UiT The Arctic University of Norway

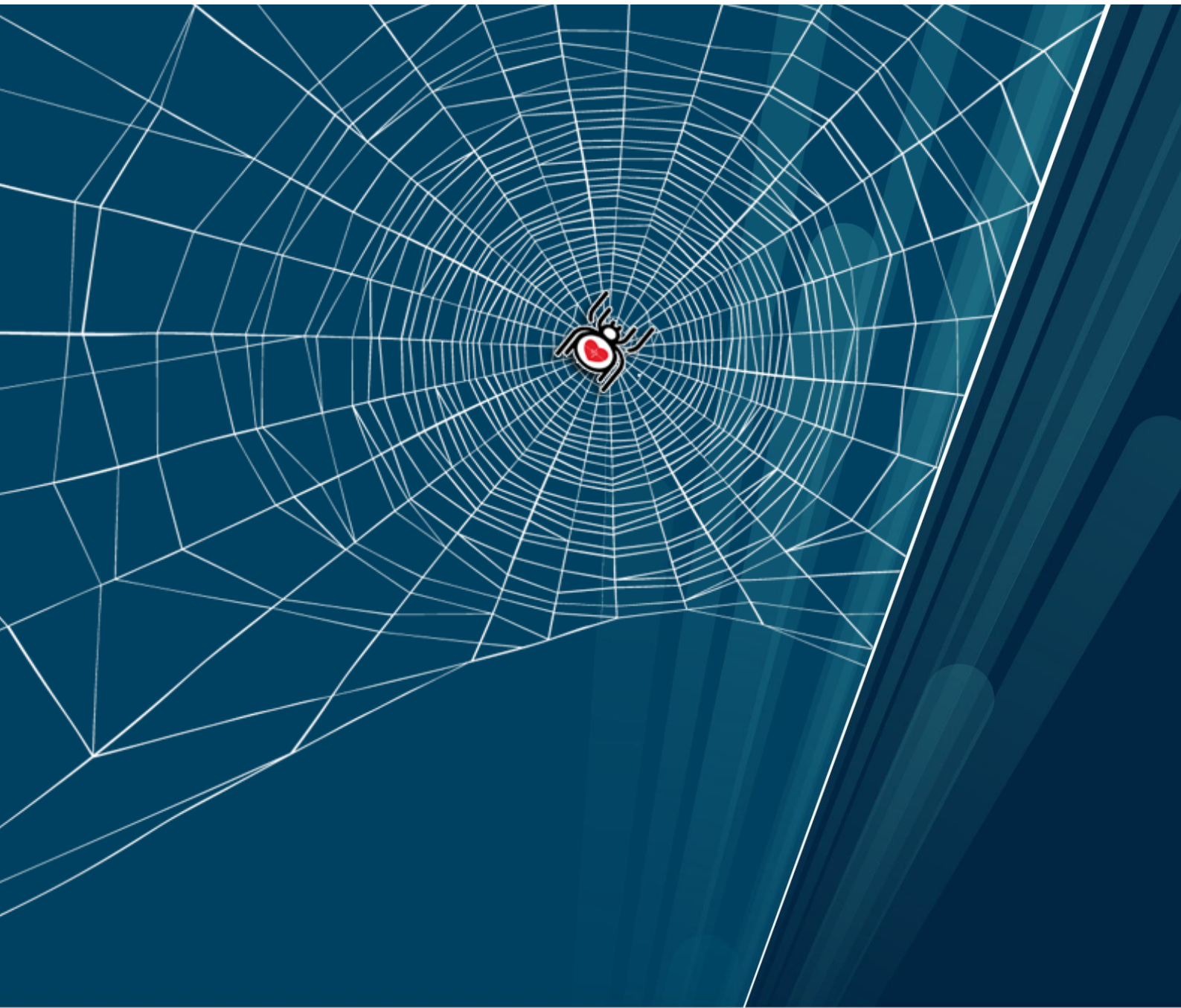
Faculty of Health Sciences
Department of Community Medicine

Continuous physical activity recording

Consumer-based activity trackers in epidemiological studies

André Henriksen

A dissertation for the degree of Philosophiae Doctor – February 2021



Preface

Insights from epidemiological studies, including population-based studies and clinical studies, contribute to new knowledge on health determinants and provide a basis for development of public health recommendations. Data collection in epidemiological studies require extensive organization and resources to plan and execute, where data on physical activity are traditionally collected using physical activity questionnaires and accelerometers, each with their own strengths and limitations.

However, study participants may already wear various mobile sensors measuring health related metrics for private use, such as physical activity trackers. Data can be recorded over several months and years, but are more unorganized and unplanned, and accuracy is often unknown. This type of data may non-the-less be an important addition to traditional methods for collection of physical activity data for use in health research.

The observed decrease in participation in population-based studies over time is a threat to the need for representative samples. In contrast, the prevalence of activity tracker ownership is steadily increasing. To monitor physical activity levels in a population over time, activity trackers, designed for long-term monitoring, stands out as an interesting additional source of physical activity data.

Therefore, this dissertation is part of an initiative to investigate the potential for using physical activity data recorded by consumer-based activity trackers as part of future epidemiological studies. Specifically, the goal was to create a new method for collecting long-term data on physical activity, to be used in the next survey of the Tromsø Study, with the goal of closing the gap between short-term objective recordings (i.e. accelerometers) and long-term subjective estimates (i.e. physical activity questionnaires) .

Acknowledgements

The work resulting in this dissertation was conducted at the Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, between October 2016 and February 2021. The PhD project was funded by UiT The Arctic University of Norway's thematic priority grant *"personalized medicine for public health"*.

I have been very lucky with my supervisors, and I cannot imagine a better team for guiding me the last four years. I would like to thank my supervisor Laila Hopstock for sharing her excellent insight into the physical activity domain and related areas, as well as for her amazing availability. You were always available for discussions, always quick to respond with reflective and thorough comments, and you were always positive and inclusive. It has been a pleasure.

Doing a PhD in health science with a background in computer science was not without its challenges. I would especially like to thank my co-supervisor Sameline Grimsgaard for her timely explanations of unknown terminology (I suspect she could detect when I made a mental note of things to google later). I am grateful for the opportunity to work with someone so experienced, who was able to see where I could contribute and make a difference. You introduced me to RESTART, giving me the opportunity to work directly with study participants and gaining personal experience with clinical trials. Also, sorry for spending all your grant money on fitness trackers.

And finally, Gunnar Hartvigsen. What an overall great person and fantastic co-supervisor. In addition to your extensive insight and experience in the health technology domain, which you happily share with everyone, you really know how to motivate your students. You have an excellent philosophy of introducing your students to your collaborators and including them in your projects. Also, and this is not trivial, we speak the same language. Your extensive experience working in the intersection between computer science and health science has been really valuable.

To all my other co-authors, Martin, Mirek, Alexander, Trygve, Anne-Sofie, Ashenafi, and Erlend. Thank you for your contributions, input, and discussions.

To my three most favourite people. To my wife Ingunn, who has been very supportive, especially the last few months when I spent too many evenings at the office, thank you. To

Håvard and Martin, who could not understand why on earth I would want to go to work on a Saturday or Sunday, I'm back now. I love you all.

Thank you to all my friends for the distractions, and thank you to friends, family, colleagues, and everyone else who contributed to my various data collections.

To all my fellow PhD-students, thank you for adding joy and laughter to my workday, as well as the great discussions about everything from whisky to statistics. Ashenafi, thank you for motivating me and being a true friend. You are exceptional and I really look forward to our continued collaboration. Marie, thank you for being the best person to share an office with. Thank you for all the laughs, your willingness to participants in all my crazy whims (e.g. going full Halloween on national television), and not being (visible) annoyed by the constant grammar policing and the constant talking. I am sure it was challenging, but at least I could help you with a *few* computer problems.

Thank you to all colleagues who have made this journey enjoyable. Especially Egil Thomassen and Jan Fuglesteg, who are the two most helpful people I know at the University. Everyone at Teknisk Kompetansegruppe, especially Jon Ivar Kristiansen who (more than once) fixed the mSpider server installations after I ran some less than clever Linux commands. Jonas with whom I feel I have connected with beyond being colleagues and Trygve who I really have enjoined working with, and who I really hope I can continue to collaborate with going forward.

André J. Henriksen, Tromsø, February 2021

Permission to use this comic strip was obtain from "Piled Higher and Deeper" by Jorge Cham, www.phdcomics.com.

Table of contents

List of Tables

List of Figures

List of Papers

This PhD dissertation is based on the four papers listed below. These are referred to in the text as Paper I, Paper II, Paper II, and Paper IV, respectively.

Paper I

Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S. **Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables**. J Med Internet Res, 2018. 20(3): p. e110. doi: 10.2196/jmir.9157.

Paper II

Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L. **Validity of the Polar M430 Activity Monitor in Free-Living Conditions: Validation Study.** JMIR Form Res, 2019. 3(3): p. e14438. doi: 10.2196/14438.

Paper III

Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L. **Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach.** BMC Public Health, 2020. 20(1): p. 1300. doi: 10.1186/s12889-020-09406-w.

Paper IV

Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock L. **Physical activity surveillance during the COVID-19 pandemic: Using consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies.** [Submitted]. doi: 10.2196/preprints.23806.

Summary

New knowledge on health and diseases from epidemiological studies, including population-based studies and clinical studies, is important for the development of more effective public health prevention strategies and interventions. The Tromsø Study is an ongoing longitudinal population-based study with seven surveys conducted to date, each survey increasing in organizational setup and complexity related to the comprehensiveness of clinical examinations, questionnaires, and biological sampling. The growing data collection yields a unique basis for research, but also generate increased cost and increased participant burden.

In contrast, many participants already wear activity trackers for self-monitoring of various health metrics. Activity trackers are designed for long-term recoding, but data collection is often more unorganized and unplanned, and accuracy is often unknown. These data may non-the-less be an important addition to data generally collected in these large population-based studies.

In the seventh survey of the Tromsø Study (Tromsø 7), data on physical activity were collected using questionnaires and accelerometers. As in previous surveys, Tromsø 7 formed a basis for several clinical follow-up studies, including intervention studies. In the upcoming survey, Tromsø 8, it is of interest to expand objective physical activity recordings using participant's privately-owned wearable mobile sensors (i.e. activity trackers), to collected long-term data on physical activity both prospectively and retrospectively.

Having access to long-term activity tracker data may provide valuable insight into how physical activity changes in a population over time. In order to assess the feasibility of collecting this type of data, there is a need to understand if and how consumer-based activity trackers can be used for this purpose and create and test the usability of a solution that can collect this large-scale source of data in a simple and manageable manner. The main aim of this dissertation was therefore to develop and explore new methods to collect data on physical activity from participants in future epidemiological studies, using smart mobile sensors (i.e. activity trackers).

In the first paper we gave an overview of the current state of wrist-worn activity trackers on the consumer-market and provided suggestions on what to consider when deciding which provider and activity tracker model to use in future research.

In the second paper we assessed the validity of an activity tracker, the Polar M430, and concluded that although it cannot be used as a replacement for current methods of physical activity data collection, it has the potential to be used as an additional source for long-term physical activity monitoring.

In the third paper we identified important factors for increasing wear time adherence and provided a list of recommendations to consider when using consumer-based activity trackers for long-term physical activity monitoring in health research. Major factors include providing satisfactory activity tracker training to participants, offer a variety of activity tracker designs, and use activity trackers with accurate measurements.

In the fourth paper we implemented a system for automatic and continuous physical activity monitoring, collected from consumer-based activity trackers. Further, to test the usability of this system, we assessed how physical activity levels changed during the COVID-19 pandemic, by retrospectively accessing activity tracker data already collected by participants who wore a tracker before-, during-, and after the Norwegian March 2020 lockdown period.

In conclusion, this dissertation provide insight into what to consider when using consumer-based activity tracker for long-term physical activity monitoring in health research and demonstrate how this type of data can be accessed and used for both retrospective and prospective study designs.

Abbreviations

AEE	Activity energy expenditure
API	Application programming interface
CPM	Counts per minute
CI	Confidence interval
CSV	Comma-separated value
DIT	Dietary induced thermogenesis
DLW	Doubly labelled water
GPS	Global positioning system
ICC	Intra-class correlation
LED	Light-emitting diode
LPA	Light physical activity
MAPE	Mean absolute percentage error
MET	Metabolic equivalent of tasks
MPA	Moderate physical activity
MVPA	Moderate-to-vigorous physical activity
PAQ	Physical activity questionnaire
PPG	Photoplethysmography
QCAT	Quality control and analysis tool
REE	Resting energy expenditure
SD	Standard deviation
SDK	Software development kit
TEE	Total energy expenditure
VM	Vector magnitude
VPA	Vigorous physical activity
WHO	World Health Organization

1 Introduction

The main topic of this PhD dissertation was to investigate new methods for measuring physical activity in epidemiological studies, using smart technology worn by participants over an extended period. The introduction gives an overview of the definition of physical activity, the current epidemiology of physical activity, current methods for measuring physical activity, and how wearable smart technology can be used to record data on physical activity over time.

1.1 Defining physical activity

Physical activity

An often referenced definition of physical activity was coined by Caspersen et al. [1] in 1985, who defined it as "*any bodily movement produced by skeletal muscles that results in energy expenditure*". This definition was the result of a need to make it easier to distinguish between physical activity, exercise, and physical fitness, terms that were party used interchangeably, resulting in difficulties when comparing studies. A almost identical definition is also currently used by the World Health Organization (WHO) [2] (i.e. *"any bodily movement produced by skeletal muscles that requires energy expenditure"*).

Further, physical activity can be considered one element in a larger framework of human movement, as suggested by Pettee Gabriel et al. in 2012 [3]. This framework classifies different aspects of human movement, where physical activity (e.g. exercising) and sedentary behaviour (e.g. sitting) are sub-elements of human movement behaviour, and energy expenditure and physical fitness (e.g. body composition, muscular strength) are sub-elements of human movement attributes.

Quantifying physical activity

The total volume of performed physical activity is a function of intensity (e.g. light, moderate, and vigorous), frequency, duration, and activity type [4, 5]. The result can be quantified using different units of measurement, where metabolic equivalents of tasks (MET), minutes in sedentary behaviour, minutes in different levels of physical activity intensity, various types of energy expenditure, and steps are common metrics.

Metabolic equivalent of tasks is the rate at which a person expends energy while performing a specific task or activity. One MET is roughly the amount of energy expended at rest, where 1 MET=1 kilocalorie/kilogram/hour (kcal/kg/h). A MET minute is the amount of energy expended during a minute at rest (i.e. 1 MET). The MET for a given minutes can be used to classify the intensity of that minute.

Sedentary behaviour is defined as inactive behaviour while awake, where energy expenditure is below 1.5 METs while sitting or lying down. Light physical activity (LPA), e.g. slow walk or standing, is defined as having a MET of 1.5-3. Moderate physical activity (MPA), e.g. brisk walk, heavy cleaning (e.g. vacuuming), or light effort bicycling, is defined as having a MET of 3-6. Vigorous physical activity (VPA), e.g. jogging, fast bicycling, or playing soccer, is defined as having a MET >6 [6].

Metabolic equivalent of tasks and minutes of activity intensity can also be converted and reported as activity energy expenditure, where total energy expenditure (TEE) is a function of physical activity energy expenditure (AEE), resting energy expenditure (REE), and thermic effect of food (a.k.a. dietary induced thermogenesis (DIT)), so that TEE = REE + AEE + DIT. Energy expenditure is given in kcal or kilojoules (kJ), where 1 kcal = 4,187 kJ.

In addition, a well-known (but disputed) goal for being active is taking 10,000 steps per day. This cut-off dates back to the 1960s where a research team lead by Dr Yoshiro Hatono calculated that the average person took between 3500 and 5000 steps per day and increasing this to 10,000 would improve their health [7]. Tudor-Locke et al. [8] have created a classification for step counts and defined 10-12,000 steps per day as being “active”.

1.2 Physical activity epidemiology

Physical inactivity is a leading risk factor for a range of non-communicable diseases (including cardiovascular diseases, diabetes, and some cancers) and death [2].

The 2020 WHO’s Guidelines on physical activity and sedentary behaviour recommends adults to engage in at least 150-300 minutes of moderate physical activity or 75-150 minutes of vigorous physical activity per week, or an equivalent combination of intensities. In addition, moderate or greater intensity resistance training using major muscle groups at least two days per week is recommended for additional health benefits. For further increased health benefits, performing more than 300 minutes of moderate- or 150 minutes of vigorous aerobic

physical activity is recommended, as well as limiting the amount of sedentary time and replace it with physical activity [9]. All levels of intensities are associated with a reduction in risk of death and some physical activity is better than no physical activity [9, 10].

Globally, 23% of adult men and 32% of adult women did not fulfil recommendations in 2016 [11, 12, 13]. Physical inactivity is an increasing challenge worldwide, especially in high-income countries [11, 12].

Most countries in the European Union have adopted, or are in the process of adopting, physical activity recommendations based on WHO's global recommendations [14, 15]. In addition to individual health benefits of achieving the recommended physical activity levels, increased physical activity also provides health and economic benefits at the population level [16], as achieving the recommended levels of physical activity can reduced both cardiovascular disease mortality and total mortality [17].

1.3 Assessing physical activity

Overview

Recording data on physical activity in large longitudinal epidemiological studies provides important insight in how physical activity behaviour changes over time in a population. Physical activity is an important variable in a range of research questions but collecting accurate data on physical activity is challenging. Physical activity, physical inactivity, sedentary behaviour, and physical activity energy expenditure can be estimated using different methods. Each method provides different metrics, at different levels of accuracy and costs, and generate different participation burden. The choice of method is therefore dependent on the study setting, the research question, and available resources.

In health research, physical activity questionnaires have been the traditional choice due to their low cost and low complexity [18]. Technological advancement has increased the use of objective measurements, commonly collected using accelerometers, heart rate sensors, and combined sensing monitors. Figure 1 (based on figure by Hildebrand and Ekelund [19]) gives an overview of common methods for physical activity- and energy expenditure assessment, which output estimates are commonly available, and how each method compares in terms of accuracy and cost, and ease of use. The figure is a simplified overview of the most common methods and not an exhaustive list.

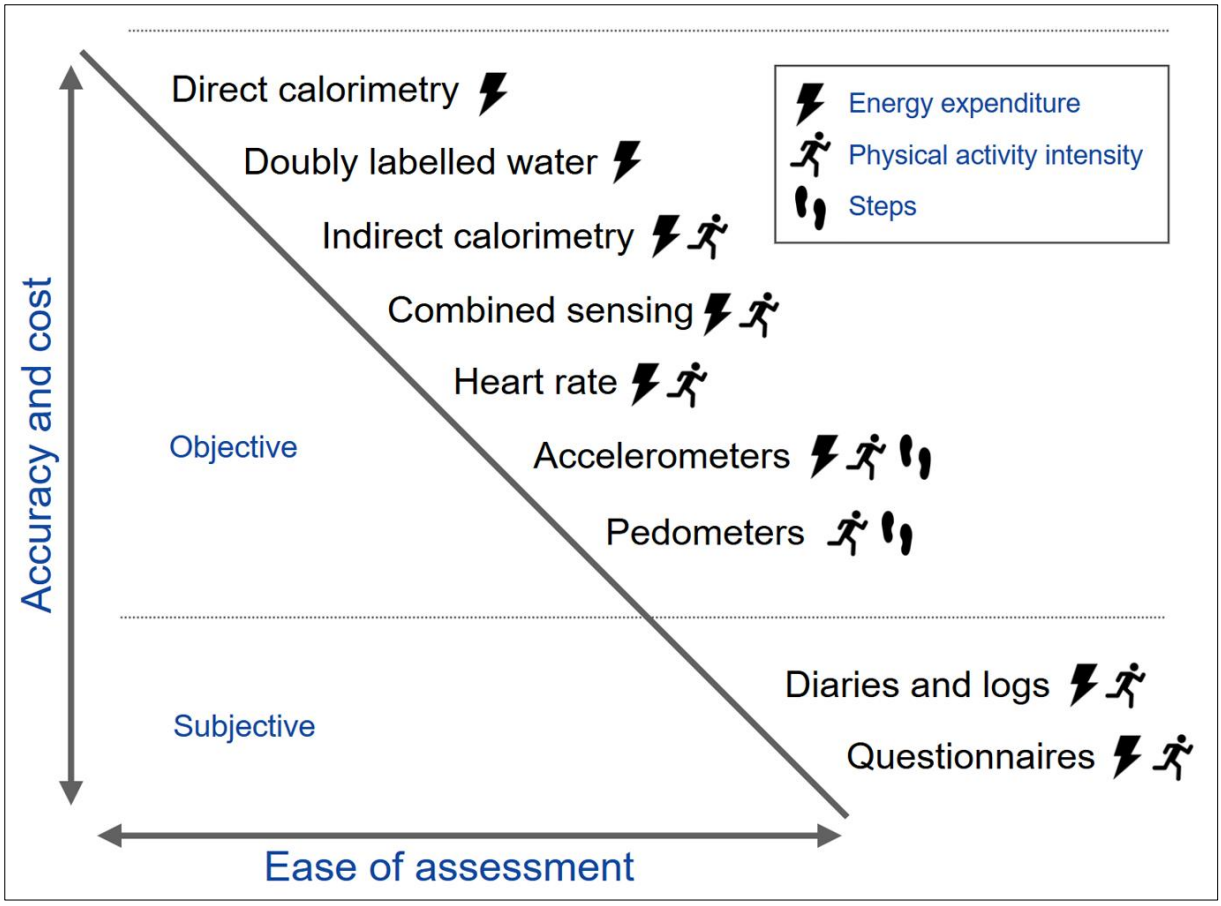


Figure 1. Methods for assessing physical activity.

Objective methods

There is a range of available methods for objective physical activity estimation. These are considered more accurate compared to subjective methods.

Direct calorimetry (a.k.a. room calorimetry) is the most accurate method for estimating energy expenditure in lab-settings. Energy expenditure is calculated by measuring the rate of heat loss from participants who are confined to a closed chamber during the measurement period. The method is both resource-demanding and complex with high participant burden [20]. Direct calorimetry gives no information about physical activity type or intensity.

The *Doubly labelled water (DLW)* method is frequently considered the gold standard for energy expenditure estimation in free-living but provides no information about physical activity type or intensity. Energy expenditure is estimated by orally introducing a known quantity of isotope-labelled water, using a heavier and stable isotope of hydrogen (^{2}H, deuterium) and oxygen (^{18}O), i.e. $^{2}H_2{}^{18}O$. The elimination rate of these two isotopes is measured in the urine over time. The rate of elimination is proportional to carbon dioxide

(CO_2) production, which is used to calculate energy expenditure. This method is costly and requires expertise, but it is safe, non-invasive, and accurate [20, 21, 22].

Indirect calorimetry is a method for estimating energy expenditure using oxygen consumption (VO_2) and carbon dioxide (CO_2) production. Gas exchange is collected through a face mask, a canopy system, or other closed-circuit solution. This method is accurate, often considered a gold standard for energy expenditure estimation in lab-settings, and is frequency used to assess energy expenditure in research- and clinical settings, as well as for validating other instruments (e.g. accelerometers). Indirect calorimetry is less resource-demanding compared to the DLW method and considered non-invasive for short time use, but it cannot be used for long-term measurements [20]. Indirect calorimetry gives no information about physical activity type but can be used to estimate physical activity intensity.

Heart rate monitoring is used to estimate energy expenditure from the linear relationship between heart rate and oxygen consumption [22]. Since heart rate is not only affected by the activity level, physical activity estimates can be improved by combining heart rate monitors and accelerometers into sensors using *combined sensing* [23].

Accelerometers estimates activity by measuring acceleration (i.e. change in velocity over time) several times per second (typically 30-100), most often in one or three axes [20, 22, 24]. Raw accelerometer data are fed to internal algorithm that outputs activity counts per epochs (e.g. counts per 10 seconds). These activity counts can be used to classify minutes of sedentary time and minutes of activity in different intensity zones, using a wide array of defined cut-points [25]. Which cut points to use is dependent on the sample characteristics, how many axes the activity counts are based on, and where the accelerometer was placed (e.g. hip, wrist). A common cut-point set for adults wearing an accelerometer on the hip, based on triaxial counts per minute (CPM) is defined by Sasaki et al. [26], where 2690–6166 CPM is classified as moderate intensity; 6167–9642 CPM is classified as vigorous intensity; and >9642 CPM is classified as very vigorous intensity. It follows that <2690 CPM is classified as light intensity unless other cut-points are used to distinguish between light intensity and sedentary time. Activity counts are also used to estimate AEE. Some accelerometers also include other sensors, e.g. gyroscope and magnetometer, for enhanced accuracy and additional outputs (e.g. body position and rotation). Accelerometers have existed since the 1980, but it was not until the late 1990 and early 2000 that it saw an increase in popularity for physical activity data collection in research [27]. Although there is no explicit gold-standard

for collecting physical activity intensity in free-living, accelerometers are generally considered the most accurate objective method in this setting [28]. However, because objective and subjective methods measures different aspects of physical activity and have different benefits and drawbacks, combining accelerometer recording with questionnaires have been suggested to achieve a more comprehensive and complementary overview [29].

Pedometers records steps taken while walking and running. Albert Stunkard was one of the first to describe how to use mechanic pedometers in research to estimate distance [30]. Using pedometers gained further popularity in the 1990s [31] as an inexpensive and objective tool for physical activity estimation in research [20, 22]. Some pedometers also report other metrics, e.g. distance travelled or energy expenditure. However, these metrics are not very accurate [32, 33]. Pedometers are the least accurate of the objective methods. Results are affected by stride length and walking speed, and several activity types are not recorded at all. Mechanical pedometers have mostly been replaced with more accurate accelerometer-based tools [34].

Subjective methods

Subjective methods, although considered less accurate, complements objective methods. They are less costly, can be used in scenarios where objective methods cannot be used, and can be used together with objective methods to give a more comprehensive understanding of participant's physical activity levels [29].

Keeping a physical activity *diary or log* can provide a very detailed overview of activity types and activity patterns. Activity type can be converted to energy cost, for instance by using the physical activity compendium which contains a large number of activity types and how these equivalates to METs [35, 36, 37]. A major challenge is the high participation burden [22].

Physical activity questionnaires (PAQ) are often used in large scale population-based studies because of the low cost and easy administration. There is a large number of different PAQs available, and which to use is dependent on participant characteristics, the preferred recall period, type of physical activity, and which dimension of physical activity is of interest [38, 39]. In a systematic review from 2012 Helmerhorst et al. [39] 96 different PAQs were presented, where most showed acceptable reliability but only moderate validity. A major limitation with PAQs is measurement error caused by information bias [39].

Summary

Assessment of physical activity is challenging, and commonly used traditional self-report and objective methods have their limitations. Objective long-term recording for monitoring of individual habitual physical activity levels is of interest. Using consumer-based activity trackers and related smart technology therefore comes forward as an interesting addition to existing methods, as it may close the gap between these two common methods. An increasing number of the population wear various activity trackers for extended periods. This create the possibility to access objective data on physical activity, retrospectively (historically) as well as in real-time.

1.4 Physical activity in the Tromsø Study: Example study

The work in this dissertation originated from the Tromsø Study context. However, findings are relevant for similar studies planning to collect long-term data on physical activity. I will use the Tromsø Study as an example study in this dissertation.

The Tromsø Study, together with the Trøndelag Health Study (HUNT) [40], are the two ongoing Norwegian population-based studies that aims to recruit representative population samples. The findings from the Tromsø Study and HUNT complement each other by recruiting both in rural and urban areas.

The Tromsø Study

The Tromsø Study was initiated in 1974 in order to better understand and prevent the high mortality of cardiovascular disease in Norway [41]. In total, seven surveys (Tromsø 1-7 1974-2016) have been conducted every five to seventh year, resulting in more than 40 years of rich longitudinal data. Each survey introduced more comprehensive questionnaires, clinical examinations, and biological samplings. More than 45,000 participants have attended one or more surveys, 18,500 have attended three or more, and 2000 have attended six or seven times [42]. The Tromsø Study is the longest running, most comprehensive (by extensiveness of clinical measurements), and best visited (by participation) population-based study in Norway. There are currently more than 230 active projects in the Tromsø Study [42].

Self-reported physical activity in the Tromsø Study

All seven surveys of the Tromsø Study used PAQs to collect self-reported levels of physical activity among participants. Both occupational physical activity and leisure time physical activity were collected using different PAQs: the Saltin-Grimby Physical Activity Level Scale

[43, 44], a leisure-time PAQ [45], and an exercise frequency, duration, and intensity PAQ [46]. A thorough summary of these PAQs used is available in Morseth et al. [47].

A recent study using questionnaire data from 40,000 Tromsø Study participants, attending between one and six repeated surveys between 1979 and 2017, showed that although there is an increase in leisure time physical activity, mainly due to increased exercise intensity and frequency, there is also an increase in occupational sedentary behaviour [47].

Objective measurements of physical activity in the Tromsø Study

In the first five surveys (Tromsø 1-5), physical activity and sedentary behaviour were collected using physical activity questionnaires only. In Tromsø 6, a subsample of 300 participants wore an ActiGraph GT1M (uniaxial) accelerometer for seven consecutive days. Results from comparing ActiGraph data with self-reported physical activity showed a discrepancy between methods [48]. In Tromsø 7, 6300 participants wore an ActiGraph wGT3X-BT (triaxial) for eight consecutive days, and 700 wore an ActiWave Cardio (accelerometer and one-lead electrocardiography) for 27 hours. Accelerometer data from Tromsø 7 have been used to determine activity levels in adults and elderly [49], to test the validity of three physical activity questionnaires [50], to test existing accelerometer non-wear time algorithms [51], and to suggest an improved algorithm for non-wear time [52]. In addition, these data have been used to assess how physical activity relates to pain sensitivity [53], as well as heart function and heart structure [54].

Current challenges and future plans

Participation in population-based studies have declined worldwide [55, 56], as well as in the Tromsø Study. Although recruitment has been comparably high in the Tromsø Study, ranging from 78.5% in Tromsø 5 to 65% in Tromsø 7 [57], greater effort was needed in recent surveys to achieve these participation proportions. Similar tendencies are present in other European health studies [58]. Attendance was lower among younger and older age groups in the Tromsø Study [57]. The increasing complexity in epidemiological research, and thus time required for participation, adds to the participant burden, which further may reduce willingness to participate [56].

Further, planning and executing large population-based studies is resource-demanding, and there is therefore a need for new strategies and non-intrusive tools for data collection in order to increase participation, collect representative samples, and reduce participant burden.

The number of sold consumer-based activity trackers and smartwatches is increasing worldwide [59, 60], and one in five Americans adults owned a fitness tracker or smartwatch in 2019 [61, 62]. This source of privately recorded objective physical activity data may be an interesting supplement to current methods of collecting physical activity. Current consumer-based activity trackers can collect a range of different physical activity variables, depending on model, and the nature of these devices allows long-term recording both prospectively and retrospectively [63, 64]. Consumer-based activity trackers are increasing being used in research to record physical activity and other health related metrics [63].

There are challenges with using these types of instruments for long-term recording in epidemiological research. In addition to privacy and ethical considerations, Wright et al. [63] especially address the problems with establishing activity tracker validity, and that activity tracker internal algorithms are unknown to the researchers. For long-term recording, using participant's own activity trackers, they also highlight challenges with selection bias (i.e. the characteristics of an owner of an activity tracker). The benefit from using this method is the potential long-term recording and large number of variables available [63]. Furthermore, consumer-based activity trackers are designed for long-term usage and are therefore less invasive and more user friendly than traditional accelerometers build for research-purposes.

The potential of health data collected from consumer-based mobile sensors such as activity trackers stretches beyond research. These data can be useful during health consultations, as they can provide a better overview of the patient's conditions and life-style choices [65, 66]. However, a review from 2017 concluded that there are no standard system allowing patient-collected health data to integrate with electronic health records and medical systems [67]. Although there are challenges on how to access and review this data in a simple way, clinicians may argue it is their responsibility to understand and use this data for patient consultations [68].

1.5 Consumer-based activity trackers

1.5.1 Introduction

A consumer-based activity tracker is a non-research device or application that records and analyses data on physical activity and other health related metrics. The terms *activity tracker, fitness tracker, activity monitor,* and *fitness monitor,* are often used interchangeably and is here understood as any wearable consumer-based smart device, capable of recording physical activity or other health related data, through integrated sensors and algorithms, with the capability of transferring this data to a connected smartphone [63, 69]. In addition, activity tracker sensors are incorporated in more advanced wearables using a range of differ names. Some are synonyms, some indicate a subtle difference in terms of features, and others indicate large differences. Some additional common terms for wearables capable of activity tracking include *smartwatch, sport watch, GPS* (global positioning system) *watch, smart band*, *smart bracelet, hybrid watch, smart ring,* and *smartphone*. I will use the term *activity tracker* or *tracker* in this dissertation as the umbrella term for all consumer-based wearables capable of collecting data on physical activity.

The state of activity trackers on the consumer market has changed over time. The first mechanical hip-worn pedometer for step counting became popular in the 1990's and were gradually replaced with more accurate accelerometer-based devices in the early 2000. Eventually, new technology made it possible to connect these devices wirelessly to smartphones using Bluetooth technology, triaxial accelerometers started to replace the uniaxial accelerometer, and additional sensors (e.g. gyroscope, heart rate sensor) became more prevalent. This allowed for more complex metrics to be calculated, and the relatively simple hip-worn pedometer have evolved into today's increasingly advanced multi sensor devices (i.e. activity trackers), which now more commonly are worn on the wrist.

In 2011, Jawbone was one of the first providers to release a wrist-worn accelerometer-based activity trackers, the Jawbone UP. Since then, new activity trackers and new providers of activity trackers appear on the consumer market every year, with increasing sensor support and, according to the suppliers, increasingly accurate internal algorithms for data analysis and more advanced connected mobile applications.

1.5.2 Sensors in consumer-based activity trackers

Where traditional research devices for physical activity tracking most often has a limited number of sensors, activity trackers and smartwatches are often packed with a wide array of sensors. The simplest models may only contain an *accelerometer*, but more high-end devices often also include a *gyroscope*, *magnetometer, barometer or altimeter*, global positioning system (*GPS*), and/or optical heart rate sensor (i.e. *photoplethysmograph)*. Other sensors also exist, e.g. electrocardiography sensors, temperature sensors, light sensors, humidity sensors, proximity sensors, and galvanic skin response sensors [70].

Accelerometers are the basic sensors in current consumer-based activity trackers [69]. Details about frequency and number of axis are not always made available by the activity tracker provider, but typically these accelerometers records data in three axes, 50-100 times per second (i.e. Hz). In contrast to accelerometers developed for research-purposes, raw accelerometer data are not commonly exposed from consumer-based activity tracker. Instead, the accelerometer data are used in internal algorithms together with data from other sensors, which produces and presents a list of calculated variables (e.g. steps, TEE, MVPA).

A *gyroscope* measures orientation (i.e. angular movement) of a device. This information is used by internal device algorithms to increase the accuracy of physical activity estimation by using the change in orientation over time to classify activity type [71]. The rotation axis is set horizontally and should point north. This rotation axis must be regularly restored as the gyroscope does not seek north, and slowly drifts away from north, i.e. gyroscopic drift.

A *magnetometer* is a digital compass that can detect the orientation of a device relative to magnetic north [72]. The magnetometer improves motion tracking accuracy by compensating for gyroscopic drift, by restoring the orientation of the gyroscope rotation axis towards north.

Barometers/altimeters are used to detects changes in altitude [72]. These sensors can further improve some activity tracker outputs. Climbing a hill or a flight of stairs increases physical activity intensity, and the additional energy expenditure can thus be added to the daily total. In addition, these sensors can be used to report additional metrics, e.g. number of stairs climbed.

Photoplethysmograph (PPG) is a low cost and non-invasive optical technique for estimating heart rate where light from a light-emitting diode (LED) is emitted onto the skin and reflected

to a photodetector. Changes in blood volume under the skin affects the wave form of the returning light, allowing the sensor to estimate heart rate and other physiological parameters (e.g. oxygen saturation) [73]. The reflected light is also affected by other factors, e.g. skin tone and deformation, blood flow dynamics, movement artefacts, ambient light and temperature, and LED colour [73, 74, 75]. The resulting signal noise reduces the accuracy of the heart rate estimation and must be cleaned, using cleaning algorithms that also include data from other sensors. The accelerometer is also used for this purpose [76], but other sensors have been suggested, including gyroscopes [77] and secondary infrared PPGs for motion detection [78].

In addition to sensors for physical activity estimation, communication hardware is needed to communicate with the user's smartphone. Most activity tracker therefore contains Bluetooth for wireless communication [69]. In addition, more expensive activity trackers sometimes also include Wi-Fi for communication through wireless local area networks, and eSIM (embedded subscriber identification module) allowing the activity tracker to make phone calls and be connected to the mobile network without being connected to a smartphone.

1.5.3 Current state of validity – physical activity

Although there are many providers and models available on the consumer market, the number of providers used in research settings is considerably lower. When we identified articles (Ovid MEDLINE) and active/planned studies (ClinicalTrials.gov) for Paper I (2018), Fitbit was by far the most popular provider, followed by Garmin, Misfit, Apple, and Polar. Provider popularity will change over time as new companies are founded, goes out of business (e.g. Jawbone), are acquired by larger companies (e.g. Pebble), or pivots away from the activity tracker domain (e.g. Microsoft). Current activity trackers use a range of sensors to estimate a range of variables, but the accuracy of these estimations varies. In order to use these devices for research purposes there is a need to continuously validate new activity trackers as they are released on the consumer marked.

A large amount of validation studies on consumer-based activity trackers have been conducted to date. Most of these were done on activity tracker models that are no longer available today. Furthermore, several providers are also no longer available or do no longer produce activity trackers. The validity of the most relevant providers, due to high market share or special relevance (e.g. Polar) in this dissertation, is discussed below. Some provider details are summarized in Table 1. Market share details varies depending on source. Apple

have had the highest worldwide market share for several years [59, 60], but has recently been surpassed by two Chinese companies (Huawei and Xiaomi) [79]. In North America (which is likely more representative for the western region), Apple and Fitbit are the two largest providers by activity tracker shipment [80].

Table 1. Provider market share summary (Paper I)

Provider	**Founded**	**First wrist/finger-worn tracker**	**Market share worldwide, Q2 2020***	**Market share, North America, Q2 2020****
Huawei	1987, China	Talk Band B1 (2014)	24.0%	
Xiaomi	2010, China	Mi band (2014)	20.4%	
Apple	1997, US	Apple Watch (2015)	17.1%	37.6%
Fitbit	2007, US	Flex (2013)	7.3%	19.3%
Garmin	1989, US	Forerunner 220 (2013)	4.5%	8.1%
Samsung	1969, Korea	Galaxy Gear (2013)		5.0%
Polar	1977, Finland	Loop (2013)		

* IDC: Wearable Devices Market Share [79]

** Canalys: North American wearables market Q2 2020 [80]

Only few validations studies are conducted on *Huawei's* activity trackers, but due to their increasing popularity, especially in Asia [59], we may see more studies in the future [79]. Degroote et al. [81] concluded that the Huawei Watch (discontinued) accurately estimates steps in free-living. Xie et al. [82] found high accuracy for steps but low accuracy for energy expenditure for the Huawei Talk Band B3 (discontinued).

Although *Xiaomi*'s Mi Band has high sales numbers worldwide, this is largely because it is very popular in Asia [59]. There are only a few available validation studies on Xiaomi, and they are all on Mi Band 1 or Mi Band 2. Mi Band is currently on its fifth generation and Xiaomi have already announced the release of Mi Band 6. Although two Xiaomi validation studies were included in the Fuller et al. [83] systematic review, their accuracy was not addressed directly due to lack of data. However, a recent study concluded that the Mi Band 2 showed high validity for counting steps [84].

Apple validation studies are included in several systematic reviews [83, 85, 86]. The largest systematic review analysing Apple validation studies was published by Fuller et al. [83] in

2020 and included 28 Apple studies. They concluded that Apple (and Samsung) showed highest validity for step counting, compared to other providers, and that Apple overestimated energy expenditure in 58% of studies. These reviews included studies conducted on Apple Watch first and second generation (six generations exist), both of which are discontinued. Several additional validation studies have been published more recently, reaching similar conclusions, also limited to the first two generations.

Fitbit is the most included provider in validation studies, and has been reviewed in several systematic reviews published between 2015 and 2020 [83, 85, 86, 87, 88, 89]. A 2020 systematic review by Fuller et al. [83] included 144 Fitbit studies. This review analysed device accuracy when estimating steps and energy expenditure. They concluded that Fitbit does seem to provide accuracy step estimates in lab-settings. Energy expenditure was found not to be accurate. Since Fuller et al. [83] conducted their search in May 2019, several new Fitbit validation studies on physical activity have been published. These studies were conducted on the same devices as previous studies, and they have similar conclusions. None of the included activity tracker models are currently being produced by Fitbit and have been replaced by newer models.

Garmin validation studies have also been included in several systematic reviews [83, 85, 90]. A study by Evenson et al. [90] published in 2020, systematically reviewed 32 Garmin validation- and reliability studies. They concluded that the validity of step counting for Garmin devices was high, but the validity of energy expenditure estimates was low. The study by Fuller et al. [83], reviewing 42 Garmin studies, concluded that Garmin had a comparable lower error with a tendency to underestimate steps, but energy expenditure was consistently underestimated with high error. Activity tracker models included in these reviews are no longer available from the provider. A few newer Garmin validation studies, not included in any systematic review, have been published. These studies are also conducted on trackers no longer produced by Garmin.

Samsung validation studies are included in two identified systematic reviews [83, 86]. Bunn et al. [86] concluded, based on two validation studies, that energy expenditure estimates was valid for Samsung Gear S (discontinued). In addition, step count had acceptable agreement, but with wide limit of agreements. Fuller et al. [83] concluded that Samsung (and Apple) showed highest validity for step counting, compared to other providers. Two more recent studies on Samsung activity trackers concludes low validity for energy expenditure in youth

[91], and high validity for steps while jogging [92]. All activity tracker models included in these studies have been discontinued.

Four identified systematic reviews include *Polar* activity trackers [83, 85, 89, 93]. Fuller et al. [83] assessed 15 Polar studies and concluded that Polar generally overestimated energy expenditure. We assessed 11 studies and concluded that step count estimates seem to be more accurate compared to energy expenditure and physical activity intensity [93]. Activity tracker models included in these reviews are now discontinued. However, a few newer validation studies have been conducted on the Polar Vantage, which is currently available from the Polar web store. Gilgen-Ammann et al. [94] concluded moderate accuracy for energy expenditure estimates during activities requiring arm movements, Düking et al [95] concluded that energy expenditure estimates were not accurate, and we concluded that although correlations were strong for steps and energy expenditure, mean error was too high [96].

1.5.4 Current state of validity – sleep and heart rate

In addition to physical activity, heart rate and sleep outputs are increasingly becoming available in new activity trackers.

The accuracy of heart rate estimates from PPG is not clear and results differ depending on tracker model, current activity levels, and which metric is considered (e.g. heart rate, heart rate variability, resting heart rate, etc.) [97]. However, a recent systematic review and meta-analysis by Zhang et al. [98] concluded acceptable validity for wrist-based PPG estimated heart rate. A general agreement is that accuracy is reduced during high intensity physical activity [97, 99, 100]. A recent validation study on the accuracy of PPG in two wrist-worn activity trackers concluded generally accurate heart rate readings for alle age groups [101].

Most current activity trackers also report sleep related variables. Although not addressed in this dissertation, sleep is worth mentioning, as sleep estimates are based on the same sensors that physical activity related estimates are based on (i.e. accelerometer and PPG). Two 2015 systematic reviews concluded low validity of sleep estimation using current wrist-worn activity trackers [87, 102]. However, included studies only assessed devices with no heart rate monitor. In a more recent systematic review and meta-analysis on Fitbit activity trackers, which also included devices with heart rate sensors (i.e. PPG), Haghayegh et al. [103] concluded promising performance when identifying sleep vs awake time, but they also stated that such devices cannot be used as a substitute for polysomnography (i.e. gold standard).

In addition to wrist-worn activity trackers, newer form-factors are also emerging. For instance, the Oura-ring, which is one of few finger-worn activity trackers available, is an activity and sleep ring that also tracks heart rate and body temperature. Oura ring shows "promising results" and strong correlation with polysomnography for sleep detection [104, 105], can potentially be used to estimate resting heart rate [96], but cannot replace research-based accelerometers for physical activity estimations [96]. The ring packs a thermometer, an accelerometer, and an optical heart rate sensor, and provides a range of estimates for the wearer.

1.5.5 Summary

Step counting is the only variable that is repeatedly found to be valid, while energy expenditure is very often considered not valid. However, the results vary between studies and models, and although there are many validation studies available, there is still a need to conduct studies on current activity trackers. Many studies suggest caution when using activity trackers, especially for energy expenditure. Although dependent on activity type and intensity, heart rate estimates using wrist-worn PPGs generally show acceptable validity.

O'Driscoll et al. [85] also published a systematic review and meta-analysis in 2020, focusing on energy expenditure. They included validation studies on activity trackers from Apple, Polar, Garmin, Misfit, Withings, and Samsung. They did not address each provider in detail, but rather gave general conclusion about energy expenditure validity. They generally agree with other reviews and add that the accuracy of energy expenditure estimates from wrist-worn activity trackers are highly dependent on the performed activity type. They also saw that activity trackers that combined accelerometer data with heart rate data achieved lower measurements error.

Most activity trackers tested in all identified reviews and subsequent validation studies have since been discontinued or replace with newer versions and is no longer produced. Some are however still available in stores.

1.6 Related solutions

Below are some relevant solutions and projects that use activity trackers to collect health data.

Apple Health, HealthKit, and ResearchKit

Apple Health (Apple Inc., CA, US) is a mobile application preinstalled on iPhones. This application analyses manual input user data (e.g. age, height, weight) and internal sensor's output (e.g. accelerometers), and estimates and presents a range of different health related variables to the iPhone user [106]. Apple HealthKit is a developer framework for accessing and updating data stored in the Health application. Apple ResearchKit was released in 2015 and is a framework that allows researchers to create mobile applications and recruit research participants among iPhone users [107].

Google Health Studies

Google recently (December 2020) announced Health Studies, a similar solution as the Apple Research Kit [108]. The first study conducted using this framework are investigating how COVID-19 is linked to a person's movements.

Open mHealth and Shimmer

Open mHealth is a non-profit organization and a mobile health data interoperability standard [109]. Their goal is to make it easier to integrate patient health data from different sources, for easier data sharing and data harmonization. Solutions created by Open mHealth is open source and adaptable and improved by a community of developers worldwide.

Shimmer is "the first open-source health data integration tool" [110]. It can collect data from popular APIs (application programming interface), including Fitbit, Google Fit, iHealth, Misfit, Runkeeper, and Withings. Collected data are stored using the Open mHealth standard format. Several of these integrations no longer work, and the last update to the Shimmer codebase was in September 2018.

Human API

Human API (Human API, CA, US) is a company and tool that provides a "customer-controlled health data platform" [111]. They have specialized in integrating data from a large number of systems used in the US health sector, including electronic health records, patient

portals, health insurers, laboratories, and pharmacies. They also have integration with activity tracker providers thru open APIs, including Fitbit, Withings, Apple, Google, and Garmin, as well as large activity tracker application companies like Strava, UnderArmour, and MyFitnessPal. Since they only operate in the US, Human API is HIPAA-compliant (US Health Insurance Portability and Accountability Act) but not GDPR-compliant (European Union General Data Protection Regulation). Human API is used by the Health eHeart study at University of California San Francisco (UCSF) to gather more participant health data [112] using data from activity trackers.

DETECT study

The DETECT health study was launched early 2020 by the Scripps Research Institute [113]. The main aim of this study is to collect activity tracker data to predict viral outbreaks, by analysing heart rates and physical activity patterns of participants over time. The DETECT team has recently shown that by using heart rate data collected from Fitbit activity trackers, it is possible to detect influenza-like outbreaks [114]. This was a research collaboration between Fitbit and Scripps Research Institute, where two years of de-identified data from 200.000 Fitbit users were analysed and compared with historic dates and areas of flu-outbreaks in five US states. In a recent study they have also looked at how sensors data and self-reported symptoms can be used for COVID-19 detection using participant data collected by Fitbit activity trackers, or collected by smartphones and stored in Google Fit and Apple Health [115].

