMI categories except for $M_{1/3}h$ ($p < 0.01$). For instance, the M30 reveals results lower than the threshold value of 160 *mg* for all BMI categories and are first met towards M15. Figure 10 illustrates further below that normal weighted participants accomplish the minimum acceleration needed to be classified as brisk walking for at least 15 minutes of a day.

Table 7. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for BMI categories (n = 3622).

	Normal (km/m²)	**Overweight (km/m²)**	**Obese (km/m²)**	**Total**
N (%)	1886 (52)	1325 (37)	411 (11)	3622 (100)
$M_{1/3}h_{mg}$	11.9 (10.8)	12.2 (10.6)	11.6 (12.5)	12.0 (10.9)
$M120_{mg}$	46.0 (25.3)*	42.9 (21.6)	38.9 (21.8)	44.1 (23.7)
$M60_{mg}$	83.9 (41.4)*	75.0 (33.5)	65.3 (29.5)	78.5 (38.0)
$M30_{mg}$	128.9 (60.2)*	113.1 (50.3)	95.3 (37.7)	119.3 (55.7)
$M15_{mg}$	165.9 (71.8)*	145.4 (62.8)	122.6 (48.4)	153.5 (67.8)
$M5_{mg}$	208.6 (83.9)*	183.2 (74.9)	155.3 (57.8)	193.3 (80.1)

BMI normal: < 24.9; BMI overweight: 25.0-29.9; BMI obese: > 29.9; $M_{1/3}h$ (M8h), M120, M60, M30, M15, M5: minimum acceleration for most active accumulated 8 hours, 120 min, 60 min, 30 min, 15 min, and 5 min of all days; M15h accel: overall daily activity based on most active continuous 15 hours of all days (mg); M15h IG: intensity distribution over the most active straight 15 hours of all day.
* P < 0.01: significant difference between BMI categories normal, overweight, and obesity

The cut point limit of 160 *mg* by overweight participants was not met except for the M5-variable (183.2 ± 74.9 *mg*) in the following analysis. Overweight participants realize the threshold limit for *at least 5 minutes of brisk walking* during a day, alongside participants classified as normal weight (208.6 ± 83.9 *mg*). On the other hand, participants classified as obese accomplished slow walking as highest activity threshold (155.3 ± 57.8 *mg*). Among BMI categories, normal (-2.47 ± 0.36) and overweight (-2.58 ± 0.35) participants had the highest reported value of intensity distribution over the most active continuous 15 hours of the day. Overall, obese participants had the lowest overall daily activity (18.8 ± 11.4 *mg*) but accomplished PA guidelines (i.e., 70 *mg* threshold) of 150 minutes of MVPA through M30-variable (95.3 ± 37.7 *mg*). Similar results for normal (83.9 ± 41.4 *mg*) and overweighted (75.0 ± 33.5 *mg*) participants were found, though for the M60-variable.

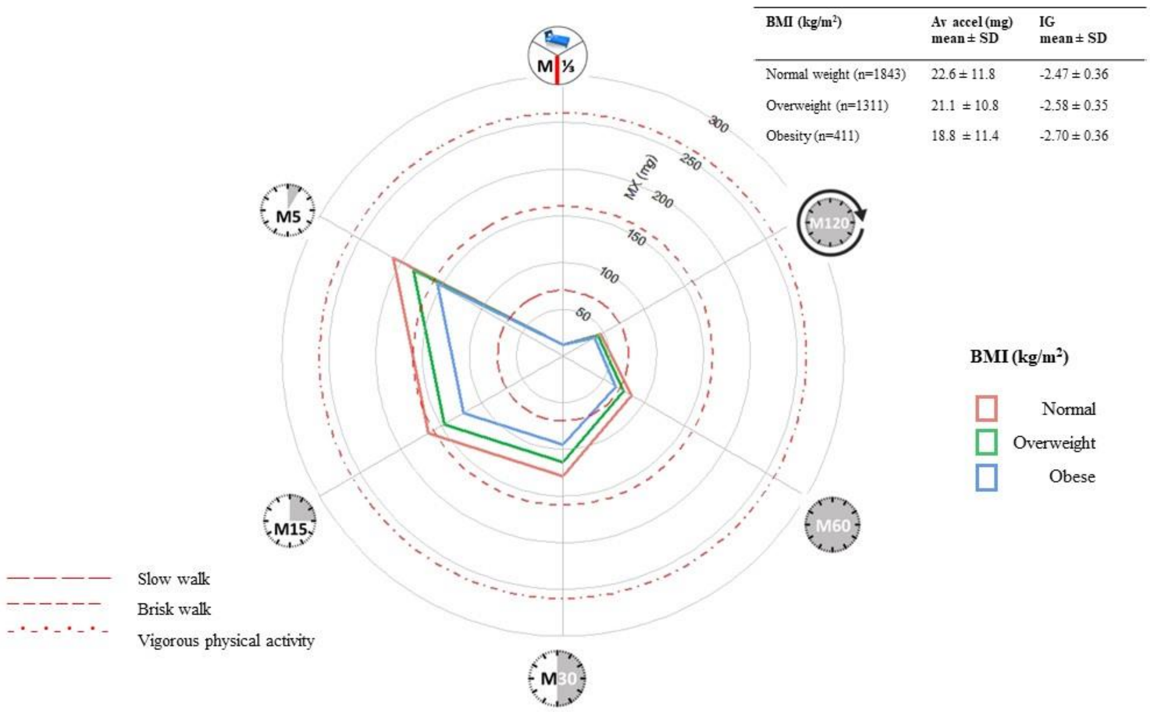

Figure 10. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different body mass index groups with unequal values for average acceleration and IG.

4.3.3 Educational attainment

Table 8 describes mean values of minimum acceleration for a given duration (*mg*) towards participant educational attainment and is illustrated in figure 11 below. Results for primary school, high school, and a university degree are illustrated with a green, red, and blue line. All data were collapsed across sex and further adjusted with sex as co-variate. Amongst education attainment, a significant difference was found for the most continuous 15 hours of all days, regarding average intensity and IG ($p < 0.01$). Primarily, participants with a university degree had the highest overall activity (22.6 ± 12.3 *mg*) and lowest IG (-2.46 ± 0.34). On the other hand, participants with the primary school as highest degree had the lowest overall activity and a higher value for IG: according to extracted values from figure 11.

Table 8. Mean values of minimum acceleration for a given duration expressed as mg-values adjusted for educational attainment (n = 3622).

	Primary	**High school**	**University**	**Total**
N (%)	311 (9)	1357 (38)	1948 (53)	3622 (100)
$M_{1/3}h_{mg}$	11.0 (8.9)	12.0 (9.9)	12.1 (11.8)	12.0 (10.9)
$M120_{mg}$	39.3 (19.3)	42.9 (21.8)	45.6 (25.4)*	44.1 (23.7)
$M60_{mg}$	67.0 (34.4)	74.5 (34.2)	83.1 (40.2)*	78.5 (38.0)
$M30_{mg}$	96.1 (49.0)	110.9 (48.8)	128.8 (59.2)*	119.3 (55.7)
$M15_{mg}$	119.4 (57.4)	141.6 (58.8)	167.2 (71.8)*	153.5 (67.8)
$M5_{mg}$	147.8 (65.1)	179.1 (70.5)	210.3 (83.8)*	193.3 (80.1)

$M_{1/3}h$ (M8h), M120, M60, M30, M15, M5: minimum acceleration for most active accumulated 8 hours, 120 min, 60 min, 30 min, 15 min, and 5 min of all days; M15h accel: overall daily activity based on most active continuous 15 hours of all days (mg); M15h IG: intensity distribution over the most active straight 15 hours of all day.
A total of 6 participants had insufficient values representing their educational attainment
* P < 0.01: significant difference between age groups

Significant results regarding educational attainment are illustrated with an asterisk in table 8. Differences across educational attainment were found for all MX-variables ($p < 0.01$), except for $M_{1/3DAY}$, where participants with university degrees had the highest measured acceleration values during a day. As for all age groups and BMI categories, any educational degree threshold for M30 is not met. Average acceleration values reveal that participants with a university degree achieve *at least* 15 minutes of brisk walking per day: as do high school for the M5-variable. Both high and primary school participants realize the threshold line representing slow walking, with a significant margin, at least for 15 minutes on all days. University-degree participants had a minimum acceleration level closer to vigorous activity (260 *mg*) than brisk walking (160 *mg*). In contrast, primary-school participants do not achieve brisk walking for at least 5 minutes on all days.