All of Us Research Program

Fitbit also has a collaboration with the US National Institute of Health, where participants can sign up to a Fitbit Bring-You-Own-Device project. The program aims to recruit one million participants in the US by 2024. In addition to questionnaires, physical measurements, and biospecimens, they also plan to access physical activity data collected by participant's Fitbit activity trackers, if they own one [116]. This is a large program and possibly the first to actively plan to use activity tracker data at this scale.

RADAR-base

The RADAR-base is an "open source platform for remote assessment using wearable devices and mobile applications" [117]. This platform can access physical activity data from Fitbit

activity trackers (and some research grade devices) by programmatically accessing the Fitbit cloud storage for participants who have granted such access [118]. This system was recently used by Sun et al. [119] to analyse change in physical activity patterns during the COVID-19 pandemic (by accessing already downloaded data among participants with chronic disease).

1.7 Rationale for this study

Physical activity is an important modifiable lifestyle factor that can improve general health and reduce the risk of disease. Epidemiological studies such as population-based studies and clinical studies cannot currently adequately monitor physical activity over time. Long-term monitoring of physical activity is important, typically for surveillance of physical activity levels in a population over time, or to study participant physical activity changes in a clinical intervention study.

The ability to objectively measure physical activity in epidemiological research is traditionally limited to providing snapshots of physical activity levels for individuals (typically a week of recording or shorter). Accessing more continuous and long-term data may provide valuable insight into how physical activity changes in a study population over time, or before, during, and after an intervention period. Consumer-based activity trackers are designed for long-term and continuous use and can therefore potentially be used for this purpose.

The rationale for this study is the need to assess the feasibility of collecting physical activity data using consumer-based physical activity trackers, as well as to create a solution that can access this diverse and large-scale data source in a simple and manageable manner.

2 Aims

The overall aim of this dissertation was to explore and develop new methods to collect data on physical activity in epidemiological studies using consumer-based activity trackers. The study settings of interest were both observational and experimental studies, i.e. monitoring of physical activity levels in a population over time, and physical activity changes among participants in a clinical intervention study.

Specific aims

Aim 1: To identify available activity trackers on the consumer market, investigate the current state of activity tracker usage in health research, and compare how different providers facilitates developer access to collected data.

Aim 2: To test the validity of a currently available activity tracker and determine which variables can be used in health research to infer physical activity levels in study participants.

Aim 3: To identify success factors for increasing wear time of activity trackers when used to collect physical activity data over a prolonged period in a clinical study.

Aim 4: To implement a system for automatic and continuous physical activity monitoring using consumer-based activity trackers, and to examine the usability of this system as a tool for long-term physical activity recording in epidemiological studies.

3 Materials and methods

3.1 Introduction

This dissertation is based on four papers, using different methods to explore the research questions and overall theme from multiple angles. In Paper I we identified and described historic characteristics for activity trackers on the consumer market (Aim 1), by searching online and offline databases and providers websites. Results were described descriptively. In Paper II we tested the accuracy of the Polar M430 activity tracker (Aim 2), with multiple reference monitors on multiple locations. We assessed validity using Pearson correlation, intraclass correlation, Bland-Altman plots, and mean absolute percentage error. In Paper III we identified success factors for increasing activity tracker wear time among participants in a clinical intervention study (Aim 3) using a mixed methods approach, combining quantitative wear time estimates with qualitative interviews. Finally, in Paper IV we implemented *mSpider* (Aim 4), a system for long-term physical activity monitoring. As an example of usability, we analysed participant's activity tracker data to study change in physical activity during the COVID-19 pandemic, with t-tests and Wilcoxon tests to compare periods. Details are described below for each paper separately.

3.2 Paper I – Analysis of consumer wrist-worn activity trackers

In Paper I we identified available activity trackers and reported findings in three sub-groups: 1) available providers (i.e. brands), activity trackers (i.e. devices), and sensors, 2) providers used in research, and 3) developer possibilities for third party data access. For each sub-group we described the search strategy and inclusion and exclusion criteria separately. We also described how activity trackers were categorized and grouped for reporting. Results were reported descriptively. The content of this chapter is based on the method section of Paper I [120].

3.2.1 Providers, activity trackers, and sensors

Search strategy

In the first sub-section, we searched five online databases containing information on various types of smart wearables: The Vandrico Wearables database (Vandrico.com) [121], GsmArena.com [122], Wearables.com [123], SpecBucket.com [124], and PrisGuide.no [125,

126]. We were also granted access to an offline wearable database: The Queen's University's Wearable Device Inventory [127].

After merging all six databases we extracted a list of providers (i.e. vendor/brands) names. For each provider we searched their website for additional activity trackers. Conflicting information between databases were resolved by accessing provider websites. If no official website existed (e.g. provider or activity tracker no longer available), we used other online sources, e.g. Wikipedia and Google searches. The search was performed between May 15th and July 1st, 2017.

Activity tracker categorization and data collection

Activity trackers were grouped into three categories: 1) smartwatches, 2) fitness trackers, and 3) hybrid watches.

An activity tracker was classified as a *smartwatch* if the provider classified it as a smartwatch and it supported smartphone notifications, or if it had a touch screen and was not specifically defined as a fitness tracker by the provider. An activity tracker was classified as a *fitness tracker* if its main function was to track physical activity, or it was specifically defined as a fitness tracker by the provider, or it did not support smartphone notifications. An activity tracker was classified as a *hybrid watch* if it had an analogue clockwork with a built-in digital accelerometer

For each tracker we collected the following 11 variables: provider name, tracker name, release year, provider country, tracker category (i.e. smartwatch, fitness tracker, or hybrid watch), and whether they had a built-in accelerometer, gyroscope, magnetometer, barometer/altimeter, GPS, and/or PPG.

Inclusion and exclusion criteria

For the "provider, activity tracker, and sensors" search, we included only wrist-worn consumer-based activity trackers that included an accelerometer to estimate physical activity. Further, activity trackers had to be designed for continuous usage and capable of sharing collected data with user's smartphone using Bluetooth technology. We included activity trackers released before July 1st, 2017. Hybrid watches were excluded.

3.2.2 Providers used in research

Search strategy

In the second sub-section, we searched Ovid MEDLINE and ClinicalTrials.gov, to assess provider usage in research so far, and planned usage in future studies.

Since we identified 132 different providers in the first sub-section, we limited this search to the most relevant providers. Relevant providers were a priori defined as 1) being one of the five most sold providers in 2015 or 2016 (i.e. Fitbit, Xiaomi, Apple, Garmin, and Samsung), or 2) had released 10 or more unique activity trackers (i.e. Garmin, No.1, MyKronoz, Samsung, and Polar). We performed a separate Ovid MEDLINE search for each identified provider. We exclude articles using out of scope activity trackers by screened the title, abstract, and method section from the resulting list of articles.

We also identified additional providers used in the included articles, to complement the list of "relevant providers", and performed a similar Ovid MEDLINE search for these additional providers. We finally defined 11 providers as most relevant. The search was performed on September 30th, 2017, and divided into validation- and reliability studies, and studies using activity trackers to collect data.

For each provider, we performed an equivalent keyword search on ClinicalTrials.gov, and screened project descriptions to identify upcoming studies where activity tracker usage was included in the protocol.

Inclusion and exclusion criteria

For the "providers used in research" search, only the 11 providers defined as "most relevant" were included. We excluded providers from companies that no longer exist or no longer produced activity trackers.

3.2.3 Provider developer possibilities

Search strategy

Different providers have different capabilities in terms of data sharing to third party systems, integration with health data clouds, mobile application developer support, and supported smartphone ecosystems, and is thus not equally relevant for all research project. In the third sub-section, we therefore reviewed the 11 identified providers to map these capabilities.

Information was gathered from the App Store (Apple), Google Play (Android), and official provider web sites. We especially focused on the availability and capabilities of the provider Application Programming Interfaces (API) and Software Development Kits (SDK). Information was collected in September 2017.

Inclusion and exclusion criteria

For the "provider developer possibility" search, we only included the identified 11 most relevant providers. We excluded providers that were not used in any of the previously identified articles from the Ovid MEDLINE search.

3.3 Paper II – Polar M430 validation study

In Paper II we performed a validation study on the Polar M430 activity tracker. We compare correlation and agreement between the Polar M430 and multiple reference monitors. The content of this chapter is based on the method section of the published Paper II [128].

3.3.1 Study sample

For the Polar M430 validation study we recruited 50 participants. We used convenience sampling to increase ranges for height, weight, body-mass-index, age, and sex. Inclusion criteria were age ≥18 years, normal physical function level, and agree to wear all instruments for one full day and night of recording. All demographic data were self-reported.

3.3.2 Protocol

Participants wore two ActiGraph wGT3X-BT (ActiGraph LLC, Pensacola, FL, USA), two Actiheart 4 (CamNtech Ltd, Cambridge, UK), and one Polar M430 (Polar Electro oy, Finland) activity tracker. Appendix A gives the study protocol detailing instrument setup.

Instruments

The ActiGraph wGT3X-BT is a triaxial accelerometer with 30-100Hz sampling rate which can be worn on multiple locations. It is extensively used in research and is considered valid for estimating sedentary behaviour [129, 130, 131], physical activity intensity [26, 129], steps [132], and energy expenditure [133].

The Actiheart is a uniaxial accelerometer with 32Hz sampling rate. It is attached to the chest and includes a 1-lead 128Hz electrocardiography sensor. The Actiheart gives valid estimations for energy expenditure, both in free-living [134] and in laboratory settings [23].

The Polar M430 is a consumer-based wrist-worn activity tracker released in 2017. It has a 50Hz triaxial accelerometer for activity tracking and an optical heart rate sensor with six light-emitting diodes for increased accuracy.

Procedure

The ActiGraph was worn on the right hip and on the wrist of the non-dominant hand. The Actiheart was attached to the chest, using two Red Dot 2238 electrodes (3M, St Paul, MN, USA) per device, in the upper- and lower position (approximately at the height of the second- and fifth intercostal space). The Polar M430 was worn on the wrist of the non-dominant hand, below the ActiGraph. Figure 2 (copied from Paper II) illustrates the wear location for each instrument. Instruments were initialized using self-reported height, weight, age, sex, and dominant hand, and set to record data for one full day and night (midnight to midnight). Accelerometers were setup with maximum sampling rates.

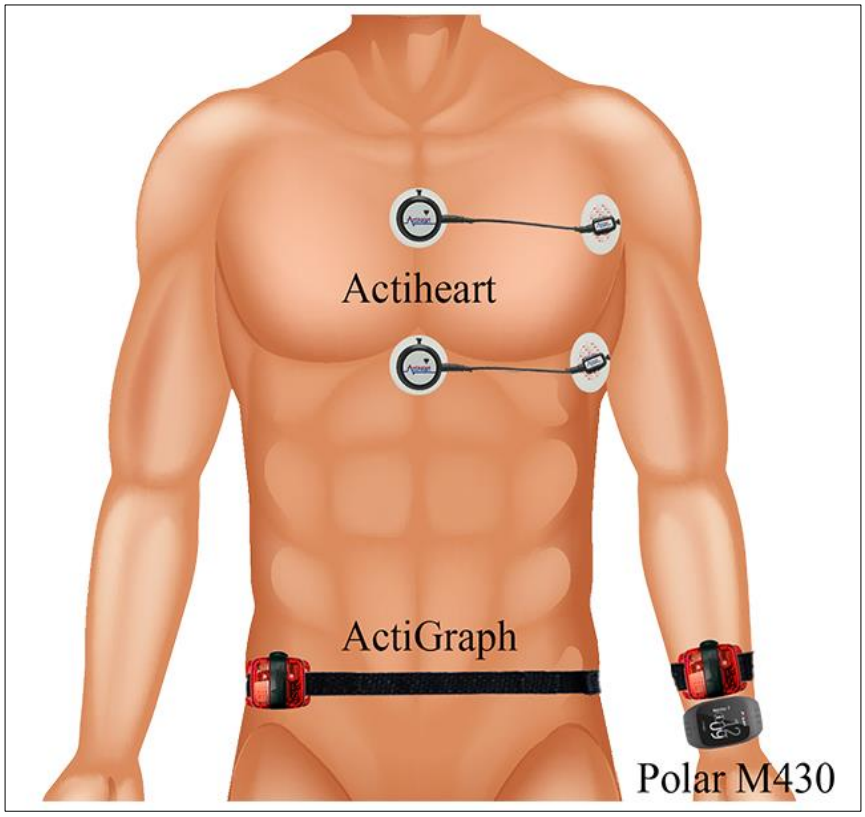


Figure 2. Instrument placement (Paper II)

Participants were asked to wear all devices for 24 hours (midnight to midnight), and only take off the ActiGraph during water activities and while showering. We collected data in May 2018. All participants received written and oral instructions and signed informed consent (See Appendix B).

3.3.3 Variable creation

The different instruments support different solutions for data export and data analysis depending on variable. ActiGraph step counting is internally calculated by the ActiGraph and were exported directly using ActiLife. Actiheart does not record steps. Energy expenditure were analysed in ActiLife and the Actiheart software. Physical activity intensity variables were analysed in an external tool for the ActiGraph and the Actiheart, using accelerometer-based activity counts.

All Polar M430 variables were calculated by Polar's internal algorithms and exported using Polar Flow. We exported total energy expenditure, steps, minutes of sitting (i.e. sedentary time), as well as minutes of low (i.e. light)-, medium (i.e. moderate)-, and high (i.e. vigorous) physical activity.

Physical activity intensity

For the ActiGraph and Actiheart, physical activity intensity was analysed using accelerometer-based activity counts. We used ActiLife to export activity counts from the ActiGraph data into CSV (comma-separated value) files. We exported activity counts for all three axes of the accelerometer (vertical, horizontal, lateral). Activity counts were exported per 10 second epochs (lowest possible setting). We similarly used the Actiheart software to export activity counts from the Actiheart data into CVS files. Actiheart is uniaxial, and activity counts were created based on vertical acceleration and exported per 15 second epochs (lowest possible setting).

There are no agreed upon cut-points for physical activity intensity classification for ActiGraphs worn on the wrist by adults [25]. Therefore, before exporting ActiGraph activity counts we applied the wrist-to-hip conversion available in ActiLife. Exported files were imported into QCAT (Quality Control and Analysis Tool), a custom-made system for analysing activity counts from accelerometer data. The 10-epoch and 15-epoch based activity

count from ActiGraph and Actiheart, respectively, were converted into 60-second epochs before further analysis, giving counts per minute (CPM).

We excluded non-valid days, where a valid day was defined as having at least 10 hours of wear time on all instruments [135]. We used the Hecht triaxial (i.e. vector magnitude (VM)) wear-time algorithm for wear time analysis [136]. The Hecht algorithm analyses VM CPM for each 1-minute epoch to answer three questions, where a given minute is defined as wear time if at least two of the following questions are affirmed (default threshold):

- Is the VM CPM above 5?
- For the following 20 minutes, is the VM CPM above 5 for at least two minutes?
- For the preceding 20 minutes, is the VM CPM above 5 for at least two minutes?

We used CPM to classify each minute of activity into five different intensity zones using separate cut-points for uniaxial ActiGraph CPM, triaxial ActiGraph CPM VM, and Actiheart CPM. We used cut-points defined by Freedson et al. [129] and Matthews et al. [137] to classify ActiGraph uniaxial CPM (vertical axis) into intensity zones. We similarly used cut-points defined by Sasaki [26], Kozey-Keadle [130], and Peterson et al. [131] to classify triaxial ActiGraph CPM VM. We classified Actiheart CPM using suggested cut-points by Schrack et al. [138], which was the only study we could identify where chest-based Actiheart cut-points in adults were suggested. For all instruments we combined vigorous and very-vigorous physical activity into one variable (i.e. vigorous), since very vigorous was not supported by the Polar M430. Table 2 from Paper II is shown in Table 2 below and shows all cut-point sets.

Table 2. Physical activity intensity zones cut points (Paper II)

Intensity zone	**ActiGraph uniaxial CPM**	**ActiGraph triaxial CPM VM**	**Actiheart CPM**
Sedentary	≤ 99	≤ 149	≤ 10
Light	100-1951	150-2689	11-95
Moderate	1952-5724	2690-6166	96-234
Vigorous	5725-9498	6167-9642	≥ 235
Very vigorous	≥ 9499	≥ 9643	

CPM: Counts per minute. VM: Vector magnitude.

Energy expenditure

QCAT only supports physical activity intensity analysis and does not support energy expenditure estimation. We therefore calculated CPM based energy expenditure using the proprietary software for each instrument (i.e. ActiLife and Actiheart).

ActiGraph AEE was estimated in ActiLife using Freedson Combination '98" (Freedson 1998 [129] + Williams Work-Energy) for uniaxial calculations and "Freedson VM3 Combination '11" (Sasaki 2011 [26] + Williams Work-Energy) for triaxial calculations. Since Hecht non-wear algorithm is unavailable in ActiLife, we used the Troiano [139] algorithm with default settings. Default settings defines episodes of non-wear where there are at least 60 consecutive minutes of zero activity counts, allowing two minutes of between zero and 100 activity counts in the 60-minute period.

The Actiheart calculates AEE using a branched model, where a combination of accelerometer derived CPMs and electrocardiogram derived hearts beat per minute defines four different energy expenditure calculations [140]. In addition to AEE, the Actiheart also report REE, dietary induced thermogenesis (DIT), and TEE. The ActiGraph and Polar M430 only reports AEE and TEE, respectively. Actiheart uses the Schofield equation [141] to calculate REE. We therefore used the Schofield equation to convert between total- and AEE for the ActiGraph and Polar M430, including subtracting or adding 10% of TEE to account for energy expended due to food digestion (i.e. dietary induced thermogenesis).

3.3.4 Statistical analysis

We presented participants characteristics using descriptive statistics. For each research grade instrument (i.e. ActiGraph and Actiheart), we compared daily values for minutes of sedentary time, minutes of light-, moderate-, and vigorous physical activity, moderate-to-vigorous physical activity (MVPA), as well as step counts (ActiGraph only), AEE, and TEE, against the Polar M430 activity tracker. Furthermore, for the ActiGraph, we generated variables using accelerometer-based activity counts from both one- and three axes.

We used the Shapiro-Wilk test to test normality. Since several variables were not normally distributed, we tested both Pearson's and Spearman's correlation, with and without bootstrapping to compare results. There were no major differences between methods, and we finally used Pearson's correlation with bootstrapping for all combinations of variable and

reference monitor, to find a more accurate confidence interval and make results comparable. We used correlation cut-offs suggested by Evans [142] to classify the strength of the association, i.e. very weak: <0.2, weak: 0.2-0.4, moderate: 0.5-0-6, strong: 0.6-0.8, and very strong: >0.8.

We used Bland-Altman limits of agreement to compare the mean difference between instruments for a given variable [143], where a positive mean difference indicates that the Polar M430 overreports that variable compared to the reference monitor, and a negative mean difference indicates an underreporting.

Furthermore, to better quantify the level of agreement, we calculated the intraclass correlation (ICC), for each combination of variable and reference monitor, using absolute agreement, 2-way random, single measures. We used the 95% confidence interval of the ICC estimates to classify agreement, using suggested cut-offs by Koo et al. [144], i.e. poor agreement: <0.5, moderate agreement: 0.5-0.75, good agreement: 0.75-0.9, and excellent agreement: >0.9

We also calculated mean absolute percentage error (MAPE) to identify the measurement error between devices for each variable. Although no cut-off is defined to indicate low error, a common practice for studies conducted in free-living is to use 5% [145] or 10% [146, 147]. MAPE was calculated using the below equation, where A is the actual measurement (i.e. value from reference monitor) and E is the estimated value (i.e. value from Polar M430). This gives a measure of accuracy of the estimated values.

$$MAPE = 100x\frac{1}{n}\sum_{t=1}^{n}\left|\frac{At - Et}{At}\right|$$

Finally, we calculated the sensitivity and specificity of the Polar M430's ability to identify participants that achieved 10,000 steps/day [8]. Analysis were done using R version 3.5.3 (R Foundation)

3.4 Paper III – Succeeding with prolonged usage of activity trackers

In Paper III we conducted a sub-study within a pilot and feasibility study for a planned complex lifestyle intervention among middle-aged and elderly inactive people with obesity and elevated risk of cardiovascular disease; the RESTART trial (*Re-inventing Strategies for healthy Ageing; Recommendations and Tools*). We used a mixed methods approach to identify potential factor for successful long-term recording using consumer-based activity

trackers in clinical research. The content of this chapter is based on the method section of the published Paper III [148].

3.4.1 Study sample

For the pilot and feasibility study [149], which the study on prolonged activity tracker usage was part of, we randomly selected and invited 75 potential participants from Tromsø 7. Inclusion criteria were: 1) 55-75 years of age, 2) self-reported sedentary lifestyle during leisure-time, 3) body-mass-index $\geq$30 kg/m^2, 4) no previous myocardial infarction (self-reported), and 5) elevated NORRISK 2 score (a 10-year risk calculation for fatal and non-fatal myocardial infarction or stroke [150]). Invitations were sent by mail to their registered home address (see Appendix C). Twenty people responded to the invitation, of which four were excluded after telephone interviews. All participants received written and oral instructions and signed informed consent (See Appendix D).

The resulting 16 participants were included in the sub-study on succeeding with prolonged usage of activity trackers in clinical studies. All 16 participants remain in the feasibility study to study-end, but two participants stopped wearing the activity tracker at intervention-end and did not contribute quantitative data in the follow-up period. Further details about participants recruitment is available in Deraas et al. [149].

3.4.2 The RESTART pilot and feasibility study – Protocol

The main goal of the RESTART pilot and feasibility study was to test the feasibility of the planned intervention, in terms of recruitment, adherence, organization, and potential side-effects of participation. Participants attended a six-month physical activity intervention, with six months of follow-up. The intervention period consisted of two instructor-led exercise sessions per week (e.g. stationary biking, aerobic hall sessions, resistance training), individual and group-sessions with a nutritionist, and group-sessions with a psychologist for habit change counselling. Participants wore a Polar M430 activity tracker for physical activity monitoring throughout the intervention and follow-up period. Additional details about the feasibility study are described in Deraas et al. [149].

3.4.3 Polar M430 activity tracker

Physical activity recording

Participants were equipped with a Polar M430 activity tracker at baseline, one week before intervention start. They were initially asked to wear the activity tracker for six months (i.e. to intervention end). They were also equipped with an ActiGraph wGT3X-BT for eight days at baseline and for eight days at the end of the intervention period (i.e. after six months). Participants thus wore the ActiGraph and Polar M430 simultaneously for up to 16 days. In Paper II we showed that the Polar M430 can be used as a valid instrument for estimating TEE. However, because average error was high for steps and MVPA, despite showing strong correlation with an ActiGraph, we should be careful when using these variables for physical activity estimation. In the present study we used the overlapping days of Polar M430 and ActiGraph usage to further test the validity of the Polar M430 in the present cohort for MVPA, steps, and TEE.

The ActiGraph was worn on the right hip. The Polar M430 was worn on the wrist of the non-dominant hand. The ActiGraph was setup with maximum sampling rate. Polar M430 and ActiGraph characteristics have been described in chapter 3.3.2.

We used ActiLife to export steps, minutes of MVPA, and energy expenditure from the ActiGraph. We estimated MVPA using VM CPM cut-off at 2690, as suggested by Sasaki et al. [26]. Similarly, we estimated AEE using the “Freedson VM3 Combination '11” (Sasaki 2011 [26] + Williams Work-Energy) method, and converted this to TEE by adding resting energy expenditure, as suggested by Schofield et al. [141], and adding energy expended from dietary induced thermogenesis (i.e. 10% of TEE) .

Setup for long-term usage

Data registered by Polar activity trackers are transferred from the device, to a smartphone via the Polar Flow mobile application, and finally uploaded and stored in Polar Flow [151], Polar’s online storage for user collected activity data. In order to download data registered by the Polar M430, we created a de-identified Polar Flow account for each participant, where we only registered gender, year of birth, height, and weight. We stored no directly identifiable information on these accounts and disabled GPS for privacy reasons and to conserve battery power. We disabled all possible notifications. Sleep feedback could not be disabled. The

standard operating procedure for setting up the Polar M430 activity tracker is given in Appendix E.

We asked participants to wear the Polar M430 all day and night during the six-month intervention period. They were asked to only take it while charging every Sunday, and if they experienced any discomfort during sleep. Participants received written and oral instructions about how to use the Polar M430 activity tracker (see Appendix F).

Because of the long recoding period, to make data collection less demanding for the research team, participants who owned a smartphone were asked to install the Polar Flow mobile app. We did not share account credentials with participants but assisted in connecting the Polar M430 to their smartphone, in addition to providing technical assistance throughout the study. These participants were also asked to initiate data synchronization every Sunday as party of the weekly charging procedure. This bring-your-own device (i.e. smartphone) approach has shown to improve the experience and engagement of participants [152]. Some participants did not own a smartphone capable of connecting to the Polar M430. For these participants we connected the activity trackers to a project smartphone and synchronized data sporadically during weekly exercise sessions.

Several participants occasionally lost connection between the activity tracker and their smartphone throughout the intervention. We resolved these issues continuously by meeting participants before or after their exercise sessions. During these sessions, some participants addressed activity tracker related issues that we found important to report for future research. These were addressed and discussed together with other relevant researcher experiences.

To quantify change in physical activity levels before, during, and after the intervention, we asked participants to continue to wear the Polar M430 in the follow-up period, totalling 12 months of wear time. During the follow-up period, we met participants who had their Polar M430 connected to the project smartphone every 2-3 month to download data.

3.4.4 Participant perspective

Qualitative methods can enrich quantitative results when we wish to access participants experiences and perceptions [153]. Therefore, in addition to collecting quantitative data, we also used a qualitative approach to gain a deeper understanding of participant’s experiences with wearing the Polar M430 for one year. We performed individual semi-structured

interviews, as described by Brinkmann [154], with all participants at two time-points; mid-way in the intervention (i.e. after three months) and at study end (i.e. six months after intervention end). Two separate interview guides were developed, with a total of seven questions related to activity tracker usage, and used during each interview. Interviews were audio-recorded and transcribed verbatim.

3.4.5 Analysis

Quantitative analysis

We used descriptive statistics to describe participant characteristics at baseline. In addition, we presented a comparison between responders and non-responders, using data recorded in Tromsø 7.

For quantitative analysis we first used Polar Open AccessLink API (application programming interface) to download daily values for MVPA, steps, TEE, and hours of wear time, as reported by the Polar M430. This API facilitates automatic data extraction from consenting Polar users, i.e. the pre-created de-identified Polar accounts. Only valid days were included in analysis. A day was defined as valid if the activity tracker was worn 10 hours or more any given day [135].

We analysed wear time descriptively, reporting the percentage of valid days for the full year of recording by participant. We also reported mean number of valid days for all 16 participants and for participants not lost to follow-up (n=14) separately. We further analysed wear time and qualitative comments given during interviews together, to identify reasons for not wearing the activity tracker during the follow-up period.

Phillips et al. [155] has suggested to test the validity of consumer-based activity trackers in the target cohort before relying on activity tracker output for outcome analysis. Therefore, although we had previously analysed the validity of this activity tracker in a different cohort, we also tested the validity of the Polar M430 in the current cohort of participants. Because we had multiple days of simultaneous recording of the Polar M430 and a reference monitor, we used repeated measures correlation [156] (with bootstrapping to ascertain a more representative confidence interval) to determine correlation between the Polar M430 and an ActiGraph wGT3X-BT accelerometer for relevant variables. We applied the following

correlation cut-offs, as suggested by Evans [142]: very weak: <0.2, weak: 0.2-0.4, moderate: 0.4-0.6, strong: 0.6-0.8, and very strong: >0.8.

Finally, we calculated MAPE, using a 10% cut-off to classify acceptable error, and Bland-Altman limits of agreement [143] to assess agreement between the Polar M430 and the ActiGraph. All statistical analyses were done using R version 3.5.3.

Qualitative analysis

We used QSR NVivo 12 Plus (QSR International, Pty Ltd) for structuring the transcribed text during the analysis phase. Thematic analysis is widely used among health researchers to identify themes and patterns in interview data [153]. We used a data-driven inductive approach to allow patterns in the data to emerge, rather than a deductive approach, since we did not know in advance how to structure our findings [157]. We also used a semantic approach, as compared to a latent approach, because we wanted to identify participant's explicit opinions, rather that underlying ideas or patterns in their responses. Further, we used the six steps for thematic analysis defined by Braun and Clarke to identify patterns (i.e. data familiarization, initial coding, generating themes, reviewing themes, defining and naming themes, and writing up report) [158]. Finally, we gave equal weight to comments mentioned by one participants as comments mentioned by multiple participants [159].

We performed coding in three iterations, where the first iteration was done using printouts and manually annotating of initial codes. This iteration resulted in a large set of codes with much overlap. In the second iteration we used NVivo to merge the paper-based list of codes into 11 themes: 1) metric inaccuracy, 2) elements that triggered irritation, 3) tracker visual design (look and feel), 4) tracker practical design (ease of use), 5) motivation for usage, 6) effect of using the tracker, 7) how tracker was used, 8) why the tracker was used, and comments on available metrics, including 9) sleep, 10) heart rate, and 11) physical activity. These 11 themes were reduced into the four following themes in the final iteration: 1) motivation, 2) activity tracker usefulness, 3) activity tracker annoyances, and 4) activity tracker improvements. The merging process is given in Table 3. The first and second author performed coding separately and harmonized codes through discussion. The first author did the initial analysis after coding, which was thoroughly reviewed by the second and last author. For each theme we finally extracted relevant quotes and tagged identified quotes with sex, age, and whether the participant owned a smartphone. The age variable was randomly

increased or decreased by one, to prevent participant identification. All quotes were translated from Norwegian.

Table 3. Theme merging process (Paper III)

Themes: Iteration 2		**Themes: Iteration 3**
Metric inaccuracy Elements that triggered irritation Sleep metric	→	Activity tracker annoyances
Tracker visual design Tracker practical design	→	Activity tracker improvements
Motivation for usage	→	Motivation
Effect of using the tracker How tracker was used Why the tracker was used Heart rate metric Physical activity metric	→	Activity tracker usefulness

3.5 Paper IV – Physical activity surveillance during COVID-19 pandemic

In Paper IV we described how we implemented a system for collecting data on physical activity from a range of different providers of consumer-based activity trackers. In addition, we use data collected from 113 participant's activity trackers to detect change in physical activity due to the Norwegian COVID-19 lockdown in March 2020. The content of this chapter is based on the method section of Paper IV [160].

3.5.1 The mSpider system

Introduction

The mSpider (Motivating continuous Sharing of Physical activity using non-Intrusive Data Extraction methods Retro- and prospectively) system ("the system") is an experimental tool designed for automatic and continuous collecting of health-related data recorded by consumer-based activity trackers. The system is designed to collect physical activity and related variables from activity trackers from a range of different providers over an extended period. Activity trackers includes any smart device with sensors capable of estimating these variables and that can transfer data from the tracker to a smartphone for persistent storage.

The aim of the system is to collect these types of data from all categories of consumer-based activity tracker, including smartphones, smartwatches, activity trackers, as well as from fitness- and health apps (Aim 4). Larger providers have cloud storage solution where user data are uploaded. Smaller providers often store data in large open cloud repositories, e.g. Google Fit and Apple Health. Which type of data are stored depends on both provider and activity tracker.

For most activity trackers, collected data are transferred to a provider-specific mobile application on the user's smartphone, and finally uploaded to the affiliated cloud storage or an open cloud storage. Some trackers do not upload data at all, and only stores data locally on the user's smartphone.

The mSpider system can access uploaded data using two methods. Most providers support Open Authorization, which allows third party systems (e.g. the mSpider system) access to user collected data through available APIs (Application Programming Interface). Some large providers (e.g. Samsung and Apple) do not exposed such APIs, but instead offer SDKs (Software Development Kits). SDKs can be used when implementing mobile apps to add support for data extraction from these providers. Instead of downloading data directly from the provider cloud, data are extracted from the provider mobile application (e.g. Samsung Health, and Apple Health), which in turn downloads data from the provider cloud. Data are then finally uploaded from the mSpider mobile application to the mSpider server backend.

There are many providers of consumer-based activity trackers. The system currently supports the following providers: Fitbit, Garmin, Oura, Polar, Samsung, Withings, Apple, and providers storing data in Google Fit or Apple Health. Figure 3 illustrates how data are transferred and stored for supported providers. Data are uploaded to a cloud storage for all providers and transferred to the mSpider system using one of the two mention methods.

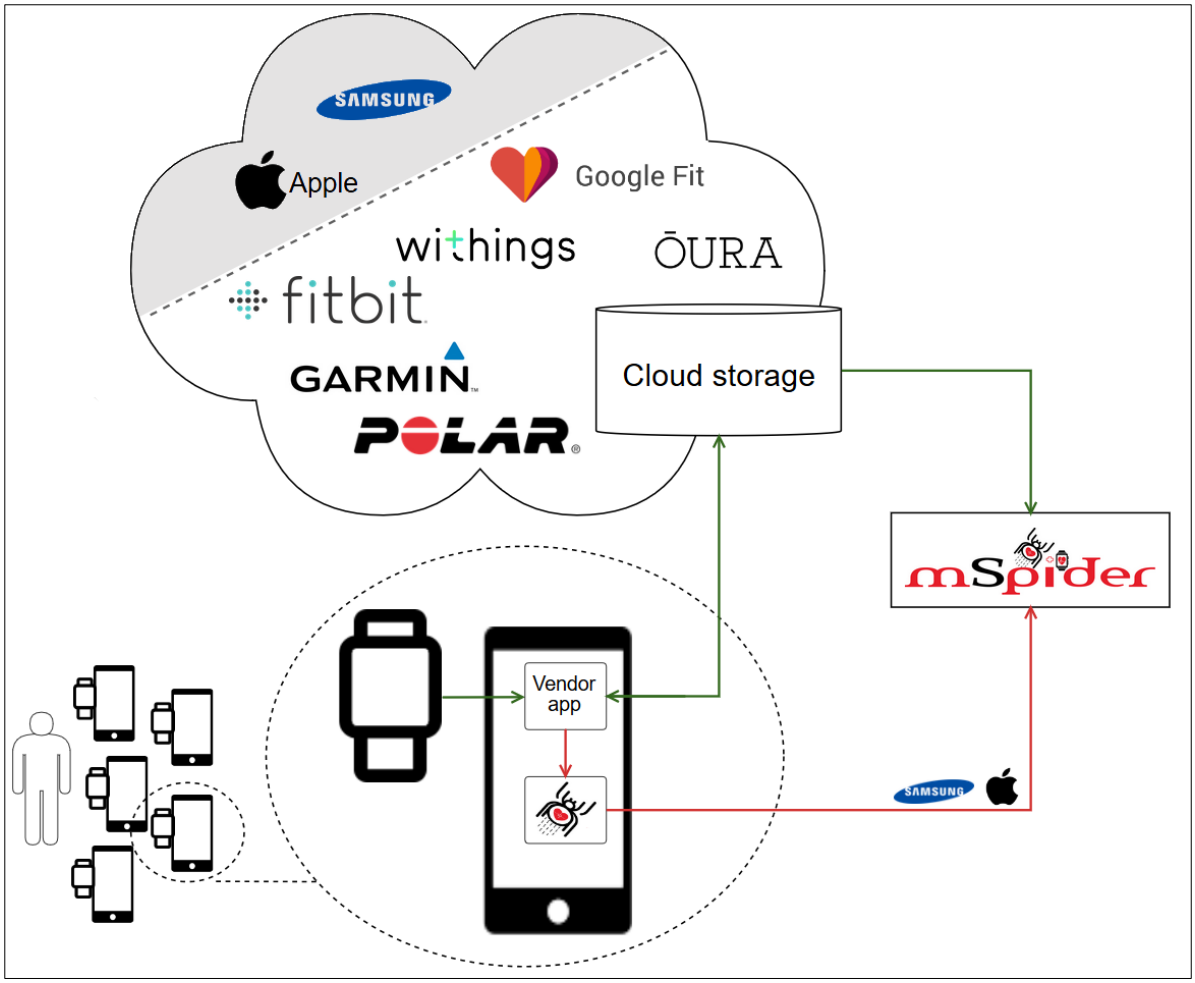


Figure 3. mSpider overview (Paper IV)

System architecture

The system has three major modules: 1) the *web frontend*, 2) the *server backend*, and 3) the *mobile application*.

The *web frontend* is used by study administrators to 1) manage studies (create, edit, delete), 2) manage participants (add/invite, remove), 3) send e-mail invitations and reminders, and 4) download participant meta data and collected physical activity data. It is used by participants to 1) register to studies (i.e. grant access to physical activity data), 2) withdraw from studies, and 3) read information about the mSpider system, including privacy statements and terms of service statement.

After recruitment, participants are registered in the web frontend and automatically assigned a unique identifier. A web link is created for each participant using this identifier, which must be shared with the participant. By clicking the web link, the participant is forwarded to a web

page where he/she can select activity tracker provider. This web page is shown in Figure 4. Finally, the participant is forwarded to the selected provider, where they authorize the mSpider system and grant access to physical activity data collected by their activity tracker.

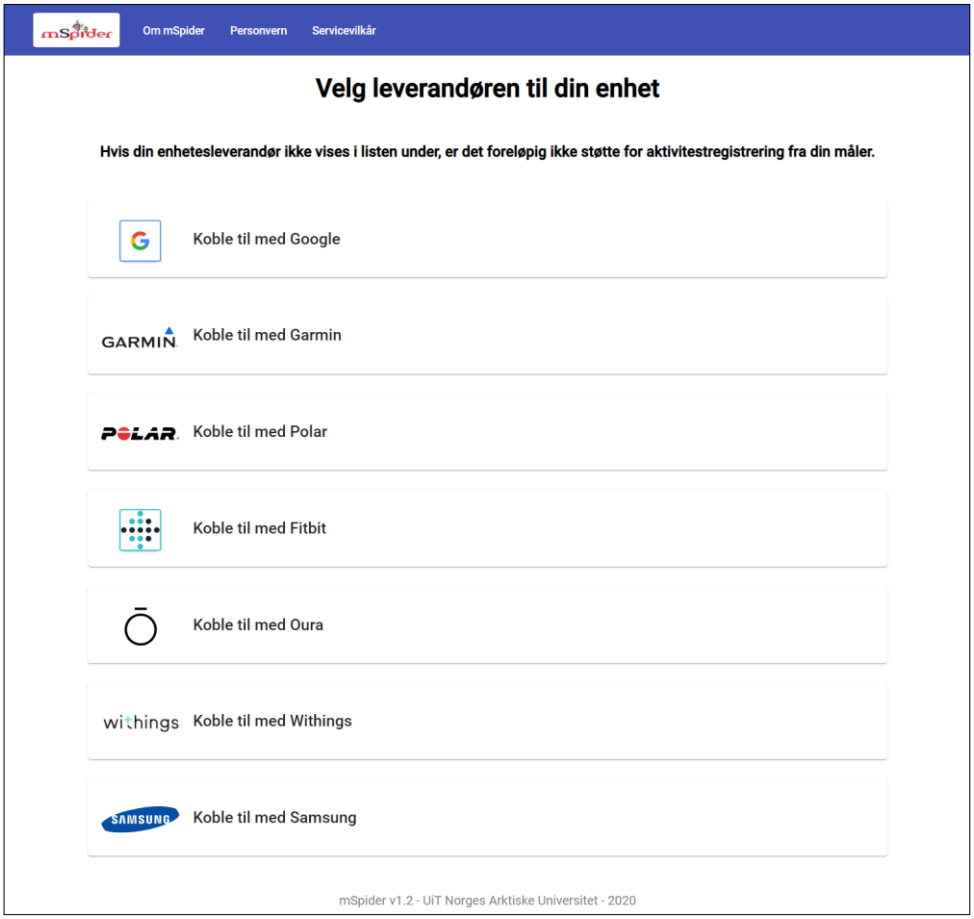


Figure 4. Provider selector web page (Paper IV)

The *server backend* handles communication with providers, stores participant authorization access information, and stores participant physical activity data. Data are stored in a database on the server. The web frontend does not have direct access to the database. All database communication is done through an interface on the mSpider backend server.

Most providers offer an API for accessing data. The *mobile application* is used for accessing physical activity data from providers that does not offer an API (e.g. Samsung and Apple). Participants using trackers from these providers, must install the mSpider mobile application to share data. This application is used by participants to give (authorize) mSpider access to their physical activity data and for automatic uploading this data to the mSpider server backend.

A detailed overview of how each provider is connected to the mSpider system is shown in Figure 5 (copied from Paper IV). Red dashed lines show lines of communication for each provider when a participant authorizes the mSpider system access to their physical activity data. Black (pull requests) and grey (push requests) lines between external systems (provider systems and mSpider mobile app) and the mSpider server backend shows lines of communication when transferring physical activity data from each system to the mSpider backend server for local storage.

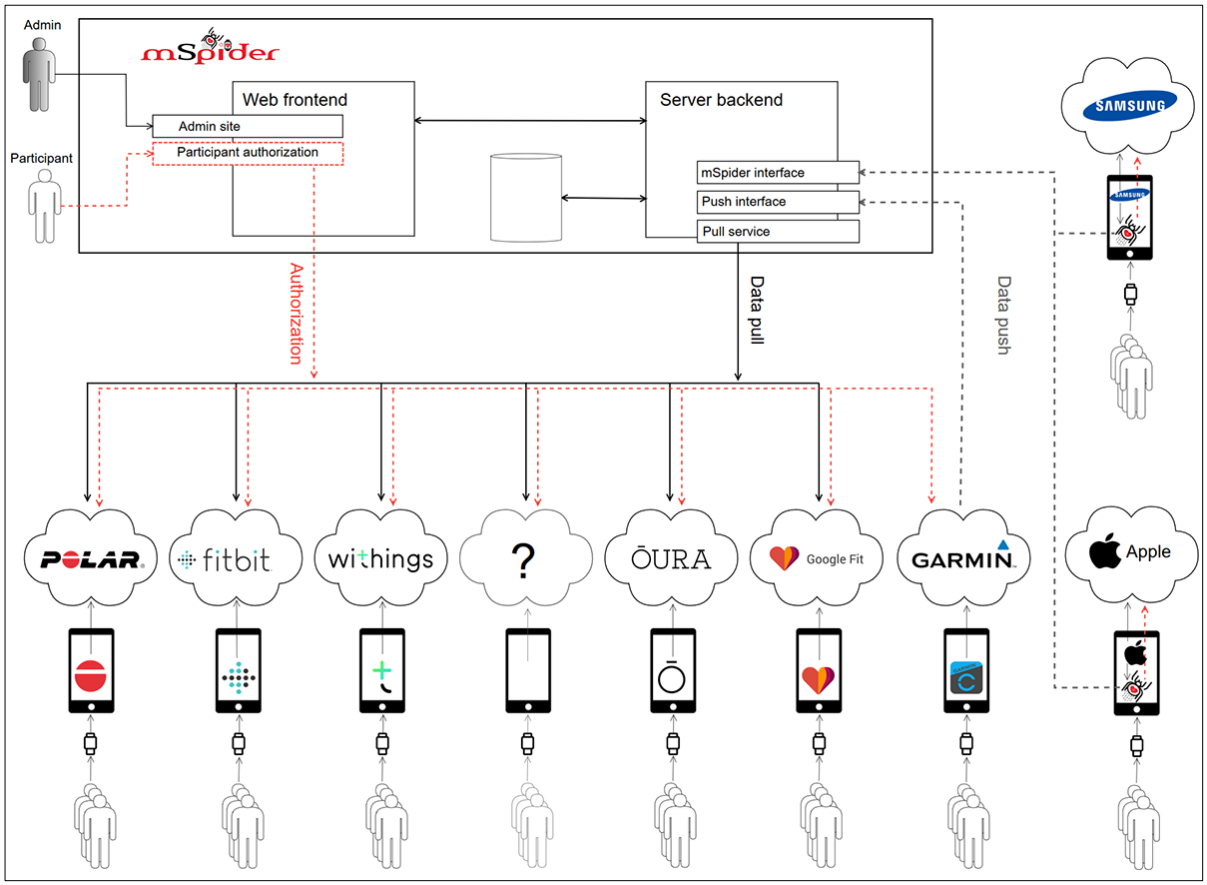


Figure 5. mSpider architectural overview (Paper IV)

Authorization

Participant data access is provided thru Open Authorization (OAuth). Open Authorization is a framework that allows users to authorize data sharing between different systems, without having to share their username or password. After authorization is granted, participant unique Tokens (identifier and secret) are exchanged. All communication related to a participant

contain these Tokens in order to identify the participant. If tokens are lost, the participant will have to re-authorize access.

Provider support

As mentioned, the current implementation supports activity trackers from Polar, Fitbit, Withings, Oura, Garmin, Samsung, Apple, as well as providers that store data in Google Fit or Apple Health open health clouds. We initially considered implementing support for Suunto activity trackers and activity tracking apps like Strava (using Strava V3 API). This was not attempted, but support for these providers (and others) should be considered before using this system for large population-based studies, in order to support all major providers.

Each provider offers different variables. An overview of accessed variables and how we defined valid days for each provider is given in Table 4 (copied from Paper IV).

Table 4. Accessed variables, by provider (Paper IV)

Provider	**Variables**	**Valid day calculation**
Apple	Steps, AEE, REE, sleep	Step>150
Fitbit	Steps, TEE, AEE, LPA, MPA, VPA, sleep	Step>150
Garmin	Steps, TEE, AEE, MPA, VPA	(Sleep + sedentary time + LPA + MPA + VPA) >10 hours
Google Fit	Steps, TEE	Step>150
Oura	Steps, TEE, AEE, sedentary time, LPA, MPA, VPA, non-wear time	Step>150
Polar	Steps, TEE, AEE, sedentary time, LPA, MPA, VPA, sleep	non-wear time<14 hours
Samsung	Steps, AEE, sleep	(Sleep + sedentary time + LPA + MPA + VPA) >10 hours
Withings	Steps, TEE, AEE, LPA, MPA, VPA, sleep	Step>150

AEE: activity energy expenditure, REE: resting energy expenditure, TEE: total energy expenditure, LPA: light physical activity, MPA: moderate physical activity, VPA: vigorous physical activity.

Technology

The mSpider system runs on an Ubuntu v18.04 (Canonical Ltd., London, UK) Linux server and was deployed using the Docker v19.03.6 (Docker Inc., CA, US) container engine. The web frontend was developed using NodeJS v10.22 (open source) and Angular v7.2 (Google Inc., CA, US) and runs using the Nginx v1.19.2 (Nginx Inc., CA, US) webserver. The backend server was written in the Go programming language v1.15 (Google Inc., CA, US) and runs using the GIN v1.3 (open source) web framework. Data are stored on a MongoDB v4.6.12 (MongoDB Inc., CA, US) server. The mobile application was developed using React v16.6.3 (Facebook Inc., CA, US) and React Native v0.57.8 (Facebook Inc., CA, US) for native cross-platform compiling. The Android app was compiled for (target version) Android v8.1 "Oreo" (API level 26) (Google Inc., CA, US), using Android Studio v3.3 (Google Inc., CA, US). The iOS app was compiled using XCode v10.2 (Apple Inc., CA, US).

3.5.2 Volunteers and study participants

Volunteers (development phase)

To test the system, we used convenience sampling to recruit volunteers throughout the development phase (2019-2020). Inclusion criteria were: 1) above 18 years of age, 2) willing to wear a provided activity tracker for several months, and 3) willing to share collected physical activity data. In this phase we recruited 35 participants who were supplied with an activity tracker from either Apple, Fitbit, Garmin, Huawei, Oura (ring), Polar, Samsung, or Withings. All participants received written and oral instructions and signed informed consent (See Appendix G).

Study participants (physical activity study)

To test the change in physical activity levels during the COVID-19 lockdown, we further recruited 130 participants who already owned an activity tracker in the lockdown period. Participants were recruited by online media advertisement in UiT local news and shared by regional and national online news sites (framtidinord.no, forskning.no, dagsavisen.no, and klikk.no), where we asked people to make contact by email if they were willing to participate. Inclusion criteria were: 1) owned an activity tracker from Garmin, Fitbit, Withings, or Oura, and 2) willing to share physical activity data prospectively and retrospectively. All

participants received written instructions (See Appendix H) and consented to participate by signing up for data sharing through their activity tracker provider.

Although the system also supports data extraction from users of activity trackers from Polar, Samsung, Apple, and trackers storing data in Google Fit, we did not include these providers when advertising for participants. Polar does not currently support historic data download. Samsung has temporary suspended accepting applications to enter their partner program, which is required to access data collected from Samsung activity trackers. Finally, we decided that the Apple and Google Fit integrations were not yet adequately tested to be included.

3.5.3 Data collection

For each participant we downloaded daily estimates for steps, AEE, light physical activity, moderate physical activity, vigorous physical activity, sedentary time, sleep duration, and non-wear time. Moderate physical activity and vigorous physical activity were further combined into a variable for MVPA. Light physical activity, sedentary time, sleep duration, and non-wear time were only used to for wear-time estimates. Data was collected for the period between January 1st, 2019, and December 31st, 2020.

Each day was tagged as valid only if the tracker was worn for at least 10 hours and the tracker had recorded at least 150 steps. The step threshold was used because some providers (see Table 4) do not share other variables for calculating wear time. The threshold was set by using the lowest step count among included participants who had at least 10 hours of wear time.

Finally, an online anonymous questionnaire was used for collecting participant self-reported data on age, height, weight, and sex.

3.5.4 Statistical analysis

Participants characteristics were presented as means, standard deviations, and range. Valid days were used to create means for daily steps, AEE, and MVPA, for each month of 2019 and 2020, as well as a separate yearly means for 2019 and 2020 (March to December). Two separate means were created for March 2020: 1st – 12th of March and 12th – 31st of March.

Depending on normality, we used a two-sided paired sample t-test or a two-sided paired Wilcoxon signed-tank test to identify periods of significant change. We compared the following periods:

1) 2019 (March-December) with 2020 (March-December)
2) March 2019 with March 1st – 12th 2020
3) March 2019 with March 13th – 31st 2020
4) April 2019 with April 2020, May 2019 with May 2020, etc.
5) March 2020, 1st – 12th with 13th – 31st

Bar plots were created to visualize differences between periods. R version 4.0.3 was used for statistical analysis.

3.6 Ethics and privacy considerations

The Regional Committee for Medical and Health Research Ethics in Northern Norway (REC North) reviewed the project description for Paper II (reference 557/2019) and Paper IV (reference 1014/2019 and reference 164780) and concluded that approval was not needed as the study did not fall under the provisions of the Health Research Act. The underlying feasibility study for Paper III, with ClinicalTrials.gov identifier NCT03807323, was approved by REC North (Reference 1100/2017). The Norwegian centre for research data (NSD) evaluated the data collection for Paper IV (reference 978318). All participants gave informed consent for in all studies.

4 Results – Summary of papers

4.1 Paper I – Analysis of consumer wrist-worn activity trackers

The aims of this study were to analyse how consumer-based activity tracker sensor support has evolved over time, identify activity tracker providers often used in research, and assess their relevance for future studies (Aim 1). We searched six activity tracker databases and visited provider websites to identify activity trackers. We further searched *Ovid MEDLINE* and *ClinicalTrials* to analyse provider usage in research.

We identified 132 different providers with a total of 423 unique activity tracker models, where the earliest models were released in 2011. Sensor support was low in 2011 for most sensors (except accelerometer) but increased steadily most years. In 2017, 71% of new activity trackers had a PPG, 50% had GPS support, 39% had a gyroscope, 34% had a magnetometer, and 32% had a barometer or altimeter.

The five most common providers used in research were Fitbit, Garmin, Misfit, Apple, and Polar. Fitbit was used in twice as many validation studies as any other providers, and Fitbit was registered 10 times as often in ClinicalTrials, compared to any other provider. Regarding developer possibilities we presented a matrix of attributes to consider when choosing provider and activity tracker. Except Apple, all providers supported both Android and iPhone smartphones. Most providers either supported upload to their own cloud storage or supported upload to Apple Health and/or Google Fit open cloud storages.

In conclusion, several new providers and activity trackers appear on the consumer market every year, with an increase in sensor support and alleged accuracy. However, tracker validation, health data cloud integration, and developer support vary, and researchers should carefully consider which activity tracker to use, depending on project needs. We provided a checklist of points to consider when making such decisions.

4.2 Paper II – Polar M430 validation study

The aim of this study was to test the validity of a recent activity tracker (Aim 2). Fifty participants wore a Polar M430, two ActiGraph (wrist, hip), and two Actiheart 4 on the chest, for one full day and night. We compared minutes of sedentary time, LPA, MPA, VPA, MVPA, TEE, AEE, and steps, between the Polar M430 and each reference monitor.

Pearson correlations between the Polar M430 and reference monitors ranged from moderate to very strong for vigorous physical activity (*r* 0.59-0.76), MVPA (*r* 0.51-0.75), steps (*r* 0.85-0.87), AEE (*r* 0.74-0.79), and TEE (*r* 0.88-0.94). For sedentary time correlations were moderate or weaker (*r 0.06-.052*), for LPA correlations were strong or weaker (*r* 0.02-0.70), and for moderate physical activity correlations were moderate or weaker (*r 0.34-0.57*). Bland-Altman plots showed that agreement was higher at higher intensities of physical activity. Only TEE showed acceptable or close to acceptable error when comparing the Polar M430 with the hip worn ActiGraph (three axes: 6.94%, one axis: 8.26%) and Actiheart (upper: 14.54%, lower: 14.37%). Steps had also close to acceptable error, but only for the wrist-worn ActiGraph (15.94%). Remaining MAPEs were 22% or higher. For most variables, Pearson correlations and ICC agreement were strongest when comparing the Polar M430 with the hip-worn triaxial ActiGraph: sedentary time (*r* 0.52, *icc* 0.10-0.51), LPA (*r* 0.70, *icc* 0.37-0.65), MPA (*r* 0.57, *icc* 0.18-0.52), VPA (*r* 0.76, *icc* 0.42-0.88), MVPA (*r* 0.75, *icc* 0.31-0.57), AEE (*r* 0.75, *icc* 0.53-0.87), TEE (*r* 0.91, *icc* 0.80-0.96), and steps (*r* 0.85, *icc* 0.49-0.75).

In conclusion, the Polar M430 has potential to be used as an additional source of physical activity data. It should not be used as a replacement for established research grade instruments, since only TEE has acceptable error and can be considered valid. It may however be suited for long-term monitoring for some variables.

4.3 Paper III – Succeeding with prolonged usage of activity trackers

The aim of this study was to identify important factors for increasing activity trackers usage for long-term monitoring among participants in health research (Aim 3). Sixteen middle-aged and elderly (55-74 years) participants from Tromsø 7 with obesity, sedentary lifestyle, and elevated cardiovascular risk, participating in a feasibility study and were equipped with a Polar M430 activity tracker for 12 months of physical activity monitoring. We used a mixed methods approach with two rounds of qualitative interviews and quantitative wear time- and validation analysis.

Mean number of valid days of recording was 292 (SD=86) for all participants over one full year of recording (i.e. 80%). In this cohort, the Polar M430 only provide close to acceptable estimates for TEE, with a moderate correlation (*r* 0.45, 95% CI 0.50-0.69), borderline acceptable MAPE (10.6%), and an under-reporting of mean 99 kcal/day. Some participants reported increased motivation to wear the activity tracker by being able to track progress, while others mainly wore the activity tracker because they were part of a study and were

asked to wear it. Activity tracker inaccuracy, limited usage training on the activity tracker, activity tracker complexity, and activity tracker appearance, were identified as important areas to address for increasing wear time. A list of recommendations to consider when using consumer-based activity trackers for long-term physical activity monitoring was also provided.

In conclusion, activity tracker wear time was high, and using an activity tracker for long-term physical activity recording was feasible in the present study. However, to achieve high wear time over a prolonged period, potential success factors includes providing satisfactory activity tracker training to participants, offer different activity tracker designs, and use activity trackers with accurate estimates.

4.4 Paper IV – Physical activity surveillance during COVID-19 pandemic

The aim of this study was to implement an automated system for collecting long-term physical activity data from consumer-based activity trackers, and to examine the usability of this system (Aim 4). We retrospectively accessed historic activity tracker data to assess change in physical activity levels due to the COVID-19 pandemic.

There was a significant reduction in daily steps and AEE (kcal) when comparing March 2019 with March 13th – 31st, 2020 (i.e. post lockdown date), with a mean reduction of 797 steps and 74 kcal per day. When comparing March 1st – 12th with March 13th – 31st, 2020, there was a significant reduction between the first and second half of March 2020, with a mean reduction of 913 steps and 85 kcal per day. Remaining step and AEE comparisons showed no significant reduction between periods, However, a a significant mean increase of 54 kcal for the September (2019 vs 2020) comparison was observed. No significant reduction was observed for MVPA, but several monthly comparisons after the lock down was lifter (i.e. May, September, October, and December) showed a significant increase in daily MVPA from 2019 to 2020. There was also an overall increase in daily MVPA in 2020 compared to 2019.

In conclusion, mSpider can be used as a tool for physical activity surveillance, by accessing historic and continuous daily physical activity data collected from participants using consumer-based activity trackers. Results showed only a temporary reduction in daily physical activity due to the COVID-19 lockdown.

5 Discussion

The discussion is divided into a methodological discussion and a discussion of main results from each paper.

5.1 Methodological discussion

In Paper I, II, and IV, we used quantitative methods for analysing data. For these papers, I will discuss potential bias. In Paper III, we used a *mixed methods* approach where quantitative wear time was analysed together with qualitative interviews. A discussion of this study design for this paper is given in a separate sub-chapter.

5.1.1 Validity

Validity, together with reliability, relates to the quality of a study. The validity of a method or test is related to its accuracy, and a valid method gives results that corresponds to the actual value of what was measured [161]. The reliability of a method or test relates to its consistency or precision, and a reliable method will give the same results when repeating a measurement under identical conditions [161].

Internal validity refers to which extent we can be confident that an outcome is trustworthy and not the results of the influence of other factors, such that findings are representative for the sample under study. Internal validity is challenged by confounding, random error (chance), and systematic error (bias) [161, 162].

A confounding factor is a factor that may affect both the exposure and the outcome under study [161, 163], i.e. other variables that may partly explain the observed outcome. We did not assess the relationship between variables in included studies, and confounding is therefore not addressed here. Systematic error, divided into information bias and selection bias, is discussed below.

External validity refers to which degree the study results can be generalized to other groups or populations [161], and is also discussed.

A note on terminology

In Paper II we referred to the research accelerometers (i.e. ActiGraph and Actiheart) as *criterion measures* and used the term *criterion validity* throughout the paper. However, a criterion measure implies a gold standard method. Although the accelerometers are validated against gold standard methods, they are not themselves gold standard methods (even though

the ActiGraph corporation refer to their devices as gold standards). Therefore, a more correct terminology would be *reference monitor* and *concurrent validity* (i.e. assessing validity using a method previous validated using a criterion measure).

5.1.2 Information bias

Information bias, a.k.a. measurement bias, occurs when the there is a systematic error when collecting, measuring, or handling information in a study [164]. We used different methods for collecting data in Paper I-IV. In Paper II and Paper III, we partly used the same method for collecting and analysing physical activity data for the validation studies. These will be discussed together. Remaining methods will be discussed for each paper separately.

Paper I – Analysis of consumer wrist-worn activity trackers

For Paper I we aimed to examine how activity tracker have change over time, in terms of sensor availability, usage in research, and provider usability in research (Aim 1). We did this by first searching for all activity trackers released on the consumer market that fit the inclusion criteria. Although we found a large list of trackers, this list was not exhaustive. If we had used more time, we would have found more devices. In the paper we addressed some limitations that may have caused potential error or misclassifications (e.g. supported sensors) in the collected data, but we did not identify any source of systematic error that would cause misclassification bias and thus affect the results.

However, one potential source of information bias could be that most sources we used to collect data from were English or Norwegian. Although Apple and Fitbit have been the two providers with most shipped devices for several years, two Chinese companies (i.e. Huawei and Xiaomi) are currently topping the list of most sold units [59, 79]. It is possible that we missed Chinese or other non-western devices popular in these regions in the 2017 search because we were limited to English and Norwegian sources.

For instance, Xiaomi, a Chinese company producing the Mi band activity tracker, is currently the second largest provider in terms of shipped units [59, 79]. Xiaomi has a Health Cloud developer platform, The platform is documented in Chinese (can be translated using Google translate) [165] and we did not identify this when doing research for Paper I. This and other platforms may therefore have been excluded. However, the final conclusions and implication for practice would likely not have changed.

Because we used several sources to collect data, partly with conflicting information, some misclassification likely exists, i.e. year and sensors support may be wrong for some devices in the final data set. Furthermore, device meta data were collected by the first author only. This was not confirmed by other authors, which would have reduced the chance of misclassification. However, there is no reason to suspect any systematic error caused by the method of collection, and any potential misclassifications would likely not affect the conclusions.

Paper II & Paper III – Polar M430 validation study

Misclassification bias

For Paper II we aimed to test the validity of the Polar M430 activity tracker (Aim 2). We also included a smaller validation study for this activity tracker in Paper III, because we wanted to assess the validity of that device in the study sample. In Paper II we use multiple reference monitors on multiple locations. In Paper III we used a hip-worn triaxial ActiGraph as reference monitor. In this section I will focus on systematic error for the reference monitor. Activity tracker misclassification will be addressed in relation to Paper IV.

Although accelerometers are objective instruments, researchers face many subjective considerations when using these instruments. Key methodological decisions include device placement, accelerometer sampling frequency (typically 30-100Hz), epoch length (i.e. time interval in seconds), non-wear algorithm, cut-point selection, valid day definition, and various signal filtering [25]. Placement and sampling frequency must be decided a priori, while remaining decisions can be finalized after data collection is completed. We used a sample rate of 100Hz (ActiGraph) and 32Hz (Actilife), which was the highest possible value. Both are adequate to catch all human movement [24].

A non-wear algorithm is used for detecting periods when the accelerometer is not worn, in order to exclude these time periods from analysis. The choice of non-wear time algorithm affects how minutes are classified. In Paper II we used the Hecht non-wear algorithm [136] as this was the algorithm available in QCAT. QCAT only support generating physical activity intensity variables and does not support generating variables for individual days of recording. For Paper III, where compared individual days of recording, we therefore used ActiLife to generate variables. ActiLife does not support Hecht and we used the Troiano non-wear algorithm [139] (using default settings). However, the choice of non-wear algorithm will

mostly affect the sedentary time variable. This was also shown in Multimedia Appendix 2 in Paper II where correlations between ActiLife (using default Troiano settings) and QCAT (using Hecht) showed very strong correlations for light-, moderate-, and vigorous PA, but strong (borderline moderate) correlation for sedentary time.

Before generating variables, all ActiGraph and Actiheart files were imported into the ActiLife the Actiheart software, respectfully. For Paper II, activity counts were further exported as CSV files, using 10 second (ActiGraph) and 15 second (Actiheart) epochs, and finally imported into QCAT.

For the *triaxial hip-worn ActiGraph* (Paper II and Paper III) we used activity cut points defined by Sasaki et al. [26] to classify minutes of physical activity intensity. In Paper II we further distinguished between light physical activity and sedentary time by using cut-points suggested by Kozey-Keadle et al. [130] and Peterson et al. [131]. Kozey-Keadle suggests using 150 counts per minute as a threshold for distinguishing between sedentary time and light physical activity. However, this is based on uniaxial (i.e. horizontal) data. The threshold of 150 counts was also suggested by Peterson et al. who used all three axes (i.e. vector magnitude). This study was conducted on university studies and may therefore not be a perfect fit for the samples in Paper II. A cut point of 100 counts per minute using only one axis is also common practice [25], which would have resulted in fewer minutes of sedentary time and more minutes of light physical activity. In Paper III we only compared the MVPA intensity variable, which should not be affected by the choice of cut-points for sedentary time, nor which non-wear algorithm we use.

For the *uniaxial hip-worn ActiGraph* (Paper II) we used activity cut-points defined by Freedson et al. [129] for classifying light-, moderate-, and vigorous physical activity. We used Matthews et al. [137] to further distinguish between light physical activity and sedentary time, classifying less than 100 counts/minute as sedentary time.

We used the same cut-points for the *wrist-worn ActiGraph* (Paper II). However, for the wrist-worn ActiGraph we applied the ActiLife “worn on wrist” option before exporting to CSV files. This option converts wrist-based counts to hip-count equivalents. In contrary to most options in ActiLife, this is not based on public scientific evidence, and the accuracy of this conversion is not proven. Some studies have tried to define cut-points for wrist-worn accelerometers, but there are no agreed upon wrist-based activity count cut points [25].

Regarding *Actiheart*, we used activity count cut-points suggested by Schrack et al. [138] to classify minutes of light-, moderate-, and vigorous physical activity, and minutes of sedentary time. Schrack et al. concluded cut-points based on a sample of 440 adults, aged 31-88 years. This is the only Actiheart activity cut point set we could identify for an adult sample. I have only identified one additional recent paper (preprint) where this cut points set was used, and that paper was co-written by Schrack [166]. The Actiheart software classifies each minute of wear time into different MET (metabolic equivalent of tasks) levels. We could alternatively have used these to classify minutes, using defined thresholds (i.e. sedentary time < 1.5 MET < LPA < 3 MET < MPA < 6 MET < VPA [6]). Using METs to classify minutes of intensity would likely have affected the output variables, which could affect the final results. Welk et al. [167] even suggest focusing on MET-minutes when validating activity trackers because it avoids the need to further categorise physical activity intensity using cut-points.

Energy expenditure was also calculated using ActiLife or Actiheart software. For ActiGraph we used ActiLife, using algorithms available in ActiLife. I.e. Freedson et al. [129] for uniaxial data and Sasaki et al. [26] for triaxial data. For Actiheart we used the Actiheart software which use a branching model where different aspects of energy expenditure is calculated based on both activity counts and heart rate. ActiGraph only report activity energy expenditure and Polar M430 only reported total energy expenditure. We therefore needed to convert between total energy expenditure and activity energy expenditure, by adding or subtracting resting energy expenditure and energy expended due to food consumption. Since the Actiheart software uses the Schofield equation [141], we used the same equation when converting ActiGraph- and Polar M430 variables. This prevents any further bias when classifying different types of energy expenditure, but it also shows the need for subjective decision making, as there are other equations available for calculating resting energy expenditure.

Further, we did not use individual calibration before placing the Actiheart (Paper II). Performing individual activity test to map hart rate at known effort levels improves accuracy of energy expenditure output, compared to using the group calibration provided by Actiheart [23]. This further adds to the potential classification bias.

In Paper II we found that wrist-worn ActiGraph reported more steps compared to the hip-worn ActiGraph. Further, The ActiGraph under-reported steps (regardless of ActiGraph placement) compared to the Polar M430 in both Paper II and Paper III. The Actiheart does not

report steps. ActiGraph step counts are internally calculated and reported directly. Different providers of accelerometers, and device placement, will report different output when performing the same activity [168]. The choice of reference monitor and placement will therefore affect the results.

When generating variables from these objective tools (i.e. accelerometers), many subjective choices must be made. Performing validation studies are therefore not straight forward, as each choice affects the output variables and may therefore affect the results. Although both the ActiGraph and the Actiheart have been previously validated in lab settings using gold standard methods, there are no gold standard for all variables of interest for studies conducted in free-living. Doubly labelled water is a gold standard for energy expenditure and is suitable for free-living studies. However, there are no equivalent gold standard method for measuring steps and physical activity intensity in free-living. Accelerometers have become the preferred tool for this setting and are non-the-less often referred to as a gold standard. It is important not to over-interpret results, as the choice of reference monitor and subsequent decisions for setup and analysis will affect the results and thus possibly the conclusions.

A major issue with activity tracker validation and comparison studies is that the large variability in how studies are conducted complicates comparisons. Welk et al. [167] have created five recommendations to alleviate this problem for future studies: 1) Use a diverse sample (sex, age, weight, height), 2) use appropriate protocol for daily behaviour (i.e. free-living or adequate simulation), 3) use appropriate criterion measure, 4) use standard protocols and wear location, and 5) include reference monitor and metrics.

In Paper II we included a diverse sample with wide range in age, weight, and height, as well as good gender balance. In Paper II this was not out aim, as the goal was to test the validity of the device in the study sample. In both Paper II and III we conducted the validation study under free-living, as recommended to capture natural daily living movement. In Paper II we used multiple reference monitors at multiple locations, but we did not use any gold standard methods for collecting energy expenditure, steps, or physical activity intensity. In Paper III we only used one reference monitor, placed at the standard location (i.e. hip).

The major limitation regarding standardization is that we did not use a gold standard method for energy expenditure calculations, as well as the overall lack of available (or practical) gold standard methods for measuring physical activity intensity and steps in free-living. The use of

reference monitors instead of gold standards criterion measures could further cause some degree of misclassification bias. In addition, a further improvement would have been to use MET for assessment instead of counts per minute.

We believe we took measures to reduce reference monitor misclassification by using suggested device setup strategies. However, because of the many subjective choices, our findings are affected by these and other studies making different choices will not result in the exact same results. From this I conclude that there may be some misclassification bias, especially for sedentary time which is affected by both non-wear algorithm and choice of cut point (Paper II), but not to the extent that it would greatly affect the final conclusions.

Statistics analysis

In Paper II we only recorded one day of measurement per participant. We calculated Pearson's product-moment correlation coefficient to compare association, as suggested by Düking et al. [169] when performing validation studies on activity trackers. Düking further states that Pearson's correlation is not sufficient, because it does not say anything about the level of agreement. We therefore calculated Intra-class correlations (ICC) and created Bland-Altman plots [143]. We further calculated mean absolute percentage error (MAPE) to assess measurement error between devices.

In Paper III we had up to 16 days of repeated measurements per participant, totalling 203 valid person-days. Because we had multiple measurements, we used repeated measures correlation, instead or calculating Pearson's correlations. We also created multiple measurement Bland-Altman plots (only limits of agreement included as table in Paper IV) and calculated the MAPE.

Comparison between studies are also challenging in terms of which statistics are used to infer conclusions. In a 2019 literature review, Welk et al. [167] found that although most studies reported "weak indicators such as correlation coefficients", only half included studies reported the MAPE summary statistics, and only one in four provided a test of agreement. Welk et al. [167] also made recommendations for how to standardize analytical methods in future validation studies. They defined three "essential" features: 1) report relevant metrics, 2) document error, and 3) focus on equivalence. In both Paper II and Paper III, we reported MAPE and Bland-Altman limits of agreements as suggested. Welk et al. [167] further suggest reporting mean percentage error (MPE), which differs from MAPE by not using absolute

values when calculating mean error. Mean percentage error (MPE) shows the direction of error for estimating error at the group-level, while MAPE shows the error for individual level estimation. We did include ICC to quantify agreement, but no test of equivalence was included.

We did not include all recommended analysis for easy comparison with future validation studies, as suggested by Welk et al. [167]. However, I believe conducted analyses together gives an appropriate image of the Polar M430 validity, albeit with the limitations already mentioned regarding reference monitor setup, usage, and variable creation (i.e. subjective decisions that will affect variable classification).

Paper IV – Physical activity surveillance during COVID-19 pandemic

Misclassification bias

For Paper IV we aimed to implement a solution for automatic physical activity monitoring using consumer-based activity trackers, and to test the usability of this system (Aim 4). We assessed usability by accessing data from participants who already owned an activity tracker.

We did not acquire information about participant's activity tracker model. Even if the activity tracker for each participant was known, most current activity trackers on the market have not been scrutinized by the research community, and the validity of most trackers currently available in stores are still unknown. Further, information about how activity trackers interpret (i.e. classifies) sensor data into physical activity, steps, energy expenditure, and other health related metrics are typically company secrets [63, 170]. A range of validation studies have also shown that when wearing different activity tracker models simultaneously on the same wrist, the output given by trackers differs enough to affect conclusion on which are considered valid (e.g. [82, 92, 171]). However, using multiple trackers on the same wrist is not in accordance with provider recommendations, which may partly explain this difference [172].

Combined, the lack of openness and lack of knowledge about accuracy makes it challenging to compare data between participants, as different activity tracker models classify sensor data differently. This difference also limits the ability to use this data to estimate absolute levels of physical activity in a groups or population.

The major challenge with the collected data is the lack of knowledge about activity tracker accuracy. Therefore, as newer activity trackers are constantly released to the consumer market, a need to find new ways to speed up validation studies emerges. There is also a need to come up with an agreed method for performing validation, as comparing existing validation studies are challenging due to large differences in study settings [167, 169].

However, for the purpose of Paper IV, we did not compare data between participants. We only compared participants with themselves between different periods. Therefore, although there are challenges with the collected data and misclassification exists, I conclude that this is not likely to affect the conclusion of this study.

Statistical analysis

In Paper IV we analysed up to two years of daily steps, AEE, and MVPA, per participant. We calculated a monthly average for each month; January 2019 to December 2020. Monthly averages were calculated from 66.274 daily measurement for each variable (i.e. steps, AEE, MVPA).

We used two-sided paired sample t-tests or Wilcoxon signed-rank test to test if there was a significant difference between each compared period (α=0.05). From the included 113 participants, each comparison was based on monthly averages from 76 to 107 participants, because participants acquired their activity tracker at different times and therefore did not contribute to the monthly average for all periods (some had no data for 2019). For the step and AEE comparisons, we also provided the mean difference per day, including the 95% confidence interval. Wilcox signed-rank test was used for MVPA-comparison, as this variable did not have a normal distribution. For MVPA we therefore provided the median difference per day between periods, as well as the inter-quartile range to show spread of values.

More sophisticated analysis could be conducted, but since the main aim of the analysis was to show that the proposed system (mSpider) could be used for detecting change in physical activity over time, we believe the selected analysis were appropriate. Further, we collected data anonymously and participant characteristics were not available. We could therefore not stratify by participant characteristics.

5.1.3 Selection bias

Selection bias occurs when there is a systematic difference between participants in a study and the population they are drawn from, i.e. they are not representative of the population [164]. Unless taken into consideration, selection bias can distort the findings and lead to errors when interpreting the results.

In Paper II (and a small part of Paper III) we performed a validation study where we compared instrument outputs. Although we recruited to increase ranges for height, weight, age, and sex, these characteristics would affect both instruments, and selection bias is therefore not a concern.

Paper III – Succeeding with prolonged usage of activity trackers

For Paper III, as part of a larger pilot and feasibility study, we randomly invited 75 people who had previously participated in the latest survey of the Tromsø Study (Tromsø 7) and were eligible for inclusion [149].

Regarding Tromsø 7 recruitment, everybody aged 40 and above living in the municipality of Tromsø were invited [173]. Attendance was 65% across all age groups, and for the 55-75 age groups, attendance was 71.7% [173]. These attendance rates are relatively high, as participation in epidemiological studies has declined worldwide the last decades [56]. In a study design article for Tromsø 6, a previous survey of the Tromsø Study (2007-08) with 65.7% attendance, the attendance rate was discussed in detail. They concluded a somewhat higher education levels among responders compared to the whole population of Tromsø, but no other major differences were reported [173]. There is no similar study for Tromsø 7, but results would likely be similar for Tromsø 7, as these studies are repeated cross-sectional studies in the same population. Tromsø 7 emerges as representative for the population and is therefore a good source of recruitment for the pilot and feasibility study.

The random selection of participants in the underlying pilot and feasibility study is a strength of the study as randomization is the best method for reducing selection bias [174]. Recruiting participants for clinical trials is challenging, especially when recruiting older adults for physical activity intervention studies [175, 176]. Of the 75 invited to the pilot and feasibility study, 20 (27%) responded and agreed to participate. Four people were excluded, resulting in 16 (21%) included participants. The resulting low response rate may have introduced a difference between those who responded and those who did not respond. We speculate that

those who responded to the invitation were likely people who already had thought about addressing their sedentary lifestyle and saw study participation as a chance to break bad habits. Because of this, included participants could be more inclined to adhere to the research protocol. For the underlying study, where the aim was to test the feasibility of a complex lifestyle intervention [149], this difference could potentially affect study outcomes because responders could be more motivated to make a lifestyle change compared to non-responders.

However, in Paper III, where we used a mixed methods approach, only the wear time analysis (and validation study which is already addressed) was analysed quantitatively. Since participants were part of an intervention study with close follow-up, the design of the study is more likely to affect adherence to the wear protocol than the characteristics of included participants. Selection bias is therefore not a major concern for Paper III.

Paper IV – Physical activity surveillance during COVID-19 pandemic

For Paper IV we recruited volunteers to test the data sharing routines during development of mSpider. We used convenience sampling to recruit volunteers. We did not analyse activity data from volunteers and did not attempt to recruit a representative sample from the population.

We further recruited participants who already owned and wore an activity tracker before-, during-, and after the Norwegian COVID-19 lockdown in March 2020. After initial recruitment, 130 volunteered to participate. After sending out invitations and instructions (and two reminder), 113 participants responded by registering. We do not know where in Norway participants live nor their characteristics. We only know the mean height, weight, sex, and age, as this was collected using an anonymous online questionnaire. Recruitment was first conducted in local online media (Tromsø) and was later picked up by regional and national online news outlets. The largest portion of participants volunteered in the earlier phase of recruitment, before it was shared nationally. It is therefore likely that most participants lived in (or read news from) the northern part of Norway. This can affect physical activity levels because seasonal weather variations affects outdoor activity negatively during winter weather [177]. However, the northern part of Norway experience heavy (some places record breaking) snow fall in the 2019/2020 winter season [178], potentially resulting in increased physical activity due to snow removal.

People who own and use an activity tracker are likely to be more physically active compared to those that do not own an activity tracker [179, 180]. There is also some evidence that people who voluntaries to participate in physical activity studies are likely to be more physically active (i.e. volunteer bias) [181, 182]. Finally, with 113 participants we cannot assume that we have a representative sample. The included sample is therefore likely more activity than the general population, which could potentially affect how active they become due to the COVID-19 lockdown. A study on Canadians residents showed that inactive participants were more likely to reduce their level of activity, compared to active participants, during the first period of nationwide restrictions in March 2020 [183].

However, since the aim of this recruitment was mainly to test the usability of the mSpider system (Aim 4), we did not aim to recruit a representative sample. In addition, since earlier research using both self-reported [183, 184] and objective data [185] have shown that national lockdowns causes temporary change in physical activity, we expected to see a change also when using objective measures on people who are more physically activity than the average population.

5.1.4 Study design – a mixed methods approach

Paper III – Succeeding with prolonged usage of activity trackers

For Paper III we aimed to identify factors that can assist in increasing participant wear-time of consumer-based activity trackers for long-term physical activity monitoring in health research (Aim 3).

Participant characteristics and results from the validation study were analysed statistically and presented separately. Results from the wear time results were analysed together with responses given during interviews. This allowed a more detailed understanding of why some participants chose to stop wearing the activity tracker after the intervention period. Further, the researcher perspective allowed a more detailed understanding of what worked and what did not work when designing a study for long-term physical activity monitoring using consumer-based activity trackers. This mixed-methods approach allows deeper insights into potential success factors as viewed from multiple angels.

In qualitative research, the required number of participants is depended on multiple factors, including study scope, prior knowledge, and data quality [186]. The included 16 participants represented 21% of invited and there may thus be a difference between included and not

included participants. Further, although we achieved rich and thorough descriptions from the interviews, the interview guide covered multiple themes. For Paper III we only included seven questions related to activity tracker usage. A more focused interview could potentially give more diverse responses.

The trustworthiness of qualitative research can be viewed in terms of credibility, dependability, and transferability [187]. The *credibility* of a study deals with the study focus, in terms of participant recruitment, data gathering, and analytical approach [187]. Participants were recruited from a population study where a representative sample of the population were included. An interview guide was created and used, allowing a focused interview with room for relevant side discussion. Further, we used an iterative approach when analysing responses in order to find meaningful themes. Decisions on final themes was achieved through discussion between authors. These measures strengthened the credibility of the study.

The *dependability* of a study deals with undesirable change in the data over time, due to the phenomenon being studies or alteration in study design [187]. To strengthen dependability, we aimed for a strong focus between phases, where we interview each participant and conducted transcription and analysis within a limited period.

The *transferability* of a study deals with whether findings are applicable in other settings or groups [187]. Participants were part of an intervention, potentially causing increased activity tracker wear time due to desirability bias. Further, since included participants potentially were more motivated than non-responders, results from interviews may have been affected. These elements may limit the transferability of results to other study designs.

5.1.5 External validity

External validity, i.e. transferability, refers to whether study findings can be applied to the source population or other populations.

The sample used in Paper III was representative for the underlying feasibility study. However, participants were older than the average population in Norway. A younger sample could potentially give other responses regarding success factors for prolonged usage. Similarly, all participants were Norwegian, and cultural difference could also potentially affect responses. These differences should be considered when applying findings from Paper III to a younger population in Norway or populations unsimilar from the Norwegian population.

Further, the study design of Paper III was a physical activity intervention with close follow-up of participants. The design is likely to affect who agrees to participate, as a physical activity intervention requires higher motivation compared to observational studies, and participants may be more inclined to adhere to the study protocol (i.e. increasing activity tracker wear time). This can also limit transferability of findings to other study designs.

As the overall aim of this dissertation was to find new methods for collecting physical activity data in future epidemiological studies, an important question related to external validity is whether this system can be used in other groups or populations.

As previously discussed, people who use an activity tracker [179, 180] and people who volunteers to participates in a physical activity study [181, 182] are likely to be more physically active compared to others. Despite this likely higher physical activity level, our findings clearly showed a temporary physical activity reduction in the first period of the lockdown. It is therefore likely that the observed reduction in physical activity also apply to the general population of Norway.

In addition, the Norwegian lockdown was less strict compared to several other European countries, and no national curfew was ordered. Although self-isolation and social distancing were encouraged, people were allowed outdoors e.g. to exercise. It is therefore likely that the observed reduction in activity levels would be more dominant in countries where stricter interventions were instigated.

Another main issue that may affect the external validity of using the proposed system for physical activity surveillance, is that the current prevalence of activity tracker ownership is not equal in all countries. Although we did not assess absolute activity levels, only change between periods, this issue is important to consider if the aim is to assess the current level of physical activity in a population.

5.2 Discussion of main results

5.2.1 Paper I – Analysis of consumer wrist-worn activity trackers

In Paper I we identified many activity tracker providers and models, where only a limited number of providers are repeatedly used for health research purposes. Because the activity tracker marked is rapidly changing with newer and improved models released every year, we also provided a list of criteria to consider when deciding which provider and model to use for

physical activity tracking in research. I will discuss these criteria below when discussing recommendation given in Paper III.

The discussion in Paper I was focused on the current state of activity trackers, in terms of provider popularity, sensors support, as well as research relevance and implication for practice. Here I will further discuss how sensor support may change over time and how this may potentially benefit future health research.

The current and future activity tracker landscape

Despite challenges in heart rate accuracy [97], especially when performing high intensity physical activity [97, 99, 100], we found PPG to be the most common sensor in wrist-worn accelerometer-based activity trackers released in 2015-2017. In 2017, more than 70% of released devices packed a PPG.

Current sensors are mostly used for detecting various metrics for physical activity, heart rate, and sleep. However, sensor technology is rapidly improving, and it is likely that more sensors will become common in future trackers. For instance, in 2018 the Apple Watch 4 was the first to release an activity tracker with a built-in electrocardiogram (ECG) which can be used for self-diagnosis of atrial fibrillation. Saghir et al. [188] found the Apple Watch 4 to produce accurate ECGs in healthy adults. However, Seshadri et al. [189] advices caution when using the watch to monitor patient with cardiac arrhythmias. Other challenges includes high price, false positives and false negatives in atrial fibrillation detection, privacy and security considerations, and that it uses a single lead system [190].

As of December 2020, there are eight activity trackers on the market with ECG support: Apple Watch 4 (2018), Apple Watch 5 (2019), Withings Move ECG (2019), Samsung Galaxy Watch Active 2 (2019), Amazfit Smartwatch (2019), Samsung Galaxy Watch 3 (2020), Fitbit Sense (2020), and Withings ScanWatch (2020). Atrial fibrillation prevalence and incidence are increasing [191], and this sensor can be used to improve patient experience and self-management [190]. This is likely to result in more activity trackers with ECG support as providers are incentivised by customer needs. Despite current challenges, there is great potential for future research in screening-, management-, and evaluation of atrial fibrillation [190].

We are also seeing a few activity trackers that includes a thermometer. The Oura-ring (2018) and the Fitbit Sense (2020) are two of few activity trackers that includes a thermometer and measures body temperature from the skin. Since change in body temperature is an important health indicator, this sensor can further assist in activity tracker enabled health assessment. For instance, long-term temperature monitoring, using sensors embedded in socks, has been identified as a promising solution for management of foot ulcers [192]. Similarly, skin temperature, together with other sensor data, can also be used to assess psychological stress and emotions [193].

However, although current and future trackers include temperature sensors and other new sensors, data from these sensors are not necessarily directly accessible and may only be used internally by the device to infer other outputs. For instance, the Oura-ring does not show current temperature, but rather the change in temperature between days. Regardless, both current and new sensors will continue to be included in future activity trackers as they become more accurate, smaller in size, and cheaper to produce, which will likely provide new and interesting opportunities for research and private health monitoring alike.

Update needed

The two major sources of activity tracker data in Paper I were the Vandrico database [121] and the offline Queens Wearable Device Inventory [127]. The Queens Wearable Device Inventory has since been published online. Neither sources have been updated since the initial search for Paper I in 2017. The other web-based sources are similarly outdated (except company web sites), where *GsmArena.com* seems to now focus solely on mobile phones, and neither *Wearables.com* nor *SpecBucket.com* are regularly updated. I found no other openly available activity tracker databases, and a follow-up paper may be timely.

5.2.2 Paper II – Polar M430 validation study

In Paper II, we discussed the validity of a specific activity tracker (i.e. Polar M430) and how it compared to previous Polar models. In this section, I will further discuss potential challenges with the current way validation studies are conducted.

In chapter 1.5 I addressed the current state of validity of activity trackers. I identified four systematic reviews that included Polar devices [83, 85, 89, 93]. Summarized, step counting is more accurate than energy expenditure and physical activity intensity estimates [93], and

Polar generally overestimated energy expenditure and underestimated steps [83]. However, there is great variation. All studies included in these systematic reviews were conducted on discontinued devices.

I further identified three validation studies conducted on the Polar Vantage, the only activity tracker included in validation studies that is currently (December 2020) available in the Polar store. Findings for this tracker is similar to earlier findings, where estimates for steps and energy expenditure is moderate at best [94, 95, 96, 128]. I identified only one study assessing the reliability of Polar Vantage heart rate estimates. This study concluded higher precision when performing low and high intensity training, compared to moderate intensity training [194].

As earlier addressed, using consumer-based activity trackers as a source of physical activity estimates has several limitations, which should be considered when discussing the results of a validation study. One of the main challenges are related to classification bias when selecting and using reference monitors to test the accuracy of an activity tracker. We showed in Paper II that the choice of reference monitor and its placement affected correlation and agreement with the Polar M430. I also discussed above how the choice of cut-points (and other options) would similarly affect the results of such studies. Therefore, due to the large difference between studies in how data are collected and analysed, it is challenging to compare result [172].

A second challenge is the short time activity trackers seems to have on the consumer market before they are replaced with newer models [170]. Some models are even replaced annually. This is a challenge, because even if new studies were conducted immediately after a device is released, the rigorous process of collecting and analysing data, and writing and publishing a paper takes time. By the time the paper is published, the vendors may already have released an updated model, and may have stopped producing the replaced model. The use of pre-prints can assist in reducing the time it takes to publish initial results. Conversely, Weisberg et al. [172] have suggested that as long as older activity trackers are supported by the provider, they could be more desirable to researchers as they have a lower cost and published validation studies are more likely to exist. Related, a further challenge is that most current activity trackers outputs (i.e. variables) have been studied only once, which is a challenge since tracker results should be confirmed by multiple studies. Combined, these challenges complicate decisions when choosing a device for data collection.

When using an activity tracker for the underlying study of Paper III, we chose to assess the accuracy of the Polar M430 in the included sample. At the time, no other Polar M430 studies had been published. Although Paper II was the first published physical activity validation study on Polar M430, data collection for Paper II was conducted after activity tracker validation data were collected for Paper III. There are now several other studies on this tracker available, showing similar results for steps [195] (abstract) and energy expenditure [196]. There is also a Master's thesis available assessing the validity of reported maximum oxygen consumption [197].

Comparing results from Paper II and Paper III, using similar methods, showed that results were not identical. For the older and less physically activity sample in Paper III, correlations and agreements were weaker compared to the sample used in Paper II. This highlights the need to conduct validation studies in the included sample, to know how accurate each variable is, before using consumer-based activity trackers to estimate physical activity.

A concluding remark is that the accuracy of consumer-based activity trackers can be challenging to assess. Therefore, greater effort should be put into conducting validation studies as soon as devices are released, studies should be conducting using comparable methods (for instance as suggested by Welk et al. [167] and Düking et al. [169]), results should be published in pre-prints, and researchers should conduct validation studies in the target sample when planning to use activity tracker for physical activity monitoring.

5.2.3 Paper III – Succeeding with prolonged usage of activity trackers

In Paper III, we discussed and provided recommendations for what to consider when planning to use consumer-based activity trackers for long-term physical activity recording. In this section, I will briefly discuss how the included validation study was affected by the sample, as well as how suggested recommendations from Paper I and Paper III compares to the current scientific literature.

Validation study

In Paper III we presented results from the validation study in a table. Although not included in the paper, Bland-Altman plots (given in Appendix I), using separate colour for each participant, showed that one participant contributed several outliers. I conducted a sub-analysis where I removed each participant one-by-one. When one specific participant was

removed, the effect size for steps changed distinctively, increasing correlations from 0.63 to 0.82, reducing MAPE from 120% to 75%, and reduced the width of the Bland-Altman limits of agreement from 17.300 steps to 9600 steps. This change in effect size was not observed for MVPA or total energy expenditure.

Although we do not know how wrist movement is interpretated by the activity tracker, increased wrist movement may be interpreted as steps. I believe this participant mentioned in passing that she was an active knitter. The increased wrist movement caused by knitting may thus register as steps by the Polar M430, resulting in misclassification and higher level of steps registered than what was actually performed. In a larger sample, knitting may not affect the effect size to the same extent. This is merely a speculation, but in addition to showing how a small sample size (even when using multiple measurements) can affect results, it also shows how activity tracker algorithms can be affected by various type of movement. Since internal algorithms for variable generation are generally company secrets [170], and may change over time, investigating to which extent specific movements affects output would be challenging.

Recommendations

In Paper III we also presented recommendations to consider when planning and executing studies where participants are asked to wear an activity tracker for an extended period. These recommendations expand on recommendations from Paper I, where we suggested criteria to consider when choosing activity tracker provider and model.

Recommendations from Paper I focused solely on considerations when selecting provider and model. Recommendations in Paper III focused on factors for increasing wear-time among study participants.

In an earlier paper, Cadmus-Bertram [170] provided a list of considerations when using consumer-based activity trackers as an intervention tool or when assessing physical activity. There were four areas of focus: 1) when is it appropriate to use activity trackers, 2) choosing provider and model, 3) ensuring wear compliance, and 4) how to extract and use data. The main conclusion of Cadmus-Bertram regarding provider and model was that a balance of features, usability, and cost is usually adequate to find the most appropriate tracker [170]. Although recommendations from Paper I were more detailed, this conclusion from Cadmus-Bertram is to the point and captures the essence of our findings. Wear compliance was also addressed, suggesting providing training on the activity tracker and handing out the user

manual, and to clearly communicate usage expectations and provide tracker support. Recommendations from Paper III are in accordance with these recommendations.

Turner-McGrievy et al. [198] provided a similar list of considerations, based on experiences from four studies. There is extensive overlap between their recommendations and recommendations by Cadmus-Bertram, as well as recommendations given in Paper I and Paper III.

Weisberg et al. [172] built on our recommendations in Paper I, Cadmus-Bertram [170], and Turner-McGrievy [198], and provided additional factors to consider when conducting research on older adults, grouped into key factors concerning 1) outcome measures, 2) protocol considerations, and 3) activity tracker features. Main overlapping recommendations from Weisberg et al. and Paper III are to perform validation studies of the selected tracker in the selected sample, give adequate training on the selected tracker, and recommendations related to tracker accuracy and usability.

Recommendations in Paper I and III fits well with other recommendations and may assist future researchers when planning to use activity tracker for physical activity monitoring in future research.

5.2.4 Paper IV – Physical activity surveillance during COVID-19 pandemic

In Paper IV, we discussed how our COVID-19 related physical activity findings compared to other studies. We also discussed how the mSpider system could be used as a method for long-term physical activity monitoring in epidemiological research. In this section, I will further discuss challenges and opportunities when using this method for long-term physical activity monitoring.

There are two ways to collect consumer-based activity tracker data from participants, 1) provide participants with an activity tracker, and 2) collected data from activity trackers already owned by participants. In Paper III, where we provided participants with an activity tracker, we believe that a major contributor to the high wear time (80%) was the study design (i.e. intervention study) with close follow-up of participants. Providing participants with an activity tracker for population-based studies (ignoring the monetary costs), where there is little to no follow-up, would present challenges which must be further addressed.

In Paper IV, where we collected data from participant's own activity trackers, wear time was higher, with an average of 94% of days contained data. Therefore, in relation to non-wear, this option stands out as the better choice for population-based settings. However, the wrist-worn activity tracker penetration in the consumer market may not be high enough to be able to recruit a representative sample.

Using data from two nationally representative surveys conducted in the United Kingdom in 2016 and 2018, Strain et al. [199] analysed smartphone and activity tracker ownership. Results showed high smartphone prevalence (79%), but lower activity tracker prevalence (14%). Further, higher age and lower social economic status were inversely correlated with both outcomes. They concluded that using activity trackers as the only source of physical activity data was premature. Similar results were observed in a 2017 study on US adults, where Omura et al. [200] concluded that activity tracker users are not representative of the population.

However, smartphones can also collect physical activity data, using internal accelerometers and other sensors. The prevalence of people owning a smartphone is higher, compared to wrist-worn activity trackers, and the method for third party data access is identical as for wrist-worn activity trackers. For instance, the mSpider tool can download physical activity estimates generated from smartphone sensor data, if these estimates are stored in Google Fit. Smartphones can therefore also be a potential source of physical activity data. Certainly, this presents other challenges, including lower wear-time.

The proposed system contributes towards finding a solution for the existing gap between traditional physical activity questionnaires, which offer long-term but self-reported and inaccurate estimates, and accelerometers, which offer objective and more accurate estimates, but only for short-term monitoring. Another benefit of implementing a solution like mSpider, compared to using existing alternatives using similar technology, is having full control of which providers to support and which variables to access from the large number of available data types.

Although this dissertation has focused on activity tracker technology, and how we can access data collected from such devices, consumer-based trackers are not presently suited to be the only source of physical activity in all settings. It is non-the-less an interesting additional data source. Some studies have also shown that people may be willing to share this type of data

[201], but they may also expect something in return, e.g. access to tailored recommendations based on their shared data [202, 203]. Therefore, although the technology may be available, work is still needed to motivate participant to share data and increase wear time.

6 Conclusion

Activity trackers has triggered broad research interest as there is potential in using these sensors as a source of digital biomarkers and biofeedback that can be used in a range of scientific fields, including physical activity interventions in healthy participants [204] and in patients with chronic disease [205], oncology [206], influenza outbreak surveillance [114], post-surgery physical activity intervention and monitoring [207, 208], and a range of other areas.

6.1 Conclusions and implication for practice

The technology to collect long-term physical activity data from a large number of participants using activity trackers currently exists. Recorded data are uploaded to provider clouds and can be accessed directly from the cloud or through the provider mobile application installed on participant's smartphones.

The main aim of this dissertation was to explore and develop new methods for collecting data on physical activity from participants in future epidemiological studies using activity trackers.