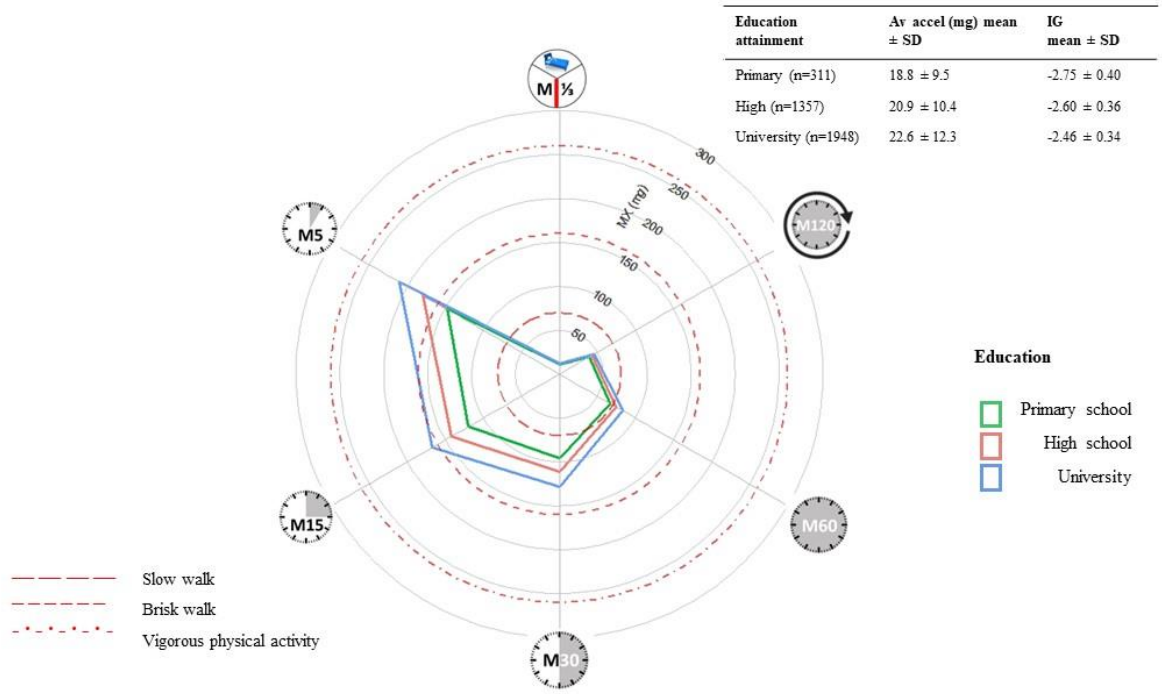

Figure 11. Radar plot illustrating MX metrics for the most active 8 h of the day (M1/3DAY), 120 min (M120), 60 min (M60), 30 min (M30), 15 min (M15), and 5 min (M5) for three different educational attainments with unequal values for average acceleration and IG.

4.4 Comparison between the various device-measured indices of PA

An examination between the average intensity of the most active continuous 15 hours of all days and CPM was conducted to assess overall activity among the participants. A Person correlation of average acceleration was .460 and illustrated a modest positive relation to the CPM variable. Figure 12 summarizes the results below. The CPM variable explained 21.2 % of the variation in average intensity. The regression coefficient (β = .46) indicated that an increase in one unit of average intensity, on average, increased CPM values by 0.46 units.

A similar assessment was performed when comparing estimates of overall activity, specifically MVPA. When comparing $MVPA_{old}$ with $MVPA_{new}$ (Figure 13), it is essential to note that data reduction criteria are not directly comparable due to the nature of the two different approaches (e.g., different wear-time criteria, epochs, and apprehending cut-points). While the new values are based on the most active 15 hours of any given day, the old data are based on at least 10 hours of valid wear time. Hence, differences are also found for epochs where the new dataset capture acceleration in 5 s epochs and old in 60 s epochs. The Pearson correlation coefficient was 0.367, $F(1, 3588) = 560.1$, $p < 0.01$), illustrate a positive correlation among the two intensity variables.

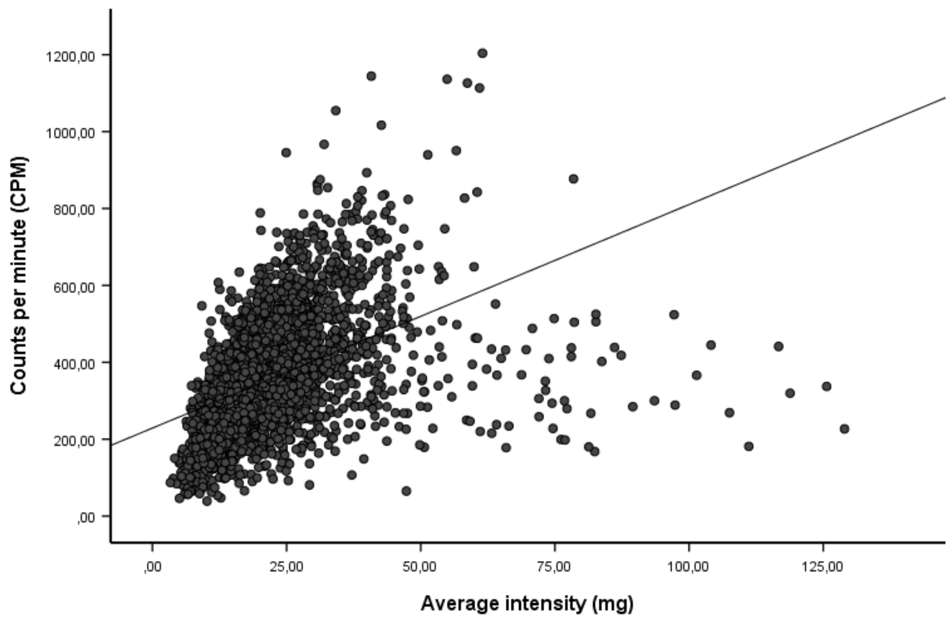

Figure 12. Scatterplot compares CPM and average intensity over the most active continuous 15 hours of all days (n = 3589).

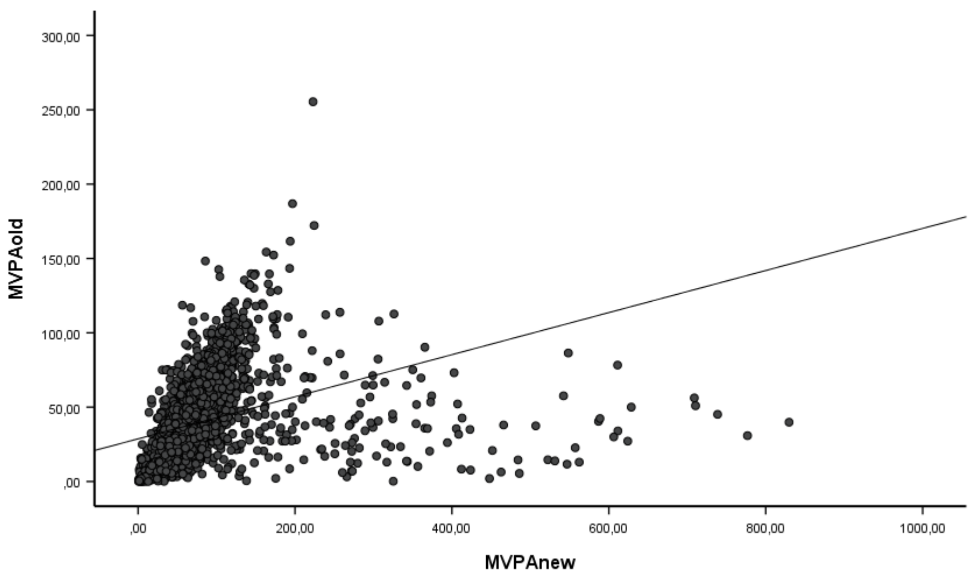

Figure 13. Scatterplot comparison between MVPAold and MVPAnew (n = 3589).

When comparing the mean values between the MVPA-variables, and sex differences between the variables, differences in significance level and percentage difference was collaborated, as described in table 9 below. An independent t-test, testing sex differences, report of no significant difference between females (M = 70.7, SD = 66.2) and males (M = 70.8, SD = 66.9; t(3587) = -.048, p = .962) according to the $MVPA_{new}$. The $MVPA_{old}$ found significant differences among female (M = 37.3, SD = 25.2) and male participants (M = 40.5, SD = 26.0; t(3587) = -3.656, p < 0.01). Furthermore, 47.2 % and 43.1 % sex differences from new and old MVPA-variable were extracted from the analysis.

Table 9. Comparison of MVPAnew and MVPAold across sex and intensity-variables: expressed as mean values (SD) (n = 3589).

				Paired sample t-test			
	Females	Males	Difference (%)	*t* value	95 % CI mean the difference		Sig (two-tailed)
$MVPA_{new}$	70.7 (66.2)	70.8 (66.9)	-.14	-.048	- 4.49	- 4.27	.962
$MVPA_{old}$	37.3 (25.5)	40.3 (26.1)	-8.04	-3.656	- 4.82	- 1.46	.001
Difference (%)	47.2	43.1					
	New	Old					
MVPA	70.7 (66.5)	38.7 (25.6)	45.3	30.976	29.97	34.02	.001

A paired sample t-test was conducted to investigate the differences between $MVPA_{new}$ (M = 70.7, SD = 66.5) and $MVPA_{old}$ (M = 38.7, SD = 25.6), regardless of sex. Significant differences were found among the intensity variables, t (3688) = 30.976, p < 0.01. The difference of 45.3 %, in favor $MVPA_{new}$, illustrates higher proportions of participants achieving the recommendations for PA during a day. The raw acceleration approach ($MVPA_{new}$) captures a higher volume of activity at higher intensities than the cut point approach ($MVPA_{old}$). Further, numerous linear regressions were calculated to predict the activity variables, alongside age, sex, and educational attainment, to predict participants' BMI (table 10). Results from these equations reveal if new acceleration metrics (average intensity and IG) are equal to or less than the CPM approach.