Towards this aim, we first mapped activity trackers on the consumer marked and assessed how the most popular providers could be used to collect long-term physical activity data in future research and provided recommendations for what to consider before choosing a provider and model.

The large number of different activity trackers available on the consumer market constitutes a challenge, because accuracy has only been assessed for a limited number of currently available models. In studies where participants wear a limited number of models, a validation study can be performed a priori, providing researchers with necessary knowledge about the validity of the tracker they use. In studies were participants share physical activity data from activity trackers they already own, the large heterogeneity in device models, offering different metrics, and with limited knowledge about device accuracy, makes it challenging to compare collected data. To contribute to alleviating this challenge, we assessed the validity of the Polar M430, and concluded that this tracker was more suited for detecting change in physical activity over time, than as an exclusive tool for physical activity assessment in a population.

An additional challenge for long-term monitoring is wear compliance, especially when participants are provided with an activity tracker. Towards successful implementation of such

study designs, we provided a list of recommendations to consider when planning and executing a study. Potential success factors especially centred around allowing participants to choose between different tracker modes with accurate measurements and to provide adequate training on tracker usage.

Finally, to test the usability of using consumer-based activity trackers as a source of long-term physical activity data, we implemented mSpider and successfully used this tool to collect historic physical activity data among participants owning a tracker before-, during-, and after the Norwegian 2020 COVID-19 lockdown. The number of people in a population currently owning an activity tracker may be rising, but this group may not currently be large enough to constitute a representative sample of the population. Similarly, people owning activity trackers are not distributed evenly among age groups, genders, and social economic statuses. Although we were able to show a clear but temporary reduction in physical activity levels among participant, this reduction is not necessarily representative for the whole population.

We have shown that the suggested method can be a valuable addition to existing methods for physical activity assessment. However, all methods have their strength and limitations. The major strength of the proposed method is the possibility for long-term monitoring. The addressed challenges should be considered before using this method. Further, this method should not be considered an alternative to existing methods, but rather an additional source of physical activity data that can contribute to closing the gap between current methods of physical activity assessment, with especial relevance for detecting change in physical activity over time.

6.2 Further perspectives

Improve support for existing providers

The current implementation for Apple and Samsung requires the participant to install a mobile application in order to share physical activity data. This is a limitation with these providers and cannot be resolved unless these providers decide to offer an API. We chose to focus on API based providers in this project, since an API (i.e. access data from the cloud) based approach required less participants burden compared to an SDK (i.e. install mobile application) based approach. However, in order to investigate challenges and estimate needed effort, we developed a prototype mobile application for Apple and Samsung.

The Samsung mSpider application was installed on a few highly motivated volunteers who owned an Android smartphone and were provided with a Samsung Galaxy watch. In order to successfully share physical activity data using an application that is not released in Google Play (i.e. under development), too many steps are required for it to be a feasible solution for most people.

We did not fully test the Apple mSpider app, but the process for installing and sharing data is similarly cumbersome. Both the Samsung and Apple mSpider apps must be improved, thoroughly tested, and put into Google Play/App Store before they can be fully integrated into the mSpider system. Apple and Samsung combined worldwide smartwatch market share was 49% in the first quarter of 2020 [60], showing the importance of also supporting these providers in future data collections.

Adding support for additional providers

In addition to finalizing Apple and Samsung application development, some additional identified provides have open APIs and should be supported in order to allow more participants to be included when using this method for physical activity monitoring:

- Misfit Cloud API (https://build.misfit.com)
- Suunto Cloud API (https://apizone.suunto.com)
- Strava v3 API (https://developers.strava.com)
- Under Armour API (https://developer.underarmour.com)

Misfit and Suunto are providers of hardware-based activity tracers, while Strava and Under Armour are software-based mobile apps for physical activity tracking.

Privacy

In this project we did not focus especially on issues related to privacy and security. This is a large field and requires special attention. In order to comply to GDPR regulations and ensure data privacy, and because of limited time, we chose to collect data for Paper IV anonymously. After participants had initiated data sharing between their provider and the mSpider tool, and the relevant physical activity data were downloaded, we removed identifying information from the server database. We also deleted e-mail communication to participants where

identifying information (including de-identified information) were included. It is thus not possible to identify the participants in the data.

We are currently working on a risk analysis for the mSpider tool, with the aim of being able to store de-identified data within the system. We have plans to use this system for long-term physical activity monitoring in several future studies, where we need to be able to identify participants.

Disease outbreak surveillance

In this dissertation I have mostly focused on activity trackers as an additional tool for collecting physical activity data in health research. As shown in Paper IV, these trackers also have the potential to detect sudden changes in physical activity behaviour in a population. An increasing number of trackers include pulse sensors and some newer trackers have also started to include thermometers. This combination of sensors could thus potentially be used to detect clusters of disease outbreaks (e.g. influence) since reduction in physical activity with increase heart rate and increased temperature is an indication of infection [114].

References

1. Caspersen CJ, Powell KE, Christenson GM, *Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research.* Public Health Rep, 1985. 100(2): p. 126-31.

2. World Health Organization. *Fact sheet: Physical activity*. 2020. [Available from: https://www.who.int/news-room/fact-sheets/detail/physical-activity, Accessed: 2020-11-31].

3. Pettee Gabriel KK, Morrow JR, Jr., Woolsey AL, *Framework for physical activity as a complex and multidimensional behavior.* J Phys Act Health, 2012. 9 Suppl 1: p. S11-8. doi: 10.1123/jpah.9.s1.s11.

4. U.S. Department of Health and Human Services. Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC; 2018.

5. Kelly P, Fitzsimons C, Baker G, *Should we reframe how we think about physical activity and sedentary behaviour measurement? Validity and reliability reconsidered.* Int J Behav Nutr Phys Act, 2016. 13: p. 32. doi: 10.1186/s12966-016-0351-4.

6. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD, Bauman A, *Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association.* Med Sci Sports Exerc, 2007. 39(8): p. 1423-34. doi: 10.1249/mss.0b013e3180616b27.

7. Tudor-Locke C, Hatano Y, Pangrazi RP, Kang M, *Revisiting "how many steps are enough?".* Med Sci Sports Exerc, 2008. 40(7 Suppl): p. S537-43. doi: 10.1249/MSS.0b013e31817c7133.

8. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, Hatano Y, Inoue S, Matsudo SM, Mutrie N, Oppert JM, Rowe DA, Schmidt MD, Schofield GM, Spence JC, Teixeira PJ, Tully MA, Blair SN, *How many steps/day are enough? For adults.* Int J Behav Nutr Phys Act, 2011. 8: p. 79. doi: 10.1186/1479-5868-8-79.

9. World Health Organization. WHO Guidelines on physical activity and sedentary behaviour. Geneva: WHO Press; 2020.

10. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, Whincup P, Diaz KM, Hooker SP, Chernofsky A, Larson MG, Spartano N, Vasan RS, Dohrn IM, Hagstromer M, Edwardson C, Yates T, Shiroma E, Anderssen SA, Lee IM, *Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis.* BMJ, 2019. 366: p. l4570. doi: 10.1136/bmj.l4570.

11. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working G, *Global physical activity levels: surveillance progress, pitfalls, and prospects.* Lancet, 2012. 380(9838): p. 247-57. doi: 10.1016/S0140-6736(12)60646-1.

12. World Health Organization. Global status report on noncommunicable diseases 2014. 2017.

13. Guthold R, Stevens GA, Riley LM, Bull FC, *Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants.* Lancet Glob Health, 2018. 6(10): p. e1077-e86. doi: 10.1016/S2214-109X(18)30357-7.

14. World Health Organization. Global recommendations on physical activity for health: WHO Press; 2010.

15. Gelius P, Tcymbal A, Abu-Omar K, Mendes R, Tribuzi Morais S, Whiting S, Breda J, *Status and contents of physical activity recommendations in European Union countries: a systematic comparative analysis.* BMJ Open, 2020. 10(2): p. e034045. doi: 10.1136/bmjopen-2019-034045.

16. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Pratt M, Lancet Physical Activity Series 2 Executive C, *The economic burden of physical inactivity: a global analysis of major non-communicable diseases.* Lancet, 2016. 388(10051): p. 1311-24. doi: 10.1016/S0140-6736(16)30383-X.

17. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, Casanova A, Swaminathan S, Anjana RM, Kumar R, Rosengren A, Wei L, Yang W, Chuangshi W, Huaxing L, Nair S, Diaz R, Swidon H, Gupta R, Mohammadifard N, Lopez-Jaramillo P, Oguz A, Zatonska K, Seron P, Avezum A, Poirier P, Teo K, Yusuf S, *The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study.* Lancet, 2017. 390(10113): p. 2643-54. doi: 10.1016/S0140-6736(17)31634-3.

18. Silfee VJ, Haughton CF, Jake-Schoffman DE, Lopez-Cepero A, May CN, Sreedhara M, Rosal MC, Lemon SC, *Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review.* Prev Med Rep, 2018. 11: p. 74-80. doi: 10.1016/j.pmedr.2018.05.003.

19. Hildebrand M, Ekelund U. The assessment of physical activity. In: Armstrong N, van Mechelen W, editors. Oxford Textbook of Children's Sport and Exercise Medicine. 3rd ed: Oxford University Press; 2017.

20. Ndahimana D, Kim EK, *Measurement Methods for Physical Activity and Energy Expenditure: a Review.* Clin Nutr Res, 2017. 6(2): p. 68-80. doi: 10.7762/cnr.2017.6.2.68.

21. Lagerros YT, Lagiou P, *Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases.* Eur J Epidemiol, 2007. 22(6): p. 353-62. doi: 10.1007/s10654-007-9154-x.

22. Hills AP, Mokhtar N, Byrne NM, *Assessment of physical activity and energy expenditure: an overview of objective measures.* Frontiers in nutrition, 2014. 1: p. 5-. doi: 10.3389/fnut.2014.00005.

23. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ, *Reliability and validity of the combined heart rate and movement sensor Actiheart.* Eur J Clin Nutr, 2005. 59(4): p. 561-70. doi: 10.1038/sj.ejcn.1602118.

24. Arvidsson D, Fridolfsson J, Borjesson M, *Measurement of physical activity in clinical practice using accelerometers.* J Intern Med, 2019. doi: 10.1111/joim.12908.

25. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, Labayen I, Ruiz JR, Ortega FB, *Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations.* Sports medicine (Auckland, NZ), 2017. 47(9): p. 1821-45. doi: 10.1007/s40279-017-0716-0.

26. Sasaki JE, John D, Freedson PS, *Validation and comparison of ActiGraph activity monitors.* J Sci Med Sport, 2011. 14(5): p. 411-6. doi: 10.1016/j.jsams.2011.04.003.

27. Troiano RP, McClain JJ, Brychta RJ, Chen KY, *Evolution of accelerometer methods for physical activity research.* Br J Sports Med, 2014. 48(13): p. 1019-23. doi: 10.1136/bjsports-2014-093546.

28. Ferrari P, Friedenreich C, Matthews CE, *The role of measurement error in estimating levels of physical activity.* Am J Epidemiol, 2007. 166(7): p. 832-40. doi: 10.1093/aje/kwm148.

29. Skender S, Ose J, Chang-Claude J, Paskow M, Bruhmann B, Siegel EM, Steindorf K, Ulrich CM, *Accelerometry and physical activity questionnaires - a systematic review.* BMC Public Health, 2016. 16: p. 515. doi: 10.1186/s12889-016-3172-0.

30. Stunkard A, *A Method of Studying Physical Activity in Man.* The American Journal of Clinical Nutrition, 1960. 8(5): p. 595-601. doi: 10.1093/ajcn/8.5.595.

31. Bassett DR, Jr., Toth LP, LaMunion SR, Crouter SE, *Step Counting: A Review of Measurement Considerations and Health-Related Applications.* Sports medicine (Auckland, NZ), 2017. 47(7): p. 1303-15. doi: 10.1007/s40279-016-0663-1.

32. Crouter SE, Schneider PL, Karabulut M, Bassett DR, Jr., *Validity of 10 electronic pedometers for measuring steps, distance, and energy cost.* Med Sci Sports Exerc, 2003. 35(8): p. 1455-60. doi: 10.1249/01.MSS.0000078932.61440.A2.

33. Schneider PL, Crouter SE, Lukajic O, Bassett DR, Jr., *Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk.* Med Sci Sports Exerc, 2003. 35(10): p. 1779-84. doi: 10.1249/01.MSS.0000089342.96098.C4.

34. Corder K, Brage S, Ekelund U, *Accelerometers and pedometers: methodology and clinical application.* Curr Opin Clin Nutr Metab Care, 2007. 10(5): p. 597-603. doi: 10.1097/MCO.0b013e328285d883.

35. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, O'Brien WL, Bassett DR, Jr., Schmitz KH, Emplaincourt PO, Jacobs DR, Jr., Leon AS, *Compendium of physical activities: an update of activity codes and MET intensities.* Med Sci Sports Exerc, 2000. 32(9 Suppl): p. S498-504. doi: 10.1097/00005768-200009001-00009.

36. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Jr., Montoye HJ, Sallis JF, Paffenbarger RS, Jr., *Compendium of physical activities: classification of energy costs of human physical activities.* Med Sci Sports Exerc, 1993. 25(1): p. 71-80. doi: 10.1249/00005768-199301000-00011.

37. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Jr., Tudor-Locke C, Greer JL, Vezina J, Whitt-Glover MC, Leon AS, *2011 Compendium of Physical Activities: a second update of codes and MET values.* Med Sci Sports Exerc, 2011. 43(8): p. 1575-81. doi: 10.1249/MSS.0b013e31821ece12.

38. Pereira MA, FitzerGerald SJ, Gregg EW, Joswiak ML, Ryan WJ, Suminski RR, Utter AC, Zmuda JM, *A collection of Physical Activity Questionnaires for health-related research.* Med Sci Sports Exerc, 1997. 29(6 Suppl): p. S1-205.

39. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U, *A systematic review of reliability and objective criterion-related validity of physical activity questionnaires.* Int J Behav Nutr Phys Act, 2012. 9: p. 103. doi: 10.1186/1479-5868-9-103.

40. HUNT research center. *The HUNT Study*. NTNU 2020. [Available from: https://www.ntnu.edu/hunt, Accessed: 2020-11-24].

41. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I, *Cohort profile: the Tromso Study.* Int J Epidemiol, 2012. 41(4): p. 961-7. doi: 10.1093/ije/dyr049.

42. The Tromsø Study. *The Tromsø Study*. 2020. [Available from: http://tromsoundersokelsen.no, Accessed: 2020-5-27].

43. Saltin B, Grimby G, *Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages.* 1968.

44. Grimby G, Börjesson M, Jonsdottir IH, Schnohr P, Thelle DS, Saltin B, *The "Saltin–Grimby Physical Activity Level Scale" and its application to health research.* Scandinavian Journal of Medicine & Science in Sports, 2015. 25(S4): p. 119-25. doi: 10.1111/sms.12611.

45. Kurtze N, Gundersen KT, Holmen J, *Selvrapportert fysisk aktivitet i 1980- og 1990-årene—Helseundersøkelsen i Nord-Trøndelag (HUNT).* Norsk Epidemiol, 2003. 13: p. 171-6.

46. Kurtze N, Rangul V, Hustvedt BE, Flanders WD, *Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study (HUNT 2).* Eur J Epidemiol, 2007. 22(6): p. 379-87. doi: 10.1007/s10654-007-9110-9.

47. Morseth B, Hopstock LA, *Time trends in physical activity in the Tromso study: An update.* PloS one, 2020. 15(4): p. e0231581. doi: 10.1371/journal.pone.0231581.

48. Emaus A, Degerstrom J, Wilsgaard T, Hansen BH, Dieli-Conwright CM, Furberg AS, Pettersen SA, Andersen LB, Eggen AE, Bernstein L, Thune I, *Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study.* Scand J Public Health, 2010. 38(5 Suppl): p. 105-18. doi: 10.1177/1403494810378919.

49. Sagelv EH, Ekelund U, Pedersen S, Brage S, Hansen BH, Johansson J, Grimsgaard S, Nordstrom A, Horsch A, Hopstock LA, Morseth B, *Physical activity levels in adults and elderly from triaxial and uniaxial accelerometry. The Tromso Study.* PloS one, 2019. 14(12): p. e0225670. doi: 10.1371/journal.pone.0225670.

50. Sagelv EH, Hopstock LA, Johansson J, Hansen BH, Brage S, Horsch A, Ekelund U, Morseth B, *Criterion validity of two physical activity and one sedentary time questionnaire against accelerometry in a large cohort of adults and older adults.* BMJ open sport & exercise medicine, 2020. 6(1): p. e000661. doi: 10.1136/bmjsem-2019-000661.

51. Syed S, Morseth B, Hopstock LA, Horsch A, *Evaluating the performance of raw and epoch non-wear algorithms using multiple accelerometers and electrocardiogram recordings.* Sci Rep, 2020. 10(1): p. 5866. doi: 10.1038/s41598-020-62821-2.

52. Syed S, Morseth B, Hopstock LA, Horsch A, *A novel algorithm to detect non-wear time from raw accelerometer data using convolutional neural networks.* medRxiv, 2020. doi: 10.1101/2020.07.08.20148015.

53. Arnes AP, Nielsen CS, Stubhaug A, Fjeld MK, Hopstock LA, Horsch A, Johansen A, Morseth B, Wilsgaard T, Steingrimsdottir OA, *Physical activity and cold pain tolerance in the general population.* Eur J Pain, 2020. doi: 10.1002/ejp.1699.

54. Heitmann KA, Lochen ML, Hopstock LA, Stylidis M, Welde B, Schirmer H, Morseth B, *Cross-sectional associations between accelerometry-measured physical activity, left atrial size, and indices of left ventricular diastolic dysfunction: The Tromso Study.* Prev Med Rep, 2021. 21: p. 101290. doi: 10.1016/j.pmedr.2020.101290.

55. Hartge P, *Participation in population studies.* Epidemiology, 2006. 17. doi: 10.1097/01.ede.0000209441.24307.92.

56. Galea S, Tracy M, *Participation rates in epidemiologic studies.* Ann Epidemiol, 2007. 17(9): p. 643-53. doi: 10.1016/j.annepidem.2007.03.013.

57. Henriksen A, Hartvigsen G, Hopstock LA, Grimsgaard S. Using mobile sensors to expand recording of physical activity and increase motivation for prolonged data sharing in a population-based study. Proceedings from The 15th Scandinavian Conference on Health Informatics 2017 Kristiansand, Norway, August 29–30, 2017: Linköping University Electronic Press, Linköpings universitet; 2017.

58. Mindell JS, Giampaoli S, Goesswald A, Kamtsiuris P, Mann C, Mannisto S, Morgan K, Shelton NJ, Verschuren WM, Tolonen H, Group HESRR, *Sample selection, recruitment and participation rates in health examination surveys in Europe--experience from seven national surveys.* BMC Med Res Methodol, 2015. 15: p. 78. doi: 10.1186/s12874-015-0072-4.

59. Canalys. Worldwide wearable band market grew 65% in Q3 2019 while Asia Pacific doubled in volume. canalys.com; 2019. Contract No.: 2020-11-01.

60. Canalys. Worldwide smartwatch shipments grew 12% in Q1 2020 despite coronavirus. canalys.com; 2020.

61. McCarthy J. One in Five U.S. Adults Use Health apps, Wearable Trackers. Gallup.com; 2019.

62. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center; 2020. Contract No.: 2020-07-01.

63. Wright SP, Hall Brown TS, Collier SR, Sandberg K, *How consumer physical activity monitors could transform human physiology research.* Am J Physiol Regul Integr Comp Physiol, 2017. 312(3): p. R358-R67. doi: 10.1152/ajpregu.00349.2016.

64. Henriksen A, Hopstock LA, Hartvigsen G, Grimsgaard S, *Using Cloud-Based Physical Activity Data from Google Fit and Apple Healthkit to Expand Recording of Physical Activity Data in a Population Study.* Studies in health technology and informatics, 2017. 245: p. 108-12.

65. Giordanengo A. Using FullFlow to manage the overwhelming flood of patients' self-collected health data: A system that addresses acceptance barriers regarding the introduction of diabetes patients' self-collected health data into electronic health records and medical consultations. UiT Norges arktiske universitet: UiT Norges arktiske universitet; 2020.

66. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, Welk GJ, *The Wild Wild West: A Framework to Integrate mHealth Software Applications and Wearables to Support Physical Activity Assessment, Counseling and Interventions for Cardiovascular Disease Risk Reduction.* Prog Cardiovasc Dis, 2016. 58(6): p. 584-94. doi: 10.1016/j.pcad.2016.02.007.

67. Giordanengo A, Bradway M, Muzny M, Woldaregay AW, Hartvigsen G, Årsand E. Systems integrating self-collected health data by patients into EHRs: a State-of-the-art review. 15th Scandinavian Conference on Health Informatics: Linköping University Electronic Press; 2017.

68. Massoomi MR, Handberg EM, *Increasing and Evolving Role of Smart Devices in Modern Medicine.* Eur Cardiol, 2019. 14(3): p. 181-6. doi: 10.15420/ecr.2019.02.

69. Aroganam G, Manivannan N, Harrison D, *Review on Wearable Technology Sensors Used in Consumer Sport Applications.* Sensors (Basel), 2019. 19(9). doi: 10.3390/s19091983.

70. Xing S, Hanghang T, Ping J, *Activity recognition with smartphone sensors.* Tsinghua Science and Technology, 2014. 19(3): p. 235-49. doi: 10.1109/tst.2014.6838194.

71. Wagenaar RC, Sapir I, Zhang Y, Markovic S, Vaina LM, Little TDC, editors. Continuous monitoring of functional activities using wearable, wireless gyroscope and accelerometer technology. Conf Proc IEEE Eng Med Biol Soc; 2011 Aug. 30 2011-Sept. 3 2011.

72. Richardson S, Mackinnon D. Left to their own Devices? Privacy Implications of Wearable Technology in Canadian Workplaces. 2017.

73. Allen J, *Photoplethysmography and its application in clinical physiological measurement.* Physiol Meas, 2007. 28(3): p. R1-39. doi: 10.1088/0967-3334/28/3/R01.

74. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H, *A review on wearable photoplethysmography sensors and their potential future applications in health care.* Int J Biosens Bioelectron, 2018. 4(4): p. 195-202. doi: 10.15406/ijbsbe.2018.04.00125.

75. Vizbara V, *Comparison of Green, Blue and Infrared Light in Wrist and Forehead Photoplethysmography.* 17th international conference Biomedical engineering 2013, 2013. 17.

76. Kim SH, Ryoo DW, Bae C, editors. Adaptive Noise Cancellation Using Accelerometers for the PPG Signal from Forehead. Conf Proc IEEE Eng Med Biol Soc; 2007 22-26 Aug. 2007.

77. Casson AJ, Vazquez Galvez A, Jarchi D, *Gyroscope vs. accelerometer measurements of motion from wrist PPG during physical exercise.* ICT Express, 2016. 2(4): p. 175-9. doi: http://dx.doi.org/10.1016/j.icte.2016.11.003.

78. Zhang Y, Song S, Vullings R, Biswas D, Simoes-Capela N, van Helleputte N, van Hoof C, Groenendaal W, *Motion Artifact Reduction for Wrist-Worn Photoplethysmograph Sensors Based on Different Wavelengths.* Sensors (Basel), 2019. 19(3). doi: 10.3390/s19030673.

79. IDC. Wearable Devices Market Share. 2020. Contract No.: 2020-11-23.

80. Canalys. North American wearables market grows 10% as average selling prices take a hit. canalys.com; 2020. Contract No.: 2020-11-01.

81. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G, *The Accuracy of Smart Devices for Measuring Physical Activity in Daily Life: Validation Study.* JMIR Mhealth Uhealth, 2018. 6(12): p. e10972. doi: 10.2196/10972.

82. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J, *Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study.* JMIR Mhealth Uhealth, 2018. 6(4): p. e94. doi: 10.2196/mhealth.9754.

83. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA, *Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review.* JMIR Mhealth Uhealth, 2020. 8(9): p. e18694. doi: 10.2196/18694.

84. Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, Vandendriessche A, Curtis RG, DeSmet A, *Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study.* JMIR Mhealth Uhealth, 2020. 8(5): p. e16674. doi: 10.2196/16674.

85. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, Finlayson G, Stubbs J, *How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies.* Br J Sports Med, 2020. 54(6): p. 332-40. doi: 10.1136/bjsports-2018-099643.

86. Bunn JA, Navalta JW, Fountaine CJ, Reece JD, *Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017.* Int J Exerc Sci, 2018. 11(7): p. 503-15.

87. Evenson KR, Goto MM, Furberg RD, *Systematic review of the validity and reliability of consumer-wearable activity trackers.* Int J Behav Nutr Phys Act, 2015. 12: p. 159. doi: 10.1186/s12966-015-0314-1.

88. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, Hamilton CB, Li LC, *Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data.* JMIR Mhealth Uhealth, 2018. 6(8): p. e10527. doi: 10.2196/10527.

89. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, Gallagher R, *The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review.* Maturitas, 2018. 112: p. 85-93. doi: 10.1016/j.maturitas.2018.03.016.

90. Evenson KR, Spade CL, *Review of Validity and Reliability of Garmin Activity Trackers.* J Meas Phys Behav, 2020. 3(2): p. 170-85. doi: 10.1123/jmpb.2019-0035.

91. LaMunion SR, Blythe AL, Hibbing PR, Kaplan AS, Clendenin BJ, Crouter SE, *Use of consumer monitors for estimating energy expenditure in youth.* Appl Physiol Nutr Metab, 2020. 45(2): p. 161-8. doi: 10.1139/apnm-2019-0129.

92. Montes J, Tandy R, Young J, Lee SP, Navalta JW, *Step Count Reliability and Validity of Five Wearable Technology Devices While Walking and Jogging in both a Free Motion Setting and on a Treadmill.* Int J Exerc Sci, 2020. 13(7): p. 410-26.

93. Henriksen A, Johansson J, Hartvigsen G, Grimsgaard S, Hopstock L, *Measuring Physical Activity Using Triaxial Wrist Worn Polar Activity Trackers: A Systematic Review.* Int J Exerc Sci, 2020. 13(4): p. 438-54.

94. Gilgen-Ammann R, Schweizer T, Wyss T, *Accuracy of the Multisensory Wristwatch Polar Vantage's Estimation of Energy Expenditure in Various Activities: Instrument Validation Study.* JMIR Mhealth Uhealth, 2019. 7(10): p. e14534. doi: 10.2196/14534.

95. Duking P, Giessing L, Frenkel MO, Koehler K, Holmberg HC, Sperlich B, *Wrist-Worn Wearables for Monitoring Heart Rate and Energy Expenditure While Sitting or Performing Light-to-Vigorous Physical Activity: Validation Study.* JMIR Mhealth Uhealth, 2020. 8(5): p. e16716. doi: 10.2196/16716.

96. Henriksen A, Svartdal F, Grimsgaard S, Hartvigsen G, Hopstock L, *Polar Vantage and Oura physical activity and sleep trackers: A validation and comparison study.* JMIR Preprints, 2021. doi: 10.2196/preprints.27248.

97. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA, *Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking.* J Sports Sci Med, 2016. 15(3): p. 540-7.

98. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW, *Validity of Wrist-Worn photoplethysmography devices to measure heart rate: A systematic review and meta-analysis.* J Sports Sci, 2020. 38(17): p. 2021-34. doi: 10.1080/02640414.2020.1767348.

99. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, Doyle Iii FJ, Clements MA, Rickels MR, Patton SR, Castle JR, Riddell MC, Jacobs PG, *Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study.* JMIR Mhealth Uhealth, 2018. 6(12): p. e10338. doi: 10.2196/10338.

100. Dondzila CJ, Lewis CA, Lopez JR, Parker TM, *Congruent Accuracy of Wrist-worn Activity Trackers during Controlled and Free-living Conditions.* International Journal of Exercise Science, 2018.

101. Chow HW, Yang CC, *Accuracy of Optical Heart Rate Sensing Technology in Wearable Fitness Trackers for Young and Older Adults: Validation and Comparison Study.* JMIR Mhealth Uhealth, 2020. 8(4): p. e14707. doi: 10.2196/14707.

102. Lee J, Finkelstein J, *Consumer sleep tracking devices: a critical review.* Studies in health technology and informatics, 2015. 210: p. 458-60.

103. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ, *Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis.* J Med Internet Res, 2019. 21(11): p. e16273. doi: 10.2196/16273.

104. de Zambotti M, Rosas L, Colrain IM, Baker FC, *The Sleep of the Ring: Comparison of the OURA Sleep Tracker Against Polysomnography.* Behav Sleep Med, 2017: p. 1-15. doi: 10.1080/15402002.2017.1300587.

105. Roberts DM, Schade MM, Mathew GM, Gartenberg D, Buxton OM, *Detecting sleep using heart rate and motion data from multisensor consumer-grade wearables, relative to wrist actigraphy and polysomnography.* Sleep, 2020. 43(7). doi: 10.1093/sleep/zsaa045.

106. Apple Inc. *Your health, from head to toe*. 2020. [Available from: https://www.apple.com/ios/health, Accessed: 2020-11-01].

107. Apple Inc. *ResearchKit and CareKit*. 2020. [Available from: https://www.apple.com/researchkit, Accessed: 2020-11-01].

108. Advancing health research with Google Health Studies [press release]. The Keyword, 2020-12-09 2020.

109. Open mHealth. *Open mHealth*. 2020. [Available from: https://www.openmhealth.org, Accessed: 2020-11-05].

110. Open mHealth. *Shimmer*. 2020. [Available from: https://www.openmhealth.org/documentation/#/data-providers/shimmer, Accessed: 2020-11-05].

111. Human API. *Human API*. 2020. [Available from: https://www.humanapi.co, Accessed: 2020-10-15].

112. USCF. *Health eHeart study*. 2020. [Available from: https://www.health-eheartstudy.org, Accessed: 2020-11-17].

113. Scripps Research Institute. *Detect Health Study*. 2020. [Available from: https://detectstudy.org, Accessed: 2020-11-17].

114. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR, *Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study.* The Lancet Digital Health, 2020. 2(2): p. e85-e93. doi: 10.1016/s2589-7500(19)30222-5.

115. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, Kheterpal V, Topol EJ, Steinhubl SR, *Wearable sensor data and self-reported symptoms for COVID-19 detection.* Nature Medicine, 2020. doi: 10.1038/s41591-020-1123-x.

116. All of Us Research Program I, Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, Jenkins G, Dishman E, *The "All of Us" Research Program.* N Engl J Med, 2019. 381(7): p. 668-76. doi: 10.1056/NEJMsr1809937.

117. RADAR-base. *RADAR-base* 2020. [Available from: https://radar-base.org, Accessed: 2020-12-01].

118. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Boettcher S, Hyve, Dobson R, Folarin A, Consortium R-C, *RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices.* JMIR Mhealth Uhealth, 2019. 7(8): p. e11734. doi: 10.2196/11734.

119. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, Cummins N, Matcham F, Dalla Costa G, Simblett S, Leocani L, Lamers F, Sorensen PS, Buron M, Zabalza A, Guerrero Perez AI, Penninx BW, Siddi S, Haro JM, Myin-Germeys I, Rintala A, Wykes T, Narayan VA, Comi G, Hotopf M, Dobson RJ, Consortium R-C, *Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19.* J Med Internet Res, 2020. 22(9): p. e19992. doi: 10.2196/19992.

120. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S, *Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables.* J Med Internet Res, 2018. 20(3): p. e110. doi: 10.2196/jmir.9157.

121. Vandrico. *The wearables database*. 2017. [Available from: http://vandrico.com/wearables, Accessed: 2017-07-12].

122. GsmArena. *GsmArena*. 2017. [Available from: http://www.gsmarena.com/results.php3?sFormFactors=8, Accessed: 2017-04-01].

123. Wearables. *Wearables.com*. 2017. [Available from: http://www.wearables.com/devices, Accessed: 2017-07-12].

124. Specbucket. *Smart Watches*. 2017. [Available from: https://smartwatches.specbucket.com, Accessed: 2017-07-12].

125. Prisguide. *Smartklokke*. 2017. [Available from: https://prisguiden.no/kategorier/smartklokke, Accessed: 2017-07-12].

126. Prisguide. *Aktivitetsarmband*. 2017. [Available from: https://prisguiden.no/kategorier/aktivitetsarmband, Accessed: 2017-07-17].

127. Richardson SM, Debra. Wearable Device Inventory. Surveillance Studies Centre, Queen's University. 2017.

128. Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L, *Validity of the Polar M430 Activity Monitor in Free-Living Conditions: Validation Study.* JMIR Form Res, 2019. 3(3): p. e14438. doi: 10.2196/14438.

129. Freedson PS, Melanson E, Sirard J, *Calibration of the Computer Science and Applications, Inc. accelerometer.* Med Sci Sports Exerc, 1998. 30(5): p. 777-81.

130. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS, *Validation of wearable monitors for assessing sedentary behavior.* Med Sci Sports Exerc, 2011. 43(8): p. 1561-7. doi: 10.1249/MSS.0b013e31820ce174.

131. Peterson NE, Sirard JR, Kulbok PA, DeBoer MD, Erickson JM, *Validation of Accelerometer Thresholds and Inclinometry for Measurement of Sedentary Behavior in Young Adult University Students.* Research in nursing & health, 2015. 38(6): p. 492-9. doi: 10.1002/nur.21694.

132. Lee JA, Williams SM, Brown DD, Laurson KR, *Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings.* J Sports Sci, 2015. 33(10): p. 991-1000. doi: 10.1080/02640414.2014.981848.

133. McMinn D, Acharya, R., Rowe, D. A., Gray, S. R. and Allan, J. L., *Measuring activity energy expenditure: accuracy of the GT3X+ and actiheart monitors.* International Journal of Exercise Science, 2013. 6(3): p. 217-29.

134. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, Wareham NJ, *Estimation of Free-Living Energy Expenditure by Heart Rate and Movement Sensing: A Doubly-Labelled Water Study.* PloS one, 2015. 10(9): p. e0137206. doi: 10.1371/journal.pone.0137206.

135. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M, *Physical activity in the United States measured by accelerometer.* Med Sci Sports Exerc, 2008. 40(1): p. 181-8. doi: 10.1249/mss.0b013e31815a51b3.

136. Hecht A, Ma S, Porszasz J, Casaburi R, *Methodology for using long-term accelerometry monitoring to describe daily activity patterns in COPD.* Copd, 2009. 6(2): p. 121-9. doi: 10.1080/15412550902755044.

137. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, Troiano RP, *Amount of time spent in sedentary behaviors in the United States, 2003-2004.* Am J Epidemiol, 2008. 167(7): p. 875-81. doi: 10.1093/aje/kwm390.

138. Schrack JA, Leroux A, Fleg JL, Zipunnikov V, Simonsick EM, Studenski SA, Crainiceanu C, Ferrucci L, *Using Heart Rate and Accelerometry to Define Quantity and Intensity of Physical Activity in Older Adults.* J Gerontol A Biol Sci Med Sci, 2018. 73(5): p. 668-75. doi: 10.1093/gerona/gly029.

139. Troiano RP, *Large-scale applications of accelerometers: new frontiers and new questions.* Med Sci Sports Exerc, 2007. 39(9): p. 1501. doi: 10.1097/mss.0b013e318150d42e.

140. Brage S, Brage N, Franks PW, Ekelund U, Wong MY, Andersen LB, Froberg K, Wareham NJ, *Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure.* J Appl Physiol (1985), 2004. 96(1): p. 343-51. doi: 10.1152/japplphysiol.00703.2003.

141. Schofield WN, *Predicting basal metabolic rate, new standards and review of previous work.* Human nutrition Clinical nutrition, 1985. 39 Suppl 1: p. 5-41.

142. Evans JD. Straightforward statistics for the behavioral sciences.: Brooks/Cole Publishing; 1996.

143. Bland JM, Altman DG, *Measuring agreement in method comparison studies.* Stat Methods Med Res, 1999. 8(2): p. 135-60. doi: 10.1177/096228029900800204.

144. Koo TK, Li MY, *A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research.* J Chiropr Med, 2016. 15(2): p. 155-63. doi: 10.1016/j.jcm.2016.02.012.

145. Fokkema T, Kooiman TJ, Krijnen WP, CP VDS, M DEG, *Reliability and Validity of Ten Consumer Activity Trackers Depend on Walking Speed.* Med Sci Sports Exerc, 2017. 49(4): p. 793-800. doi: 10.1249/MSS.0000000000001146.

146. Roos L, Taube W, Beeler N, Wyss T, *Validity of sports watches when estimating energy expenditure during running.* BMC Sports Sci Med Rehabil, 2017. 9: p. 22. doi: 10.1186/s13102-017-0089-6.

147. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH, *Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types.* Med Sci Sports Exerc, 2016. 48(8): p. 1619-28. doi: 10.1249/MSS.0000000000000933.

148. Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L, *Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach.* BMC Public Health, 2020. 20(1): p. 1300. doi: 10.1186/s12889-020-09406-w.

149. Deraas TS, Hopstock L, Henriksen A, Morseth B, Sand A-S, Njølstad I, Pedersen S, Sagelv E, Johansson J, Grimsgaard S, *Complex lifestyle intervention among inactive*

older adults with elevated cardiovascular disease risk and obesity. Research Square, 2020. doi: 10.21203/rs.3.rs-39292/v1.

150. Selmer R, Igland J, Ariansen I, Tverdal A, Njølstad I, Furu K, Tell GS, Klemsdal TO, *NORRISK 2: A Norwegian risk model for acute cerebral stroke and myocardial infarction.* 2017. 24(7): p. 773-82. doi: 10.1177/2047487317693949.

151. Polar Electro. *Polar Flow*. 2018. [Available from: https://flow.polar.com, Accessed: 2018-03-05].

152. Pugliese L, Woodriff M, Crowley O, Lam V, Sohn J, Bradley S, *Feasibility of the "Bring Your Own Device" Model in Clinical Research: Results from a Randomized Controlled Pilot Study of a Mobile Patient Engagement Tool.* Cureus, 2016. 8(3): p. e535. doi: 10.7759/cureus.535.

153. Braun V, Clarke V, *What can "thematic analysis" offer health and wellbeing researchers?* Int J Qual Stud Health Well-being, 2014. 9: p. 26152. doi: 10.3402/qhw.v9.26152.

154. Kvale S., Brinkmann S. Interviews. Learning the craft of Qualitative Research Interviewing. 2nd ed. Los Angeles, London, New Delhi, Singapore: SAGE publications; 2009.

155. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM, *Wearable Technology and Physical Activity in Chronic Disease: Opportunities and Challenges.* Am J Prev Med, 2018. 54(1): p. 144-50. doi: 10.1016/j.amepre.2017.08.015.

156. Bakdash JZ, Marusich LR, *Repeated Measures Correlation.* Front Psychol, 2017. 8: p. 456. doi: 10.3389/fpsyg.2017.00456.

157. Boyatzis RE. Transforming Qualitative Information. SAGE Publications, Inc 1998.

158. Braun V, Clarke V, *Using thematic analysis in psychology.* Qualitative Research in Psychology, 2006. 3(2): p. 77-101. doi: 10.1191/1478088706qp063oa.

159. Fereday J, Muir-Cochrane E, *Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development.* International Journal of Qualitative Methods, 2016. 5(1): p. 80-92. doi: 10.1177/160940690600500107.

160. Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA, *Physical activity surveillance during the COVID-19 pandemic: Using consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies.* JMIR Preprints, 2021. doi: 10.2196/preprints.23806

161. Porta M, Greenland S, Burón A, International Epidemiological A. A dictionary of epidemiology. Oxford, England: Oxford University Press; 2014.

162. Morseth B. Physical activity, osteoporosis, and fracture risk: UiT The Arctic University of Norway; 2011.

163. Pourhoseingholi MA, Baghestani AR, Vahedi M, *How to control confounding effects by statistical analysis.* Gastroenterol Hepatol Bed Bench, 2012. 5(2): p. 79-83.

164. Grimes DA, Schulz KF, *Bias and causal associations in observational research.* The Lancet, 2002. 359(9302): p. 248-52. doi: 10.1016/s0140-6736(02)07451-2.

165. Xiaomi. *Xiaomi Health Cloud Open Platform User Documentation*. 2020. [Available from: https://dev.mi.com/docs/micloud/health, Accessed: 2020-12-08].

166. McDonnell EI, Zipunnikov V, Schrack JA, Goldsmith J, Wrobel J, *Scale-invariant time registration of 24-hour accelerometric rest-activity profiles and its application to human chronotypes.* bioRxiv, 2020. doi: 10.1101/2020.10.13.337550.

167. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L, *Standardizing Analytic Methods and Reporting in Activity Monitor Validation Studies.* Med Sci Sports Exerc, 2019. 51(8): p. 1767-80. doi: 10.1249/MSS.0000000000001966.

168. John D, Morton A, Arguello D, Lyden K, Bassett D, *"What Is a Step?" Differences in How a Step Is Detected among Three Popular Activity Monitors That Have Impacted Physical Activity Research.* Sensors (Basel), 2018. 18(4). doi: 10.3390/s18041206.

169. Duking P, Fuss FK, Holmberg HC, Sperlich B, *Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity.* JMIR Mhealth Uhealth, 2018. 6(4): p. e102. doi: 10.2196/mhealth.9341.

170. Cadmus-Bertram L, *Using Fitness Trackers in Clinical Research: What Nurse Practitioners Need to Know.* J Nurse Pract, 2017. 13(1): p. 34-40. doi: 10.1016/j.nurpra.2016.10.012.

171. Maganja SA, Clarke DC, Lear SA, Mackey DC, *Formative Evaluation of Consumer-Grade Activity Monitors Worn by Older Adults: Test-Retest Reliability and Criterion Validity of Step Counts.* JMIR Form Res, 2020. 4(8): p. e16537. doi: 10.2196/16537.

172. Weisberg A, Campelo AM, Bhaidani T, Katz L, *Physical Activity Tracking Wristbands for Use in Research With Older Adults: An Overview and Recommendations.* Journal for the Measurement of Physical Behaviour, 2020. 3(4): p. 265-73. doi: 10.1123/jmpb.2019-0050.

173. The Tromsø Study. *The seventh survey of the Tromsø Study*. 2020. [Available from: https://uit.no/research/tromsostudy/project?pid=708909, Accessed: 2020-03-12].

174. Kahan BC, Rehal S, Cro S, *Risk of selection bias in randomised trials.* Trials, 2015. 16(1). doi: 10.1186/s13063-015-0920-x.

175. McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, Elbourne DR, Francis D, Garcia J, Roberts I, Snowdon C, *What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies.* Trials, 2006. 7: p. 9. doi: 10.1186/1745-6215-7-9.

176. Harris TJ, Carey IM, Victor CR, Adams R, Cook DG, *Optimising recruitment into a study of physical activity in older people: a randomised controlled trial of different approaches.* Age Ageing, 2008. 37(6): p. 659-65. doi: 10.1093/ageing/afn159.

177. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A, *Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants.* Sci Rep, 2018. 8(1): p. 7961. doi: 10.1038/s41598-018-26174-1.

178. Electroverse. *Norway Breaks Snowfall Records*. 2020. [Available from: https://electroverse.net/norway-breaks-snowfall-records, Accessed: 2020-12-04].

179. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, Peng W, *The Use of Wearable Activity Trackers Among Older Adults: Focus Group Study of Tracker Perceptions, Motivators, and Barriers in the Maintenance Stage of Behavior Change.* JMIR Mhealth Uhealth, 2019. 7(4): p. e9832. doi: 10.2196/mhealth.9832.

180. Macridis S, Johnston N, Johnson S, Vallance JK, *Consumer physical activity tracking device ownership and use among a population-based sample of adults.* PloS one, 2018. 13(1): p. e0189298. doi: 10.1371/journal.pone.0189298.

181. Barreto PdS, Ferrandez A-M, Saliba-Serre B, *Are Older Adults Who Volunteer to Participate in an Exercise Study Fitter and Healthier Than Nonvolunteers? The Participation Bias of the Study Population.* Journal of Physical Activity and Health, 2013. 10(3): p. 359-67. doi: 10.1123/jpah.10.3.359.

182. van Heuvelen MJ, Hochstenbach JB, Brouwer WH, de Greef MH, Zijlstra GA, van Jaarsveld E, Kempen GI, van Sonderen E, Ormel J, Mulder T, *Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons.* Aging Clin Exp Res, 2005. 17(3): p. 236-45. doi: 10.1007/BF03324603.

183. Lesser IA, Nienhuis CP, *The Impact of COVID-19 on Physical Activity Behavior and Well-Being of Canadians.* Int J Environ Res Public Health, 2020. 17(11). doi: 10.3390/ijerph17113899.

184. Castaneda-Babarro A, Arbillaga-Etxarri A, Gutierrez-Santamaria B, Coca A, *Physical Activity Change during COVID-19 Confinement.* Int J Environ Res Public Health, 2020. 17(18). doi: 10.3390/ijerph17186878.

185. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, Olgin JE, *Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study.* Ann Intern Med, 2020. doi: 10.7326/M20-2665.

186. Sand A-S. Overweight, body size perceptions, lifestyle changes and health concerns in young adults. The Tromsø Study, Fit Futures . A combined approach using qualitative interviews and a population-based cross-sectional study: UiT The Arctic University of Norway; 2017.

187. Graneheim UH, Lundman B, *Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness.* Nurse education today, 2004. 24(2): p. 105-12. doi: 10.1016/j.nedt.2003.10.001.

188. Saghir N, Aggarwal A, Soneji N, Valencia V, Rodgers G, Kurian T, *A comparison of manual electrocardiographic interval and waveform analysis in lead 1 of 12-lead ECG and Apple Watch ECG: A validation study.* Cardiovascular Digital Health Journal, 2020. 1(1): p. 30-6. doi: 10.1016/j.cvdhj.2020.07.002.

189. Seshadri DR, Bittel B, Browsky D, Houghtaling P, Drummond CK, Desai M, Gillinov AM, *Accuracy of the Apple Watch 4 to Measure Heart Rate in Patients With Atrial Fibrillation.* IEEE J Transl Eng Health Med, 2020. 8: p. 2700204. doi: 10.1109/JTEHM.2019.2950397.

190. Isakadze N, Martin SS, *How useful is the smartwatch ECG?* Trends Cardiovasc Med, 2020. 30(7): p. 442-8. doi: 10.1016/j.tcm.2019.10.010.

191. Kornej J, Borschel CS, Benjamin EJ, Schnabel RB, *Epidemiology of Atrial Fibrillation in the 21st Century: Novel Methods and New Insights.* Circ Res, 2020. 127(1): p. 4-20. doi: 10.1161/CIRCRESAHA.120.316340.

192. Reyzelman AM, Koelewyn K, Murphy M, Shen X, Yu E, Pillai R, Fu J, Scholten HJ, Ma R, *Continuous Temperature-Monitoring Socks for Home Use in Patients With Diabetes: Observational Study.* J Med Internet Res, 2018. 20(12): p. e12460. doi: 10.2196/12460.

193. Kamisalic A, Fister I, Jr., Turkanovic M, Karakatic S, *Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review.* Sensors (Basel), 2018. 18(6). doi: 10.3390/s18061714.

194. Climstein M, Alder JL, Brooker AM, Cartwright EJ, Kemp-Smith K, Simas V, Furness J, *Reliability of the Polar Vantage M Sports Watch when Measuring Heart Rate at Different Treadmill Exercise Intensities.* Sports (Basel), 2020. 8(9). doi: 10.3390/sports8090117.

195. Kim H, Sun W, Malaska M, B. M, Han H, *Validation of Wearable Activity Monitors for Real-Time Cadence.* International Journal of Exercise Science, 2019. 2(11).

196. Janicki GJ, Porcari JP, Jagim AR, Emineth K, Doberstein S, Schwab E, Foster C, *The Accuracy of Heart Rate and Energy Expenditure Measurements using the PolarM430 Watch during Exercise.* Int J Res Ex Phys, 2020. 15(2): p. 91-105.

197. Miller KE. Validity of the Polar M430 Among Females of Varying Fitness Levels, Body Fat Percentage, and Reported Physical Activity. Boise State University Theses and Dissertations2019.