Table 10 summarizes results from multiple linear regressions regarding calculating activity variables, age, sex, and educational attainment's prediction of BMI. Firstly, a Pearson correlation found average intensity (1) as moderately negative correlated, *r*(3583) = -.021, and not a significant predictor of BMI with p = .550. Significant regression equation was further developed, (F(4, 3578) = 34.890 p < 0.01, with an R^2 of .038. According to the following results, average intensity explained 3.8 % of the actual BMI value. Furthermore, participants predicted BMI is equal to 24.664 + 0.01 (age) + 1.230 (sex) – 0.611 (education) = 25.293, where (b_1) is the actual age, (b_2) sex is coded as 1 = male and 2 = female, and (b_3) educational attainment are coded 1 = primary school, 2 = high school, and 3 = university degree. Participant's BMI decreased with 0.611 in educational attainment and enhanced 0.01 with age. Both sex and educational attainment were significant predictors of BMI.

Table 10. Comparison of activity variables, age, sex, educational attainment, and BMI prediction (n = 3583).

Model	Dependent variable	β	Age (b_1)	Sex (b_2)	Education (b_3)	95 % CI mean difference for β		R^2
1	BMI	24.664	.010	1.230	-.611	23.7	25.6	.038
2	BMI	19.553	-.002	1.255	-.407	18.3	20.8	.065
3	BMI	24.830	.008	1.235	-.592	24.0	25.7	.041
4	BMI	26.235	.005	1.296	-.576	25.3	27.1	.062
5	BMI	25.367	.006	1.324	-.493	24.5	26.2	.066

1: Average intensity; 2: Intensity gradient; 3: $MVPA_{new}$; 4: Counts per minute; 5: $MVPA_{old}$

Similar linear regression was executed with IG (2), age, sex, and educational attainment, and their estimate of BMI score. The Pearson correlation retrieved a modest negative relation, $r(3583) = -.194$, $p < 0.01$, but a strongly significant predictor of BMI. A significant regression equation report of the same, ($F(4, 3578) = 62.514$, $p < 0.01$, with an R^2 of 0.65. This R^2 value refers to 6.5 % of the actual BMI value that can be accounted for with IG as an activity variable. According to following regression, participant's projected BMI is equal to 19.553 – 0.002 (age) + 1.255 (sex) -0.407 = 20.399. The results classify sex, educational attainment, and IG as significant predictors of BMI ($p < 0.01$). Furthermore, IG's prediction of actual BMI is as good as CPM results (3) below.

Regarding activity intensities, linear regressions were performed with $MVPA_{new}$ (threshold 70 *m*, min/day), age, sex, and educational attainment and used to calculate BMI. Output from Pearson correlation report of an insignificant negative relation and significant predictor of BMI, $r(3583) = -.073$, $p < 0.01$. The regression equality report of similar with an R^2 value of .041, ($F(4, 3578) = 37.947$, $p < 0.01$. Interpret, the new intensity variable accounts for a marginal percentage of predicted BMI with 4.1 %. The predicted BMI is thereby 24.83 + 0.008 + 1.235 – 0.592 = 25.481: a predicted value of overweight towards the sample's participants. Nevertheless, the intensity variable is a strong predictor of BMI despite a lower R2 value than $MVPA_{old}$.

The fourth linear regression was executed with similar contextual variables with CPM (4) and its prediction of the participant's BMI value. The Pearson correlation classify the equation as modestly negative with a R^2 value of .062, $r(4, 3578) = -.163$, $p < 0.01$. The R^2 value, small but percentages difference to average intensity, report that the CPM approach can describe a minimum of 6.2 % of predicted BMI. With that in mind, the BMI for this linear regression equivalent is predicted to 26.235 + 0.005 + 1.296 – 0.576 = 26.96. Considering results from previous regressions, CPM can report the highest predicted BMI value for all participants and can be determined as overweighted by a decent margin. Additionally, apart from age, every included variable is considered a significant predictor for BMI ($p < 0.01$).

Finally, a fifth regression equation was performed with age, sex, educational attainment, $MVPA_{old,}$ and actual BMI expectations. The equation is, through a Pearson correlation, modestly negative correlated but significant, $r(4, 3578) = -.175$, $p < 0.01$. The R^2 value report that 6.6 % of actual BMI can be predicted with $MVPA_{old}$ as an activity variable. The predicted BMI of 26.114 (25.367 + 0.006 + 1.324 – 0.493) indicate an overweighted sample of participants, alongside with previous regression analysis: except for IG. Moreover, the intensity variable is considered a significant predictor of BMI: accompanied by sex and educational attainment. In the interest of prediction score, report $MVPA_{old}$ higher than $MVPA_{new}$ and similar R^2 value as IG.

5. Discussion

This thesis has provided population estimates of PA based on numerous activity variables and compared new metrics to the traditional CPM approach. Results show significant differences between age, BMI, and educational groups and the proportion of participants who meet current recommendations for PA ($p < 0.01$). Significantly, participants aged 18-34 years old are normal weight according to BMI classification, have a university degree, have considerably higher overall activity and total MVPA, compared to the other groups of participants. Furthermore, the selected activity metrics (IG and CPM) comparability was equally good in predicting BMI and the traditional accelerometer metrics. Thus, the results emphasize the feasibility of using the new accelerometer metrics to aid comparability across studies. Furthermore, the combination of analytical and translational metrics provides functional and user-friendly indices of PA for both researchers and end-users.

5.1 Results discussion

Results from this thesis are primarily based on an analytical and translational metric except for CPM in subchapter 4.4. While it might be interesting to re-run existing CPM values with this thesis inclusion/exclusion criteria, Kan2 contains the same proportion of participants and accelerometer signal. Results from 2015, based on CPM values, were considered representative enough in comparison with other accelerometer signals. It was the same participants, and the results for the CPM variable have already been analyzed. The inclusion of these other activity variables (e.g., $MVPA_{new}$, $MVPA_{old}$, MX metrics, IG, and average acceleration) was considered necessary to *provide normative new metrics for population estimates of PA*. However, the result differences between Kan2 to this thesis are 1) descriptive, 2) valid wear time, and 3) comparisons between accelerometer signals.

First, the results for descriptive are based on age, height (cm), weight (kg), overall BMI and categorized (kg/m^2), and educational attainment. While Kan2, one average, operated with five age groups (i.e., 18-34 yr, 35-44 yr, 45-54 yr, 55-64 yr, and 65 and older) selected this thesis three (i.e., 18-34 yr, 35-64 yr, and 65 and older). According to both analytical and MX metric results, the inclusion of the 34-64 age group was based on non-significant differences between participants between ages 20 and 64. Furthermore, height and weight were combined through BMI (kg/m^2), which has several disadvantages.

BMI is used to capture body fat and differencing populations into categorized groups: underweight, normal weight, overweight and obese. Rothman (2008) exemplifies disadvantages as percentages of body fat increase biological with age and the incapableness to differential muscle mass from actual fat. In addition, BMI is an indirect measurement for body fat, not directly as DEXA, and self-reported data could contribute to an over-and under-estimation of actual BMI for adults (Rothman, 2008; Yoshitake et al., 2012). For instance, females often underestimate actual weight compared to males (Allison et al., 2020; Yoshitake et al., 2012). Considering Kan2 used self-reported data to determine the participant’s weight and further BMI, the data for this thesis is hampered by its credibility. BMI results found for females could be higher than reported, and well-trained participants with high percentage muscle mass could be classified as overweight. However, BMI is used as a health variable to classify the participant's health status, compare different accelerometer signals, and predict health variables.

A considerable number of Kan2-participants have a high socio-economic background (i.e., educational attainment, income, and health status). Therefore, the population levels of PA derived from Kan2 might be considered somewhat overestimated because of the well-known phenomena as the healthy participant's bias and social desirability bias. Such assumptions are equally applicable to this thesis. Regardless of sex, the proportions of participants with high educational attainment are significantly different from those with primary school (51 % vs. 8 %). The descriptive data reveals high proportions of elderly with primary school, whereas most participants aging 18-34 have a university degree. As the age groups, respectively, consist of 627 (18-34) and 868 participants (>64), there is a chance of some sort of selection bias. Hence, while the Kan2-sample had fewer participants (n = 3173 vs. 3622) than this thesis, the percentage difference in educational attainment was the same.

Secondly, the results for overall activity (i.e., average acceleration and IG) and accumulated time in intensities during an average day (i.e., $MVPA_{new}$, and VPA_{new}) reveal significant differences between age groups in subchapter 4.2. As found in Kan2, participants aging 18-34 years old had the highest overall activity level and time within MVPA and VPA. Hence, from these results, it is not possible to determine if the participants had higher activity levels on weekdays than on weekends. The results are entirely based on activity captured for all days. Activity variables differencing a week’s overall activity were available in autumn 2020 but excluded in the final results. The exclusion is based on a subjective opinion and the fact that this thesis's first assignment was to provide an estimate of PA with new metrics. From Kan2 is emphasized a modest increase in activity level for females, from weekends to weekdays, whereas males aging 65 years and older had significant differences in activity level (Helsedirektoratet, 2015). As the participants in the thesis are subtracted from the Kan2 survey, it might be possible to draw a similar conclusion regardless of dataset (e.g., raw accelerometry or CPM). Though the results from Kan2 are based on CPM values, it would be interesting to investigate the differences using other average acceleration variables. The estimates for weekdays could be even higher than found in Kan2.

Thirdly, the values describing average acceleration during a day provide only volume and cannot alone be used to describe time within intensities. Migueles et al. (2021) described that estimating a 24-h movement profile needs more than one analytical metric to capture volume and intensities. Results from the IG variable reveals a steeper drop in intensity (i.e., increase in negative values) with increasing age, and later found to be an equally good predictor for BMI as the CPM variable. Hence, the results are based on hip-worn ActiGraph GT3X+. Compared to Axivity and GENEActiv, researchers have found that the metrics are wear-site specific, and the monitor might underestimate the magnitude of acceleration (Rowlands et al., 2019a). The specificity of metrics hampers comparability for different attachment sites (e.g., wrist vs. hip) and could underestimate acceleration values compared to other brands. Therefore, it is plausible that the results might misjudge actual overall activity and time within MVPA and VPA. However, this is a subjective opinion and does not leave out the many benefits of the analytical approaches compared to the CPM approach.

Intensity estimates for all days were primarily conducted using the MVPA-variable (i.e., a threshold value of 70 *mg*) with some results describing VPA. As results for vigorous activity are noteworthy in meeting PA guidelines, the VPA variable is left out in comparisons of subchapter 4.4. The MVPA-variables were more exciting and subjectively selected to estimate participants who met the PA guidelines. The differences in results for $MVPA_{new}$ among age groups was like Kan2, but with higher estimates were reported. As the $MVPA_{new}$ compares acceleration values towards a suitable threshold and not time above or below as found for the CPM approach, it is questionable if the variables are comparable. Considering the differences found in data management (e.g., 10-h instead of 12-h valid wear-time during a day) and methodological approaches (e.g., raw accelerometry vs. CPM), there is no surprise to this thesis that $MVPA_{new}$ reports more minutes at higher intensities. The number of participants apprehending determined PA recommendations are therefore higher than reported in Kan2.

The analytical metrics are further distributed and translated with selected MX metrics in subchapter 4.3. The results regarding age (i.e., 18-34 yr, 35-64 yr, > 64 yr), BMI (i.e., normal, overweight, obese), and educational attainment (i.e., primary, high school, university) are based on acceleration above which a person's most active non-consecutive X minutes spent over the day. The results are based on average values and do not consider the number of participants with values above the average. For example, figure 9 imply that neither age group, on average, performs brisk walking *for at least* 30 minutes during an average day. The reality is that a substantial proportion of participants have higher accumulated acceleration values than illustrated. The assumption is applicable for every contextual variable (i.e., age, BMI, educational attainment) and might have unfortunate impacts in interpreting the results, even though the thresholds represent activities. Especially, a researcher with little or no knowledge in interpreting the MX metrics structure could have difficulty understanding the results.

In comparing accelerometer signals (i.e., analytical and translational metrics vs. CPM-approach) with multiple linear regressions, IG and CPM were equally good predictors of BMI. As emphasized by Migueles et al. (2021), is it possible to establish a relationship among different explanatory variables and a health outcome. The result implies similarities among IG and CPM, but not for average acceleration and CPM. Such inequality was also found between $MVPA_{new}$ and $MVPA_{old}$. However, the R^2 and predicted beta coefficient had minor to modest differences between the metrics and might questionnaire if more than one analytical metric is equally good in predicting BMI. The question *as good* will require new metrics to be as good as the old metric. Considering the assignments for this thesis and minor to no differences between the accelerometer signals, average acceleration, and $MVPA_{new}$ are as good as CPM and $MPVA_{old}$.

In comparison to previous research, Buchan and McLellan (2019) conducted similar comparisons between raw and count processing methods using hip-worn ActiGraph GT3X+, though using children as participants. The researchers found significant time differences in intensities were estimated for MPA was considerably higher for raw accelerometry than CPM-approach (Buchan & McLellan, 2019). Hence, when comparing raw and count-based methods (e.g., MPA, VPA, MVPA), equivalence tests showed insufficient similarities amongst the variables and were considered not comparable (Buchan & McLellan, 2019). In addition, the authors exemplified differences in calibration techniques (e.g., auto-calibration from GGIR) and implementation of conversion factors (e.g., the Rosetta Stone equations) to facilitate the comparability between studies that process and report raw and count activity data (Buchan & McLellan, 2019). Hence, as these comparisons are of interest to this thesis, the comparison was purely based on intensity variables and the use of different cut-points to distinguish differences. As this might be the case of the MVPA-variables used in this thesis, is it uncertain if these results are representative towards other activity variables (e.g., average acceleration, IG, MX metrics).

The result from this thesis and Buchan and McLellan (2019) are of interest as they imply the differences between the accelerometer signals: except for IG and CPM ($R^2 =$ 0.65-0.62). Furthermore, linear regressions for MX metrics were available in the prediction of BMI values. However, they were absent due to 1) the interest of analytical metrics comparability towards CPM, 2) insecurity of which MX metric should be used in the analysis, and 3) the risk of collinearity among the explanatory variables (e.g., age, sex, educational attainment). Collinearity can be defined as a phenomenon that occurs when predictor variables are highly correlated and creates a problem when exploring the explanatory variables (e.g., MX metrics) towards an outcome variable (Migueles et al., 2021). Therefore, it is suggested by Migueles et al. (2021) to use multivariate pattern analysis (i.e., approach that handles an unlimited number of multicollinear explanatory variables) to handle the risk of collinearity: by comparing MX metrics towards an outcome variable: instead of linear regression equations. However, this approach was neither considered nor used in this thesis.

5.2 Method discussion

This thesis is based on a cross-sectional study examining the use of new and old metrics on PA levels and their comparability. As it is cross-sectional, the estimate for PA is suitable to estimate the prevalence of PA among Norwegian adults. However, a disadvantage with this design is the incapableness of explaining variables over time (Kolle & Grydeland, 2018). For example, this thesis does not know if low activity levels cause participants' weight over time or if the low activity level is the cause of their BMI status. However, the methodology behind this thesis is based on two different datasets using a similar accelerometer brand, attachment site of accelerometer, and participants from Kan2. The inclusion of raw accelerometry and CPM approaches makes it feasible to compare accelerometer signals (e.g., IG vs. CPM) and discover differences from Kan2 towards this thesis. The results classified them as equally good in predicting a health outcome through multiple linear regressions with several accelerometer signals. In addition, a higher percentage of participants met the PA recommendations compared to the Kan2 survey.

In apprehending results from Kan2, using two different sets of activity variables, this thesis encountered several advantages and disadvantages in data management. Among benefits, analysis and translation of raw accelerometry were considered easy to generate and appliable using the free, open-source software from R, GGIR (Rowlands et al., 2019a). In combination with specific algorithms and RStudio, analytical and translational metrics were produced and presented using radar plots. Regardless of the dataset, sampling frequency was collected at 30 Hz, and participants used ActiGraph GT3X+ over seven consecutive days. Rowlands et al. (2019a) presented that the epoch should be equal between datasets as more significant epochs tend to omit higher intensity activity. For example, CPM results could miss out on levels of activity recorded by the new metrics (Rowlands et al., 2019a). The difference between 5 and 60-second epochs from Kan2 creates uncertainties if the different datasets are comparable. The difference in epochs can be expressed through results found for total MVPA-variables.

The $MVPA_{old}$-variable report of approximately 38 minutes, while $MVPA_{new}$ implies 70 minutes per day ($p < 0.01$). The differential of threshold values, reduction criteria (e.g., *mg* or counts/min), and determination of valid wear time (e.g., 10 or 12 hours) are, with epochs, possible reasons for the difference in MVPA results. *First*, the cut point approach uses an absolute threshold value of 2020 counts/min to capture MVPA. While interpretation and readability are easily current for researchers, activity close to the threshold will be classified as light intensity and may give an unrepresentative reality of overseeing applicable recommendations. Second, while results from the Kan2 report of 1 out of 3 adults met the PA recommendations, new metrics in this thesis found 2 out of 3 adults. The higher proportions of adults meeting recommendations might be due to capturing more valid wear-time during a day and comparing to a threshold instead of above-or-below-principle from the CPM approach.