198. Turner-McGrievy G, Jake-Schoffman DE, Singletary C, Wright M, Crimarco A, Wirth MD, Shivappa N, Mandes T, West DS, Wilcox S, Drenowatz C, Hester A, McGrievy MJ, *Using Commercial Physical Activity Trackers for Health Promotion Research: Four Case Studies.* Health Promot Pract, 2019. 20(3): p. 381-9. doi: 10.1177/1524839918769559.

199. Strain T, Wijndaele K, Brage S, *Physical Activity Surveillance Through Smartphone Apps and Wearable Trackers: Examining the UK Potential for Nationally Representative Sampling.* JMIR Mhealth Uhealth, 2019. 7(1): p. e11898. doi: 10.2196/11898.

200. Omura JD, Carlson SA, Paul P, Watson KB, Fulton JE, *National physical activity surveillance: Users of wearable activity monitors as a potential data source.* Prev Med Rep, 2017. 5: p. 124-6. doi: 10.1016/j.pmedr.2016.10.014.

201. Hyde ET, Omura JD, Fulton JE, Weldy A, Carlson SA, *Physical Activity Surveillance Using Wearable Activity Monitors: Are US Adults Willing to Share Their Data?* Am J Health Promot, 2020. 34(6): p. 672-6. doi: 10.1177/0890117119900587.

202. Woldaregay AZ, Henriksen A, Issom DZ, Pfuhl G, Sato K, Richard A, Lovis C, Arsand E, Rochat J, Hartvigsen G, *User Expectations and Willingness to Share Self-Collected Health Data.* Studies in health technology and informatics, 2020. 270: p. 894-8. doi: 10.3233/SHTI200290.

203. Woldaregay AZ, Issom DZ, Henriksen A, Marttila H, Mikalsen M, Pfuhl G, Sato K, Lovis C, Hartvigsen G, *Motivational Factors for User Engagement with mHealth Apps.* Studies in health technology and informatics, 2018. 249: p. 151-7.

204. Coughlin SS, Stewart J, *Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies.* J Environ Health Sci, 2016. 2(6). doi: 10.15436/2378-6841.16.1123.

205. Franssen WMA, Franssen G, Spaas J, Solmi F, Eijnde BO, *Can consumer wearable activity tracker-based interventions improve physical activity and cardiometabolic health in patients with chronic diseases? A systematic review and meta-analysis of randomised controlled trials.* Int J Behav Nutr Phys Act, 2020. 17(1): p. 57. doi: 10.1186/s12966-020-00955-2.

206. Low CA, *Harnessing consumer smartphone and wearable sensors for clinical cancer research.* NPJ Digit Med, 2020. 3: p. 140. doi: 10.1038/s41746-020-00351-x.

207. Stienen MN, Rezaii PG, Ho AL, Veeravagu A, Zygourakis CC, Tomkins-Lane C, Park J, Ratliff JK, Desai AM, *Objective activity tracking in spine surgery: a prospective feasibility study with a low-cost consumer grade wearable accelerometer.* Sci Rep, 2020. 10(1): p. 4939. doi: 10.1038/s41598-020-61893-4.

208. Schwab M, Brindl N, Studier-Fischer A, Tu T, Gsenger J, Pilgrim M, Friedrich M, Frey PE, Achilles C, Leuck A, Burgel T, Feisst M, Klose C, Tenckhoff S, Dorr-Harim C, Mihaljevic AL, *Postoperative complications and mobilisation following major abdominal surgery with vs. without fitness tracker-based feedback (EXPELLIARMUS): study protocol for a student-led multicentre randomised controlled trial (CHIR-Net SIGMA study group).* Trials, 2020. 21(1): p. 293. doi: 10.1186/s13063-020-4220-8.

Paper I

Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S.

Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables.

J Med Internet Res, 2018. 20(3): p. e110. doi: 10.2196/jmir.9157.

Original Paper

Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables

André Henriksen[1], MSc (Comp Sci), MBA; Martin Haugen Mikalsen[2], BSc (Comp Sci); Ashenafi Zebene Woldaregay[2], MSc (Comp Eng), MSc (Telemed & e-Health); Miroslav Muzny[3,4], MSc (Comp Sci); Gunnar Hartvigsen[2], MSc (Comp Sci), PhD; Laila Arnesdatter Hopstock[5], RN, CRNA, MSc (Nursing), PhD; Sameline Grimsgaard[1], MPH, MD, PhD

[1]Department of Community Medicine, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

[2]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

[3]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[4]Spin-Off Company and Research Results Commercialization Center, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

[5]Department of Health and Care Sciences, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
André Henriksen, MSc (Comp Sci), MBA
Department of Community Medicine
University of Tromsø – The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 77644000
Email: andre.henriksen@uit.no

Abstract

Background: New fitness trackers and smartwatches are released to the consumer market every year. These devices are equipped with different sensors, algorithms, and accompanying mobile apps. With recent advances in mobile sensor technology, privately collected physical activity data can be used as an addition to existing methods for health data collection in research. Furthermore, data collected from these devices have possible applications in patient diagnostics and treatment. With an increasing number of diverse brands, there is a need for an overview of device sensor support, as well as device applicability in research projects.

Objective: The objective of this study was to examine the availability of wrist-worn fitness wearables and analyze availability of relevant fitness sensors from 2011 to 2017. Furthermore, the study was designed to assess brand usage in research projects, compare common brands in terms of developer access to collected health data, and features to consider when deciding which brand to use in future research.

Methods: We searched for devices and brand names in six wearable device databases. For each brand, we identified additional devices on official brand websites. The search was limited to wrist-worn fitness wearables with accelerometers, for which we mapped brand, release year, and supported sensors relevant for fitness tracking. In addition, we conducted a Medical Literature Analysis and Retrieval System Online (MEDLINE) and ClinicalTrials search to determine brand usage in research projects. Finally, we investigated developer accessibility to the health data collected by identified brands.

Results: We identified 423 unique devices from 132 different brands. Forty-seven percent of brands released only one device. Introduction of new brands peaked in 2014, and the highest number of new devices was introduced in 2015. Sensor support increased every year, and in addition to the accelerometer, a photoplethysmograph, for estimating heart rate, was the most common sensor. Out of the brands currently available, the five most often used in research projects are Fitbit, Garmin, Misfit, Apple, and Polar. Fitbit is used in twice as many validation studies as any other brands and is registered in ClinicalTrials studies 10 times as often as other brands.

Conclusions: The wearable landscape is in constant change. New devices and brands are released every year, promising improved measurements and user experience. At the same time, other brands disappear from the consumer market for various reasons. Advances in device quality offer new opportunities for research. However, only a few well-established brands are frequently used in research projects, and even less are thoroughly validated.

(J Med Internet Res 2018;20(3):e110) doi:10.2196/jmir.9157

KEYWORDS

motor activity; physical activity; fitness trackers; heart rate; photoplethysmography

Introduction

Background

The World Health Organization recommends 150 min of moderate intensity physical activity (PA) each week for adults and 60 min for children and adolescents [1]. However, 25% of adults and more than 80% of adolescents do not achieve the recommended PA targets [1]. Results from the Tromsø Study, the longest running population study in Norway, shows that only 30.4% of women and 22.0% of men reach the recommended target [2].

Low PA is currently the fourth leading risk factor for mortality worldwide [3]. Even though there is limited evidence that using wearable fitness trackers will improve health [4,5], these devices are still popular, and new fitness devices appear on the consumer market regularly. In 2016, vendors shipped 102 million devices worldwide, compared with 82 million in 2015 [6]. Fifty-seven percent of these devices were sold by the top five brands: Fitbit, Xiaomi, Apple, Garmin, and Samsung. The first quarter of 2017 shows an increase of 18% in devices sold, compared with the same period in 2016 [7]. With a large number of available devices and brands, it is difficult to navigate through an ever-growing list of brands and devices with different capabilities, price, and quality.

Available sensors and internal interpreting algorithms determine device output. Sensor data are, in most devices, reduced to a limited set of metrics before being transferred to the user's mobile phone. In addition, limited space affects how long the device can collect data before such a transfer is needed. Data are stored locally, and in many cases, uploaded to brand specific or open cloud–based health repositories. Accessing these data by third-party apps and comparing them is not always possible. These interoperability challenges were recently identified in a study by Arriba-Pérez et al [8]. They suggested ways to handle these issues, but they did not make any brand or device recommendations. Several studies have compared activity-tracking wearables. As an example, Kaewkannate and Kim [9] did a comparison of four popular fitness trackers in 2016. They compared devices objectively and subjectively. Data were thoroughly collected, but because of the rapid release of new devices, these four devices will be among the most popular only for a relatively short time. A comparison of brands is also of interest because brands from larger companies are, compared with small start-ups and crowd funded brands, likely to survive longer. In addition, it is of interest to know which brands support the various available programming options. Sanders et al [10] did a literature review on articles using wearables for health self-monitoring and sedentary behavior and PA detection. They reviewed various aspects of these devices, but they gave no details about device sensor support and suitability in research.

The objective of this study was to examine how the consumer market for wearables has evolved, and analyze and summarize available devices that can measure PA and heart rate (HR). Moreover, we aim to identify brands that are used extensively in research projects, and compare and consider their relevance for future studies.

Sensors

A plethora of devices promises to measure PA in new and improved ways. These devices use different sensors and algorithms to calculate human readable metrics based on sensor output. Traditional step counters use pedometers to detect daily step counts. Although cheap and energy efficient, pedometers are not as accurate as accelerometers, which is the current standard for collecting PA data [11]. All modern fitness trackers and smartwatches have an accelerometer. Compared with research tools (eg, ActiGraph [12]), these devices are considered less accurate for some measurements [13,14]. However, they are generally less invasive, cheaper, have more functionality, are more user-friendly, and are increasingly being used in research. Most accelerometer-based fitness wearables measure acceleration in three directions [15] and can be used to estimate type of movement, count steps, calculate energy expenditure (EE) and energy intensity, as well as estimate sleep patterns and more. The validity and reliability of these metrics varies. Evenson et al [14] did a review in 2015 and found high validity for steps but low validity for EE and sleep. Furthermore, they found reliability for steps, distance, EE, and sleep to be high for some devices.

In addition, some wearables have gyroscopes, magnetometers, barometers, and altimeters. A gyroscope can potentially increase device accuracy by measuring gravitational acceleration, that is, orientation and angular velocity, and better estimate which activity type a person is performing [16]. A magnetometer is a digital compass [15] and can improve motion tracking accuracy by detecting the orientation of the device relative to magnetic north. Magnetometers improve accuracy by compensating for gyroscope drift, a problem with gyroscopes where the rotation axis slowly drifts from the actual motion and must be restored regularly. Accelerometers, gyroscopes, and magnetometers are often combined into an inertial measurement unit (IMU). Most mobile phones use IMUs to calculate orientation, and an increasing number of fitness wearables include this unit to give more accurate metrics. Barometers or altimeters detect changes in altitude [15] and can be used to improve some metrics (eg, EE), as well as report additional metrics (eg, climbed floors).

Photoplethysmography (PPG) is a relatively new technique in wearables. PPG is an optical technique to estimate HR by monitoring changes in blood volume beneath the skin [17]. A light-emitting diode projects light onto the skin, which is affected by the HR and reflected back to the sensor. However, movement, ambient light, and tissue compression affect the light, resulting in signal noise, and cleaning algorithms often use accelerometer data to assist HR estimation [18]. There is some evidence that gyroscopes could be used [19] to reduce

PPG signal noise, so we are likely to see more devices in the future equipped with PPG sensors. To further enrich the PA data collection, some devices have a built in global positioning system (GPS) receiver. This is especially true for high-end fitness trackers and sports watches specifically targeting physically active people. With a GPS, it is possible to track more data, including position, speed, and altitude.

Algorithms and Mobile Apps

Raw data from sensors must be converted into readable metrics to be meaningful for the user. Many devices only display a limited set of metrics directly on the device (eg, today's step count or current HR) and rely on an accompanying mobile app to show the full range of available metrics (eg, historic daily step count and detailed HR data). Although the physical sensors in these devices are very similar, the algorithms that interpret sensor output are unique for most vendors. These algorithms are often company secrets, and they can be changed without notice. In addition, the quality and supported features of the accompanying mobile apps varies, and the total user experience will therefore differ. Each additional sensor included in a device can be used to add additional types of metrics for the user or supply internal algorithms with additional data to improve accuracy of already available metric types. However, additional sensors affect price and power consumption.

Device Types

There are many similarities between different types of devices, and they may be difficult to categorize. We will use the term wearable in this paper as a common term for wrist-worn devices that can track and share PA data with a mobile phone.

A smartwatch is a wrist-worn device that, mostly, acts as an extension to a mobile phone and can show notifications and track PA and related metrics. Modern smartwatches often include a touch screen and can support advanced features and display high resolution activity trends [15]. Fitness trackers (ie, smart band or fitness band), normally worn on the wrist or hip, are devices more dedicated to PA tracking. A fitness tracker is typically cheaper than a smartwatch because of less expensive hardware and often fewer sensors. Due to this, it generally also has better battery life and a limited interface for displaying tracking results [15].

Other terms are also used, for example, sports watch and GPS watch, which can be considered merges between smartwatches and fitness trackers. In addition, there are hybrid watches (ie, hybrid smartwatches) that have a traditional clockwork and analogue display that have been fitted with an accelerometer. An accompanying mobile app is needed to access most data, but daily step counts are often represented as an analogue gauge on the watch face.

Wearable Usage Scenario

Wearables come forward as a new alternative to tracking PA in research (compared with, eg, ActiGraph), especially when it is desired to collect measurements for a prolonged period of time. In an intervention study, continuous data collecting from wearables would allow researchers to better track changes in PA and adjust the intervention accordingly. Wearables can also be used in epidemiological research as a tool for tracking PA for an extended period. This could reveal detailed PA changes in a population over time. In both scenarios, there are several potential important requirements to consider when choosing a device for the study, including usability, battery life, price, accuracy, durability, look and feel, and data access possibilities.

Methods

Search Strategies

Brands, Devices, and Sensors

We searched six databases to create a list of relevant wearable devices: The Queen's University's Wearable Device Inventory [20], The Vandrico Wearables database [21], GsmArena [22], Wearables.com [23], SpecBucket [24], and PrisGuide [25,26]. We only used publicly available information when comparing devices. We did the search from May 15, 2017 to July 1, 2017.

We identified wearables in two steps. In step one, we identified and searched the six defined databases. In step two, we extracted all brands from the list of devices identified in step one and examined brand websites for additional devices. If we found the same device in several databases with conflicting information, we manually identified the correct information from the device's official website or other online sources (eg, Wikipedia and Google search). We removed duplicates and devices not fitting the inclusion criteria.

Brand Usage in Research

We searched Ovid MEDLINE on September 30, 2017 to determine how often the most relevant brands were used in previous studies. For each search, we performed a keyword search with no limitations set. We divided our findings into validation and reliability studies and data collection studies.

To decide which brand to consider most relevant, we did two sets of searches. In the first set, we created a brand-specific keyword search for brands that were (1) One of the five most sold brands in 2015 or 2016 or (2) Had released 10 or more unique devices. From the resulting list of articles, we screened title, abstract, and the method section. This screening was done to (1) Exclude articles out of scope and (2) To identify additional brands used in these studies. We compiled a list of these brands and performed a second set of searches, one for each new identified brand. Eleven brands were finally included. The specific keyword search used for each brand is given in the Results section where we summarize our findings.

We also searched the US National Library of Medicine database of clinical studies through the ClinicalTrials website, using the same 11 keyword searches, to determine brand usage in ongoing projects. One author did the articles screening, as well as the projects description screening in ClinicalTrials.

Brand Developer Possibilities

To determine how relevant a specific brand is when planning a new research project, we reviewed the 11 identified brands and considered available developer options, supported mobile phone environments, and options for health data storage. We especially reviewed availability of an application programming

interface (API) and a software development kit (SDK). Information was collected from Google Play, Apple's App Store, and official brand websites. Information retrieval was done in September 2017.

Inclusion and Exclusion Criteria

Brands, Devices, and Sensors

The study is limited to wrist-worn consumer devices that utilize accelerometers to measure PA. Devices capable of collecting HR from the wrist using an optical sensor were tagged as PPG devices. Devices were tagged as GPS devices only if they had a built-in GPS tracker. We only included devices meant for personal use, designed to be worn continuously (24/7), and were capable of sharing data with mobile phones through Bluetooth. The wrist-worn limitation was added because hip-worn devices are not normally worn during the night (ie, not 24/7). Only devices released before July 1, 2017 were included. We excluded hybrid watches because most hybrid vendors make a large number of watch variations, with what seems to be the same hardware. In addition, these watches are mostly available through high-end suppliers of traditional watches, at a price point that would prevent researchers from considering their use in a large study.

Brand Usage in Research

Due to the large number of available brands, we limited our search to include only the 11 brands already identified as relevant. We excluded brands that are no longer available (ie, company shut down). Review studies were also excluded.

Brand Developer Possibilities

When reviewing brand relevance in research, we only reviewed developer capabilities for the 11 brands we had already included in the list of relevant brands. We set the additional limitation that the brand was used in at least one article in Ovid MEDLINE.

Device Categorization, Data Collection, and Reporting Categories

When collecting information about wearables, we categorized them into three groups:

1. Smartwatches: a device was tagged as a smartwatch if
   - It supported mobile phone notifications, and the vendor described it as a smart watch, or if
   - It had a touch screen and was not explicitly described as a fitness tracker by the vendor.
2. Fitness trackers: we classified a device as a fitness tracker if
   - Its main purpose was to track PA, or if
   - The vendor called it a fitness tracker, or if
   - The device did not support notifications from the connected mobile phone (eg, incoming calls or texts).
3. Hybrid watches: to be considered a hybrid watch, the device had to have an analogue clockwork with a built-in digital accelerometer.

We collected the following variables for each device: brand name, device name, year of release, country of origin, device type (eg, fitness tracker), and whether they had a built-in accelerometer, gyroscope, magnetometer, barometer or altimeter, GPS, and PPG.

We looked at three aspects of the devices we identified and reported under three categories:

1. Metrics and trends: in this category, we described the status for available brands, devices, and sensors, as well as reviewed trends in sensor availability over time.
2. Brand usage in research: in this category, we searched Ovid MEDLINE and ClinicalTrials and determined which brands are most used in a research setting.
3. Brand developer possibilities: in this category, we reviewed software integration platforms and mobile platform support for the most relevant brands.

Results

Relevant Devices

An overview of the device search process is given in Figure 1. We found 572 devices by searching online and offline databases and 131 additional devices by visiting the official websites for each identified brand, totaling 703 devices. Removing duplicates left 567 unique devices. These were screened for variation, that is, the same device with different design. After excluding 41 because of variation, 526 remained and were screened for eligibility. We removed 103 devices for not fitting the inclusion criteria. The remaining 423 devices were included in the study.

Brands, Devices, and Sensors

Brands

We identified 423 unique wearables, distributed between 132 different brands. Almost half the brands (47.0%, 62/132) had only one device. Moreover, 75.0% (99/132) of brands had three or fewer devices, and 83.3% (110/132) had five or fewer devices. Brands originated from 23 different countries, but the United States (43.2%, 57/132) and China (16.7%, 22/132, mainland China; 19.0%, 25/132, including Taiwan) represented the largest number of brand origin. Each remaining country represented between 0.8% (1/132) and 5.3% (7/132) of brands.

As the market has grown and wearable technology has become increasingly popular, a number of new brands have appeared on the market. In 2011, there were only three brands available. There was a small increase in brand count in 2012 and 2013, but in 2014, we saw the largest increase with 41 new brands. The number of new brands started to decrease in 2015, with 36 new brands in 2015 and 23 in 2016. Only three new brands have been introduced in 2017, but this number only represents the first 6 months of 2017. The final count for 2017 will likely be higher. An overview of the number of new brands that appeared on the market between 2011 and 2017 is given in Figure 2. Note that some companies are no longer active and, for 17 devices, we could not determine release year.

Most brands only had a small number of wearables, but some produced a lot more. The brand with most unique wearables was Garmin (United States) with 40 different devices. No.1 (China) introduced the second highest number of wearables

with 19 devices. An overview of the release year of the 22 (out of 132) brands that have released more than five devices is given in Table 1. Seven out of these 22 brands originated in the United States, five (six including Taiwan) originated in China, and two originated in South Korea. All other countries are represented only once. Some of these brands are no longer active (eg, Pebble and Jawbone).

Devices

Three devices were released in 2011 (earliest year), seven in 2012, 30 in 2013, and 87 in 2014. The year with the highest number of new wearables was 2015, with 121 new devices. In 2016, 120 new devices were released; the first year with a decreasing number of new wearables. The number of new and accumulated devices from 2011 to 2017 is summarized in Table 2. The last column (unknown) represents devices where we could not identify the release year. The above numbers represent the total number of new devices. If grouped into fitness trackers and smartwatches, there is a small overrepresentation among new smartwatches. Up until 2014, about half of devices were smartwatches. In 2015 and 2016, smartwatches represented 59.3% (143/241) of new devices, whereas fitness trackers represented 40.6% (98/241).

Sensors

The number of sensors included in new devices have increased in the last few years. Since 2015, the order of the most common sensors has consistently been PPG, GPS, gyroscope, magnetometer, and barometer or altimeter. In addition, these sensors have had a steady increase in availability in the same period. For 2017, 71% (27/38) of new devices included a PPG sensor, 50% (19/38) included a GPS, 39% (15/38) included a gyroscope, 34% (13/38) included a magnetometer, and 32% (12/38) included a barometer or altimeter. Figure 3 gives an overview of the number of devices each year that includes each sensor, in percent of total number of released devices that year. Devices with more than one sensor are represented once for each sensor it includes.

In total, since 2011, 38.5% (163/423) of wearables have only been equipped with one sensor (accelerometer). Moreover, 29.8% (126/423) of devices had two sensors, 12.1% (51/423) had three sensors, 11.1% (47/423) had four sensors, and 6.4% (27/423) had five sensors. Only 2.1% (9/423) of devices had all six sensors. In Table 3, these numbers are broken down by sensor combination and year. Some sensor combinations do not exist and are excluded.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart.

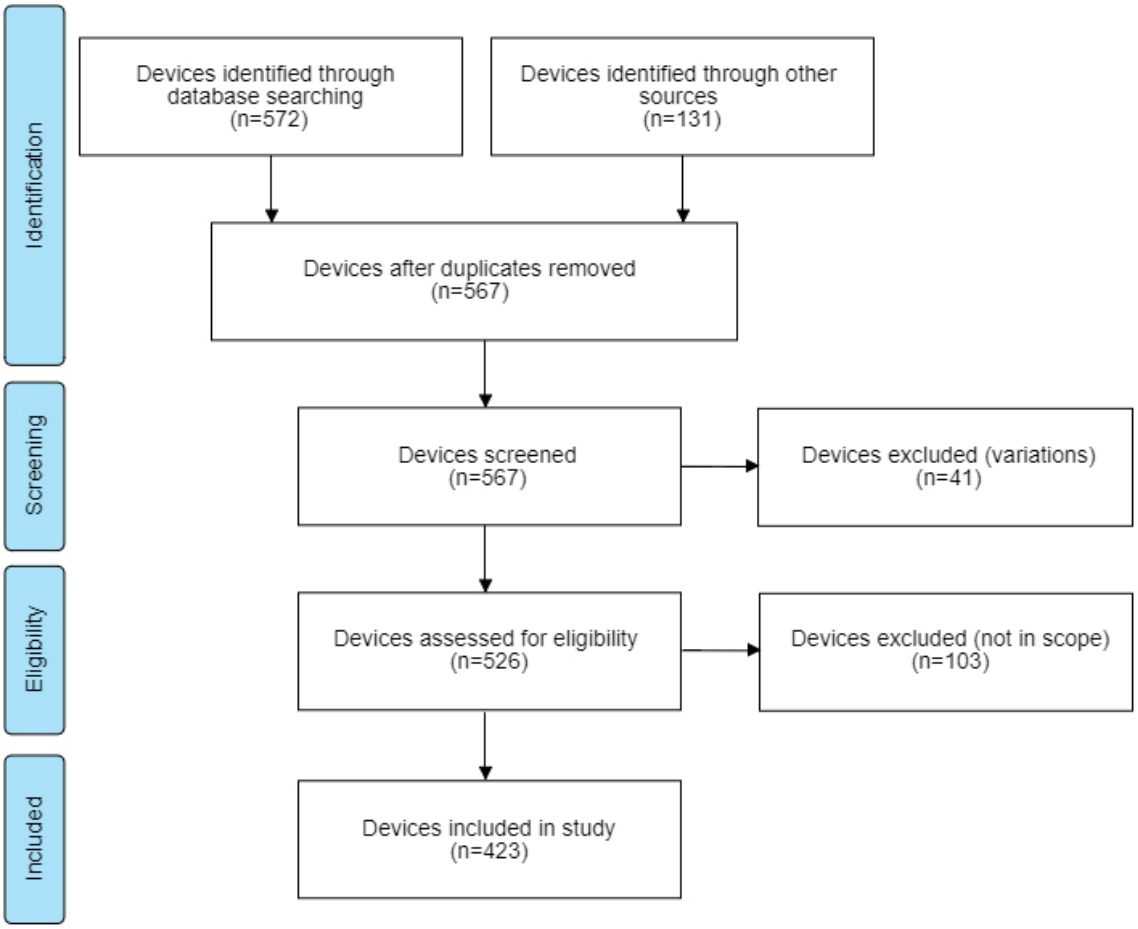


Figure 2. Number of new and aggregated available brands by year.

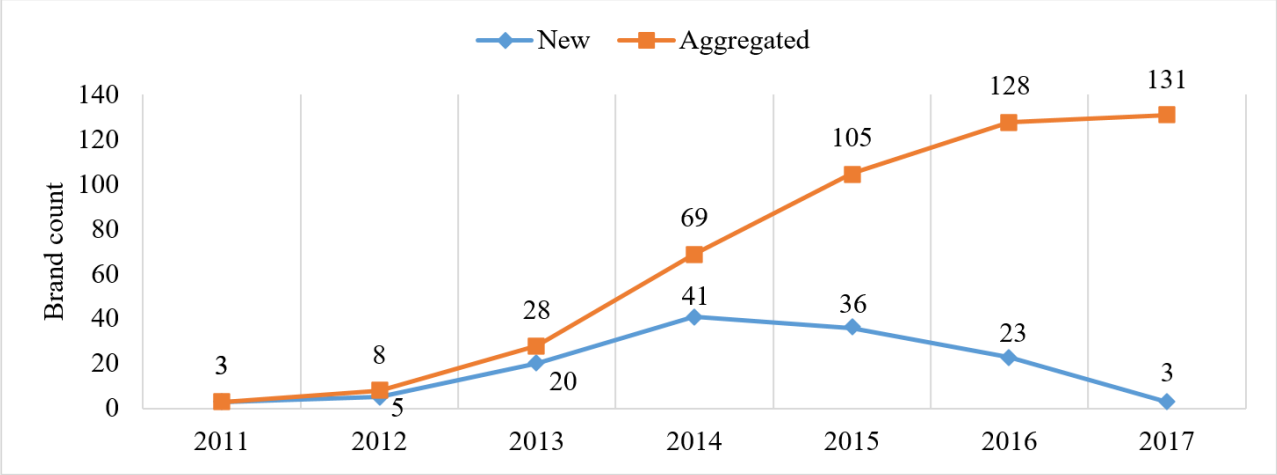


Table 1. Device count per year for brands with six or more wearables.

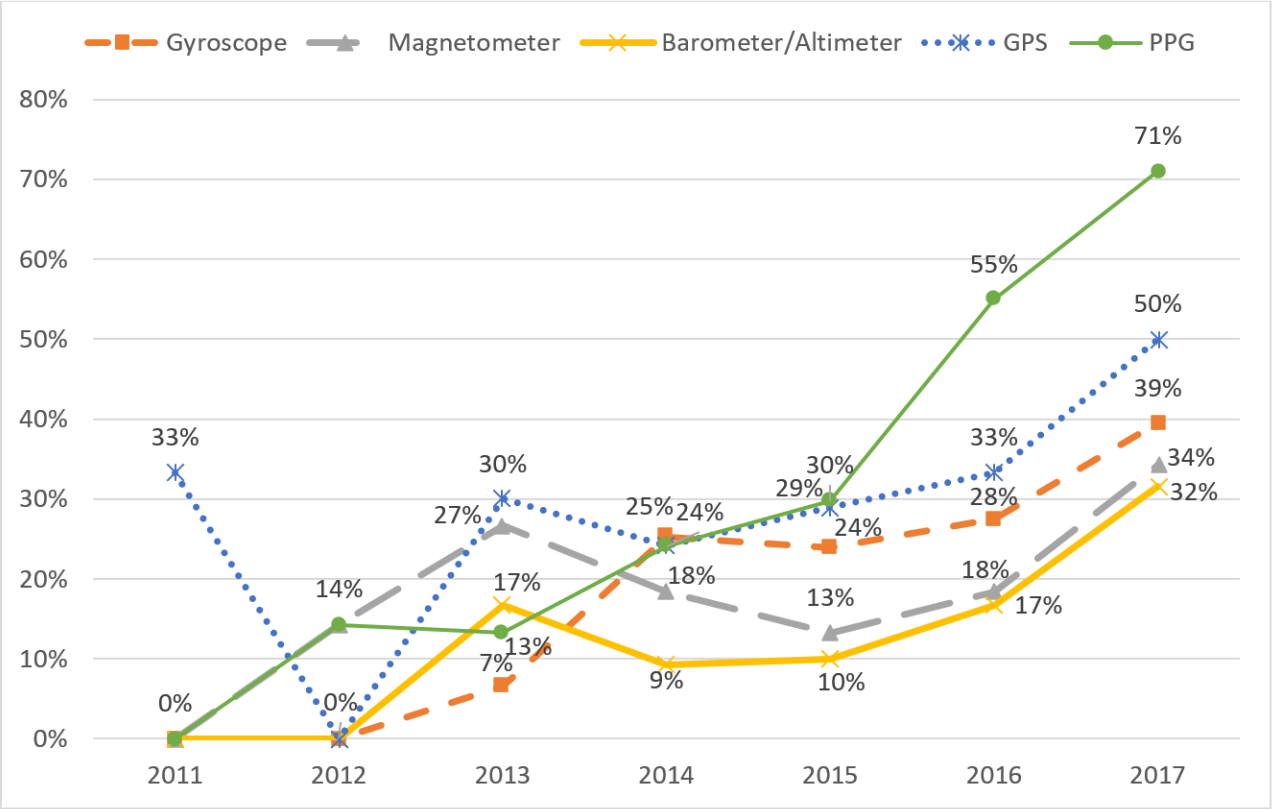
Brand	Country	2011	2012	2013	2014	2015	2016	2017	Unknown	Total[a]
Garmin	United States		1	5	6	11	13	4		40
Fitbit	United States			1	1	2	4	1		9
Misfit	United States			1	1	3	1	2		8
LifeTrak	United States			1	5		1			7
iFit	United States				1	4	1			6
Jawbone	United States	1		1	1	3				6
Pebble	United States			1	1	3	1			6
No. 1	China					5	9	5		19
Omate	China				2	5	2			9
Zeblaze	China					2	5	2		9
Huawei	China				1	3	3	1		8
Oumax	China				1	2	2	1	1	7
Mobile Action	Taiwan					2	2		4	8
Samsung	South Korea			1	6	1	4			12
LG	South Korea				3	1	1	2		7
WorldSim	England					1	1		5	7
Polar	Finland			1	2	4	2	2		11
Technaxx	Germany				4		2			6
Awatch	Italy						3	4		7
Epson	Japan				2	5				7
TomTom	Netherlands				2	1	4			7
MyKronoz	Switzerland				4	6	7	1		18

[a]Total brand count for the United States=7, China and Taiwan=6, and South Korea=2. All other countries are represented only once.

Table 2. Number of new and accumulated devices by year.

Devices	2011	2012	2013	2014	2015	2016	2017	Unknown
New	3	7	30	87	121	120	38	17
Accumulated	3	10	40	127	248	368	406	423

Figure 3. Percentage of devices released each year, supporting each sensor. GPS: global positioning system; PPG: photoplethysmography.



Brand Usage in Research

The top five vendors in 2015 [27] and 2016 [6], in sold units, were Fitbit, Xiaomi, Apple, Garmin, and Samsung. Brands with more than 10 unique wearables include Garmin, No.1, MyKronoz, Samsung, and Polar. These eight, and additional brands identified during the MEDLINE search and ClinicalTrials search, were considered. We did not find any publications or active clinical trials that used devices from No.1 or MyKronoz. Devices from Basis, BodyMedia, Pebble, Jawbone, Microsoft, and Nike were also used in some of the identified studies, but these brands do no longer produce wearables within the scope of this paper and were excluded from further analysis.

The MEDLINE search resulted in 81 included studies that we divided into two groups: (1) validation and reliability studies and (2) data collection studies. Studies where wearable output was compared with existing research instruments known to give accurate results (eg, ActiGraph) or with direct observation, as well as studies where several wearables were compared with each other for accuracy or reliability, were classified as validation and reliability studies. Studies where wearables were used as a tool for intervention or observation, to collect data on PA, HR, EE, sleep, or other available metrics, were classified as data collection studies. Out of these 81 studies, 61 were classified as validation and reliability studies, whereas 20 were classifies as data collection studies.

Fitbit devices were used in 54 studies [9,13,28-79]. Out of these, 40 studies were validation or reliability studies. In 22 of the studies, one or more Garmin devices were used [32,33,46,49,50,62,77-92]. Of these, 18 were validation or reliability studies. Eight studies used Apple devices [29,30,35,49,62,79,93,94]. Six of these were validation or reliability studies. All studies using devices from Misfit, Polar, Withings, Mio, Samsung, PulseOn, TomTom, and Xiaomi were validation or reliability studies. Misfit devices were used in 12 studies [9,36,42,43,46,61-63,85,95-97]; Polar devices were used in 6 studies [36,43,46,62,98,99]; Withings [63,85,89,100,101], Mio [29,30,54,102,103], and Samsung [29,30,58,62,96] devices were used in 5 studies; PulseOn devices were used in 4 studies [29,104-106]; TomTom devices were used in 2 studies [54,79]; and Xiaomi devices were used in 1 study [96].

From ClinicalTrials, we found that the vast majority of ongoing projects use, or are planning to use, Fitbit devices. All other devices were mentioned in three or less projects, whereas Fitbit devices were mentioned in 31 studies. A summary of these studies and projects is given in Table 4. We further grouped the validation and reliability studies into five categories. A total of 31 studies focused on step counts or distance, 15 studies researched EE, 15 studies measured HR, 10 studies measured sleep, and 7 studies collected other metrics. Multimedia Appendix 1 gives an overview of articles found in MEDLINE, which brands they included in the study, and which of the five categories they are grouped into.

Brand Developer Possibilities

Next, we considered developer possibilities for the 11 brands already identified as most relevant in research: Apple, Fitbit, Garmin, Mio, Misfit, Polar, PulseOn, Samsung, TomTom, Withings, and Xiaomi. All brands had an app in the Apple App Store and could connect to the iPhone. Except for the Apple Watch, all other brands had an app in Google Play and could be used with Android phones.

Table 3. Number and percentage of devices supporting a specific group of sensors, by year.

Sensors	2011	2012	2013	2014	2015	2016	2017
Accelerometer (Acc), n (%)	2 (67)	5 (71)	16 (53)	40 (46)	50 (41.3)	37 (30.8)	4 (11)
Acc + 1 sensor, n (%)							
PPG[a]		1 (14)	1 (3)	9 (10)	11 (9.1)	27 (22.5)	10 (26)
GPS[b]	1 (33)		2 (7)	9 (10)	15 (12.4)	3 (2.5)	
Gyroscope (Gyro)			1 (3)	3 (3)	9 (7.4)	4 (3.3)	1 (3)
Magnetometer (Mag)		1 (14)	2 (7)	1 (1)	3 (2.5)		
Barometer (Bar)				1 (1)	1 (0.8)		2 (5)
Acc + 2 sensors, n (%)							
GPS + PPG			1 (3)		7 (5.8)	6 (5)	3 (8)
Gyro + PPG				4 (5)	5 (4.1)	5 (4.2)	1 (3)
Gyro + GPS				1 (1)	2 (1.7)	2 (1.7)	
Bar + PPG			1 (3)		1 (0.8)	2 (1.7)	
Gyro + Mag				2 (2)	1 (0.8)		
Mag + GPS			1 (3)	1 (1)	1 (0.8)		
Mag + PPG						1 (0.8)	
Gyro + Bar				1 (1)			
Bar + GPS				2 (2)			
Acc + 3 sensors, n (%)							
Gyro + Mag + GPS			1 (3)	3 (3)	3 (2.5)	2 (1.7)	1 (3)
Gyro + Mag + PPG				4 (5)	2 (1.7)	3 (2.5)	1 (3)
Mag + Bar + GPS			3 (10)	2 (2)		4 (3.3)	1 (3)
Gyro + GPS + PPG				1 (1)		6 (5)	1 (3)
Bar + GPS + PPG						2 (1.7)	2 (5)
Mag + GPS + PPG						1 (0.8)	1 (3)
Gyro + Bar + PPG					2 (1.7)		
Gyro + Mag + Bar						1 (0.8)	
Acc + 4 sensors, n (%)							
Mag + Bar + GPS + PPG			1 (3)		3 (2.5)	4 (3.3)	
Gyro + Mag + GPS + PPG				1 (1)		3 (2.5)	3 (8)
Gyro + Bar + GPS + PPG					2 (1.7)	4 (3.3)	1 (3)
Gyro + Mag + Bar + GPS						1 (0.8)	2 (5)
Gyro + Mag + Bar + PPG				1 (1)	1 (0.8)		
Acc + 5 sensors, n (%)							
All sensors				1 (1)	2 (1.7)	2 (1.7)	4 (11)
Total, n	3	7	30	87	121	120	38

[a]PPG: photoplethysmography.

[b]GPS: global positioning system.

Table 4. Number of identified articles in Medical Literature Analysis and Retrieval System Online (MEDLINE) and ClinicalTrials.

Brand	MEDLINE[a] search term	MEDLINE		ClinicalTrials	
		Validation or reliability studies[b] (total article count=61)	Data collection studies[c] (total article count=20)	Validation or reliability studies[d]	Data collection studies[e]
Fitbit	Fitbit AND (Alta OR Blaze OR Charge OR Flex OR Surge)	40	14	1	30
Garmin	Garmin AND (Approach OR D2 OR Epix OR Fenix OR Forerunner OR Quatix OR Swim OR Tactix OR Vivo*)	18	4	1	2
Misfit	Misfit AND (Flare OR Flash OR Link OR Ray OR Shine OR Vapor)	12	0	0	1
Apple	Apple watch	6	2	1	1
Polar	Polar AND ("Polar Loop" OR M200 OR M4?0 OR M600 OR V800 OR A3?0)	6	0	1	3
Withings	Withings	5	0	0	2
Mio	Mio Alpha OR Mio Fuse OR Mio Slice	5	0	1	2
Samsung	Samsung Gear NOT "Gear VR" NOT Oculus	5	0	0	2
PulseOn	PulseOn	4	0	0	1
TomTom	TomTom	2	0	1	
Xiaomi	Xiaomi	1	0	0	1

[a]MEDLINE: Medical Literature Analysis and Retrieval System Online.

[b]Number of validation or reliability studies in MEDLINE.

[c]Number of data collection studies in MEDLINE.

[d]Number of validation or reliability studies in ClinicalTrials.

[e]Number of data collection studies in ClinicalTrials.

Three brands supported Windows Phone: Fitbit, Garmin, and Misfit. Apple Health and Google Fit are the two most common open cloud health repositories. Mio, Misfit, Polar, Withings, and Xiaomi, were the only brands that automatically synchronized fitness data to both of these repositories through these open APIs. The Apple Watch only synchronized automatically to the Apple Health repository. Seven out of 11 brands had a private cloud repository with an accompanying API, which allows third-party apps to access these data. Five brands had an SDK, which makes it possible to create custom programs to communicate with the device or create watch faces that can run on the device.

The Apple Watch was the only device running on watchOS. Three brands had at least one device running on Android Wear. The remaining seven brands used a custom system. A summary of all attributes for each brand is given in Table 5. Not all devices for a specific brand support all features. In addition, this is a snapshot of the status of these attributes, which are likely to change over time as new devices and brands expand their capabilities. The Apple Watch development environment is called WatchKit SDK and can be used to write apps for the Apple Watch [107]. Apple's health storage solution is called Apple Health. A variety of different data types can be stored here and accessed by third-party developers through the HealthKit API [108]. Access to any of these services requires enrollment in the Apple Developer Program, which currently costs US $99 per year.

Fitbit offers three major SDKs (Device API, Companion API, and Settings API) for developing apps for Fitbit devices. In addition, Fitbit offers the Web API that can be used to access Fitbit cloud-stored fitness data. The Web API exposes six types of data: PA, HR, location, nutrition, sleep, and weight [109]. Fitbit also has a solution for accessing high-resolution step and HR data (ie, intraday data), granted on a case by case basis. There is no cost for developing with the Fitbit SDKs or API.

There are two generations of programmable Garmin wearables [110]. The Connect IQ SDK can be used by both generations, but devices using the newer Connect IQ 2 generation support more features. Development with this SDK is free. Garmin also offers a cloud-based Web API, Garmin Connect, which allows third-party apps to access users' cloud-based fitness data. Access to this API costs US $5000 (one-time license). In addition, Garmin maintains a separate Health API intended to be used by companies for wellness improvement of their employees. This API is free but requires a manual approval from Garmin.

Table 5. Brand environment, integration, and development support.

Feature	Apple	Fitbit	Garmin	Mio	Misfit	Polar	PulseOn	Samsung	TomTom	Withings	Xiaomi
Supported platform											
Android		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
iPhone	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Windows phone		✓	✓		✓						
Integration											
Automatic synchronization to Apple Health	✓			✓	✓	✓				✓	✓
Automatic synchronization to Google Fit				✓	✓	✓				✓	✓
Private cloud storage		✓	✓		✓	✓		✓	✓	✓	
Cloud storage API[a]	✓	✓	✓		✓	✓		✓	✓	✓	
Developer SDK[b]	✓	✓	✓		✓			✓			
Watch system											
Android Wear					✓	✓					✓
watchOS (Apple)	✓										
Custom		✓	✓	✓			✓	✓	✓	✓	

[a]API: application programming interface.

[b]SDK: software development kit.

The Misfit developer ecosystem consists of three SDKs (Sleep SDK, Link SDK, and Device SDK) [111]. The Misfit Device SDK is the major SDK for developing apps for and communication with Misfit devices. This SDK is only available on request. Misfit also offers the Misfit Scientific Library that can be used to access Misfits proprietary sensor algorithms directly. This library is also only available on request. In addition, the Misfit Cloud API is used to access users' data from the Misfit cloud server. All SDKs and the API are free.

Polar does not offer a separate SDK. Polar devices can integrate with Google Fit and Apple Health and deposits collected data there [112]. This data are accessed using Google Fit APIs and Apple HealthKit APIs. In addition, data are uploaded to Polar's cloud storage, which is accessible by third-party developers through the AccessLink API. Besides PA data (steps, EE, and sleep), basic training data are also stored here. Access to AccessLink is free.

Development for a Samsung smartwatch is done using the Tizen SDK (Samsung smartwatch operating system is called Tizen). The Samsung Health SDK platform consists of two parts: Data SDK and Service SDK. Together these can be used to store and access health data collected from internal and external sensors, as well as third-party apps running on a Samsung watch or a mobile phone. Development using any of these services is free [113].

TomTom offers the Sports Cloud API for accessing data collected from TomTom devices. The API provides four types of data: PA (eg, exercises bouts), HR, tracking (eg, steps and EE), and physiology (eg, weight). Access to the API is free [114].

Nokia acquired Withings in 2016, and the original Withings API is now available as the Nokia Health API. Besides PA and sleep measurements, the API also gives access to intraday PA data. Nokia must manually approve access to this high-resolution activity API. The API is free [115].

Summarizing Results

Which features are most important when considering devices for a research project will depend on the purpose and design of the study. It is therefore not possible to identify one brand as the best brand in all circumstances. However, we have tried to quantify various aspects of a brand to identify and summarize their benefits.

Table 6. Brand summary.

Brand	Fitbit	Garmin	Misfit	Apple	Polar	Samsung	Withings	Mio	PulseOn	TomTom	Xiaomi	MyKronoz	No. 1
Devices[a]	9	40	8	3	11	12	2	3	1	7	3	18	19
MEDLINE[b]	54	22	12	8	6	5	5	5	4	2	1		
Validation or reliability[c]	40	18	12	6	6	5	5	5	4	2	1		
Steps	21	10	6	1	2	2	4				1		
Energy expenditure	10	4	3	4	3	1		2	2				
Heart rate	7	4	1	4	1	2		5	4	2	1		
Sleep	8	1	4		1		2						
Other	3	4	2		1								
ClinicalTrials[d]	31	3	1	2	4	2	2	3	1	1	1		
SDK[e]	✓	✓	✓	✓	✓	✓							
API[f]	✓	✓	✓	✓	✓	✓	✓			✓			
Apple Health[g]			✓	✓	✓		✓	✓			✓		
Google Fit[h]			✓		✓		✓	✓			✓		

[a]Number of unique devices.

[b]MEDLINE: Medical Literature Analysis and Retrieval System Online. Number of articles in MEDLINE.

[c]Number of validation or reliability studies in MEDLINE, grouped by metric (step, EE, HR, sleep, and others).

[d]Number of active projects in ClinicalTrials.

[e]Supports an SDK for third-party software implementation.

[f]API: application programming interface. Supports an API for developer access to data cloud.

[g]Supports automatic synchronization to Apple Health data cloud.

[h]Supports automatic synchronization to Google Fit data cloud.

We used eight categories in this custom comparison, which we suggest to consider before deciding on a brand for any research project:

1. Device count: a higher number of available devices make it possible to pick a device that is more tailored to the study.
2. Article count: a higher number of articles in Ovid MEDLINE indicate usage in previous studies.
3. Validation or reliability count: a high number of validation or reliability studies provides knowledge about device and brand accuracy.
4. ClinicalTrials count: a high number of active projects in ClinicalTrials indicate brand relevance.
5. SDK support: brands that allows third-party programs to run on their devices or communicate directly with the device, by offering an SDK, adds more possibilities for customization.
6. API support: brands that allows third-party programs to access the data cloud repository, by offering API access, adds more possibilities for health data collection and retrieval.
7. Apple Health: brands supporting automatic synchronization to Apple Health allow usage of Apple HealthKit API.
8. Google Fit: brands supporting automatic synchronization to Google Fit allow usage of Google Fit API.

A consensus between authors was reached to include these specific categories because we think together they indicate how often a specific brand has been used in the past and will be used in the future, and they show which options are available for data extraction. These are not the only possible categories, and each category will not be equally important for all studies.

Table 6 gives a summary of these categories for each brand. A transposed Excel (Microsoft) version for dynamic sorting is given in Multimedia Appendix 2. We have divided MEDLINE validation and reliability studies into subgroups, making it easier to compare brands for specific study purposes.

Discussion

Availability and Trends

The number of new brands increased every year from 2011 to 2014, but from 2015 to 2016, we saw a decrease in the number of new brands. The number of new devices also increased from 2011 to 2015, with a slight reduction in 2016. Many new and existing companies have tried to enter the wearable market during these years. Some have become popular, whereas others are no longer available. The number of new devices in the first two quarters of 2017 seems low, and there is a small indication that the number of new brands and devices released each year is declining. During the data collection phase, we also identified

a large number of hybrid watches. Although we did not report on these, this relatively new branch of wearables has grown in popularity. The Fossil group, representing 19 brands, recently announced they would launch more than 300 hybrid watches and smartwatches in 2017 [116]. Most of these will be hybrids, and 2017 may see the highest number of new hybrids released to date.