Secondly, when the cut point approach reduces raw accelerometry into count values, $MVPA_{new}$ utilizes all accumulated acceleration above 70 *mg* as min/day. However, the new variable is not flawless in terms of structure. The threshold value was determined by Alex V. Rowlands upfront of this thesis in the autumn of 2020; the same author behind selected analytical and translational metrics used in the results. The threshold value of 70 mg is the same intensity limit set for slow walking in subchapter 4.3 and maybe questioning the $MVPA_{new}$ representativity towards higher intensity groups. Hence, the threshold value derives from Hildebrand et al. (2014), who investigated different accelerometer brands and attachment sites. Results regarding adults using ActiGraph GT3X+ located at their hip found an acceleration intensity threshold (*mg*) of 69.1 as sufficiently representative of moderate PA (Hildebrand et al., 2014). Thereby, 70 *mg* was used in the calculation of participant's activity levels at higher intensity groups.

Thirdly, researchers behind Kan2 determined that a minimum of 600 minutes, or 10 hours with registered minutes, must exist to be classified as a valid day (Helsedirektoratet, 2015). Determination of a valid day is similar (10 $h \cdot d^{-1}$) or differs (16 $h \cdot d^{-1}$) from recently published studies regarding measures of PA objectively, and the second dataset includes in this thesis: 12 hours (Fairclough et al., 2019a; Rowlands et al., 2019b; Watson et al., 2014). Furthermore, in determining valid days, activity below 3 and above 7 days was considered unrepresentative and removed from the thesis dataset (Rowlands et al., 2019a). Unfortunately, the exclusion criteria were only used for the 12 hours variable and not for the 10 hours, resulting in the elimination of 874 participants. While results within subchapters 4.2, 4.3, and 4.4 are adjusted for both validation criteria, is it possible that the exclusion was unnecessary and could have been prevented. Without knowing, this decision might influence the anthropometric values for all participants, and exclude participants who originally overseeded the validation criteria regarding the 10-hours variable.

Fourth, the data management might have increased the possibility of selection bias from 5052 to 3622 participants. For example, initially from Kan2 accomplished 32 % of all Norwegian adult's (n=3020) recommendations set for PA at least 150 minutes of moderate PA (Helsedirektoratet, 2015). Compared to this thesis, 71.8 % of all PA recommendations were obtained for the participants, based on a similar sample of participants retrieved from Kan2. Furthermore, given a minimum of 21.40 minutes MVPA each day, regardless of CPM or the raw accelerometry approach, the results were considerably higher. Thus, the MVPA results might be questioning the credibility of the results and should be considered when interpreting this thesis results. The significant difference between the MVPA variables is though interesting. It is possible that the difference in data management favors $MVPA_{new}$ due to its possibility to capture more activity during a day, number of registrations per minute, and avoids collapsing the raw acceleration into determined cut-point thresholds. Such assumption is, though, a subjective opinion.

Fifth, while the accelerometer is presented as the future for PA surveillance nationally and worldwide, it is hampered by several biases in capturing and monitoring activity (Guthold et al., 2018; Hansen et al., 2018b). As mentioned by Steene-Johannessen et al. (2018), the device-monitor is hampered by 1) lack of contextual information (i.e., collected with IPAQ in Kan2), 2) increased activity due to the feeling of being monitored (i.e., Hawthorne-effect), and 3) numerous activities are either underestimated (i.e., walking with backpack or up-hills) or not recorded (i.e., bicycling) (Steene-Johannessen et al., 2018). Without contextual information of accumulated PA, there is impossible to classify the participant's surroundings and weather conditions when executing activity while, for example, bicycling and swimming. However, PA is considered a *desired behavior,* and self-reported data consisting of mentioned activities rely on cognitive ability to remember intensity, duration, and frequency (Hansen et al., 2018b).

In addition, deployment strategies (e.g., attachment site, number of days classified as a representative, and determination of wear-time as day- or night time) and data reduction strategies (i.e., raw accelerometry and CPM) differ between studies. The differences in device measurements require additional consensus than the GRANDA consensus by Migueles et al. (2021). Steene-Johannessen et al. (2018) highlight other research question is to be considered whenever using an accelerometer: 1) definition of non-wear-time, 2) the number of axis in use (e.g., vertical, mediolateral, and anteroposterior), and 3) distribution of accelerometer to participants, and, finally, 4) assess participants sufficient information regarding the use of accelerometer.

The difference in data management is further found between accelerometer signals (i.e., CPM and MX metrics) and their way of retrieving results for overall activity and PA guidelines. Concerning PA guidelines, researchers behind the GRANADA consensus concluded that interpretation of average acceleration values is of no relativeness unless IG is included

(Migueles et al., 2021). Similar was described towards IG alone, Migueles et al. (2021) termed insecurity whether the activity variable is adherence to PA recommendations. More research is needed for the analytical metrics, although MX metrics were considered suitable for enabling public health messaging (Migueles et al., 2021). As CPM stands for counts per minute, results for overall activity in Kan2 are based on converted data: which may favor readability for the researcher. Nevertheless, the overall activity level from CPM is affected by cut-point assumptions on energy expenditure and could differ between datasets (Migueles et al., 2021)

As this thesis's first assignment phrase, providing normative new metrics for population estimates of PA, radar plots were used as suitable translational charts for average acceleration, IG, and MX metrics (Rowlands et al., 2018). Through specific algorithms and mean values from IBM SPSS 25, radar plots were enabled and used; see Attachments. The comprehending of mean values in RStudio might seem incomprehensible for inexperienced researchers and requires specific knowledge. For instance, without guidance from Alex V. Rowlands in interpreting the new metrics and algorithms, this thesis would never have reached the delivery deadline of 28th of May 2021. The algorithms take time to detain, and minor flaws as missing parenthesis or commas in creating radar plots might result in error messages. However, the RStudio has a tutorial available for beginners and makes radar plots feasible and understandable.

Rowlands et al. (2019a) have prompted radar plots to translate analytical metrics into interpretable illustrations. An advantage for further population-based research is the possibility to use other accelerometer signals (i.e., MVPA and IG) as a dependent activity variable for radar plots. Rowlands et al. (2019a) exemplified the use of different accelerometer signals, where MPVA exemplifies the proportion of participants meeting the PA recommendations. Additionally, IG can be appropriate to use to illustrate differences in intensity during a day. Both suggestions were considered in the analysis but omitted in the interest of other variables (e.g., age, BMI, and educational attainment). As this thesis focuses only on between-group comparisons, it was possible to illustrate 1) within-groups comparisons, 2) differences across periods (i.e., week vs. weekends), and 3) to different wear sites (Rowlands et al., 2019a). Rowlands et al. (2019a) highlight that while the metrics (i.e., analytical and translational) are appliable to any wear-site, and results appear similar between datasets, the wrist-worn accelerometer could report higher MX values than other attachment sites. Therefore, it is advised to use the same wear-site.

In the AC approach (i.e., CPM), accelerometer brands, protocols, and determination of threshold values have been differently treated and affected the number of participants overseeing recommendations for PA (Migueles et al., 2019b). For instance, the proportion who oversee recommendations can differ by up to 80 %: depending on the MVPA threshold (i.e., 200 mg to 250 mg) used (Rowlands et al., 2019a). It is determined by the participant's ability to score above or below the given threshold. These are protocol and population-specific, making comparability between surveys impossible: unless researchers use the same methodological approach. For instance, the reality of such assumptions between Kan2 (ActiGraph) and HUNT surveys (Axivity) are experienced in Norway. Differentially in apprehending accelerometer-assessed PA have hampered the possibility to compare PA results among regional and national surveys.

Estimations for overall activity differ in results and interpretation depending on the accelerometer signal. For instance, CPM and MX metrics levels are entirely different from each other in cut points, translational charts, and structure. MX metrics represent the acceleration of most active non-consecutive x minutes during a day, and, as mention above, cut-points can be post-hoc applied for public messaging (Migueles et al., 2021). Although, it is highly insecure whether selected MX metrics (e.g., $M_{1/3}$h, M120, M60, M30, M15, and M5) in this thesis are the preferred choice among researchers. Rowlands et al. (2019a) suggest a standard norm for the duration of time to secure comparability between studies. However, to this thesis knowledge, there is no such norm available up to 2021. Still, when the GRANADA consensus was published, Migueles et al. (2021) left a common consensus regarding this problem. Therefore, is it unclear if M6h, M4h, M45, M20, M10, and M2 should be implemented in this thesis.

Several studies have compared activity estimates from hip-mounted ActiGraph GT3X+ based on raw and count-based data (Buchan et al., 2019; Fairclough et al., 2019a). Buchan et al. (2019) indicate that cut points' choice significantly impacts activity estimates and comparability. For instance, the Hildebrand et al. cut points results resulted in considerably fewer participants achieving recommended 60 min.d^{-1} of MVPA (Buchan et al., 2019). Considering the use of a similar cut point in this thesis, results for the PA guidelines might imply an underestimation of reality. Even so, the Hildebrand cut point of 70 *mg* was considered suitable and used in accessing representative results for this thesis (Hildebrand et al., 2014). Differentially between MVPA-variables (70.7 ± 66.5 vs. 38.7 ± 25.6 min^{-1}) discovered in this thesis are similar to MVPA-comparability performed in previously published research: though only for children (Fairclough et al., 2019a).

Due to the lack of equivalence and standard agreement between raw and count-based metrics, Buchan et al. (2019) imply the comparability as unrealistic. Additional calibration, correction factors, and equations are proposed as necessary to facilitate the comparison of findings between raw and count-based data (Buchan et al., 2019). Finally, though, the authors from the GRANDA consensus exemplify a possibility: depending on activity metric used to predict health outcomes. The GRANDA consensus states that it is possible to examine PA relationships of volume and intensity with analytical metrics and their prediction of health: in this thesis case BMI (Migueles et al., 2021).

6. Conclusion

This thesis has provided normative values of new metrics for device-based indices of physical activity (average acceleration, intensity gradient, MX-variables) in a representative sample of Norwegian adults and older, showing the feasibility and usefulness of this approach. Furthermore, the new approach provides a way of presenting such data in a way that is interpretable to a broader audience. The more or less similar properties further strengthen the appropriateness of the new metrics as the traditional metric its predictive abilities. Given the new approach's ability to overcome significant limitations of device-based measurements of physical activity, it might provide a feasible way forward in order to aid comparability between studies of physical activity and health.

7. Future research

This thesis has successfully provided PA estimates for a selected Norwegian adult sample while analyzing and interpreting acceleration values with different approaches (i.e., new and old metrics). Estimates suggest that analytical and translational metrics are feasible to use and increases the chance of comparability between datasets. Researchers and public-health personnel should note the practical limitations of the new metrics, as addressed in the discussion, and investigate the possibility of further consensus within the field of device measurements. While the GRANADA report increases a researcher’s understanding and might assist in selecting analytical approaches, there is a need for further comparisons between the selected metrics and the use of other physical behaviors (i.e., sedentary and sleep behavior). Further research should investigate the possibility of standardization of deployment and reduction strategies for increased comparability between device-based measurements of PA.

Sources

E. Aadland, & E. Ylvisåker. (2015). Reliability of the Actigraph GT3X+ accelerometer in adults under free-living conditions. *PloS one, 10*(8), 1-10. doi:10.1371/journal.pone.0134606

C. Allison, S. Colby, A. Opoku-Acheampong, T. Kidd, K. Kattelmann, M. D. Olfert, & W. Zhou. (2020). Accuracy of self-reported BMI using objective measurement in high school students. *9*(35), 1-8. doi:10.1017/jns.2020.28

D. R. Basset, A. V. Rowlands, & S. G. Trost. (2012). Calibration and validation of wearable monitors. *Med. Sci. Sports Exerc., 44*, 32-38. doi:10.1249/MSS.0b013e3182399cf7

E. Bentlage, A. Ammar, D. How, M. Ahmed, K. Trabelsi, H. Chtourou, & M. Brach. (2020). Practical Recommendations for Maintaining Active Lifestyle during the COVID-19 Pandemic: A systematic Literature Review. *International Journal of Environmental Research and Public Health, 17*(6265). doi:10.3390/ijerph17176265

S. Berntsen, & S. A. Anderssen. (2018). Et historisk blikk på fysisk aktivitet. In M. K. Torstveit, H. Lohne-Seiler, S. A. Anderssen, & S. Berntsen (Eds.), *Fysisk aktivitet og helse: Fra begrepsforståelse til implementering av kunnskap* (pp. 21-36). Oslo: Cappelen Damm AS 2018.

C. Bouchard, A. Tremblay, C. Leblanc, G. Lortie, R. Savard, & G. Thériault. (1983). A method to assess energy expenditure in children and adults. *The American Journal of Clinical Nutrition, 37*(3), 461-467. doi:10.1093/ajcn/37.3.461

D. S. Buchan, & G. McLellan. (2019). Comparing physical activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods. *Journal of Sports Sciences, 37*(7), 779-787. doi:https://doi.org/10.1080/02640414.2018.1527198

D. S. Buchan, G. McLellan, S. Donnelly, & R. Arthur. (2019). The use of the intensity gradient and average acceleration metrics to explore associations with measures of health and wellbeing in children. *Journal of Sports Sciences, 37*(23), 2751-2758. doi:https://doi.org/10.1080/02640414.2019.1664536

C. J. Caspersen, K. E. Powell, & G. M. Christenson. (1985). Physical Activity, Exercise, and Physical Fitness: Definitions and Distinctions for Health-Related Research. *Public Health Reports, 100*(2), 126-131.

K. Y. Chen, & D. R. Bassett. (2005). The Technology of Accelerometry-Based Activity Monitors: Current and Future. *Medicine & Science in Sports & Exercise 37*(11), 490-500. doi:10.1249/01.mss.0000185571.49104.82

Departementene. (2020). *Sammen om aktive liv: Handlingsplan for fysisk aktivitet 2020–2029*. (Handlingsplan for fysisk aktivitet 2020-2029). Retrieved from https://www.regjeringen.no/no/dokumenter/sammen-om-aktive-liv/id2704955/

S. J. Fairclough, A. V. Rowlands, & S. T. L. M. Boddy. (2019a). Cut‐point‐free accelerometer metrics to assess children's physical activity: An example using the school day. *Scand J. Med. Sci. Sports. , 30*(2020), 117-125. doi:10.1111/sms.13565

S. J. Fairclough, A. V. Rowlands, S. Taylor, & L. M. Boddy. (2019b). Cut‐point‐free accelerometer metrics to assess children's physical activity: An example using the school day. *Scand J. Med. Sci. Sports. , 30*(2020), 117-125. doi:10.1111/sms.13565

V. Ferrahi, M. Niemlä, M. Kangas, R. Korpelainen, & T. Jämsä. (2019). Calibration and validation of accelerometer-based activity monitors: A systematic review of machine-learning approaches. *Gait & Posture, 68*(2019), 285-299. doi:https://doi.org/10.1016/j.gaitpost.2018.12.003
P. S. Freedson, E. Melanson, & J. Sirard. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Medicine & Science in Sports & Exercise, 30*(5), 777-781. doi:https://doi.org/10.1097/00005768-199805000-00021
M. L. V. H. Greenfield, J. E. Kuhn, & E. M. Wojtys. (1998). A Statistics Primer. Validity and reliability. *Am. J. Sports. Med. , 26*(3), 483-485. doi:https://doi.org/10.1177/03635465980260032401
M. Grydeland, B. H. Hansen, M. Ried-Larsen, E. Kolle, & S. A. Anderssen. (2014). Comparison of three generations of ActiGraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports, Medicine, and Rehabilitation, 6*(26). doi:http://biomedcentral.com/2052-1847/6/26
R. Guthold, G. A. Stevens, L. M. Riley, & F. C. Bull. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. *Lancet Glo. Health, 6*(2018), 1077-1086. doi:http://dx.doi.org/10.1016/S2214-109X(18)30357-7
B. H. Hansen, E. Kolle, J. Steene-Johannessen, K. E. Dalene, U. Ekelund, & S. A. Andersen. (2018a). Monitoring population levels of physical activity and sedentary time in Norway across the lifespan. *Scand J. Med. Sci. Sports., 29*(2019), 105-112. doi:10.1111/sms.13314
B. H. Hansen, J. Steene-Johannessen, & E. Kolle. (2018b). Status fysisk aktivitet og fysisk form i befolkningen. In M. K. Torstveit, H. Lohne-Seiler, S. A. Anderssen, & S. Berntsen (Eds.), *Fysisk aktivitet og helse: fra begrepsforståelse til implemtering av kunnskap* (pp. 92-110). Oslo: Cappelen Damm 2018.
T. J. Harris, C. G. Owen, C. R. Victor, R. Adams, U. Ekelund, & D. G. Cook. (2008). A Comparison of Questionnaire, Accelerometer, and Pedometer: Measures in Older People. *Medicine & Science in Sports & Exercise, 41*(7), 1392-1402. doi:10.1249/MSS.0b013e31819b3533
W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, & A. Bauman. (2007). Physical Activity and Public Health: Updated Recommendation for Adults From the American College of Sports Medicine and the American Heart Association. *Medicine & Science in Sports & Exercise, 39*(8), 1423-1434. doi:10.1249/mss.0b013e3180616b27
Helse-og-omsorgsdepartmentet. (2018-2019). *Meld. St. 19 (2018–2019) Folkehelsemeldinga — Gode liv i eit trygt samfunn*. ((Meld. St. 19 (2018-2019))). Retrieved from https://www.regjeringen.no/no/dokumenter/meld.-st.-19-20182019/id2639770/
Helsedirektoratet. (2015). *Fysisk aktivitet og sedat tid blant voksne og eldre i Norge - Nasjonal kartlegging 2014-2015*. (IS-2367). Retrieved from https://www.helsedirektoratet.no/rapporter/fysisk-aktivitet-kartleggingsrapporter
M. Hildebrand, V. v. Hees, B. H. Hansen, & U. Ekelund. (2014). Age group comparability of raw accelerometer output from wrist-and hip-worn monitors|. *Medicine & Science in Sports & Exercise, 46*(9), 1816-1824. doi:10.1249/MSS.0000000000000289

M. Kang, S. J. Marshall, T. V. Barreira, & J.-O. Lee. (2009). Effect of pedometer-based physical activity interventions: a meta-analysis. *Research quarterly for exercise and sport, 80*(3), 648-655.