We only found nine devices that support all five sensors considered in this study. Among the 11 most relevant brands, only Fitbit Surge, Garmin Forerunner 935, Garmin Quatix 5, Samsung Gear S, and TomTom Adventure fall in this category. Most devices (68%) support only one sensor, in addition to the accelerometer. These numbers indicate that sensor count is not the main argument when choosing a device for personal use. In addition to the accelerometer, the most common sensors are PPG and GPS, regardless of sensor count. One reason for this may be that the added benefit of having these sensors, in a fitness setting, is very clear. Accelerometers can be used for step counting, PA intensity, exercise detection, and other well-understood metrics, whereas the added benefit of a gyroscope may be less intuitive. The added convenience of using a PPG compared with a pulse chest strap, or no HR detection at all, is also easy to understand. Adding a GPS also adds some easy-to-understand benefits, where tracking progress on a map and the possibility to detect speed is the most obvious. Magnetometers and barometers or altimeters may not be sensors that most people consider relevant for PA, although they can be used to enhance accuracy of EE and other metrics.

Brand Usage in Research

In the MEDLINE literature search, we found 81 studies that used one or more of the 11 brands we identified as most relevant in research. Out of these, 61 were validation or reliability studies. The remaining 20 studies used wearable devices as data collection instruments to measure PA, HR, EE, sleep, or other metrics. Fitbit was used in twice as many validation or reliability studies as any other brand. This has likely contributed to the high number of studies where Fitbit was used as the only instrument for health data collection. The same trend will likely continue in future publications because numbers from ClinicalTrials for active projects shows an overrepresentation of Fitbit-enabled projects. Of the brands currently available, the five most often used in research projects are Fitbit, Garmin, Misfit, Apple, and Polar. In addition, these brands have all existed for several years and have either released a large number of unique devices or shipped a large number of total devices. As such, they are likely to stay on the market for the near future.

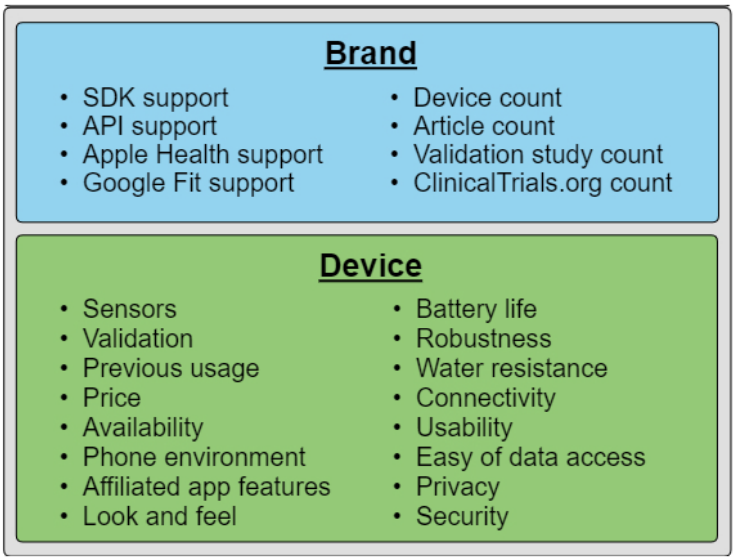
A high article count, high number of validation or reliability studies, or high number of studies in ClinicalTrials for a specific brand does not automatically imply validity or reliability. It does, however, show researcher interest in these brands.

Implication for Practice

Table 6 is a good starting point when considering brands for a new research project. Article count, validation or reliability study count, and ClinicalTrials count together indicate brand dependability. Larger numbers indicate how relevant, usable, and valid previous researchers have found each brand to be. In projects where it is relevant, SDK support allows programmatic interaction directly with the device. API support allows storage in, and access to, a brand-specific cloud-based health data repository. Apple Health and Google Fit support are alternative solutions for storing and accessing health data in an open cloud repository. For projects that require multiple brand support, using open solutions reduces the need to implement specific software for each brand. SDK, API, Apple Health, and Google Fit must be supported on both the brand and device level, however.

A high brand device count makes it easier to find a device that best supports the study needs. In addition to available sensors (ie, metrics), validation, and previous usage in research, several other potential relevant criteria exist, including price, availability, phone environment support, affiliated app features, look and feel, battery life, build quality or robustness, water resistance, connectivity, and usability.

Figure 4. Criteria to consider when choosing brand or device. API: application programming interface; SDK: software development kit.



Furthermore, projects that need programmatic access to the wearable or stored health data should especially consider SDK or API features and ease of use, as well as privacy and security. Figure 4 gives a summary of criteria to consider when selecting brand and device.

Limitations

We visited all the brands' websites to find additional devices, but several sites did not contain any information about discontinued devices. The release year of a device was rarely available on device webpages, and we had to search for reviews and other sources to find this information. The level of detail in device hardware specifications varied. Some vendors did not specify which sensor they included in their devices and only mentioned which features the device had. In some cases, the sensor could be derived from this information, but in other cases, we had to find this information elsewhere. Wikipedia was also used to collect sensor support and release year for some devices. This open editable encyclopedia is not necessarily always updated with correct information. For these reasons, there may be some inaccuracies in reported sensor support and release year. We did not collect information about device discontinuation. Reported numbers for total available devices does, therefore, not reflect the numbers of devices that currently can be store bought but rather the number of unique devices that have existed at some point.

Conclusions

In the last few years, we have seen a large increase in available brands and wearable devices, and more devices are released with additional sensors. However, for activity tracking, some sensors are more relevant than others are. In this study, we have focused on sensor support, health data cloud integration, and developer possibilities; because we find these to be most relevant for collection of PA data in research. However, deciding which wearable to use will depend on several additional factors.

The wearable landscape is constantly changing as new devices are released and as new vendors enter or leave the market, or are acquired by larger vendors. What currently are considered relevant devices and brands will therefore change over time, and each research project should carefully consider which brand and device to use. As a tool for future research, we have defined a checklist of elements to consider when making this decision.

Acknowledgments

The authors would like to thank Vandrico Solutions Inc for giving them API access to their database. They would also like to thank Steven Richardson and Debra Mackinnon at Queen's University for giving them an excerpt from their offline wearable database. The publication charges for this study have been funded by a grant from the publication fund of the University of Tromsø – The Arctic University of Norway.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of MEDLINE articles included in the results for "Brand usage in research".

[XLSX File (Microsoft Excel File), 24KB - jmir_v20i3e110_app1.xlsx]

Multimedia Appendix 2

Summary of the most important categories to consider when selecting a wearable brand for research.

[XLSX File (Microsoft Excel File), 13KB - jmir_v20i3e110_app2.xlsx]

References

1. World Health Organization. Physical activity URL: http://www.who.int/mediacentre/factsheets/fs385/en [accessed 2017-09-11] [WebCite Cache ID 6tOYSnXaX]
2. Emaus A, Degerstrøm J, Wilsgaard T, Hansen BH, Dieli-Conwright CM, Furberg AS, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. Scand J Public Health 2010 Nov;38(5 Suppl):105-118. [doi: 10.1177/1403494810378919] [Medline: 21062845]
3. World Health Organization. 2017. Global Strategy on Diet, Physical Activity and Health URL: http://www.who.int/dietphysicalactivity/pa/en [accessed 2017-09-11] [WebCite Cache ID 6tOXoFxbF]
4. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
5. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. J Am Med Assoc 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]

6. IDC. Wearables Aren't Dead, They're Just Shifting Focus as the Market Grows 16.9% in the Fourth Quarter, According to IDC URL: https://www.idc.com/getdoc.jsp?containerId=prUS42342317 [accessed 2018-03-01] [WebCite Cache ID 6xZyWGaxT]
7. IDC. Xiaomi and Apple Tie for the Top Position as the Wearables Market Swells 17.9% During the First Quarter URL: https://www.idc.com/getdoc.jsp?containerId=prUS42707517 [accessed 2018-03-01] [WebCite Cache ID 6xZypbatQ]
8. de Arriba-Perez F, Caeiro-Rodríguez M, Santos-Gago JM. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. Sensors (Basel) 2016 Sep 21;16(9) [FREE Full text] [doi: 10.3390/s16091538] [Medline: 27657081]
9. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
10. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
11. Corder K, Brage S, Ekelund U. Accelerometers and pedometers: methodology and clinical application. Curr Opin Clin Nutr Metab Care 2007 Sep;10(5):597-603. [doi: 10.1097/MCO.0b013e328285d883] [Medline: 17693743]
12. ActiGraph Corp. ActiGraph URL: http://actigraphcorp.com/ [accessed 2017-09-11] [WebCite Cache ID 6tOYrA9Ke]
13. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
14. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
15. Richardson S, Mackinnon D. Left to their own devices? Privacy implications of wearable technology in Canadian workplaces URL: http://www.sscqueens.org/publications/left-to-their-own-devices [accessed 2018-03-01] [WebCite Cache ID 6xZzwLAN6]
16. Wagenaar RC, Sapir I, Zhang Y, Markovic S, Vaina LM, Little TD. Continuous monitoring of functional activities using wearable, wireless gyroscope and accelerometer technology. Conf Proc IEEE Eng Med Biol Soc 2011;2011:4844-4847. [doi: 10.1109/IEMBS.2011.6091200] [Medline: 22255423]
17. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
18. Kim SH, Ryoo DW, Bae C. Adaptive noise cancellation Using accelerometers for the PPG signal from forehead. Conf Proc IEEE Eng Med Biol Soc 2007;2007:2564-2567. [doi: 10.1109/IEMBS.2007.4352852] [Medline: 18002518]
19. Casson AJ, Vazquez Galvez A, Jarchi D. Gyroscope vs. accelerometer measurements of motion from wrist PPG during physical exercise. ICT Express 2016 Dec;2(4):175-179. [doi: 10.1016/j.icte.2016.11.003]
20. Richardson SM, Mackinnon D. Queen's University. 2017. Wearable Device Inventory, Queen's University.
21. Vandrico. 2017. The wearables database URL: http://vandrico.com/wearables/ [accessed 2017-07-08] [WebCite Cache ID 6roNvEjdE]
22. GSM Arena. 2017. URL: http://www.gsmarena.com/results.php3?sFormFactors=8 [accessed 2017-07-08] [WebCite Cache ID 6roO1WCdv]
23. Wearables. 2017. Wearables.Com: Helping you make sense of wearable tech URL: http://www.wearables.com/devices/ [accessed 2017-07-08] [WebCite Cache ID 6roO94rOu]
24. Smartwatches. 2017. SpecBucket: Smart Watches URL: https://smartwatches.specbucket.com/ [accessed 2017-07-08] [WebCite Cache ID 6roOBKrxr]
25. Prisguiden. 2017. Smartklokke URL: https://prisguiden.no/kategorier/smartklokke [accessed 2017-07-08] [WebCite Cache ID 6roOHtXzf]
26. Prisguiden. 2017. Aktivitetsarmband URL: https://prisguiden.no/kategorier/aktivitetsarmband [accessed 2017-07-08] [WebCite Cache ID 6roOJt6HJ]
27. Business Wire. 2016. The Worldwide Wearables Market Leaps 126.9% in the Fourth Quarter and 171.6% in 2015 URL: https://www.businesswire.com/news/home/20160223005496/en/Worldwide-Wearables-Market-Leaps-126.9-Fourth-Quarter [accessed 2008-03-01] [WebCite Cache ID 6xa0wRrvV]
28. Balto JM, Kinnett-Hopkins DL, Motl RW. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. Mult Scler J Exp Transl Clin 2016;2:2055217316634754 [FREE Full text] [doi: 10.1177/2055217316634754] [Medline: 28607720]
29. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2) [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
30. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]

31. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
32. Chen MD, Kuo CC, Pellegrini CA, Hsu MJ. Accuracy of wristband activity monitors during ambulation and activities. Med Sci Sports Exerc 2016 Dec;48(10):1942-1949. [doi: 10.1249/MSS.0000000000000984] [Medline: 27183123]
33. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
34. Sprint G, Cook D, Weeks D, Dahmen J, La Fleur A. Analyzing sensor-based time series data to track changes in physical activity during inpatient rehabilitation. Sensors (Basel) 2017 Sep 27;17(10) [FREE Full text] [doi: 10.3390/s17102219] [Medline: 28953257]
35. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
36. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
37. Gaudet J, Gallant F, Bélanger M. A bit of fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
38. Jacobsen RM, Ginde S, Mussatto K, Neubauer J, Earing M, Danduran M. Can a home-based cardiac physical activity program improve the physical function quality of life in children with fontan circulation? Congenit Heart Dis 2016;11(2):175-182. [doi: 10.1111/chd.12330] [Medline: 26879633]
39. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the Raising Awareness of Physical Activity (RAW-PA) Study. BMC Public Health 2017 Jan 04;17(1):6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
40. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
41. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. J Med Eng Technol 2016 Aug;40(6):325-329. [doi: 10.1080/03091902.2016.1197978] [Medline: 27280592]
42. Bellone GJ, Plano SA, Cardinali DP, Chada DP, Vigo DE, Golombek DA. Comparative analysis of actigraphy performance in healthy young subjects. Sleep Sci 2016;9(4):272-279 [FREE Full text] [doi: 10.1016/j.slsci.2016.05.004] [Medline: 28154740]
43. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
44. Lee HA, Lee HJ, Moon JH, Lee T, Kim MG, In H, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. Psychiatry Investig 2017 Mar;14(2):179-185 [FREE Full text] [doi: 10.4306/pi.2017.14.2.179] [Medline: 28326116]
45. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
46. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
47. Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol 2017 Feb;264(2):316-326. [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
48. Morhardt DR, Luckenbaugh A, Goldstein C, Faerber GJ. Determining resident sleep during and after call with commercial sleep monitoring devices. Urology 2017 Aug;106:39-44. [doi: 10.1016/j.urology.2017.03.059] [Medline: 28502597]
49. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
50. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the Future Patient Trial, Phase I. Sensors (Basel, Switzerland) 2017;17(1):1-11.
51. Bian J, Guo Y, Xie M, Parish AE, Wardlaw I, Brown R, et al. Exploring the association between self-reported asthma impact and Fitbit-derived sleep quality and physical activity measures in adolescents. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e105 [FREE Full text] [doi: 10.2196/mhealth.7346] [Medline: 28743679]
52. Coughlin SS, Hatzigeorgiou C, Anglin J, Xie D, Besenyi GM, De Leo G, et al. Healthy lifestyle intervention for adult clinic patients with type 2 diabetes mellitus. Diabetes Manag (Lond) 2017;7(2):197-204 [FREE Full text] [Medline: 28794802]

53. Åkerberg A, Koshmak G, Johansson A, Lindén M. Heart rate measurement as a tool to quantify sedentary behavior. Stud Health Technol Inform 2015;211:105-110. [Medline: 25980855]
54. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):1-7 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106]
55. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How accurate is your activity tracker? A comparative study of step counts in low-intensity physical activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106 [FREE Full text] [doi: 10.2196/mhealth.6321] [Medline: 28801304]
56. Winslow BD, Nguyen N, Venta KE. Improved mental acuity forecasting with an individualized quantitative sleep model. Front Neurol 2017;8:160 [FREE Full text] [doi: 10.3389/fneur.2017.00160] [Medline: 28487671]
57. Spiotta AM, Fargen KM, Denham SL, Fulton ME, Kellogg R, Young E, et al. Incorporation of a physical education and nutrition program into neurosurgery: a proof of concept pilot program. Neurosurgery 2016 Oct;79(4):613-619. [doi: 10.1227/NEU.0000000000001358] [Medline: 27465847]
58. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile device accuracy for step counting across age groups. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e88 [FREE Full text] [doi: 10.2196/mhealth.7870] [Medline: 28659255]
59. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
60. Schoenfelder E, Moreno M, Wilner M, Whitlock KB, Mendoza JA. Piloting a mobile health intervention to increase physical activity for adolescents with ADHD. Prev Med Rep 2017 Jun;6:210-213 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.003] [Medline: 28373931]
61. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
62. Fokkema T, Kooiman TJ, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
63. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 Dec 05;16(5) [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
64. Chen JL, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017 Aug 02;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
65. Reichardt LA, Aarden JJ, van Seben R, van der Schaaf M, Engelbert RH, Bosch JA, Hospital-ADL study group. Unravelling the potential mechanisms behind hospitalization-associated disability in older patients; the Hospital-Associated Disability and impact on daily Life (Hospital-ADL) cohort study protocol. BMC Geriatr 2016 Mar 05;16:59 [FREE Full text] [doi: 10.1186/s12877-016-0232-3] [Medline: 26945587]
66. Laranjo L, Lau AY, Martin P, Tong HL, Coiera E. Use of a mobile social networking intervention for weight management: a mixed-methods study protocol. Br Med J Open 2017 Jul 12;7(7):e016665 [FREE Full text] [doi: 10.1136/bmjopen-2017-016665] [Medline: 28706104]
67. Pumper MA, Mendoza JA, Arseniev-Koehler A, Holm M, Waite A, Moreno MA. Using a Facebook group as an adjunct to a pilot mHealth physical activity intervention: a mixed methods approach. Stud Health Technol Inform 2015;219:97-101. [Medline: 26799887]
68. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017 Aug 01;217:299-305. [doi: 10.1016/j.jad.2017.04.030] [Medline: 28448949]
69. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
70. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. J Gerontol A Biol Sci Med Sci 2016 Jun 02;72(2):229-236. [doi: 10.1093/gerona/glw098] [Medline: 27257217]
71. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Feb 23;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
72. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, et al. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. BMC Res Notes 2016 Apr 12;9:213 [FREE Full text] [doi: 10.1186/s13104-016-2020-8] [Medline: 27068022]

73. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
74. Kang SG, Kang JM, Ko KP, Park SC, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009] [Medline: 28606497]
75. Voss C, Gardner RF, Dean PH, Harris KC. Validity of commercial activity trackers in children with congenital heart disease. Can J Cardiol 2017 Jun;33(6):799-805. [doi: 10.1016/j.cjca.2016.11.024] [Medline: 28347581]
76. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
77. Treacy D, Hassett L, Schurr K, Chagpar S, Paul S, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Phys Ther 2017;97(5):581-588.
78. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025]
79. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
80. Woodman JA, Crouter SE, Bassett Jr DR, Fitzhugh EC, Boyer WR. Accuracy of consumer monitors for estimating energy expenditure and activity type. Med Sci Sports Exerc 2017;49(2):371-377. [Medline: 27580155]
81. Bronikowski M, Bronikowska M, Glapa A. Do they need goals or support? A report from a goal-setting intervention using physical activity monitors in youth. Int J Environ Res and Public Health 2016;13(9):1-12. [Medline: 27649219]
82. Jones AP, Coombes EG, Griffin SJ, van Sluijs EM. Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems. Int J Behav Nutr Phys Act 2009;6:42. [Medline: 19615073]
83. Mooney R, Quinlan LR, Corley G, Godfrey A, Osborough C, Olaighin G. Evaluation of the Finis Swimsense® and the Garmin SwimTM activity monitors for swimming performance and stroke kinematics analysis. PloS One 2017;12(2):e0170902. [Medline: 28178301]
84. An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
85. Ehrler F, Weber C, Lovis C. Influence of pedometer position on pedometer accuracy at various walking speeds: a comparative study. J Med Internet Res 2016 Oct 06;18(10):e268 [FREE Full text] [doi: 10.2196/jmir.5916] [Medline: 27713114]
86. Ammann R, Taube W, Neuhaus M, Wyss T. The influence of the gait-related arm swing on elevation gain measured by sport watches. J Hum Kinet 2016;51:53-60. [doi: 10.1515/hukin-2015-0170]
87. Eyre EL, Duncan MJ, Birch SL, Cox V, Blackett M. Physical activity patterns of ethnic children from low socio-economic environments within the UK. J Sports Sci 2015;33(3):232-242. [Medline: 24998418]
88. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. Physiother Theory Pract 2017;33(10):788-796. [Medline: 28777710]
89. O'Connell S, ÓLaighin G, Kelly L, Murphy E, Beirne S, Burke N, et al. These shoes are made for walking: sensitivity performance evaluation of commercial activity monitors under the expected conditions and circumstances required to achieve the international daily step goal of 10,000 steps. PLoS One 2016;11(5):e0154956 [FREE Full text] [doi: 10.1371/journal.pone.0154956] [Medline: 27167121]
90. Jennings CA, Berry TR, Carson V, Culos-Reed SN, Duncan MJ, Loitz CC, et al. UWALK: the development of a multi-strategy, community-wide physical activity program. Transl Behav Med 2017;7(1):16-27. [doi: 10.1007/s13142-016-0417-5]
91. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. J Med Eng Technol 2017;41(3):208-215. [Medline: 27919170]
92. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. J Med Eng Technol 2017;41(6):480-485. [doi: 10.1080/03091902.2017.1333166]
93. O'Brien A, Schlosser RW, Shane HC, Abramson J, Allen AA, Flynn S, et al. Brief report: just-in-time visual supports to children with autism via the Apple watch®: a pilot feasibility study. J Autism Dev Disord 2016 Dec;46(12):3818-3823. [doi: 10.1007/s10803-016-2891-5] [Medline: 27573856]
94. Cook S, Stauffer JC, Goy JJ, Graf D, Puricel S, Frobert A, et al. Heart rate never lies: interventional cardiologist and Braude's quote revised. Open Heart 2016;3(1):e000373 [FREE Full text] [doi: 10.1136/openhrt-2015-000373] [Medline: 26835145]
95. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR mHealth uHealth 2016;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]

96. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
97. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
98. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
99. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 2016 Mar;116(3):563-571 [FREE Full text] [doi: 10.1007/s00421-015-3303-9] [Medline: 26708360]
100. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Jun;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]
101. O'Connell S, ÓLaighin G, Quinlan LR. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One 2017;12(1):e0169616 [FREE Full text] [doi: 10.1371/journal.pone.0169616] [Medline: 28085918]
102. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
103. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
104. Parak J, Uuskoski M, Machek J, Korhonen I. Estimating heart rate, energy expenditure, and physical performance with a wrist photoplethysmographic device during running. JMIR mHealth uHealth 2017 Jul 25;5(7):e97 [FREE Full text] [doi: 10.2196/mhealth.7437] [Medline: 28743682]
105. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:430-433. [doi: 10.1109/EMBC.2015.7318391] [Medline: 26736291]
106. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:8099-8102. [doi: 10.1109/EMBC.2015.7320273] [Medline: 26738173]
107. Apple. 2017. WatchOS - Apple Developer URL: https://developer.apple.com/watchos [accessed 2017-09-19] [WebCite Cache ID 6tamfJinU]
108. Apple. 2017. HealthKit - Apple Developer URL: https://developer.apple.com/healthkit/ [accessed 2017-09-13] [WebCite Cache ID 6tRfiSKVw]
109. Fitbit. 2017. Fitbit Developer URL: https://dev.fitbit.com/ [accessed 2017-09-22] [WebCite Cache ID 6tfGRUrmR]
110. Garmin. Garmin Developers URL: https://developer.garmin.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRfk6phg]
111. Misfit. Misfit Developer Toolkit URL: https://build.misfit.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRgRKOAN]
112. Polar. Polar AccessLink URL: https://www.polar.com/en/connect_with_polar/polar_accesslink [accessed 2017-09-19] [WebCite Cache ID 6tamdLu6Q]
113. Samsung. Samsung Health - Samsung Developers URL: http://developer.samsung.com/health [accessed 2017-09-13] [WebCite Cache ID 6tRftLAHj]
114. TomTom. TomTom Developer Portal URL: https://developer.tomtom.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRgHGW0A]
115. Nokia. Nokia Health API Developer Documentation URL: https://developer.health.nokia.com/api [accessed 2017-09-13] [WebCite Cache ID 6tRgHmHhR]
116. PR Newswire Association. Fossil Group Reimagines the Watch URL: https://www.prnewswire.com/news-releases/fossil-group-reimagines-the-watch-300427943.html [accessed 2018-03-01] [WebCite Cache ID 6xa1Z8fI6]

Abbreviations

API: application programming interface
EE: energy expenditure
GPS: global positioning system
HR: heart rate
IMU: inertial measurement unit
MEDLINE: Medical Literature Analysis and Retrieval System Online
PA: physical activity
PPG: photoplethysmography
SDK: software development kit

Edited by G Eysenbach; submitted 17.11.17; peer-reviewed by J Sanders, P Wark, K Winfree, R Fallahzadeh, C Fernández; comments to author 07.12.17; revised version received 18.12.17; accepted 06.01.18; published 22.03.18

Please cite as:
Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S
Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables
J Med Internet Res 2018;20(3):e110
URL: http://www.jmir.org/2018/3/e110/
doi:10.2196/jmir.9157
PMID:

Paper II

Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L.

Validity of the Polar M430 Activity Monitor in Free-Living Conditions: Validation Study.

JMIR Form Res, 2019. 3(3): p. e14438. doi: 10.2196/14438.

Original Paper

Validity of the Polar M430 Activity Monitor in Free-Living Conditions: Validation Study

André Henriksen[1], MSc, MBA; Sameline Grimsgaard[1], MPH, MD, PhD; Alexander Horsch[2], Dr rer nat, Dr med habil; Gunnar Hartvigsen[2], MSc, PhD; Laila Hopstock[1], RN, MSc, PhD

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway

[2]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
André Henriksen, MSc, MBA
Department of Community Medicine
UiT The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 77645214
Email: andre.henriksen@uit.no

Abstract

Background: Accelerometers, often in conjunction with heart rate sensors, are extensively used to track physical activity (PA) in research. Research-grade instruments are often expensive and have limited battery capacity, limited storage, and high participant burden. Consumer-based activity trackers are equipped with similar technology and designed for long-term wear, and can therefore potentially be used in research.

Objective: We aimed to assess the criterion validity of the Polar M430 sport watch, compared with 2 research-grade instruments (ActiGraph and Actiheart), worn on 4 different locations using 1- and 3-axis accelerometers.

Methods: A total of 50 participants wore 2 ActiGraphs (wrist and hip), 2 Actihearts (upper and lower chest position), and 1 Polar M430 sport watch for 1 full day. We compared reported time (minutes) spent in sedentary behavior and in light, moderate, vigorous, and moderate to vigorous PA, step counts, activity energy expenditure, and total energy expenditure between devices. We used Pearson correlations, intraclass correlations, mean absolute percentage errors (MAPEs), and Bland-Altman plots to assess criterion validity.

Results: Pearson correlations between the Polar M430 and all research-grade instruments were moderate or stronger for vigorous PA (*r* range .59-.76), moderate to vigorous PA (*r* range .51-.75), steps (*r* range .85-.87), total energy expenditure (*r* range .88-.94), and activity energy expenditure (*r* range .74-.79). Bland-Altman plots showed higher agreement for higher intensities of PA. MAPE was high for most outcomes. Only total energy expenditure measured by the hip-worn ActiGraph and both Actiheart positions had acceptable or close to acceptable errors with MAPEs of 6.94% (ActiGraph, 3 axes), 8.26% (ActiGraph, 1 axis), 14.54% (Actiheart, upper position), and 14.37% (Actiheart, lower position). The wrist-worn ActiGraph had a MAPE of 15.94% for measuring steps. All other outcomes had a MAPE of 22% or higher. For most outcomes, the Polar M430 was most strongly correlated with the hip-worn triaxial ActiGraph, with a moderate or strong Pearson correlation for sedentary behavior (*r*=.52) and for light (*r*=.7), moderate (*r*=.57), vigorous (*r*=.76), and moderate to vigorous (*r*=.75) PA. In addition, correlations were strong or very strong for activity energy expenditure (*r*=.75), steps (*r*=.85), and total energy expenditure (*r*=.91).

Conclusions: The Polar M430 can potentially be used as an addition to established research-grade instruments to collect some PA variables over a prolonged period. However, due to the high MAPE of most outcomes, only total energy expenditure can be trusted to provide close to valid results. Depending on the variable, the Polar M430 over- or underreported most metrics, and may therefore be better suited to report changes in PA over time for some outcomes, rather than as an accurate instrument for PA status in a population.

(JMIR Form Res 2019;3(3):e14438) doi: 10.2196/14438

KEYWORDS

actigraphy; fitness trackers; motor activity; validation studies

Introduction

Background

Lack of physical activity (PA) is the fourth-leading risk factor for global mortality, and the World Health Organization recommends at least 150 minutes weekly of moderate-intensity PA (eg, 30 minutes of moderate PA, 5 times per week) or 75 minutes weekly of vigorous-intensity PA for adults, and 60 minutes weekly of moderate to vigorous PA (MVPA) for children and adolescents [1]. However, worldwide, these recommendations are not reached by 80% of adolescents and 31% of adults (ranging from 17% in Southeast Asia to 43% in the eastern Mediterranean and the Americas) [2]. Two national reports from the Norwegian Directorate of Health show that, in the Norwegian population, these recommendation were reached by only 20% in 2009 [3] and 32% in 2015 [4].

Accelerometers and combined sensing (ie, accelerometers and heart rate) are used to track PA. Research-grade instruments are often expensive and have limited battery capacity, limited storage, and high participant burden. Consumer-based activity trackers, on the other hand, are designed for long-term wear, equipped with similar technologies, generally cheaper, and less intrusive, and can potentially track PA for research purposes.

Consumer-based activity trackers are increasingly being evaluated for use in research. Recent examples includes Lawrie et al [5] and Beukenhorst et al [6], who included smart watches in their research protocols. The major limitation of these devices is the limited knowledge of device validity. Due to the rapid growth of new devices, high-quality validation studies of emerging models are needed [7]. Specifically, to our knowledge, no validation study on the Polar M430 has been conducted to date. Most previous validation studies have compared multiple consumer devices with 1- or 2-criterion instruments (eg, [8,9]). In this study, we compared 1 consumer device with multiple criteria, placed on multiple locations, and analyzing 1 and 3 axes of the accelerometer.

Objective

The aim of this study was to assess the criterion validity of time (in minutes) spent in various PA intensity zones, step counts, and energy expenditure (EE) between the Polar M430 and 2 extensively used research-grade instruments (ActiGraph and Actiheart) worn on 4 different locations using uniaxial and triaxial measurements in free-living conditions. We used multiple criteria because we wanted to show how the choice of criterion and placement affects outcomes. The ActiGraph can be considered a reference standard for PA intensity in free-living people, but because the Actiheart also has a heart rate sensor, it can be an attractive alternative in many cases.

Methods

Sample

We recruited 50 participants, who were eligible for inclusion if they were 18 years of age or older with normal physical function. We used convenience sampling to maximize ranges for weight, height, body mass index, age, and sex.

Instruments

The Polar M430 (Polar Electro Oy, Kempele, Finland), released in 2017, is a sport watch with a 6–light-emitting diode wrist-based optical heart rate sensor and a 50-Hz triaxial accelerometer for tracking PA. It weighs 51 g, with 20 days of battery life.

ActiGraph wGT3X-BT (ActiGraph LLC, Pensacola, FL, USA) is a 19-g triaxial accelerometer with a 30- to 100-Hz sampling rate, to be worn on the wrist, hip, ankle, or thigh, with 25 days of battery life. ActiGraph has been previously validated for sedentary behavior [10-12], PA intensity for both uniaxial [12] and triaxial [13] acceleration, step counting [14], and EE [15].

The Actiheart (CamNtech Ltd, Cambridge, UK) is a 10-g uniaxial accelerometer with 32-Hz sampling rate and additional electrocardiography with 128-Hz sampling rate, to be worn on the chest, with 21 days of battery life. The Actiheart is extensively used to measure EE, and has been shown by Brage et al to produce valid estimations for EE both in laboratory settings [16] and under free-living conditions [17].

Procedure

We used self-reported information on height, weight, age, sex, and dominant hand to initialize the devices. The Polar M430 and an ActiGraph (attached with an elastic band) were placed on the wrist of the nondominant hand. One ActiGraph was placed on the right hip (attached with an elastic band). One Actiheart was placed approximately at the level of the second intercostal space at the sternum (medial part) and to the left (lateral part). The second Actiheart was placed approximately at the level of the fifth intercostal space at the sternum (medial part) and to the left (lateral part). The Actihearts were attached with 2 Red Dot 2238 electrodes (3M, St Paul, MN, USA) each. Table 1 [18] gives the setup used for all instruments and Figure 1 shows the placement of each instrument.

Devices were attached by 1 of 2 researchers after agreement of method in accordance with manufacturer recommendations. Participants were instructed to wear all instruments at all times except for temporarily removing the ActiGraph for showering and water activities. Participants wore all instruments for 1 full day (24 hours). We collected data in May 2018. Participants received written and oral instructions on how to wear the devices. All participants signed an informed consent form.

Table 1. Device setup and output variables.

Variables	Instrument		
	ActiGraph	Actiheart	Polar M430
Hardware and setup			
Epoch length (lowest available)	10 s	15 s	24 h
Accelerometer sample rate	100 Hz	32 Hz	50 Hz
Wear location	Nondominant wrist, right hip	Chest (V_2), chest (V_5)	Nondominant wrist
Parameters	Height, weight, sex, age, wear location	Height, weight, sex, age	Height, weight, sex, age, wear location
Software for setup and download	ActiLife 6.13.3	Actiheart 4.0.122	Polar Flow [18]
Software for analysis	QCAT[a]/ActiLife	QCAT/Actiheart	Polar Flow
Device model	wGT3X-BT	4	2P
Device firmware version	1.9.2	H90.65	1.1.34
Output variables			
Sitting or sedentary behavior	Yes	Yes	Yes
Light physical activity	Yes	Yes	Yes
Moderate physical activity	Yes	Yes	Yes
Vigorous physical activity	Yes	Yes	Yes
Activity energy expenditure	Yes	Yes	No
Total energy expenditure	No	Yes	Yes
Steps	Yes	No	Yes

[a]QCAT: Quality Control and Analysis Tool.

Figure 1. Instrument placement and Polar M430 illustrations.

Variable Creation

Using the proprietary software of ActiGraph and the Actiheart, we exported activity counts into comma-separated values files, using the lowest possible epoch setting, that is, 10- (ActiGraph) and 15- (Actiheart) second epochs. From the ActiGraph, triaxial (vertical, horizontal, lateral) counts and steps per epoch were exported. From the Actiheart, uniaxial (vertical) counts were exported. We extracted precalculated variables from the Polar M430 from Polar Flow directly. Table 1 details the software and epochs.

Due to no agreed-upon cut points for calculating PA intensity from the wrist-worn ActiGraph in adults, we applied a conversion function provided by ActiLife version 6.13.3 (after ActiLife export; ActiGraph) to the wrist-worn ActiGraph data before further analysis (Multimedia Appendix 1).

Exported comma-separated values files with epoch data were imported into the custom-made Quality Control and Analysis Tool (QCAT) developed at UiT The Arctic University of Norway and Technical University of Munich. We converted activity counts into 60-second epochs before doing further analysis. We used counts per minute (CPM) to calculate minutes in the various PA intensity zones, using several algorithms. By

using QCAT, data from ActiGraph and Actiheart were analyzed by the same program and comparable variables were created. We included only valid days, a priori defined as all instruments worn at least 10 hours per day [19], in the analysis. We identified nonwear time using the triaxial wear-time algorithm of Hecht et al [20]. Multimedia Appendix 2 shows correlations between QCAT and ActiLife.

For the ActiGraph data (wrist and hip), we calculated 5 PA intensity zones using cut points defined by Freedson et al [12] and Matthews et al [21], using only the vertical axis. In addition, we used a combination of the methods of Sasaki et al [13], Kozey-Keadle et al [10], and Peterson et al [11] to generate the same PA intensity zones using all 3 axes, or vector magnitude (VM). To our knowledge, there are no agreed-upon cut points for chest-based PA counts in adults using an Actiheart. However, we used cut points identified in a study by Schrack et al [22]. We combined minutes spent in vigorous and very vigorous intensity into 1 variable. Table 2 gives an overview of each cut point set.

As QCAT does not support EE calculation, we calculated this variable from the proprietary software tools ActiLife and Actiheart. We calculated EE from ActiLife using the Freedson combination 1998 formula (Freedson et al [12] plus Williams work-energy equation) for uniaxial calculation and the Freedson VM3 combination 2011 formula (Sasaki et al [13] plus Williams work-energy equation) for triaxial calculation. We analyzed nonwear time using the default Troiano [23] settings. Actiheart uses a branched model where recorded activity and heart rate from the electrocardiogram are used together to improve EE calculations [24].

While the Actiheart reports resting EE (REE), activity EE (AEE), diet-induced thermogenesis, and total EE (TEE), the ActiGraph reports only AEE, and the Polar M430 reports only TEE. Since Actiheart used the Schofield equation [25] when calculating REE, we used the same equation to convert between AEE and TEE for the Polar M430 and the Actiheart. Furthermore, we subtracted or added, respectively, 10% of TEE to account for diet-induced thermogenesis.

The Polar M430 exports data for TEE, steps, and 5 PA intensity zones: minutes in (1) rest, (2) sitting, (3) low-intensity PA, (4) medium-intensity PA, and (5) high-intensity PA. We did not know the algorithm used by Polar to assign PA in 1 of these 5 categories, but we used the following conversion between the Polar M430 and other instruments: sitting = sedentary, low = light, medium = moderate, and high = vigorous + very vigorous PA. We did not use “minutes in rest” from the Polar M430. We compared steps only between the Polar M430 and the 2 ActiGraph locations, as this variable is not available in the exported Actiheart data. We did not compare heart rate outcomes in this analysis, as our aim was to investigate PA measures. We will address heart rate measures in a separate analysis.

Table 2. Alternative cut-point sets for physical activity intensity zones.

Intensity zone	ActiGraph uniaxial CPM[a]	ActiGraph triaxial CPM vector magnitude	Actiheart CPM
Sedentary	≤99	≤149	≤10
Light	100-1951	150-2689	11-95
Moderate	1952-5724	2690-6166	96-234
Vigorous	5725-9498	6167-9642	≥235
Very vigorous	≥9499	≥9643	N/A[b]

[a]CPM: counts per minute.

[b]N/A: not applicable.

Statistical Analysis

We investigated Polar M430 validity for the following variables: sedentary behavior minutes per day, light PA minutes per day, moderate PA minutes per day, vigorous PA minutes per day, MVPA minutes per day, steps per day, AEE per day, and TEE per day. We used the Shapiro-Wilk test to test normality. We calculated and compared Pearson and Spearman correlations, with and without bootstrapping. Finally, we used the Pearson correlation coefficient, with bootstrapping, to assess the association between all instrument outcomes.

We used correlation cutoffs suggested by Evans [26]: very weak, less than .2; weak, .2-.4; moderate, .4-.6; strong, .6-.8; and very strong, greater than .8. We also calculated the intraclass correlation coefficient (ICC) to test agreement between instruments (absolute agreement, 2-way random, and single measures), which is not indicated by Pearson. We used the 95% confidence intervals of the ICC estimate to indicate poor (<.5), moderate (.5-.75), good (.75-.9), and excellent (>.9) agreement [27]. Mean absolute percentage error (MAPE) was used to calculate measurement error between devices for each outcome. There is no agreed-upon cutoff for MAPE, but previous validation studies have used a MAPE of less than 5% [9] or 10% [28,29] to indicate low error.

We also used Bland-Altman plots to assess the agreement between instrument outcomes [30]. Bland-Altman limits of agreement (LoA) indicate the mean difference between 2 instruments, when comparing the mean for each outcome. A positive mean value indicates an overreporting from the Polar M430. The width of the upper and lower LoA indicates the agreement between instruments, where a narrower range indicates a higher agreement.

For each variable, we present (as a figure or multimedia appendix) a scatterplot and a Bland-Altman plot for each criterion. In the scatterplot, the blue straight line shows the fit

line for the Pearson correlations. The black dashed line shows how a perfect correlation and agreement would appear, and can be used, together with the ICC, to see how much the Polar M430 over- or underreported the variables. In the Bland-Altman plot, the blue line indicates the mean difference between the Polar M430 and each criterion. Red lines show the upper and lower LoA.

Finally, we performed sensitivity and specificity tests to evaluate the ability of the Polar M430 to identify a target of 10,000 steps/day [31]. We did not report sensitivity and specificity for the recommended 30 minutes of MVPA per day, because the Polar M430 recorded at least 30 minutes of MVPA for all participants. All statistical analysis were performed using R version 3.5.3 (R Foundation).

Ethics Approval and Consent to Participate

The Norwegian regional committees for medical and health research ethics reviewed the study (2019/557/REK nord). All participants gave informed and written consent. This study was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments.

Results

Participant Demographics and Wear Time

Table 3 presents participants' height, weight, body mass index, age, and sex.

All ActiGraphs had a wear time of at least 10 hours and were included in the analysis. Recording on 1 Actiheart in the upper position failed, and we excluded it from the analysis. Two Actihearts were incorrectly initialized and were excluded from the TEE and AEE analyses. Although 7 Actihearts in the upper position and 5 Actihearts in the lower position had less than 10 hours of wear time, we did not exclude these because the participants informed us that they did not remove the device and manual review of the activity data indicated misclassification of nonwear and sleep.

Polar M430 Validity and Agreement

Multimedia Appendix 3 shows all outcomes for all criteria. Table 4 gives an overview of group data for all variables from the Polar M430. The tables in Multimedia Appendix 4 present all outcomes and group variables for each variable and all criteria.

Table 3. Participant characteristics (N=50).

Variable	Value	Range
Height (cm), mean (SD)	173.7 (10.1)	152-193
Weight (kg), mean (SD)	75.3 (16.4)	49-125
Body mass index (kg/m^2), mean (SD)	24.7 (3.6)	19.0-33.6
Age (years), mean (SD)	45.1 (15.5)	19-74
Females, n (%)	24 (48)	N/A[a]

[a]N/A: not applicable.

Table 4. Data of exported variables from the Polar M430 (N=50).

Variable	Value
Sedentary behavior (minutes), mean (SD)	500.61 (110.78)
Light physical activity (minutes), mean (SD)	308.45 (96.40)
Moderate physical activity (minutes), mean (SD)	98.10 (48.71)
Vigorous physical activity (minutes), mean (SD)	25.55 (37.27)
Moderate to vigorous physical activity (minutes), mean (SD)	123.65 (67.50)
Total energy expenditure (kcal), mean (SD)	2591.5 (619.1)
Activity energy expenditure (kcal), mean (SD)	N/A[a]
Steps, n (%)	13,426 (4775)

[a]N/A: not applicable.

Sedentary Behavior

Only the hip-worn ActiGraph VM gave a moderate Pearson correlation with the Polar M430. The remaining criteria gave a weak or very weak correlation. All ICC agreements were poor. The Bland-Altman LoA indicated underreporting of sedentary behavior by the Polar M430 compared with the hip-worn ActiGraph, and overreporting of the remaining criteria. All MAPEs were high. Table A in Multimedia Appendix 4 provides details of all criteria. Multimedia Appendix 5 gives correlations and Bland-Altman plots for the Polar M430 against each criterion.

Light Physical Activity

The hip-worn ActiGraph and both Actihearts gave a strong Pearson correlation with the Polar M430. The highest ICC agreement was shown for the hip-worn ActiGraph CPM, with a poor to moderate ICC. The Bland-Altman LoA indicated an overreporting of light PA by the Polar M430 compared with the hip-worn ActiGraph CPM and both Actihearts, and an underreporting for the remaining criteria. All MAPEs were high. Table B in Multimedia Appendix 4 provides details of all criteria. Multimedia Appendix 6 gives correlations and Bland-Altman plots for the Polar M430 against each criterion.

Moderate Physical Activity

All criteria except the Actiheart in the upper position gave a moderate Pearson correlation with the Polar M430. The highest ICC agreement was shown for the Actiheart in the lower position, with a poor to moderate ICC. The Bland-Altman LoA indicated an overreporting of moderate PA by the Polar M430 compared with the hip-worn ActiGraph CPM and both Actihearts, and an underreporting for the remaining criteria. All MAPEs were high. Table C in Multimedia Appendix 4 provides details of all criteria. Multimedia Appendix 7 gives correlations and Bland-Altman plots for the Polar M430 against each criterion.

Vigorous Physical Activity

The hip-worn ActiGraph gave a strong Pearson correlation with the Polar M430. The wrist-worn ActiGraph reported 0 minutes in vigorous PA and were therefore excluded from analysis. The Actiheart in the upper and lower position gave a strong and moderate correlation, respectively. The highest ICC agreement was shown for the hip-worn ActiGraph VM, with a poor to good ICC. The Bland-Altman LoA indicated an overreporting of vigorous PA by the Polar M430 compared with the hip-worn ActiGraph, and an underreporting for both Actihearts. All MAPEs were high. Table D in Multimedia Appendix 4 provides details of all criteria. Multimedia Appendix 8 gives correlations and Bland-Altman plots for the Polar M430 against each criterion.

Moderate to Vigorous Physical Activity

All criteria, regardless of cut points and number of axes considered, gave a moderately or strongly significant Pearson correlation when comparing MVPA for the Polar M430. The hip-worn ActiGraph VM had the strongest correlation. The highest ICC agreement was shown for the Actiheart in the lower position, with a poor to good ICC. The Bland-Altman LoA indicated an overreporting of MVPA by the Polar M430 compared with the hip-worn ActiGraph, a minor underreporting for the Actiheart in the upper position, and an underreporting for the wrist-worn ActiGraph and the Actiheart in the lower position. All MAPEs were high. Table E in Multimedia Appendix 4 provides details of all criteria. Figures 2 and 3 show correlations and Bland-Altman plots, respectively, for the Polar M430 against each criterion.