G. P. Kenny, S. R. Notley, & D. Gagnon. (2017). Direct calorimetry: a brief historical review of its use in the study of human metabolism and thermoregulation. *Eur J Appl Physiol, 117*(9), 1765-1785. doi:10.1007/s00421-017-3670-5

E. Kolle, & M. Grydeland. (2018). Begrepsavklaringer. In M. K. Torstveit, H. Lohne-Seiler, S. A. Anderssen, & S. Berntsen (Eds.), *Fysisk aktivitet og helse - fra begrepsforståelse til implementering av kunnskap* (pp. 40-54). Oslo: Cappelen Damm 2018.

S. R. LaMunion, E. C. Fitzhugh, & S. E. Crouter. (2020). Challenges and opportunities related to the objective assessment of physical activity within U.S. health surveys. *Annals of Epidemiology, 43*(2020), 1-10. doi:https://doi.org/10.1016/j.annepidem.2020.01.011

K. Lyden, N. Petruski, S. Mix, J. Staudenmayer, & P. Freedson. (2014). Direct Observation is a Valid Criterion for Estimating Physical Activity and Sedentary Behavior. *Journal of Physical Activity and Health, 11*, 860-863. doi:http://dx.doi.org/10.1123/jpah.2012-0290

D. Martinez-Gomez, M. A. Puertollano, J. Wärnberg, M. Calabro, G. Welk, M. Sjöström, O. Veiga, & A. Marcos. (2009). Comparison of the ActiGraph acceloremeter and Bouchard diary to estimate energy expenditure in Spanish adolescents. *Nutricion hospitalaria, 24*(6), 701-710. doi:10.3305/nh.2009.24.6.4555

J. J. Mcclain, S. B. Sisson, & C. Tudor-Locke. (2007). Actigraph accelerometer interinstrument reliability during free-living in adults. *Medicine and science in sports and exercise, 39*(9), 1509-1514. doi:10.1249/mss.0b013e3180dc9954

T. L. McKenzie, D. A. Cohen, A. Sehgal, S. Williamson, & D. Golinelli. (2006). System for Observing Play and Recreation in Communities (SOPARC): reliability and feasibility measures. *Journal of Physical Activity and Health, 3*(s1), S208-S222.

J. H. Migueles, E. Aadland, L. B. Andersen, J. C. Brønd, S. F. Chastin, B. H. Hansen, K. Konstabel, O. M. Kvalheim, D. E. McGregor, A. V. Rowlands, S. Sabia, V. v. Hees, R. Walmsley, & F. B. Ortega. (2021). GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies. *Br. J. Sports. Med.*, 1-9. doi:10.1136/bjsports-2020-103604

J. H. Migueles, C. Cadenas-Sanchez, U. Ekelund, C. D. Nyström, J. Mora-Gonzalez, M. Löf, I. Labayen, J. R. Ruiz, & F. B. Ortega. (2017). Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. *Sports medicine, 47*(9), 1821-1845. doi:http://dx.doi.org/10.1007/s40279-017-0716-0

J. H. Migueles, C. Cadenas-Sanchez, A. V. Rowlands, P. Henriksson, E. J. Shiroma, F. M. Acosta, M. Rodriguez-Ayllon, I. Estaban-Cornejo, A. Plaza-Florido, J. J. Gil-Cosano, U. Ekelund, V. T. v. Hees, & F. B. Ortega. (2019a). Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. *Scientific Reports, 9*(18235). doi:https://doi.org/10.1038/s41598-019-54267-y

J. H. Migueles, A. V. Rowlands, F. Huber, S. Sabia, & V. T. v. Hees. (2019b). GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. *Journal* *for the Measurement of Physical Behaviour, 2*(2019), 188-196. doi:https://doi.org/10.1123/jmpb.2018-0063

N. C. Ministers. (2014). Nordic Nutrition Recommendations 2012: Integrating nutrition and physical activity. In *Nordic Council Ministers 2014* (pp. 195-216). Copenhagen: Nordic Council Ministers.

A. J. Nihiser, S. M. Lee, H. Wechsler, M. McKenna, E. Odom, C. Reinold, D. Thompson, & L. Grummer-Strawn. (2009). BMI Measurement in Schools. *Journal of School Healthe, 124*, 89-99. doi:https://doi.org/10.1542/peds.2008-3586L

K. J. Rothman. (2008). BMI-related errors in the measurement of obesity. *International Journal of Obesity, 32*, 56-59. doi:https://doi.org/10.1038/ijo.2008.87

A. V. Rowlands, N. P. Dawkins, B. Maylor, C. L. Edwardson, S. J. Fairclough, M. J. Davies, D. M. Harrington, K. Khunti, & T. Yates. (2019a). Enhancing the value of accelerometer-assessed physical activity: meaningful visual comparisons of data-driven translational accelerometer metrics. *Sports Medicine - Open, 5*(47), 1-11. doi:https://doi.org/10.1186/s40798-019-0225-9

A. V. Rowlands, C. L. Edwardson, M. J. Davies, K. Khunti, D. M. Harrington, & T. Yates. (2018). Beyond Cut Points: Accelerometer Metrics that Capture the Physical Activity Profile. *Medicine & Science in Sports & Exercise, 50*(6), 1323-1332. doi:10.1249/MSS.0000000000001561

A. V. Rowlands, E. M. Mirkes, T. E. Yates, S. A. Clemes, M. J. Davies, K. Khunti, & C. L. Edwardson. (2017). Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? *Med. Sci. Sports Exerc., 50*(2), 257-265. doi: 10.1249/MSS.0000000000001435

A. V. Rowlands, L. B. Sherar, S. J. Fairclough, T. Yates, C. L. Edwardson, D. M. Harrington, M. J. Davies, F. Munir, K. Khunti, & V. H. Stiles. (2019b). A data-driven, meaningful, easy to interpret, standardised accelerometer outcome variable for global surveillance. *Journal of Science and Medicine in Sport, 22*(2019), 1132-1138. doi:https://doi.org/10.1016/j.jsams.2019.06.016

A. V. Rowlands, T. Yates, M. Davies, K. Khunti, & C. L. Edwardson. (2016). Raw Accelerometer Data Analysis with GGIR R-package: Does accelerometer Brand Matter? *Medicine & Science in Sports & Exercise, 48*(10), 1935-1941. doi:10.1249/MSS.0000000000000978

J. Steene-Johannessen, M. Grydeland, & B. H. Hansen. (2018). Måling av fysisk aktivitet og fysisk form. In M. K. Torstveit, H. Lohne-Seiler, S. A. Anderssen, & S. Berntsen (Eds.), *Fysisk aktivitet og helse - fra begrepsforståelse til implementering av kunnskap* (pp. 62-86). Oslo: Cappelen Damm 2018.

J. Steene-Johannessen, B. H. Hansen, K. E. Dalene, E. Kolle, K. Northstone, N. C. Møller, A. Grøntved, N. Wedderkopp, S. Kriemler, A. S. Page, J. J. Puder, J. J. Reilly, L. B. Sardinha, E. M. F. v. Sluijs, L. B. Andersen, H. v. d. Ploeg, W. Ahrens, C. Flexeder, M. Standl, H. Shculz, L. A. Moreno, S. D. Henauw, N. Michels, G. Cardon, F. B. Ortega, J. Ruiz, S. Aznar, M. Fogelholm, A. Decelis, L. G. Olesen, M. F. Hjorth, R. Santos, S. Vale, L. B. Christiansen, R. Jago, L. Basterfield, C. G. Owen, C. M. Nightingale, G. Eiben, A. Polito, F. Lauria, J. Vanhelst, C. Hadjigeorgiou, K. Konstabel, D. Molnár, O. Sprengeler, Y. Manios, J. Harro, A. Kafatos, S. A. Anderssen, & U. Ekelund. (2020). Variations in accelerometry measured physical activity and sedentary time across Europe – harmonized analyses of 47,497 children and adolescents. *International Journal of Behavioral Nutrition and Physical Activity, 17*(38), 1-14. doi:https://doi.org/10.1186/s12966-020-00930-x

R. P. Troiano, D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert, & M. McDowell. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise, 40*(1), 181-188. doi:https://doi.org/10.1249/mss.0b013e31815a51b3

v. H. Vincent, Z. Fang, J. Langford, F. Assah, A. Mohammad, I. C. M. d. Silva, M. I. Trenell, T. White, N. J. Wareham, & S. Brage. (2014). Autocalibration of accelerometer data for free-living physical activity
assessment using local gravity and temperature: an evaluation on
four continents. *117*(7), 738-744. doi:https://doi.org/10.1152/japplphysiol.00421.2014

K. B. Watson, S. A. Carlson, D. D. Carroll, & J. E. Fulton. (2014). Comparison of accelerometer cut points to estimate physical activity in US adults. *Journal of Sports Sciences, 32*(7), 660-669. doi:http://dx.doi.org/10.1080/02640414.2013.847278

G. J. Welk. (2005). Principles of Design and Analyses for the Calibration of Accelerometry-Based Activity Monitors. *Medicine & Science in Sports & Exercise, 37*(11), 501-511. doi:10.1249/01.mss.0000185660.38335.de

WHO. (2004). The global burden of disease : 2004 update. Retrieved from https://www.who.int/publications/i/item/9789241563710

WHO. (2010). *Global Recommendations on Physical Activity for Health* [Brochure]. Switzerland: World Health Organization 2010.

WHO. (2018). *Global action plan on physical activity 2018–2030: more active people for a healthier world* [Brochure]. Switzerland: World Health Organization 2018.

WHO. (2020). Public consultation on the draft WHO Guidelines on physical activity and sedentary behaviour for children and adolescents, adults and older adults 2020. Retrieved from https://www.who.int/news-room/articles-detail/public-consultation-on-the-draft-who-guidelines-on-physical-activity-and-sedentary-behaviour-for-children-and-adolescents-adults-and-older-adults-2020

N. Yoshitake, M. Okuda, S. Sasaki, I. Kunitsugu, & T. Hobara. (2012). Validity of self-reported body mass index of Japanese children and adolescents. *Pediatrics International, 54*, 397-401. doi:10.1111/j.1442-200X.2011.03541.x

Attachments

```
# 1. Description ---------------------------------------------------------

# Title:      Radar plot generator
# Version:    1.0 (Using Rstudio version 1.3.959)
# Date:       13-09-2020
# Author:     Ben Maylor (bm259@leicester.ac.uk)
# Contributors: Alex Rowlands, Nathan Dawkins and Charlotte Edwardson
# URL:        https://github.com/Maylor8/RadarPlotGenerator

# Description
# A tool used for generating radar plots to assist with the interpretation of
accelerometry data.
# When using Radar plot generator, please cite Rowlands et al. (2019) - DOI:
10.1186/s40798-019-0225-9.
# Please contact Ben Maylor (bm259@leicester.ac.uk) for any enquiries.

# 2. Library setup ------------------------------------------------------

# Packages required to generate Radar plots
library(ggplot2) # Verified using V3.3.1
library(scales)  # Verified using V1.1.1


# 3. Data setup ---------------------------------------------------------

Group.no    <- 3                              # Number of groups to be compared
Group.names <- c("Healthy",         "Overweight",          "Obese")  # Names of
comparison groups
Metric.no   <- 5                              # Number of different MX metrics used
Metriclist  <- c("M60", "M30", "M10", "M5", "M2")   # Quote each MX metric used in
descending order. Report MX values in minutes to ensure they are ordered correctly.

# Make the data frame (Skip if you are importing an existing data frame)
Metric  <- c(rep(c(Metriclist), each = Group.no))
Group   <- c(rep(c(Group.names), times = Metric.no))


# Paste values where the first MX metric is listed for each Group in the order they
appear above (e.g. in the provided example this would be M60-Healthy, Overweight,
Obese, M30-Healthy etc..)
Mg    <- c(105,    101,    120.2,    # Data in Mg.
         148.7,  140.4,  127,
         196.4,  185,    168.2,
         240.1,  225.6,  197.7,
         329.3,  290.2,  234)

# Make data frame
Data <- data.frame(Metric, Group, Mg)
```

```
# Inspect data frame for accuracy
print(Data)

# Ordering of MX metrics for the plots
temp <- vector(mode="numeric", length=length(Metriclist))

for (i in 1:length(Metriclist)) {
   temp[i] <- as.numeric(substring(Metriclist[i],2,nchar(Metriclist[i])))}

temp <- sort(temp, decreasing=T)
temp <- paste("M", temp, sep="")
Data$Metric <- factor(Data$Metric, levels = temp)
Data$Group <- factor(Data$Group, levels = Group.names)

# Optional alternative method of specifying the order of Mx metrics - deactivate lines
48-56 first.
# Data$Metric <- factor(Data$Metric,levels = c("M60", "M30", "M10", "M5", "M2))

# 4. Generate plot -------------------------------------------------------

# The code first generates a polygon plot in which the data and main aesthetics are
specified by the user  ##
# WARNING: When adding/removing lines of code here, ensure that symbols at the
end of each line remain the ##
# same (e.g. + or ,). This is typically the most common cause for the code tripping up.
##
# The plot viewer in R will not reflect the true rendering quality applied to the
exported image file      ##

# Set Y axis parameters suitable for your data
Ystart <- 0 # Y axis start value
Yend <- 400 # Y axis end value
Yint <- 50  # Y axis interval

# Make plot
Plot <- ggplot(data=Data, aes(x = Metric, y = Mg, group = Group)) +
   geom_polygon(aes(group = Group, colour = Group), fill = NA, size = 1.1) +

   # OPTIONAL: Specify group colours/shade - Activate and edit one only at a time
   # A comprehensive list of colours can be found here:
http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf
   # Example for 5 groups
   # scale_colour_manual(values = c("grey10", "grey30", "grey50", "grey70",
"grey90"),name = "Group") +
   # scale_colour_manual(values = c("black", "green", "blue","red","yellow"),name =
"Group") +

   # Set Y axis gridlines
```

```
   geom_hline(yintercept = seq(Ystart,Yend, by = Yint), alpha = 0.8, colour = "grey70", size = 0.2) + # Automatically uses the Y axis interval value specified previously

   # Add vertical X lines for metrics
   geom_segment(aes(x="M60", y=Ystart, xend="M60", yend=Yend), colour = "grey70", size = 0.2) +
   geom_segment(aes(x="M30", y=Ystart, xend="M30", yend=Yend), colour = "grey70", size = 0.2) +
   geom_segment(aes(x="M10", y=Ystart, xend="M10", yend=Yend), colour = "grey70", size = 0.2) +
   geom_segment(aes(x="M5", y=Ystart, xend="M5", yend=Yend), colour = "grey70", size = 0.2) +
   geom_segment(aes(x="M2", y=Ystart, xend="M2", yend=Yend), colour = "grey70", size = 0.2) +

   # Add Acceleration cut-points
   geom_hline(yintercept = 100, colour = alpha("red", 0.7), size = 0.9, linetype = 5) + # add as many of these cutpoint as you want by pasting this line.
   geom_hline(yintercept = 200, colour = alpha("red", 0.7), size = 0.9, linetype = 4) + # For different line types: http://www.cookbook-r.com/Graphs/Shapes_and_line_types/
   geom_hline(yintercept = 400, colour = alpha("red", 0.7), size = 0.9, linetype = 6)   # Ensure The last line does not end in "+" or the plot will not be finalised.

# OPTIONAL: Quality check when trying to identify aesthetic and layout problems.
# Plot

# Make radar plot (wrap ggplot around a coordinate)
Plot + coord_polar(start = -((180/Metric.no)*(pi/180))) + # Start argument ensures that the first MX metric begins directly vertical of the centre of the radar plot.
   theme_classic(base_size = 16) +
   theme(axis.ticks = element_blank(), # Theme edits. For examples of additional theme edits, see here: https://ggplot2.tidyverse.org/reference/theme.html
        axis.title = element_blank(),
        axis.line = element_blank(),
        #axis.text.x = element_blank(), Remove MX labels if you are adding MX symbols instead. Recommended that you do not activate this for the first run to ensure metrics are in the right order and the plot makes sense.
        axis.text.y = element_blank(),
        panel.grid = element_blank()) + # Remove various elements from plot

   scale_y_continuous(limits = c(Ystart, Yend),breaks = seq(Ystart,Yend,Yint)) + #set y axis limits

   # Add Y axis labels. currently takes a bit of manual adjustment. One of the main aims for future releases is to make this less cumbersome.
   geom_text(x = 0.75, y = (Yend/2), label = expression(paste("MX (m",italic("g"),")")), angle = 93, colour = "grey20", hjust = "left", size = 5) + # Y Axis title
```

```
   geom_text(x = 1, y = 100, label = "100", angle = 0, colour = "grey20", hjust = "right")
+
   geom_text(x = 1, y = 200, label = "200", angle = 0, colour = "grey20", hjust = "right")
+
   geom_text(x = 1, y = 300, label = "300", angle = 0, colour = "grey20", hjust = "right")
+
   geom_text(x = 1, y = 400, label = "400", angle = 0, colour = "grey20", hjust = "right")
#geom_text(x = 1, y = 500, label = "500", angle = 0, colour = "grey20", hjust = "right") +
#geom_text(x = 1, y = 600, label = "600", angle = 0, colour = "grey20", hjust = "right") +
#geom_text(x = 1, y = 700, label = "700", angle = 0, colour = "grey20", hjust = "right")


# 5. Export Radar plot ----------------------------------------------------

# this step is deactivated initially to save time until you are happy with the plot.

# Export plot at high rendering quality to you current working directory
# ggsave("PlotTest.jpg", width = 25, height = 20, units = "cm", dpi = 1500) # Input fill
format here

# where is the plot?
getwd()

### End ###
```