Figure 2. Correlation between the Polar M430 and all criterion measures for moderate to vigorous physical activity (MVPA). CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

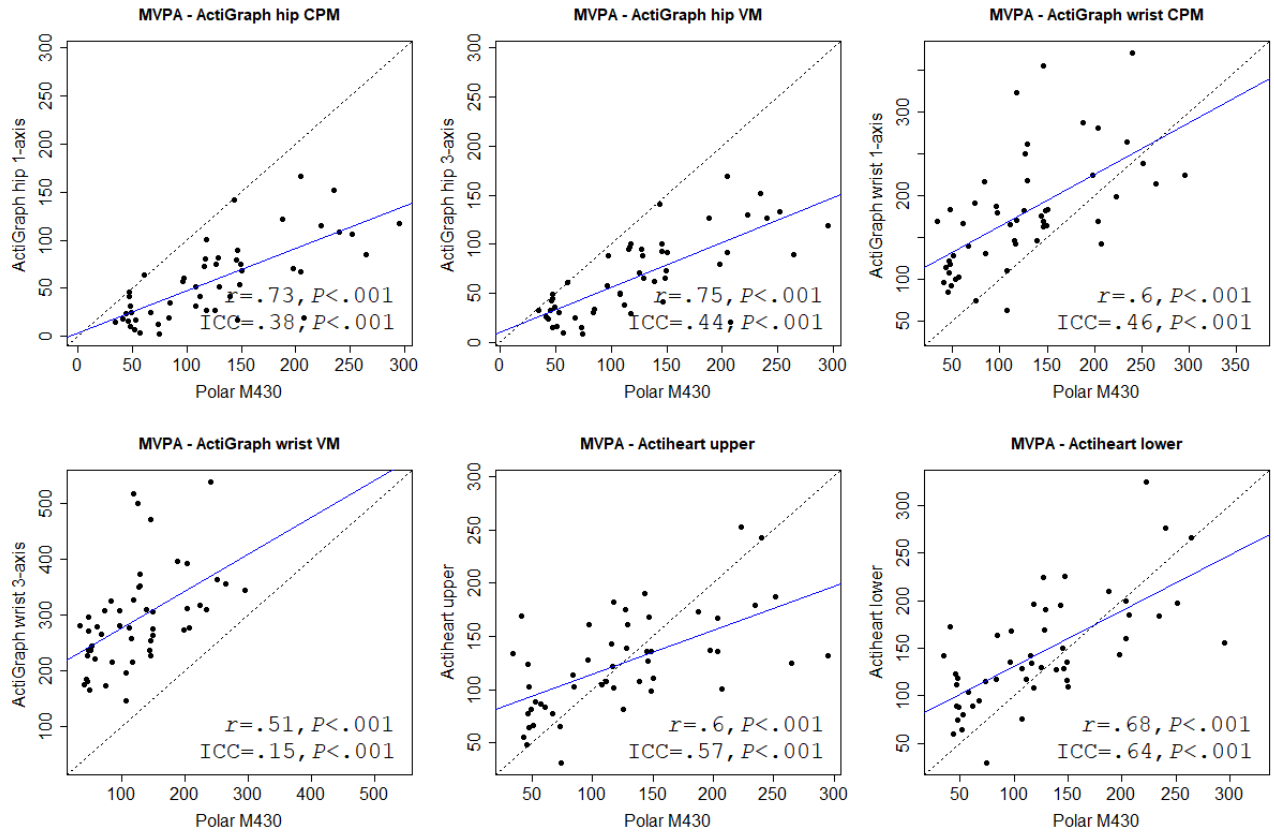
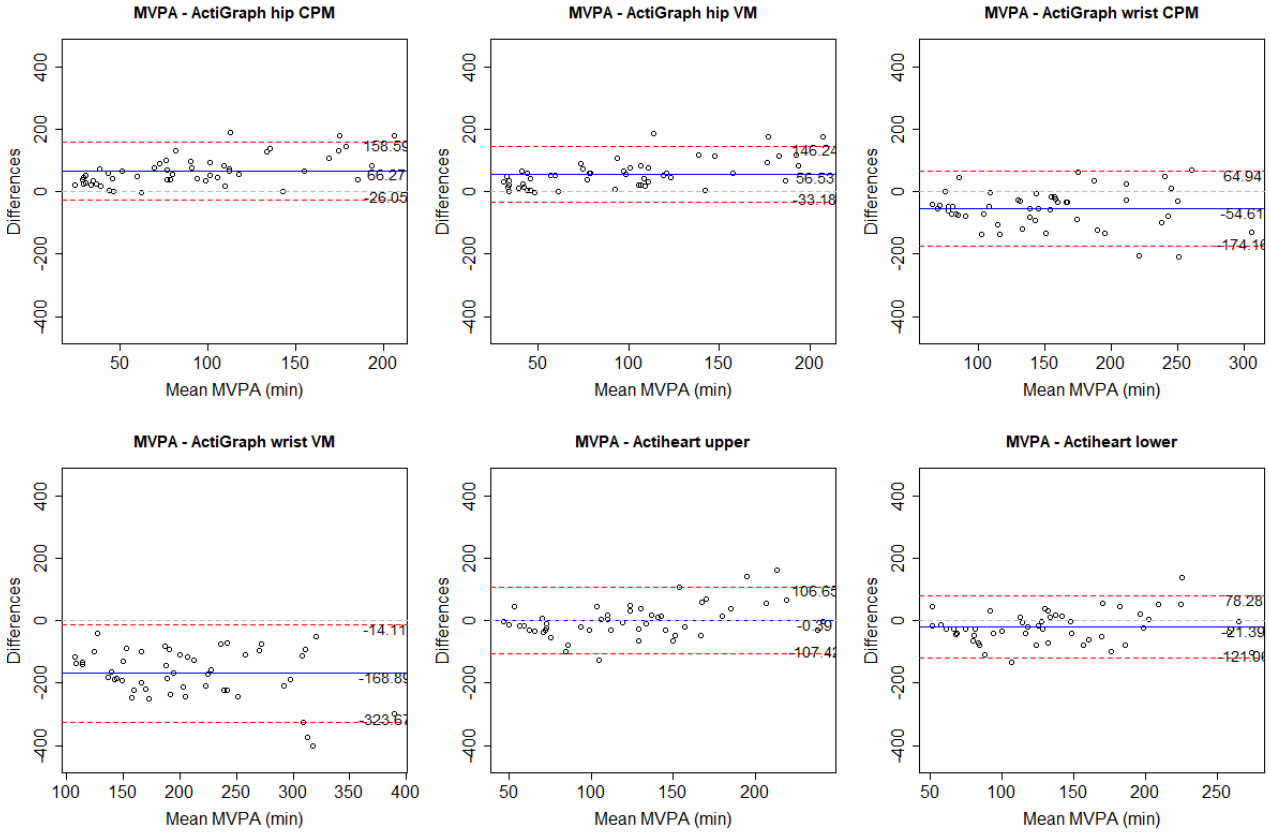


Figure 3. Bland-Altman plots for the Polar M430 and each criterion measure for moderate to vigorous physical activity (MVPA). Numbers are mean difference and upper and lower limits of agreement (95% CI). CPM: counts per minute; VM: vector magnitude.



Activity Energy Expenditure

All criteria showed a strong and significant Pearson correlation for AEE, where the wrist-worn ActiGraph VM was marginally stronger than the other criteria. ICC agreement was highest for the hip-worn ActiGraph VM with a moderate to good agreement. The Bland-Altman LoA showed an overreporting of AEE by the Polar M430 compared with the hip-worn ActiGraph and an underreporting for the wrist-worn ActiGraph and both Actihearts. All MAPEs were high. Table F in Multimedia Appendix 4 provides details of all criteria. Multimedia Appendix 9 gives correlations and Bland-Altman plots for the Polar M430 against each criterion. Multimedia Appendix 10 gives a combined plot for AEE and TEE.

Total Energy Expenditure

All criteria showed a very strong and significant Pearson correlation for TEE. The correlation for wrist-worn ActiGraph CPM was marginally stronger than other ActiGraphs. ICC agreement was highest for the hip-worn ActiGraph VM, with good to excellent agreement. The Bland-Altman LoA showed an overreporting of TEE by the Polar M430 compared with the hip-worn ActiGraph, and an underreporting for remaining criteria. The hip-worn ActiGraph had an acceptable MAPE of 6.94% (VM) and 8.26% (CPM). the remaining criteria had a high MAPE. Table G in Multimedia Appendix 4 provides details of all criteria. ActiGraph does not report TEE, and group data are therefore not available. Figures 4 and 5 show correlations and Bland-Altman plots, respectively, for the Polar M430 against each criterion.

Figure 4. Correlation between the Polar M430 and each criterion measure for total energy expenditure (TEE). CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

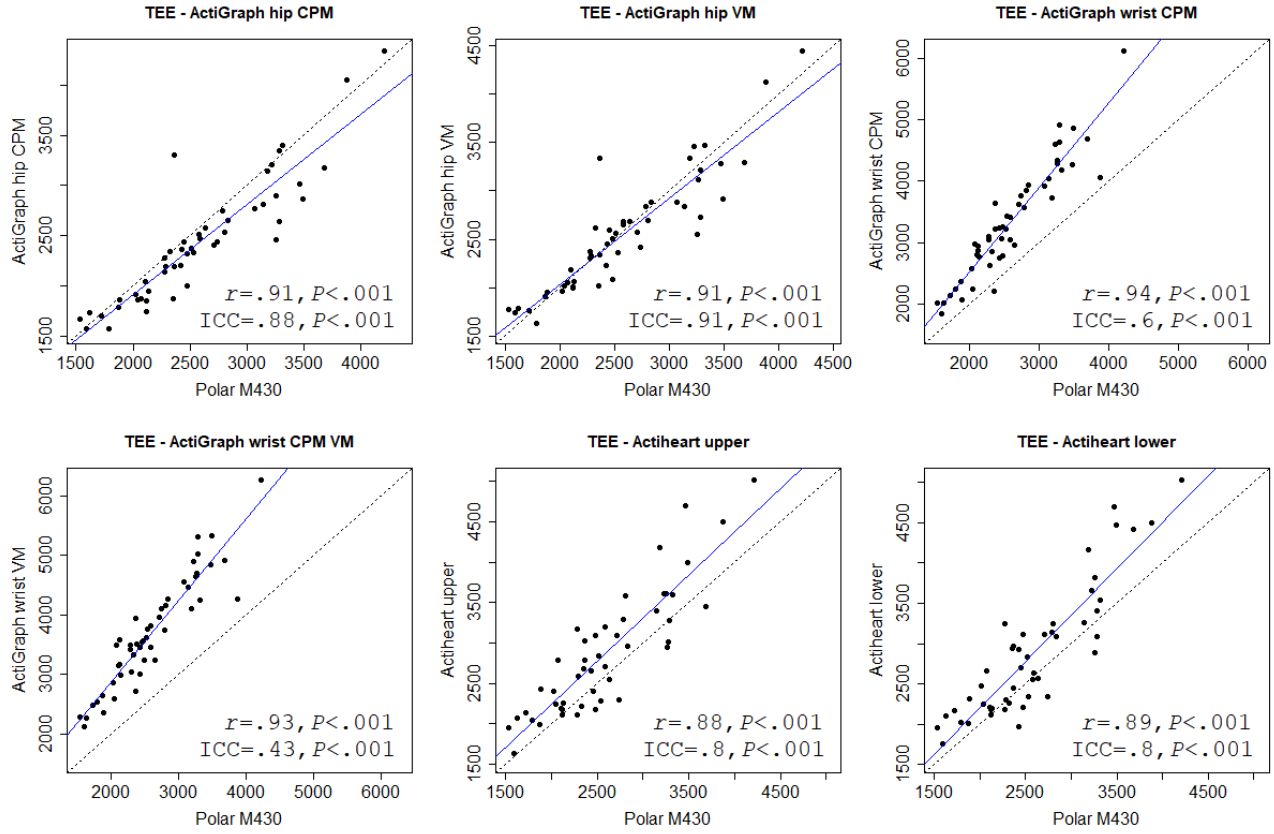
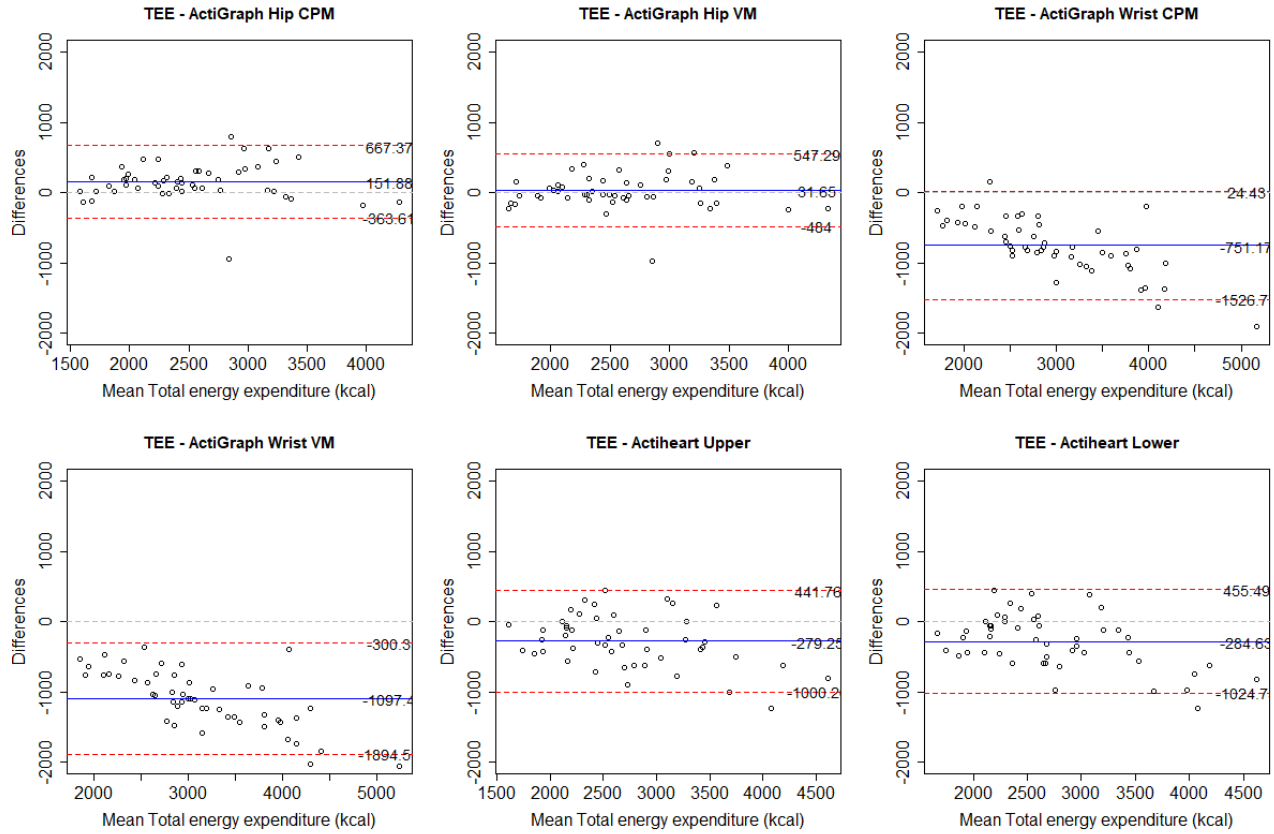


Figure 5. Bland-Altman plots for the Polar M430 and each criterion measure for total energy expenditure (TEE). Numbers are mean difference and upper and lower limits of agreement (95% CI). CPM: counts per minute; VM: vector magnitude.

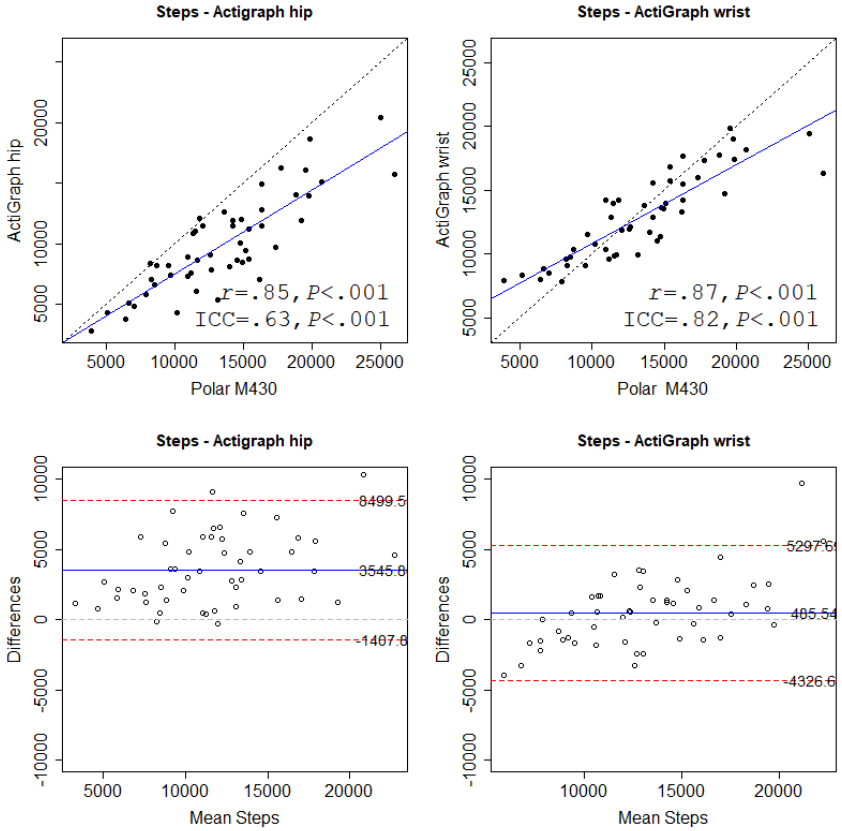


Steps

There was a very strong significant, and approximately equal, correlation between the Polar M430 and both the wrist-worn and hip-worn ActiGraph when measuring steps. ICC agreement was moderate to good for both locations. The Bland-Altman plot showed that the Polar M430 overreported steps for both placements of the ActiGraph, but at a higher rate on the hip-worn ActiGraph. Both MAPEs were high, but the hip-worn ActiGraph had the lowest MAPE. Table H in Multimedia Appendix 4 provides details of all criteria. Figure 6 shows correlations and Bland-Altman plots for the Polar M430 against both criteria.

Sensitivity (true-positive) analysis showed that the Polar M430, compared with the hip-worn ActiGraph, identified all cases in which a participant achieved 10,000 steps/day. For the wrist-worn ActiGraph, sensitivity was .94. Specificity, the ability of the Polar M430 to correctly identify those not achieving the 10,000 step/day target was .43 for the hip-worn ActiGraph and .71 for the wrist-worn ActiGraph.

Figure 6. Correlations and Bland-Altman plots for the Polar M430 and the wrist-worn and hip-worn ActiGraph for steps. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI). ICC: intraclass correlation coefficient.



Discussion

Principal Findings

We have shown how the available variables correlate and agree between the Polar M430 and 6 different combinations of device, placement, and number of accelerometer axes. For most outcomes, the Polar M430 showed the strongest correlation with the hip-worn triaxial ActiGraph (VM). Similarly, agreement was most often highest (or almost as high) when we compared the Polar M430 with this criterion. Exceptions are MVPA and moderate PA, where the Actiheart in the lower position showed a somewhat higher agreement.

A previous study by Tudor-Locke et al [32] showed that the hip-worn ActiGraph had a higher accuracy for step counting than the wrist-worn ActiGraph in laboratory settings. Under free-living conditions, the same study showed that the ActiGraph detected more steps when placed on the wrist. It is therefore possible to conclude that, although our study showed that the wrist-worn ActiGraph had a higher correlation, higher agreement, and lower MAPE, the true step counts may be closer to the numbers reported by the hip-worn ActiGraph. Similarly, studies comparing how wear location affected PA intensity [33] and EE [34,35] outcomes showed that the hip-worn ActiGraph is more accurate than the wrist-worn ActiGraph.

When compared with the hip-worn ActiGraph VM, the Polar M430 had a very strong correlation for TEE and steps, a strong correlation for AEE, MVPA, light PA, and vigorous PA, and a moderate correlation for sedentary behavior and moderate PA. Bland-Altman plots showed that the mean agreement was higher for higher intensities of PA, with underreporting by the Polar M430 for sedentary behavior and light PA, and overreporting for the remaining variables. Sensitivity analysis also indicated that the Polar M430 overreported the number of steps. However, MAPE was high for most variables, and only TEE had an acceptable MAPE of 6.9%. Multimedia Appendix 11 and Multimedia Appendix 12 give correlations and Bland-Altman

plots, respectively, for the Polar M430 and the hip-worn ActiGraph VM for all variables.

MVPA was strongly correlated for all criteria except 1 (ie, wrist-worn ActiGraph VM), and all criteria gave a strong correlation for AEE and a very strong correlation for TEE and steps. For the hip-worn ActiGraph, most outcomes showed a stronger correlation when using the triaxial variable than the uniaxial variable. For all criteria, all correlations for TEE were stronger and all MAPEs were smaller than for AEE. This is expected, as REE constitutes between 60% and 75% of TEE [36]. Except for sedentary behavior and moderate PA, outcomes were similar for the upper and lower position of the Actiheart. This is in accordance with a previous study by Brage et al [37], in which position did not affect outcome significantly.

Comparison With Previous Studies

We identified 12 previous studies that compared wrist-worn Polar devices with an objective criterion measure for measuring steps, PA intensity, and EE. These studies tested 5 different Polar models: the Polar Loop (released in 2013), Polar V800 (released in 2014), Polar A300 (released in 2015), Polar A360 (released in 2015), and Polar M600 (released in 2016). We found no studies on the Polar M430 (released in 2017). The validity of EE, steps, and PA intensity levels for the Polar devices in these studies varied, and correlations ranged from weak to strong, depending on the study setting (laboratory vs free-living), device, and criterion measure.

We found 3 previous Polar validation studies on EE in laboratory settings showed a very weak to weak Pearson correlation for the Polar Loop (r range .02-.3) [38] and Polar A360 (r=.28) [39], and a very strong correlation for the Polar V800 (r range .63-.85) [28]. In free-living study participants, the Polar Loop [40], Polar A300 [41], and Polar V800 [42], showed a very strong (r=.9), strong (r=.83), and weak to moderate (r range .34-.69) correlation, respectively, for EE.

We found 3 studies on PA intensity levels in free-living populations showed poor agreement for the Polar A300 (ICC=.36) [41], strong to very strong Pearson correlation for the Polar V800 (r range .84-.93) [42], and moderate Spearman correlation and poor agreement for MVPA on the Polar M600 (ρ=.53, ICC=.38) [43]. We found no studies comparing PA intensity levels conducted in laboratory settings.

A total of 5 studies compared steps in laboratory settings. The Polar Loop was tested in 4 studies, where Wahl et al (r range .06-.83) [38], Wang et al (correlation not given) [44], and Fokkema et al (r range .08-.26) [9], showed low validity for steps, with a higher validity for higher walking speeds in 1 study (Wahl et al [38]). An et al [45], on the other hand, found higher validity for this device (r range .4-.7). Bunn et al [46] tested the Polar A360 and also found it to have low validity (r range –.24 to .49). In addition, 4 studies compared steps in free-living populations. The Polar Loop showed a strong to very strong Pearson correlation (r range .7-.89) [47], the Polar A300 showed a very strong correlation (r=.99) [41], the Polar V800 showed a very strong correlation (r range .89-.92) [42]), and the Polar M600 showed good agreement (ICC=.7) and a strong Spearman correlation (ρ=.85) [43].

The results from previous studies showed that the validity of Polar devices, ranging from the Polar Loop, released in 2013, to the Polar M600, released in 2016, was highly dependent on the study setting. Studies conducted in free-living populations seem to agree that EE was reasonably valid, but not always. Our study also showed a strong correlation for AEE and a very strong correlation for TEE, for some criteria. The correlations for MVPA were stronger in our study than in all other studies. Results from previous research on step counting in free-living populations showed similar strong correlations to those found in our study.

With the exception of the Polar Loop, there are a limited number of studies for each device. For all other devices, only 1 or 2 studies were available for a given device, and at most 1 per device in free-living populations. In addition, previous studies used a range of criteria, and as we found in our study, correlations between the Polar M430 and each criterion can be very different depending on which criterion is used for comparison. It is therefore difficult to compare our results with previous validation studies. However, because all of the previous validation studies were conducted on older devices, it is reasonable that our results showed stronger correlations and higher agreements, as modern devices are likely to be more accurate than older devices.

Other studies on non-Polar consumer-based activity trackers generally agreed that the validity of step was high, but validity for EE was lower. In a 2015 systematic review, Evenson et al [48] concluded that, for consumer-based activity trackers such as Jawbone and Fitbit, validity of steps was high, but validity for EE was lower. Similarly, Feehan et al [49] systematically reviewed Fitbit devices and found that validity for EE was low, but validity for measuring steps was higher. Bunn et al [50] systematically reviewed validation studies testing devices by Fitbit, Garmin, Apple, Misfit, Samsung Gear, TomTom, and Lumo, and found a tendency for devices to underestimate EE and steps, but step validity was higher at higher intensities. This is partly in contrast to our study. Compared with step counting, TEE showed higher correlations for all ActiGraph outcomes. For AEE, on the other hand, step counting was more strongly correlated.

Strengths and Limitations

The strengths of this study include the large sample size, with a wide range for participant weight, height, body mass index, and age. We compared the Polar M430 against multiple criterion measures, showing that the outcomes were highly dependent on instrument type and placement. Furthermore, we used 1 tool (QCAT) to convert all activity counts into PA intensity variables, thereby limiting the number of unknowns introduced when using multiple software packages.

Limitations are mainly related to uncertainties in cutoffs and conversions. We compared TEE and AEE between instruments because the Polar M430 did not report AEE and the ActiGraph did not report TEE. We used the same algorithm for adding and removing REE, but since we did not know how Polar calculates REE, we did not know the conversion's accuracy. No agreed-upon cut points for PA intensity exist for the Actiheart or the wrist-worn ActiGraph, so the accuracy of related

outcomes was also somewhat uncertain. We did not individually calibrate Actiheart devices, which could have given a more accurate EE measure. Finally, the Hecht 2009 nonwear time algorithm was not created for uniaxial accelerometer CPM. This likely caused misclassification between nonwear time and sedentary behavior, and lower correlation for this outcome.

Conclusion

This first validation study of the Polar M430 indicated higher validity for MVPA, steps, and EE than with previous Polar devices. The Polar M430 can potentially be used as an addition to established research-grade instruments to collect some PA variables over a prolonged period. Depending on the variable, the Polar M430 over- or underreported most metrics and may therefore be better suited to report changes in PA over time for some outcomes, rather than as an accurate instrument for PA status in a population. Due to the high MAPE of most outcomes, only TEE or activity tracking in large samples can be trusted to provide close to valid results. Before using any consumer activity tracker or smart watch in research, we suggest piloting the selected device in the population under study. In a future study, we will attempt to create a function for converting Polar M430 reported steps, MVPA, and EE into the ActiGraph hip-worn VM equivalent, in order to determine whether such an approach can be used to better track PA status in a population over time.

Acknowledgments

The publication charges for this study were granted from the publication fund at UiT, the Arctic University of Norway. This work was funded by UiT, the Arctic University of Norway thematic priority grant Personalized medicine for public health and by Troms County, grant TFK 2016-058.

Authors' Contributions

A Henriksen, SG, LH, and GH conceived the study. All authors contributed to the planning of the study. A Henriksen and LH collected the data. A Henriksen, LH, SG, and A Horsch analyzed the data. A Henriksen wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

ActiGraph wrist-to-hip activity-count conversion table.
[DOCX File, 12KB-Multimedia Appendix 1]

Multimedia Appendix 2

Correlations between ActiLife and the Quality Control and Analysis Tool.
[DOCX File, 12KB-Multimedia Appendix 2]

Multimedia Appendix 3

Group data and outcomes for all variables.
[XLSX File (Microsoft Excel File), 23KB-Multimedia Appendix 3]

Multimedia Appendix 4

Tables (A to H) of group data for each criterion compared with the Polar M430 for all outcomes.
[DOCX File, 26KB-Multimedia Appendix 4]

Multimedia Appendix 5

Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for sedentary behavior. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 56KB-Multimedia Appendix 5]

Multimedia Appendix 6

Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for light physical activity. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 57KB-Multimedia Appendix 6]

Multimedia Appendix 7

Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for moderate physical activity. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 56KB-Multimedia Appendix 7]

Multimedia Appendix 8

Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for vigorous physical activity. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 38KB-Multimedia Appendix 8]

Multimedia Appendix 9

Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for activity energy expenditure. Numbers in the Bland-Altman plots are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 56KB-Multimedia Appendix 9]

Multimedia Appendix 10

Combined scatterplots for energy expenditure with Pearson correlations and intraclass correlations for activity energy expenditure and total energy expenditure.
[PNG File, 46KB-Multimedia Appendix 10]

Multimedia Appendix 11

Correlations for the Polar M430 and hip-worn triaxial ActiGraph for all variables.
[PNG File, 35KB-Multimedia Appendix 11]

Multimedia Appendix 12

Bland-Altman plots for the Polar M430 and hip-worn triaxial ActiGraph, for all variables. Numbers are mean difference and upper and lower limits of agreement (95% CI).
[PNG File, 31KB-Multimedia Appendix 12]

References

1. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010. URL:https://www.who.int/dietphysicalactivity/global-PA-recs-2010.pdf [accessed 2019-08-12]
2. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
3. Fysisk aktivitet blant voksne og eldre i Norge Resultater fra en kartlegging i 2008 og 2009. Report NoIS-1754. Oslo, Norway: Helsedirektoratet; 2009. URL:https://www.helsedirektoratet.no/rapporter/fysisk-aktivitet-kartleggingsrapporter/Fysisk%20aktivitet%20blant%20voksne%20og%20eldre.pdf [accessed 2019-08-12]
4. Fysisk aktivitet og sedat tid blant voksne og eldre i Norge-Nasjonal kartlegging 2014-2015. Report No. IS-2367. Oslo, Norway: Helsedirektoratet; 2015. URL:https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/991/Fysisk%20aktivitet%20og%20sedat%20tid%20blant%20voksne%20og%20eldre%20i%20Norge%202014-15.pdf [accessed 2019-08-12]
5. Lawrie S, Dong Y, Steins D, Xia Z, Esser P, Sun S, Smart Watch Activity Feedback Trial Committee (SWAFT). Evaluation of a smartwatch-based intervention providing feedback of daily activity within a research-naive stroke ward: a pilot randomised controlled trial. Pilot Feasibility Stud 2018;4:157. [doi: 10.1186/s40814-018-0345-x] [Medline: 30323946]
6. Beukenhorst AL, Parkes MJ, Cook L, Barnard R, van der Veer SN, Little MA, et al. Collecting symptoms and sensor data with consumer smartwatches (the Knee OsteoArthritis, Linking Activity and Pain Study): protocol for a longitudinal, observational feasibility study. JMIR Res Protoc 2019 Jan 23;8(1):e10238 [FREE Full text] [doi: 10.2196/10238] [Medline: 30672745]
7. Henriksen A, Haugen MM, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
8. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]

9. Fokkema T, Kooiman TJM, Krijnen WP. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
10. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
11. Peterson NE, Sirard JR, Kulbok PA, DeBoer MD, Erickson JM. Validation of accelerometer thresholds and inclinometry for measurement of sedentary behavior in young adult university students. Res Nurs Health 2015 Dec;38(6):492-499 [FREE Full text] [doi: 10.1002/nur.21694] [Medline: 26444969]
12. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [Medline: 9588623]
13. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011 Sep;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
14. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
15. McMinn D, Acharya R, Rowe DA, Gray SR, Allan JL. Measuring activity energy expenditure: accuracy of the GT3X+ and Actiheart monitors. Int J Exerc Sci 2013;6(3):217-229.
16. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor Actiheart. Eur J Clin Nutr 2005 Apr;59(4):561-570. [doi: 10.1038/sj.ejcn.1602118] [Medline: 15714212]
17. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. PLoS One 2015;10(9):e0137206. [doi: 10.1371/journal.pone.0137206] [Medline: 26349056]
18. Polar Flow. Kempele, Finland: Polar Electro Oy; 2018. URL:https://flow.polar.com [accessed 2019-08-08]
19. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
20. Hecht A, Ma S, Porszasz J, Casaburi R, COPD Clinical Research Network. Methodology for using long-term accelerometry monitoring to describe daily activity patterns in COPD. COPD 2009 Apr;6(2):121-129 [FREE Full text] [doi: 10.1080/15412550902755044] [Medline: 19378225]
21. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
22. Schrack JA, Leroux A, Fleg JL, Zipunnikov V, Simonsick EM, Studenski SA, et al. Using heart rate and accelerometry to define quantity and intensity of physical activity in older adults. J Gerontol A Biol Sci Med Sci 2018 Apr 17;73(5):668-675 [FREE Full text] [doi: 10.1093/gerona/gly029] [Medline: 29509832]
23. Troiano RP. Large-scale applications of accelerometers: new frontiers and new questions. Med Sci Sports Exerc 2007 Sep;39(9):1501. [doi: 10.1097/mss.0b013e318150d42e] [Medline: 17805080]
24. Brage S, Brage N, Franks PW, Ekelund U, Wong M, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol (1985) 2004 Jan;96(1):343-351 [FREE Full text] [doi: 10.1152/japplphysiol.00703.2003] [Medline: 12972441]
25. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr 1985;39 Suppl 1:5-41. [Medline: 4044297]
26. Evans JD. Straightforward Statistics for the Behavioral Sciences. Pacific Grove, CA: Brooks/Cole Publishing; 1996.
27. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
28. Roos L, Taube W, Beeler N, Wyss T. Validity of sports watches when estimating energy expenditure during running. BMC Sports Sci Med Rehabil 2017;9:22 [FREE Full text] [doi: 10.1186/s13102-017-0089-6] [Medline: 29296281]
29. Nelson MB, Kaminsky LA, Dickin DC, Montoye AHK. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
30. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999 Jun;8(2):135-160. [Medline: 10501650]
31. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
32. Tudor-Locke C, Barreira TV, Schuna JM. Comparison of step outputs for waist and wrist accelerometer attachment sites. Med Sci Sports Exerc 2015 Apr;47(4):839-842. [doi: 10.1249/MSS.0000000000000476] [Medline: 25121517]
33. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and wrist accelerometer algorithms for free-living behavior classification. Med Sci Sports Exerc 2016 May;48(5):933-940 [FREE Full text] [doi: 10.1249/MSS.0000000000000840] [Medline: 26673126]

34. Hildebrand M, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc 2014 Sep;46(9):1816-1824. [doi: 10.1249/MSS.0000000000000289] [Medline: 24887173]
35. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
36. Poehlman ET. A review: exercise and its influence on resting energy metabolism in man. Med Sci Sports Exerc 1989 Oct;21(5):515-525. [Medline: 2691813]
37. Brage S, Brage N, Ekelund U, Luan J, Franks PW, Froberg K, et al. Effect of combined movement and heart rate monitor placement on physical activity estimates during treadmill locomotion and free-living. Eur J Appl Physiol 2006 Mar;96(5):517-524. [doi: 10.1007/s00421-005-0112-6] [Medline: 16344938]
38. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
39. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
40. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
41. Boeselt T, Spielmanns M, Nell C, Storre JH, Windisch W, Magerhans L, et al. Validity and usability of physical activity monitoring in patients with chronic obstructive pulmonary disease (COPD). PLoS One 2016;11(6):e0157229 [FREE Full text] [doi: 10.1371/journal.pone.0157229] [Medline: 27305105]
42. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
43. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e10972 [FREE Full text] [doi: 10.2196/10972] [Medline: 30545810]
44. Wang L, Liu G, Wang Y, Li Q, Yi J, Inoue Y. Evaluation on Step Counting Performance of Wristband Activity Monitors in Daily Living Environment. IEEE Access 2017;5:13020-13027. [doi: 10.1109/ACCESS.2017.2721098]
45. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
46. Bunn JA, Jones C, Oliviera A, Webster MJ. Assessment of step accuracy using the Consumer Technology Association standard. J Sports Sci 2019 Feb;37(3):244-248. [doi: 10.1080/02640414.2018.1491941] [Medline: 29958058]
47. Simunek A, Dygryn J, Gaba A, Jakubec L, Stelzer J, Chmelik F. Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults. Acta Gymnica 2016 Sep 30;46(3):129-135. [doi: 10.5507/ag.2016.014]
48. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
49. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
50. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current state of commercial wearable technology in physical activity monitoring 2015-2017. Int J Exerc Sci 2018;11(7):503-515 [FREE Full text] [Medline: 29541338]

Abbreviations

AEE: activity energy expenditure
CPM: counts per minute
EE: energy expenditure
ICC: intraclass correlation coefficient
LoA: limits of agreement
MAPE: mean absolute percentage error
MVPA: moderate to vigorous physical activity
PA: physical activity
QCAT: Quality Control and Analysis Tool
REE: resting energy expenditure
TEE: total energy expenditure
VM: vector magnitude

Edited by G Eysenbach; submitted 22.04.19; peer-reviewed by P Düking, L Feehan; comments to author 14.05.19; revised version received 01.07.19; accepted 07.07.19; published 15.08.19

Please cite as:
Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L
Validity of the Polar M430 Activity Monitor in Free-Living Conditions: Validation Study
JMIR Form Res 2019;3(3):e14438
URL: http://formative.jmir.org/2019/3/e14438/
doi: 10.2196/14438
PMID:

Multimedia Appendix 1 – ActiGraph wrist-to-hip activity-count conversion table

The table below gives an overview of the cut-points used by ActiLife to convert wrist-worn ActiGraph activity counts to the hip-worn ActiGraph equivalent.

Wrist CPM	Equivalent counts
0	0
0 – 644	0.5341614 * Wrist count
645 – 1272	1.7133758 * Wrist count - 759.414013
1273 – 3806	0.3997632 * Wrist count + 911.501184
3807 - infinite	0.0128995 * Wrist count + 2383.904505

Multimedia Appendix 2 – Correlations between ActiLife and the QCAT

The correlation between QCAT and ActiLife was found to be strong or very strong for all activity intensity zones, when using the sample defined in this paper, and when using the default setting for wear-time validation in ActiLife (i.e. Troiano [20]).

The table below gives the Pearson correlation for each PA variable when comparing triaxial hip-worn ActiGraph, using the Troiano wear-time algorithm in ActiLife and the Hecht 2009 wear-time algorithm in QCAT. Except for sedentary behaviour, correlations were very strong.

Variable	Pearson's r	Lower CI	Upper CI	Significance level
Steps	1.000	1.000	1.000	$P < .001$
Sedentary	.606	.394	.757	$P < .001$
Light	.977	.959	.987	$P < .001$
Moderate	.935	.887	.963	$P < .001$
Vigorous	.993	.988	.996	$P < .001$
MVPA	.957	.924	.975	$P < .001$

CI: Confidence interval (95%)

Multimedia Appendix 4 - Tables (A to H) of group data for each criterion compared with the Polar M430 for all outcomes

Table A. Group data for each criterion compared with the Polar M430: output for sedentary behavior and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	50	50	50	50	49	50

Minutes in activity, mean (SD)	738 (105)	620 (109)	477 (114)	399 (106)	398 (110)	416 (126)
Pearson *r* (95% CI)	.49 (.1 to .7)[a]	.52 (.15 to .73)[a]	.09 (–.16 to .32)	.06 (–.21 to .3)	.05 (–.19 to .29)	.3 (.02 to .57)[b]
Intraclass correlation coefficient (95% CI)	.14 (.05 to .26)[a]	.33 (.1 to .51)[a]	.09 (0 to .31)	.04 (0 to .21)	.04 (0 to .23)	.24 (.01 to .5)[b]
Mean absolute percentage error (%)	52.68	29.24	25.11	28.90	25.74	22.22
Mean difference	–237.23	–119.59	23.19	101.97	98.05	84.29
Upper limit of agreement	–22.99	91.00	320.66	394.00	392.14	359.83
Lower limit of agreement	–451.47	–330.18	–274.28	–190.06	–196.04	–191.25

[a]*P*≤.001.

[b]*P*≤.05.

Table B. Group data for each criterion compared with the Polar M430: output for light physical activity and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	50	50	50	50	49	50
Minutes in activity, mean (SD)	291 (99)	399 (111)	536 (81)	501 (73)	245 (85)	272 (87)
Pearson *r* (95% CI)	.62 (.46-.75)[a]	.7 (.53-.81)[a]	.41 (.12-.65) [b]	.02 (–.23 to .29)	.69 (.52-.8)[a]	.62 (.44-.76)[a]
Intraclass correlation coefficient (95% CI)	.62 (.47-.75)[a]	.5 (.37-.65)[a]	.1[b] (.03-.18)	.01 (0-.08)	.55 (.4-.68)[a]	.58 (.42-.72)[a]
Mean absolute percentage error (%)	22.75	38.09	92.40	83.44	25.41	24.24
Mean difference	17.11	–90.79	–227.93	–192.43	65.52	36.71

Upper limit of agreement	182.99	69.79	–37.85	42.46	207.03	193.71
Lower limit of agreement	–148.77	–251.37	–418.01	–427.32	–75.99	–120.29

[a]*P*≤.001.

[b]*P*≤.05.

Table C. Group data for each criterion compared with the Polar M430: output for moderate physical activity and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	50	50	50	50	49	50
Minutes in activity, mean (SD)	49 (36)	56 (35)	178 (69)	293 (88)	57 (25)	71 (37)
Pearson *r* (95% CI)	.52 (.25-.66)[a]	.57 (.27-.70)[a]	.53 (.33-.72)[a]	.53 (.33-.72)[a]	.34 (.03-.59)[b]	.56 (.31-.74)[a]
Intraclass correlation coefficient (95% CI)	.31 (.16-.45)[a]	.36 (0.18-0.52)[a]	.27 (.15-.42)[a]	.1 (.06-.17)[a]	.18 (.02-.34)[b]	.45 (.25-.62)[a]
Mean absolute percentage error (%)	49.72	40.89	109.11	250.02	43.93	40.58
Mean difference	48.72	41.86	–80.16	–194.44	42.02	27.36
Upper limit of agreement	133.01	121.83	37.20	–48.39	133.52	108.56
Lower limit of agreement	–35.57	–38.11	–197.52	–340.49	–49.48	–53.84

[a]*P*≤.001.

[b]*P*≤.05.

Table D. Group data for each criterion compared with the Polar M430: output for vigorous physical activity and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts	Hip vector	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest

	per minute	magnitude				
Number	50	50	-	-	49	50
Minutes in activity, mean (SD)	8 (21)	11 (23)	-	-	68 (40)	74 (40)
Pearson *r* (95% CI)	.6 (.25-.82)[a]	.76 (.52-.85)[a]	-	-	.62 (.4-.78)[a]	.59 (.37-.75)[a]
Intraclass correlation coefficient (95% CI)	.44 (.14-.83)[a]	.62 (.42-.88)[a]	-	-	.39 (.21-.59)[a]	.33 (.17-.52)[a]
Mean absolute percentage error (%)	82.59	79.53	-	-	833.33	953.60
Mean difference	17.55	14.67	-	-	–42.41	–48.75
Upper limit of agreement	76.16	63.40	-	-	24.11	20.00
Lower limit of agreement	–41.06	–34.06	-	-	–108.92	–117.50

[a]*P*≤.001.

Table E. Group data for each criterion compared with the Polar M430: output for moderate to vigorous physical activity and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	50	50	50	50	49	50
Minutes in activity, mean (SD)	57 (41)	67 (41)	178 (69)	293 (88)	124 (47)	145 (58)
Pearson *r* (95% CI)	.73 (.53-.82)[a]	.75 (.54-.84)[a]	.6 (.44-.73)[a]	.51 (.34-.66)[a]	.6 (.39-.76)[a]	.68 (.51-.79)[a]
Intraclass correlation coefficient (95% CI)	.38 (.26-.5)[a]	.44 (.31-.57)[a]	.46 (.31-.61)[a]	.15 (.09-.23)[a]	.57 (.36-.74)[a]	.64 (.48-.77)[a]

Mean absolute percentage error (%)	53.53	43.49	79.40	198.61	44.79	53.87
Mean difference	66.27	56.53	–54.61	–168.89	–0.39	–21.39
Upper limit of agreement	158.59	146.24	64.94	–14.11	106.65	78.28
Lower limit of agreement	–26.05	–33.18	–174.16	–323.67	–107.42	–121.06

[a]*P*≤.001.

Table F. Group data for each criterion compared with the Polar M430: output for activity energy expenditure and each criterion measure.

Measure	ActiGraph				Actiheart	
	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	50	50	50	50	48	49
Kcal, mean (SD)	603 (348)	711 (353)	1415 (543)	1727 (558)	987 (488)	991 (525)
Pearson *r* (95% CI)	.75 (.5-.87)[a]	.75 (.54-.87)[a]	.78 (.65-.86)[a]	.79 (.63-.87)[a]	.74 (.57-.85)[a]	.79 (.63-.87)[a]
Intraclass correlation coefficient (95% CI)	.69 (.51-.84)[a]	.75 (.53-.87)[a]	.31 (.24-.41)[a]	.2 (.15-.27)[a]	.57 (.45-.69)[a]	.59 (.49-.69)[a]
Mean absolute percentage error (%)	27.71	24.01	103.08	152.65	54.38	53.38
Mean difference	136.69	28.48	–676.05	–987.66	–252.65	–257.58
Upper limit of agreement	600.63	492.57	21.99	–270.27	396.16	408.60
Lower limit of agreement	–327.25	–435.60	–1374.09	–1705.06	–901.46	–923.76

[a]*P*≤.001.

Table G. Group data for each criterion compared with the Polar M430: output for total energy expenditure and each criterion measure.

Measure	ActiGraph	Actiheart

	Hip counts per minute	Hip vector magnitude	Wrist counts per minute	Wrist vector magnitude	Upper chest	Lower chest
Number	-	-	-	-	48	49
Kcal, mean (SD)	-	-	-	-	2864 (763)	2866 (806)
Pearson *r* (95% CI)	.91 (.75-.95)[a]	.91 (.78-.95)[a]	.94 (.88-.97)[a]	.93 (.86-.96)[a]	.88 (.8-.93)[a]	.89 (.81-.94)[a]
Intraclass correlation coefficient (95% CI)	.88 (.78-.94)[a]	.91 (.8-.96)[a]	.6 (.52-.68)[a]	.44 (.35-.53)[a]	.8 (.71-.86)[a]	.8 (.73-.86)[a]
Mean absolute percentage error (%)	8.26	6.94	28.71	42.54	14.54	14.37
Mean difference	151.88	31.65	–751.17	–1097.40	–279.25	–284.63
Upper limit of agreement	667.37	547.29	24.43	–300.30	441.76	455.49
Lower limit of agreement	–363.61	–484.00	–1526.77	–1894.51	–1000.26	–1024.75

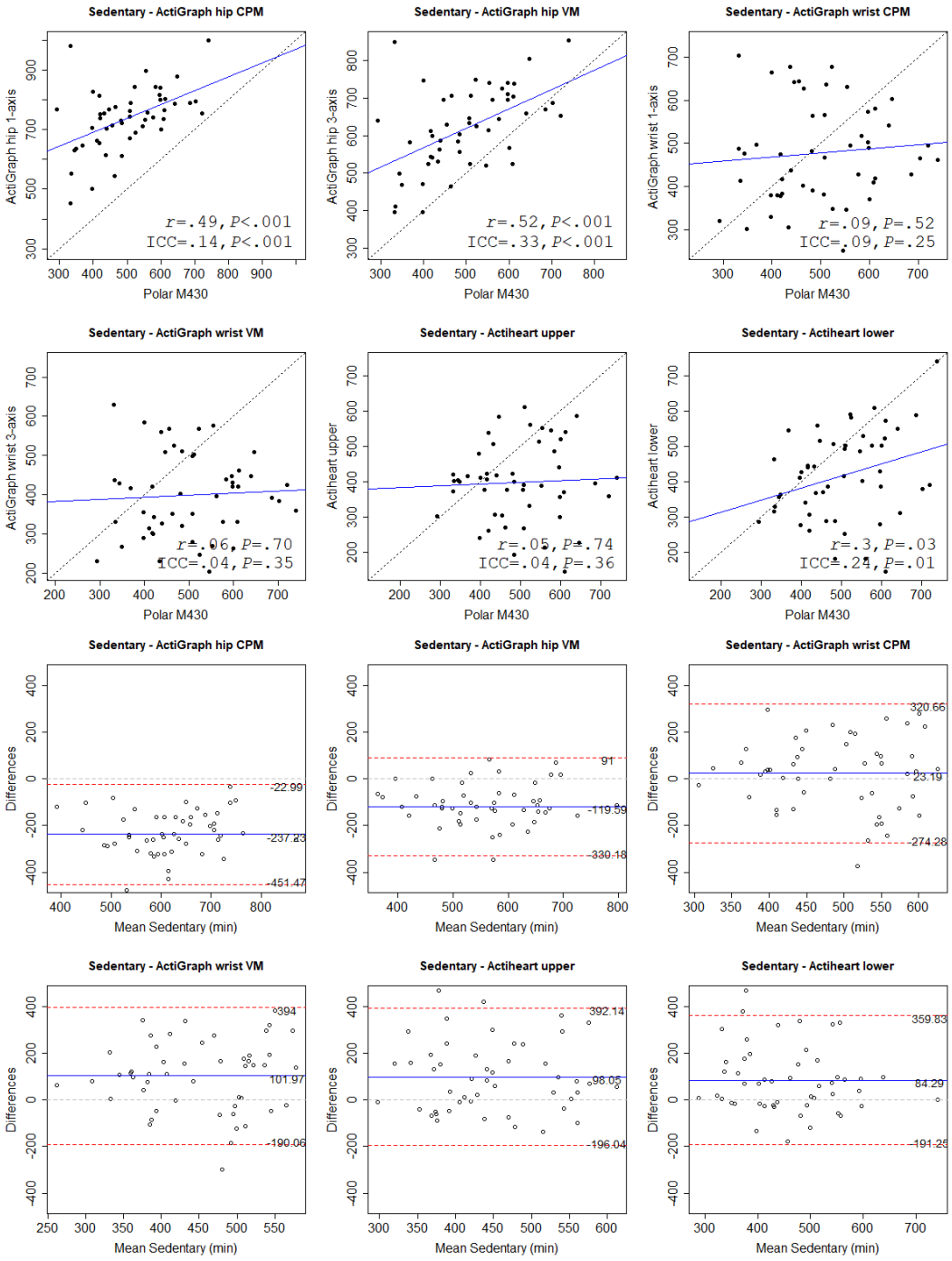
[a]*P*≤.001.

Table H. Group data for each criterion compared with the Polar M430: output for steps and both criteria reporting steps.

Measure	ActiGraph hip CPM	ActiGraph wrist CPM
Number	50	50
Step count, mean (SD)	9880 (3913)	12940 (3381)
Pearson *r* (95% CI)	.85 (.75-.91)[a]	.87 (.79-.92)[a]
Intraclass correlation coefficient (95% CI)	.63 (.49-.75)[a]	.82 (.7-.88)[a]
Mean absolute percentage error (%)	25.98	15.94
Mean difference	3546	486
Upper limit of agreement	8500	5298
Lower limit of agreement	–1408	–4327

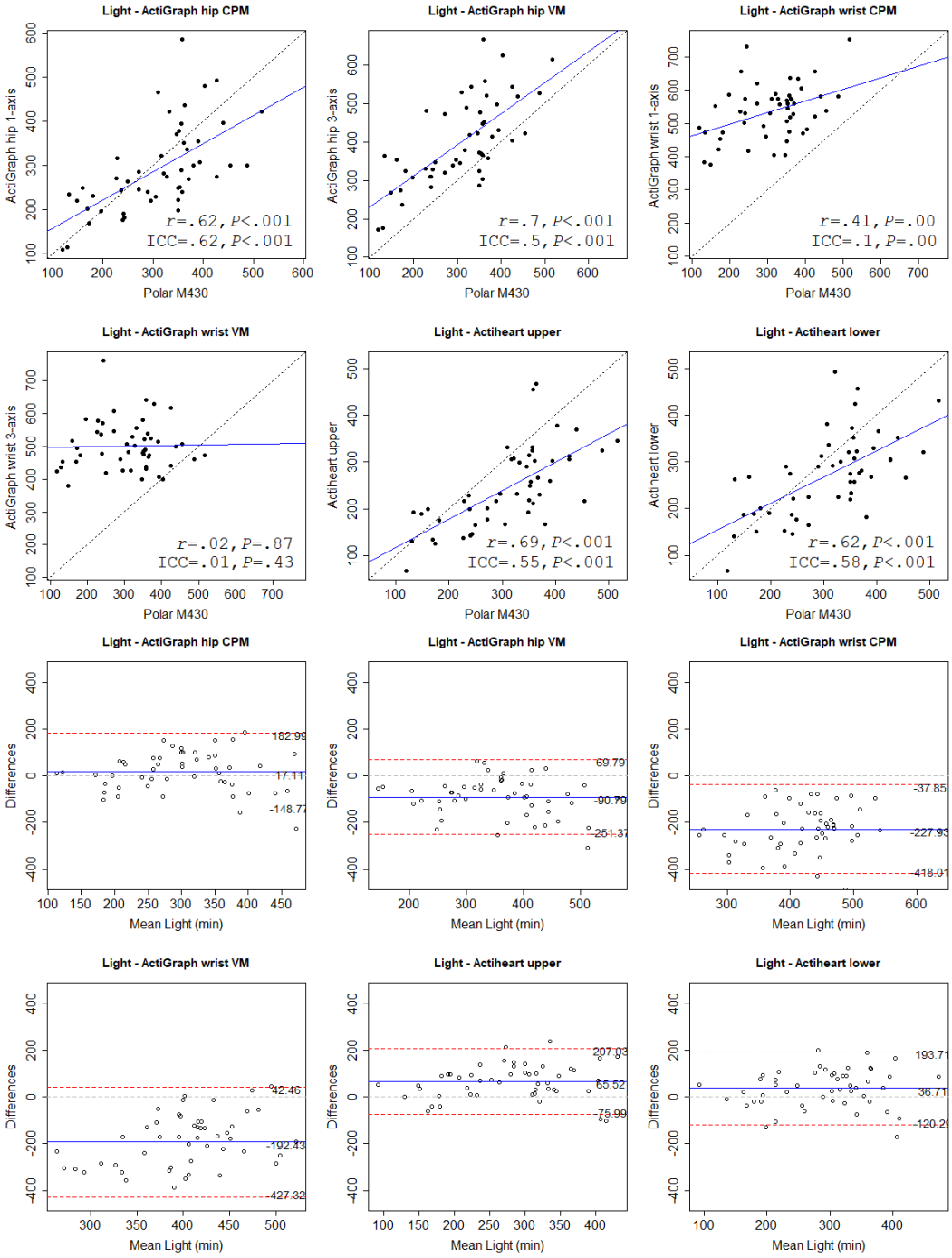
[a]*P*≤.001.

Multimedia Appendix 5 – Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for sedentary behavior.



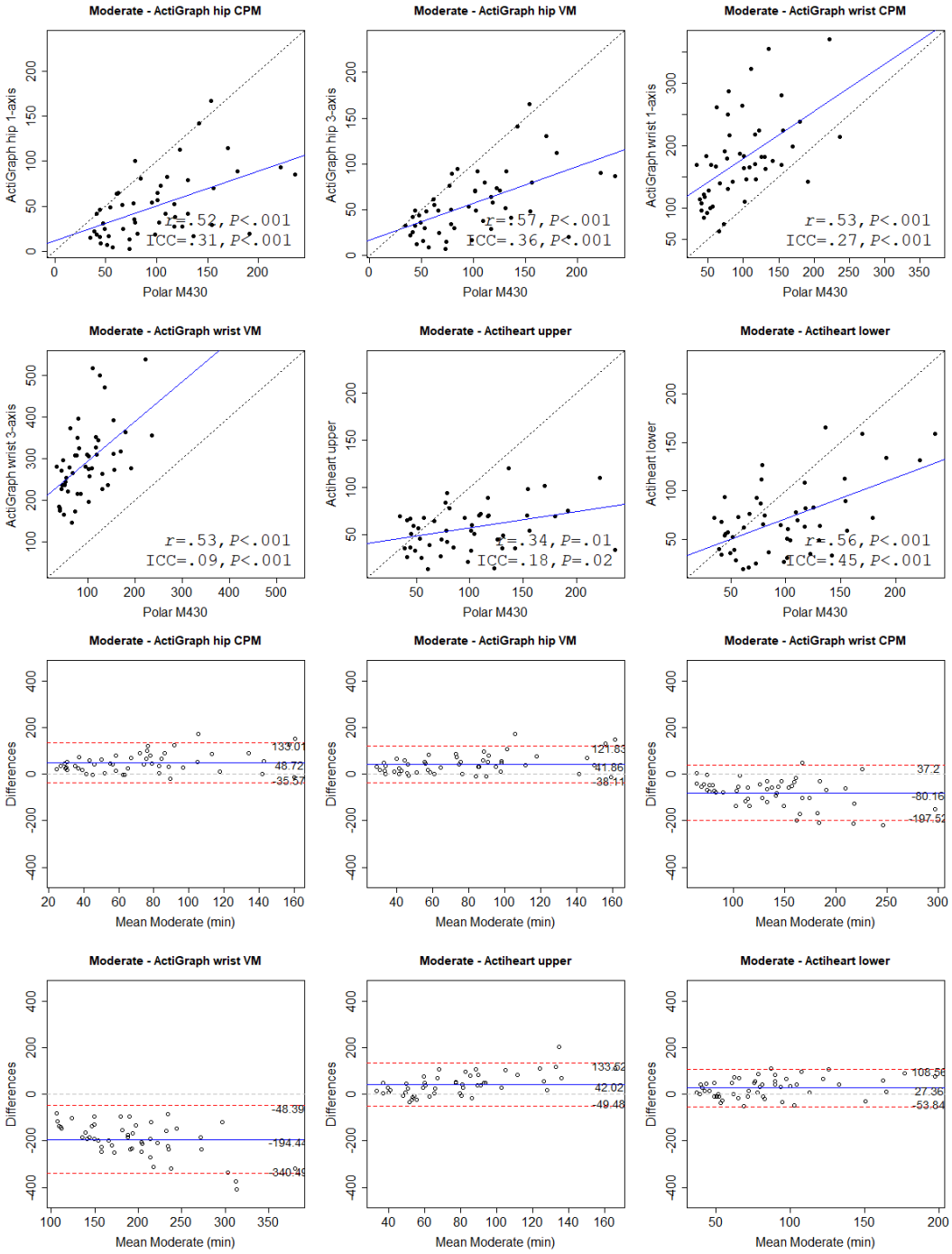
CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 6 – Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for light physical activity.



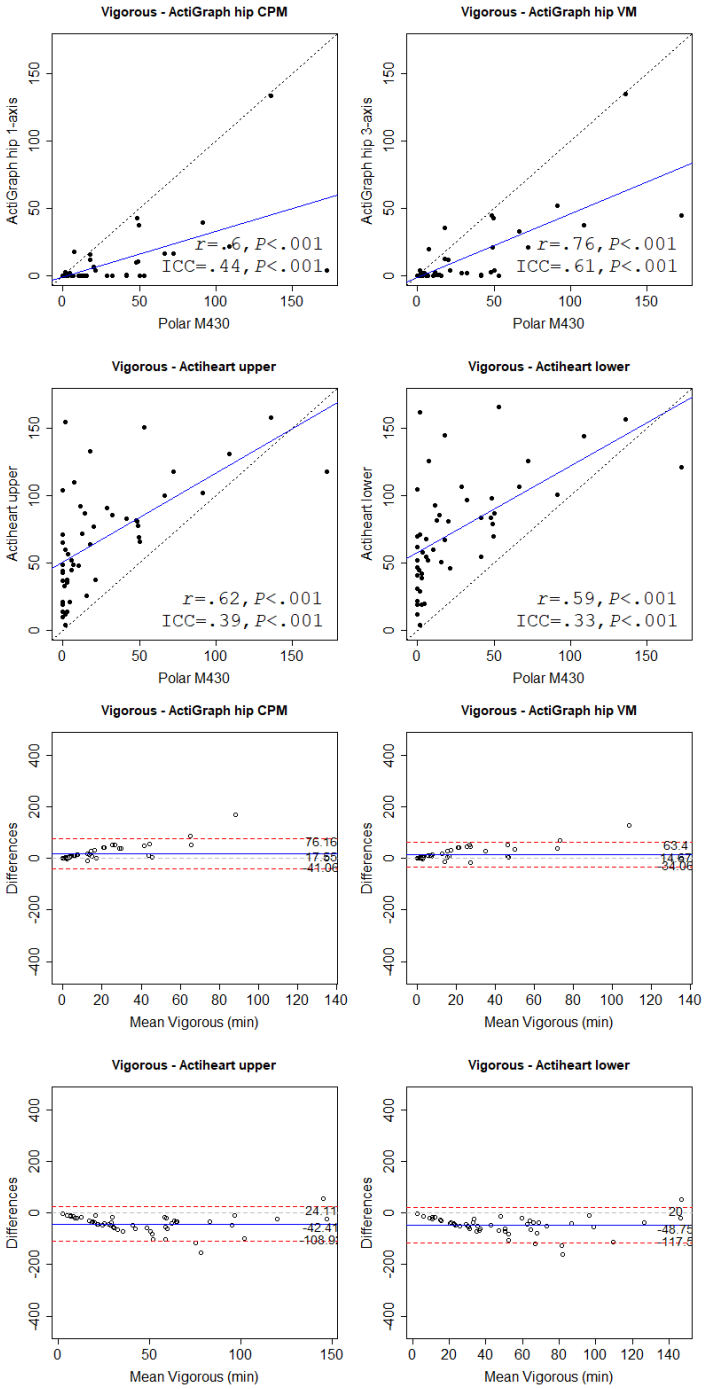
CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 7 – Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for moderate physical activity.



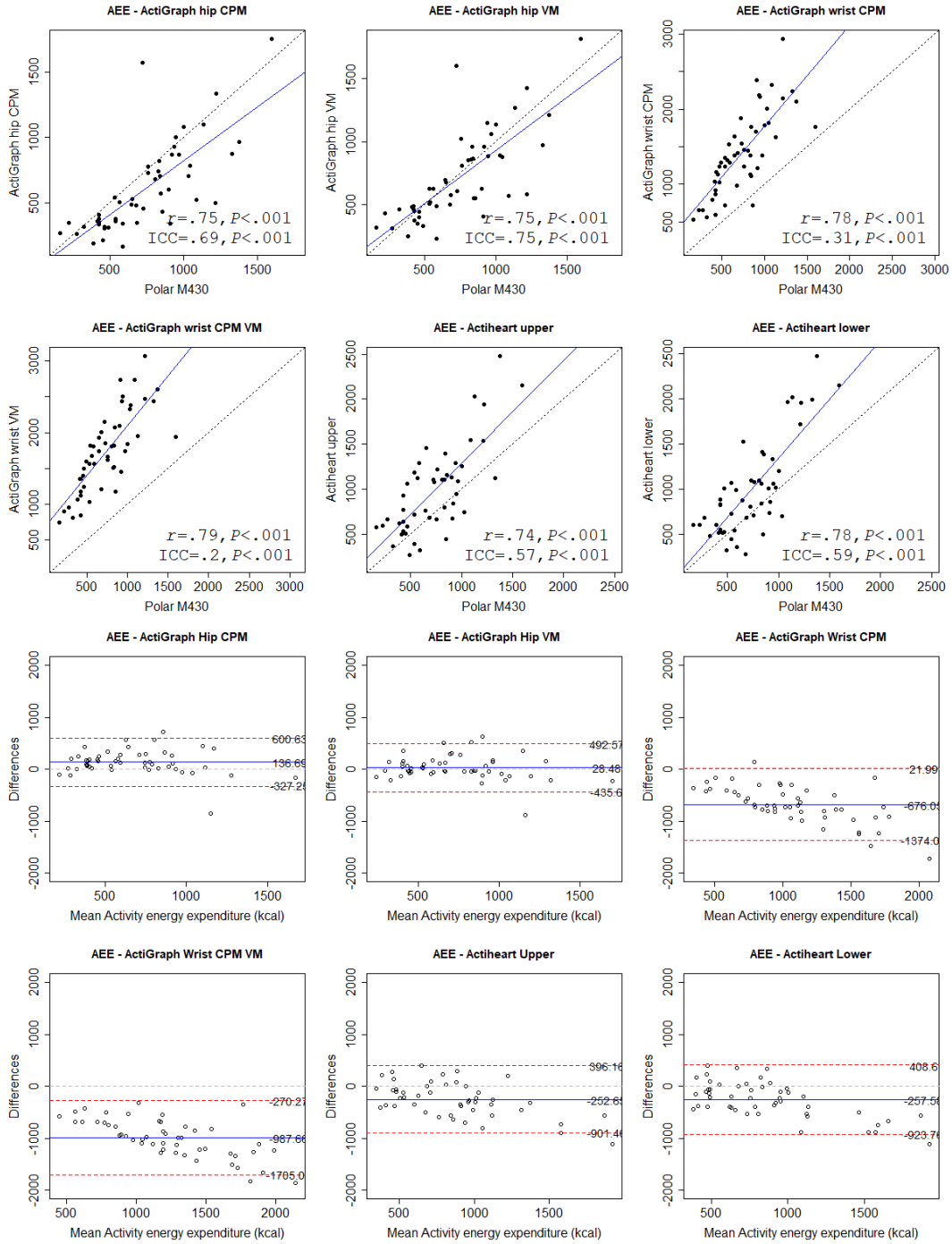
CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 8 – Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for vigorous physical activity.



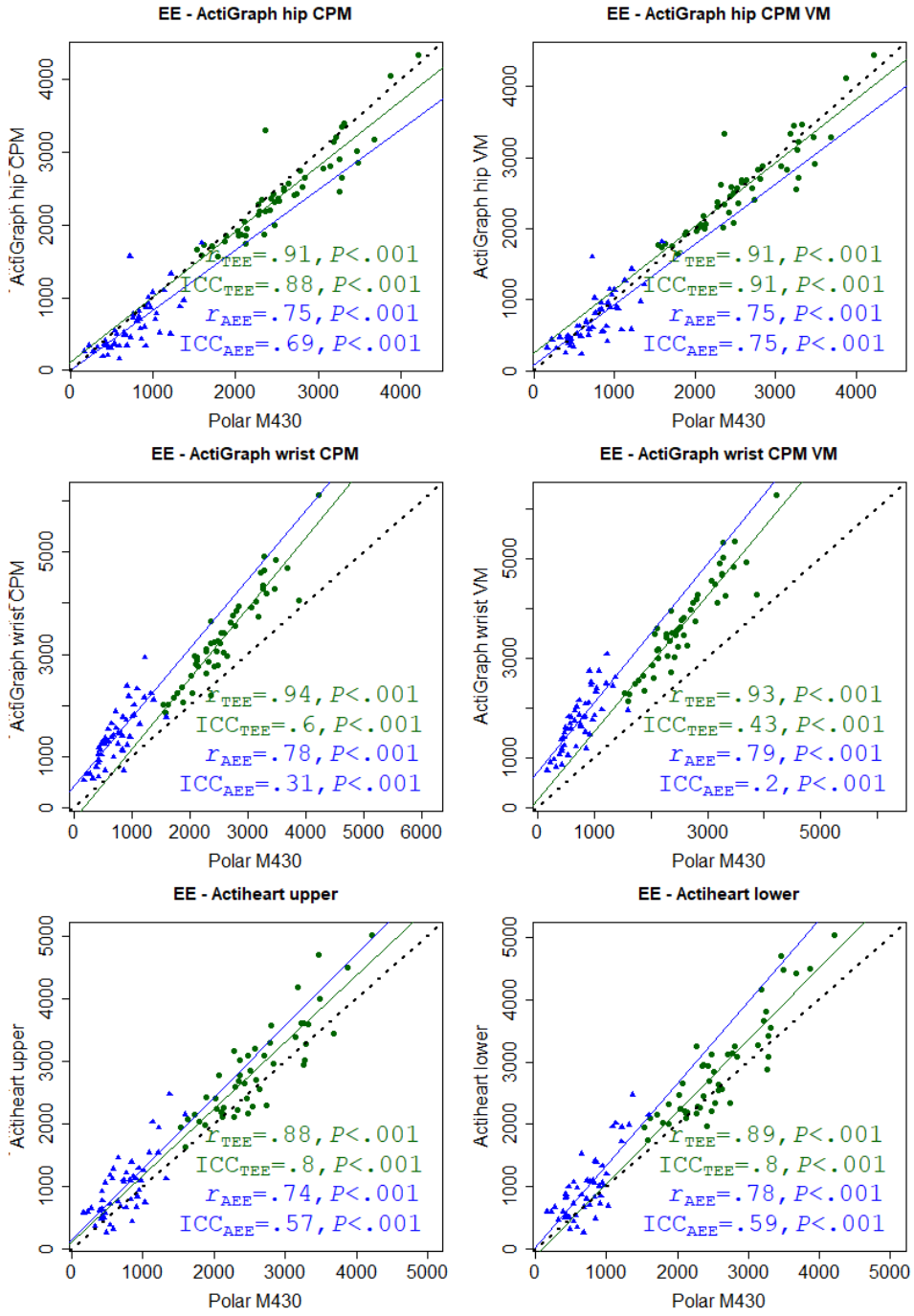
CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 9 – Correlations and Bland-Altman plots for the Polar M430 and each criterion measure for activity energy expenditure.



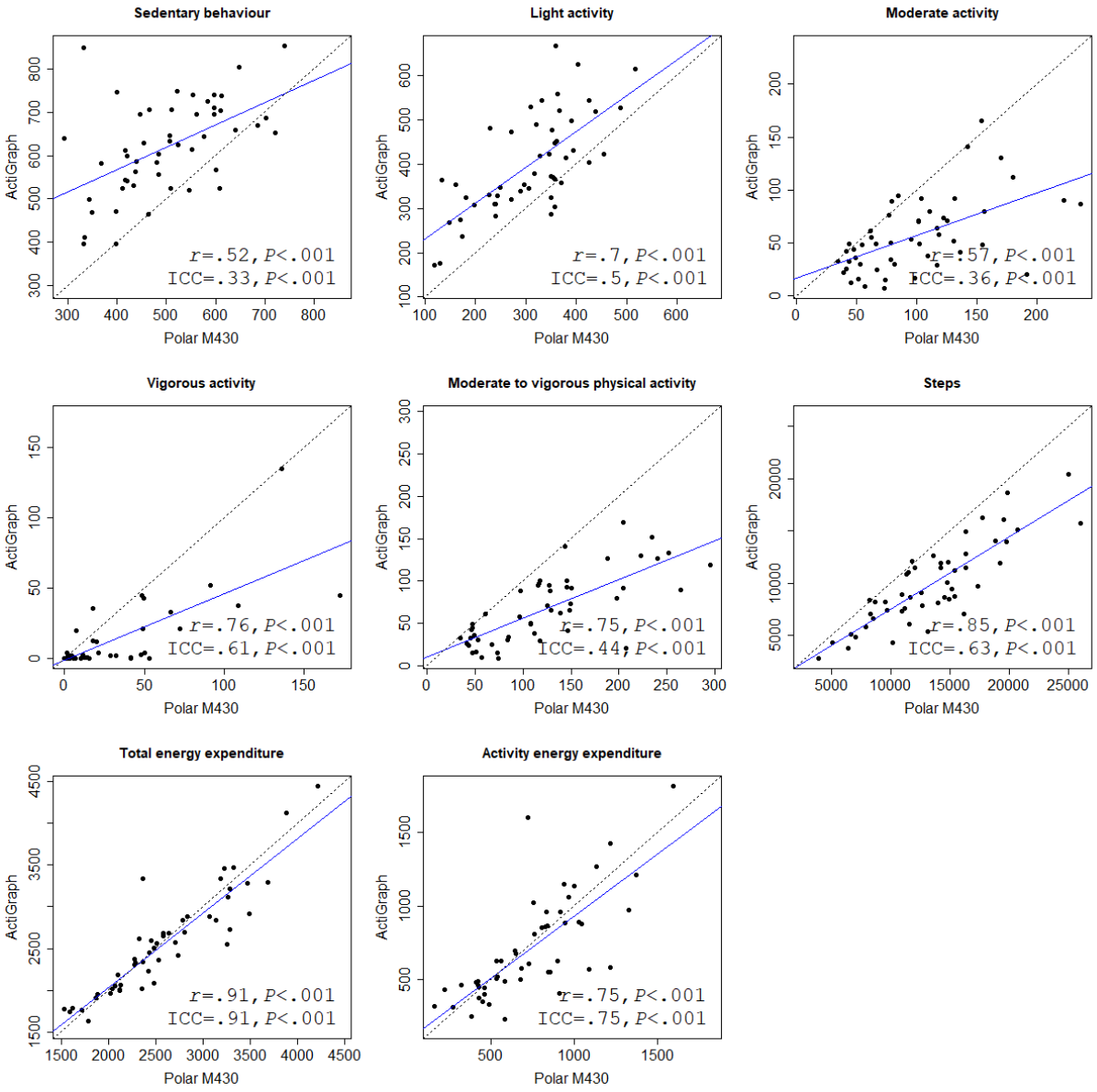
AEE: activity energy expenditure; CPM: counts per minute; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 10 – Combined scatterplots for energy expenditure with Pearson correlations and intraclass correlations for activity energy expenditure and total energy expenditure.

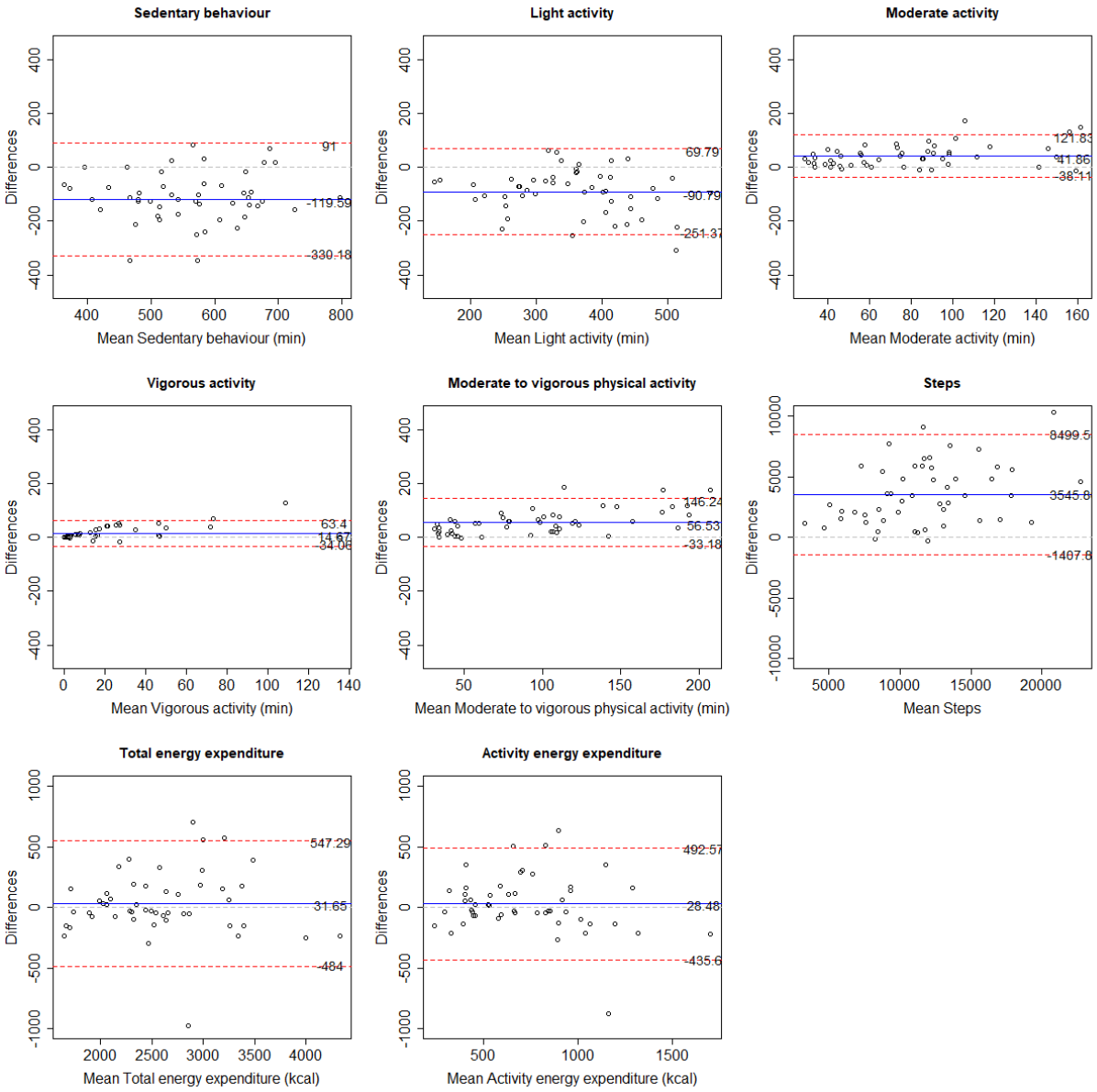


EE: energy expenditure; CPM: counts per minute; TEE: total energy expenditure; AEE: activity energy expenditure; ICC: intraclass correlation coefficient; VM: vector magnitude.

Multimedia Appendix 11 – Correlations for the Polar M430 and hip-worn triaxial ActiGraph for all variables.



Multimedia Appendix 12 – Bland-Altman plots for the Polar M430 and hip-worn triaxial ActiGraph, for all variables.



Paper III

Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L.

Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach.

BMC Public Health, 2020. 20(1): p. 1300. doi: 10.1186/s12889-020-09406-w.

Henriksen et al. BMC Public Health (2020) 20:1300
https://doi.org/10.1186/s12889-020-09406-w

BMC Public Health

RESEARCH ARTICLE Open Access

Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach

André Henriksen[1*], Anne-Sofie Sand[2], Trygve Deraas[1], Sameline Grimsgaard[1], Gunnar Hartvigsen[3] and Laila Hopstock[1]

Abstract

Background: Lack of physical activity (PA) is a risk factor for death and non-communicable disease. Despite this, more than one fourth of adults worldwide do not follow PA guidelines. As part of a feasibility study to test a complex intervention for increasing PA, we included a consumer-based activity tracker (AT) as a tool to measure PA outcomes and to track heart rate during exercise sessions. The aim of the present study was to identify factors that increase wear time when using a consumer-based AT for monitoring of participants in clinical research.

Methods: Sixteen participants aged 55–74 years, with obesity, sedentary lifestyle, and elevated cardiovascular risk were recruited to a 12-month feasibility study. Participants wore a Polar M430 AT to collect continuous PA data during a six-month intervention followed by 6 months of follow-up. We performed quantitative wear time analysis, tested the validity of the AT, and completed two rounds of qualitative interviews to investigate how individual wear-time was linked to participant responses.

Results: From 1 year of tracking, mean number of valid wear days were 292 (SD = 86), i.e. 80%. The Polar M430 provides acceptable measurements for total energy expenditure. Motivations for increased wear time were that participants were asked to wear it and the ability to track PA progress. Perceived usefulness included time keeping, heart rate- and sleep tracking, becoming more conscious about day-to-day activity, and improved understanding of which activity types were more effective for energy expenditure. Sources of AT annoyance were measurement inaccuracies and limited instruction for use. Suggestions for improvement were that the AT was big, unattractive, and complicated to use.

Conclusions: Adherence to wearing a consumer-based AT was high. Results indicate that it is feasible to use a consumer-based AT to measure PA over a longer period. Potential success factors for increased wear time includes adequate instruction for AT use, allowing participants to choose different AT designs, and using trackers with accurate measurements. To identify accurate trackers, AT validation studies in the target cohort may be needed.

Trial registration: U.S. National Library of Medicine, Clinical Trial registry: NCT03807323; Registered 16 September 2019 – Retrospectively registered.

Keywords: Actigraphy, Human activity, Activity trackers, Motor activity, Intervention study, Clinical trial, Polar M430

* Correspondence: andre.henriksen@uit.no
[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
Full list of author information is available at the end of the article

Background

The World Health Organization recommends at least 150 min of moderate physical activity (PA) or 75 min of vigorous PA each week for adults [1]. Physical inactivity is a leading risk factor for death and a range of non-communicable diseases, including cardiovascular disease, diabetes, and some cancers [2]. Worldwide, in the adult population, 23% of men and 32% of women were physically inactive in 2016 [3–5]. Physical inactivity is more prevalent in high-income countries, and together with obesity, they are increasing globally [3, 4]. At the population level, increased PA provides health and economic benefits [6], and achievement of the PA recommendations has shown to reduce both cardiovascular disease mortality and total mortality [7].

Most lifestyle intervention studies use traditional research instruments (e.g. accelerometers, pedometers, doubly labelled water, and calorimetry) for objective PA and energy expenditure (EE) data collection [8], but the number of studies using consumer-based activity trackers are increasing [9]. Validation studies on such activity trackers show different results, but several recent reviews show that some metrics for some activity trackers are accurate enough to measure PA in research settings [9–13]. In addition, a recent meta-analysis by Brickwood et al. [14] indicates that including an activity tracker as part of a PA intervention may increase PA participation through self-monitoring as well as assist researchers in participant monitoring. This is also supported by earlier systematic reviews, where De Vries et al. [15] found an increase of PA in adults with overweight and obesity, and Lewis et al. [16] found similar findings among intervention studies on adults.

However, few studies utilizing activity trackers use tracker output as outcomes, and recording time is mostly limited to the intervention period [8]. Exceptions include Schrager et al. [17] who used a Fitbit Flex to collect PA over 1 month (secondary outcomes), Carmichael et al. [18] who used a Garmin Vivofit 3 to collect PA (primary outcome) for up to 1 month of follow-up, and Patel et al. [19] who used a Fitbit Flex in a 12-week intervention, with 12-weeks of follow-up. Although long time follow-ups with consumer-based activity trackers are uncommon, such studies are likely to increase in frequency going forward. For instance, Halse et al. [20] are planning an RCT where participants will be asked to wear an activity tracker for 6 months as part of an intervention, with six additional months of follow-up. Similarly, Maxwell-Smith et al. [21] planned a 12-week RCT, where participants would wear a Fitbit Alta, with 12 additional weeks of follow-up. Although Fitbit-results for the follow-up period are not yet reported, results from this intervention period have been published [22].

Measuring long-term effects of a PA intervention by requesting participants to return for additional measurements several months after intervention end, can be expensive, time consuming, and add to the participant burden. To understand the long-term effect of PA interventions better, future research should include activity trackers and collect PA data during- and beyond- the intervention period. There is a need to identify success factors that can contribute to the adaptation of this approach. Phillips et al. [23] identified a range of challenges associated with using activity trackers in research. They grouped challenges into participants' challenges, challenges with the research setting, and challenges with the activity tracker.

In the planning of a randomized controlled trial (RCT), the RESTART trial, with a complex lifestyle intervention for lasting lifestyle changes, we conducted a feasibility study that included a Polar M430 (Polar oy, Finland) activity tracker to track PA for 1 year. The Polar M430 was chosen because it was recently released (2017), claimed high pulse sensor accuracy, and had an acceptable price.

Having access to both quantitative and qualitative data from the same study gives an opportunity to gain a more complete understanding of the research topic by comparing and combining different perspectives [24]. To look further into some of the areas identified by Phillips et al. [23], we used a qualitatively driven mixed methods approach where we analysed qualitative participant interviews together with an analysis of relevant quantitative PA recordings. In this paper, we describe our findings and provide recommendations for future research.

The aim of the present study was to identify factors that increase wear time, in terms of daily wear adherence and prolonged usage, when using a consumer-based activity tracker for participant PA monitoring in clinical research.

Method

Participant characteristics

Sample

For the feasibility study we invited 75 randomly selected participants from the seventh wave of the Norwegian population based Tromsø Study [25]. Inclusion criteria were age ≥ 55 years, body mass index ≥30 kg/m^2, self-reported sedentary lifestyle, and increased cardiovascular risk. Sixteen participants (participation 21%) responded and were recruited for a 12-month feasibility study on lasting life-style change, comprising a six-month exercise intervention with 6 months of follow-up.

RESTART feasibility study

Participants in the feasibility study were exposed to a 22-week intervention of two 1-h group-sessions per week with instructor-led gradually intensified exercise

sessions (endurance and strength), three 2-h group counselling sessions with nutritionist (Nordic Nutritional Recommendations [26]) and psychologist (Implementation Intention-based strategies [27]). Participants wore a Polar M430 activity tracker during the intervention period and for 6 months of follow-up. The activity tracker was used for participant monitoring and to allow participants to self-monitor heart rate during training sessions. The activity tracker was *not* used as a tool for behaviour change. The primary aim of the feasibility study was to examine whether the intervention was feasible to progress to a definitive RCT, regarding recruitment, adherence, and side effects. Participants received written and oral instructions on how to wear the activity tracker. Details about the feasibility study are described elsewhere [28].

Polar M430

Physical activity recording

We equipped participants with a Polar M430 activity tracker 1 week before intervention start and instructed them to wear it for the duration of the intervention study (i.e. 6 months). Participants also wore an ActiGraph for 8 days at baseline and 8 days at the end of the six-month intervention. For each participant, we therefore recorded up to 16 days of simultaneous measurements with the ActiGraph and the Polar M430. ActiGraph output was used to monitor change in PA and to test the validity of the Polar M430 in the present cohort for relevant variables (i.e. MVPA, steps, and total energy expenditure (TEE)).

Instruments

The Polar M430 was released in 2017. It has a six LED (light-emitting diode) wrist-based photoplethysmography sensor, i.e. optical pulse sensor, and a 50 Hz triaxial accelerometer for tracking PA. It is waterproof, weighs 51 g, has up to 20 days of battery life, and cost 150 USD. In a previous study we have shown that the Polar M430 gives valid results for TEE in a wider age- and weight-range, when compared to a hip-worn ActiGraph wGT3X-BT accelerometer (ActiGraph, Pensacola, FL, USA) [29]. The same study shows that although correlations are strong for moderate-to-vigorous physical activity (MVPA) and steps, average error is high, and researchers should be careful to use these variables to infer PA levels.

The ActiGraph is extensively used in PA research and is considered valid for PA intensity [30], step counting [31], and EE recording [32]. The ActiGraph (firmware version 1.9.2) was setup using ActiLife version 6.13.3 (ActiGraph, Pensacola, FL, USA). Output variables were generated using ActiLife. MVPA variables were calculated using triaxial activity count cut-offs at 2690 or above, as suggested by Sasaki et al. [13]. Steps were internally calculated by the ActiGraph and exported directly (through ActiLife). Activity EE variables were calculated using "Freedson VM3 Combination '11" (Sasaki 2011 [13] + Williams Work-Energy), and converted to TEE by adding resting energy expenditure (using the Schofield equation [33]) and 10% of TEE to account for dietary induced thermogenesis.

Polar M430 setup and usage

For each participant we created a de-identified account on Polar Flow [34], Polar's online cloud storage solution, containing only demographic data (i.e. sex, year of birth, weight, and height). No identifiable information was stored on the accounts, and participants did not have access to account credentials. Since we did not want activity tracker feedback to affect participant behaviour, all notifications and feedback messages were disabled, except sleep, which was impossible to disable. The Global positioning system (GPS) was disabled to reduce battery consumption and for privacy reasons. We initially asked participants to wear the activity tracker for the duration of the study (i.e. 6 months) and to wear the tracker all day and night (24 h/day). They were told to take the activity tracker off during sleep if they experienced any discomfort.

Due to the long recording period, we asked participants who owned a smartphone to install the Polar Flow mobile application on their private smartphones. Polar Flow is used to transfer data between the activity tracker, a smartphone, and Polar's online cloud storage. We assisted participants with connecting the activity tracker to their smartphone and aided in any issues related to the activity tracker throughout the study period.

We instructed participants to initiate data synchronization (between activity tracker and smartphone) and charging every Sunday. This bring-your-own device (i.e. smartphone) approach has shown to improve the experience and engagement of participants [35]. For participants who did not own a smartphone, we linked their activity tracker to a project smartphone. Since only five activity trackers had to be connected to the project smartphone, we did not encourage pairing with other private devices (e.g. laptop). Data synchronization between the project smartphone and activity trackers were initialized every few weeks during the weekly exercise sessions.

The first author met with participants regularly to assist in connectivity issues with the activity tracker. During these sessions, spontaneous discussions between the researcher and participants about the activity tracker occurred. Relevant information from these discussions is reported and addressed in the discussion together with other experiences from the researcher perspective.

After intervention end, we asked participants to continue to wear the Polar M430 for an additional 6 months, for a total wear time of 12 months. Participants without a smartphone meet with a researcher every 2–3

months in the follow-up period to download data from the activity tracker. After study end, we offered the Polar M430 to the participants for their private use. Participants were only informed after study end that they would receive the activity tracker. We collected no further data after the handover.

Participant perspective

To gain a deeper understanding of participant experiences with the Polar M430, we used a qualitative approach as qualitative methods are well suited for accessing participants experiences and perceptions [36]. We performed semi-structured interviews, as described by Kvale and Brinkmann [37]. All participants took part in individual interviews at two time-points, mid-way in the intervention and 6 months after intervention end. Interview guides were developed and used during the interviews to secure that all relevant aspects were covered. An excerpt of the interview guides, with questions related to activity tracker experiences, is given in in Table 1.

Analysis

Participant characteristics were described descriptively. In addition, we included a comparison of responders and non-responders, using data registered at the seventh wave of the Tromsø Study. We downloaded daily values for steps, TEE, MVPA, and hours of wear time from the Polar M430, and analysed hours of wear time to define valid days for the full year of recording. A day was considered valid if the activity tracker had at least 10 h of wear time [38]. Wear time was analysed descriptively, reporting valid days (percentage of 1 year) for each participant, mean number of valid days, and number of valid days for participants who used the activity tracker for the whole 12 months of recording. In addition, wear time was analysed with participant comments.

As suggested by Phillips et al. [23], we also tested the validity of the Polar M430 to check whether it was valid in the current cohort of participants. We used repeated measures correlations [39], with bootstrapping, to calculate correlations between the Polar M430 activity tracker and the ActiGraph wGT3X-BT accelerometer. We used correlation cut-offs suggested by Evans [40], i.e. very weak: < 0.2, weak: 0.2–0.4, moderate: 0.4–0.6, strong: 0.6–0.8, and very strong: > 0.8. We also calculated mean absolute percentage error (MAPE) for each variable, using 10% error as cut-off for acceptable error in free-living studies. Finally, we used Bland-Altman limits of agreements to assess consistency between instrument outcomes [41]. Statistical analyses were performed using R version 3.5.3.

The second author performed the verbatim transcriptions of the *mid-way interview* audiotaped sessions, while a professional firm (Digforsk AS) performed the transcriptions of the *six-months after* audiotaped sessions. We used the computer software QSR NVivo 12 Plus (QSR International, Pty Ltd) as a tool for structuring data in the analysis process. We used thematic analysis when identifying and reporting themes and patterns in the data, a widely used method among health researchers [36]. We used an inductive and semantic approach to identify themes, to allow the themes to emerge from the data and to identify participant's opinions. To identify patterns in the text, we used the six steps defined by Braun and Clarke [42] for thematic analysis: data familiarization, initial coding, generating themes, reviewing themes, defining and naming themes, and writing up report. Comments mentioned by only one participant were given equal weight as comments mentioned by multiple participants [43]. Coding was done by the first and second author and later harmonized through discussion. Analysis was done by the first author, and thoroughly reviewed by the second and last author. Quotes used in the manuscript were translated from Norwegian.

Coding was done in three iterations. The first iteration was done on paper and resulted in many partly overlapping codes. The second iteration was done in NVivo, where we merged initial codes into the following 11 themes: 1) metric inaccuracy, 2) elements that triggered irritation, 3) tracker visual design (look and feel), 4) tracker practical design (ease of use), 5) motivation for usage, 6) effect of using the tracker, 7) how tracker was used, 8) why the tracker was used, and comments on available metrics, including 9) sleep, 10) pulse, and 11) PA. These were further refined into the four final themes: motivation, activity tracker usefulness, activity

Table 1 Excerpt from interview guides with questions related to the activity tracker

Interview	Number	Question
Mid-way	1	How was your experience with using the activity tracker?
6 months after	2	How did you use the activity tracker? (Only during workouts, or also other times? Pulse zones? As a watch?)
6 months after	3	Did the activity tracker motivate you to work out more often? Harder?
6 months after	4	Was there anything special about the activity tracker that made you more motivated?
6 months after	5	What motivated you to wear the activity tracker (for an extended period)?
6 months after	6	Is there anything you wish was possible with the activity tracker, which could have motivated you to wear it longer?

tracker annoyances, and activity tracker improvements. Quotes are tagged with sex, age group, and whether they owned a smartphone or not.

Results

Participant characteristics

Among the 16 participants, 11 (70%) owned a smartphone and could connect their phone to the activity tracker. Participant characteristics at baseline are given in Table 2. Activity tracker recording was performed between October 2017 and September 2018. An overview of responders and non-responders, using recorded data from the seventh wave of the Tromsø Study, held approximately 2 years before the RESTART feasibility study, is given in Table 3.

Polar M430

Wear time

From the available 365 days of tracking, when including all 16 participants, mean number of valid days was 292 (SD = 86), i.e. 80%. Half of the participants had 30 or less non-valid days for the whole year of recording. Two participants (number 14 and 15 in Fig. 1) stopped using their activity tracker at the end of the intervention (after 6 months of wear time). Mean number of valid days for the whole year, when excluding these participants, was 313 (SD = 69), i.e. 88%. We observed no difference in wear time between the different months, except one participant (8) who stopped using the activity tracker during the summer holiday (July), and one participant (13) who mostly stopped using the activity tracker after the intervention but resumed wearing it after the summer months. An overview of valid days for all participants for the whole year of recording is given is Fig. 1.

The two participants who terminated use of the activity tracker after 6 months reported similar reasons for this. One participant reported being very conscious about wearing the activity tracker during the intervention and said that she became more disciplined by wearing it, which resulted in an increase in motivation. However, after the intervention ended, she "*just felt done with it*" *(Participant 10, female 70–80, smartphone).* Two specific reasons were that it was too complicated, and she had trouble with the connected smartphone, and therefore did not have easy access to all the metrics. As stated in the interviews, "*I did not see the results I wanted on my iPad … my daughter has a watch I like better … it is simpler*" *(Participant 10, female 70–80, smartphone).* The other participant reported mainly using the activity tracker as a tool to keep track of pulse zones during instructor led exercise sessions: "*you were told to increase your heart rate by an amount, and then you could look at the watch*" *(Participant 14, male 60–70, no smartphone).* In addition, he did not have a connected smartphone, and felt the activity tracker was too complicated, especially without access to the instruction manual. "*When you don't know … how to use the watch … if I had the instruction manual I could see [how to use it]" (Participant 14, male 60–70, no smartphone).*

Table 2 Participant characteristics at baseline. The RESTART feasibility study 2017–18

Characteristics	Value
Age in years, mean (SD)	66.1 (5.8)
Smartphone owner, mean age (SD)	65.2 (4.8)
Not smartphone owner, mean age (SD)	68.2 (7.8)
Male sex, % (number)	68.8 (11)
Body mass index, kg/m^2, mean (SD)	35.6 (5.3)
Current smoking[a], % (number)	13 (2)
High total cholesterol, % (number)	50 (8)
Low HDL (high-density lipoprotein) cholesterol, % (number)	25 (4)
Hypertension, % (number)	19 (3)

Current smoking Self-reported daily smoking, *High total cholesterol* Total cholesterol ≥5 mmol/L; *Low HDL cholesterol* HDL cholesterol < 1.3 (women) or < 1.0 (men) mmol/L, *Hypertension* Blood pressure ≥ 140/90 mmHg, *SD* Standard deviation. [a]missing values: 1 participant

Polar M430 validity

We used output from overlapping days of Polar M430 and ActiGraph usage to test the validity of the Polar M430 in the present study. One participant did not wear both devices simultaneously and were excluded from analysis. Remaining participants had 8 to 16 valid days of simultaneous recordings. All analyses are based on data from 203 days of measurements distributed among 15 participants.

We found a strong correlation between the ActiGraph and the Polar M430 for step count, and a moderate correlation for MVPA and TEE. On average, the Polar M430 over-reports steps and time in MVPA, and under-reports TEE. Only TEE had a borderline acceptable MAPE. Details for each variable are given in Table 4.

Participant perspective

We grouped comments into four themes: Motivation, activity tracker usefulness, activity tracker annoyances, and activity tracker improvements.

Motivation

This theme explores if and how participants were motivated by wearing an activity tracker during- and after the intervention. Some participants mentioned the activity feedback from the activity tracker, and the possibility of directly observing progress, as the primary motivation to wear it for such a long period. For these participants, this feedback was an opportunity to push themselves

Table 3 Descriptive characteristics by attendance. The seventh wave of the Tromsø study

Characteristics	Attended the pilot	
	No	Yes
Number of participants	59	16
Age in years, mean (SD)	65.3 (5.7)	64.1 (5.8)
Male sex, % (number)	76.3 (45)	68.8 (11)
Body mass index, kg/m^2, mean (SD)	34.0 (3.5)	36.2 (5.8)
Current smoking, % (number)	27.1 (16)	18.8 (3)
Total cholesterol, mmol/L, mean (SD)	5.8 (1.1)	5.5 (1.2)
HDL (high-density lipoprotein) cholesterol, mmol/L, mean (SD)	1.3 (0.5)	1.2 (0.3)
Systolic blood pressure, mmHg, mean (SD)	151.7 (18.3)	144.4 (15.4)

Current smoking Self-reported daily smoking, *SD* Standard deviation

harder, especially during the instructor led exercise session.

One participant stated that being able to measure progress, when she did not think there would be any progress, was very motivating and gratifying, and stated: *"I reached my goals … it was very gratifying … I did not think I would [reach my goals]" (Participant 12, female 70–80, no smartphone).* Another participant highlighted that the ability of using the activity tracker to push himself into working harder and harder each session was motivating, and said that *"It was interesting to follow progress, … I have never used this [technology] before, … nice to observe that … yes, now I have pushed myself" (Participant 2, male 50–60, smartphone).*

During ad-hoc conversations throughout the intervention period, many participants stated that they were happy with being invited to the project and wanted to contribute to the research by sharing their data. This was also confirmed in the interviews, where several participants indicated that an important reason for wearing the activity tracker for such a long period was that they were asked to do it. This willingness to share was expressed by several participants: *"We were asked to wear it … I though it is only fair [for the benefit of the study]" (Participant 4, female 60–70, smartphone), "I know how important research is … so that you will get reliable data … I was willing to make the 'sacrifice' for you and the research" (Participant 13, male 60–70, smartphone),* and *"No, I didn't (when asked if he reviewed recorded data), I just let it [the activity tracker] do what it was supposed to do [record data] and I just did what I was supposed to do [share data]" (Participant 3, male 70–80, smartphone).*

Activity tracker usefulness

This theme encapsulates how and why participants used the activity tracker, as well as their perceived effects of using it. Most participants reported mainly using the

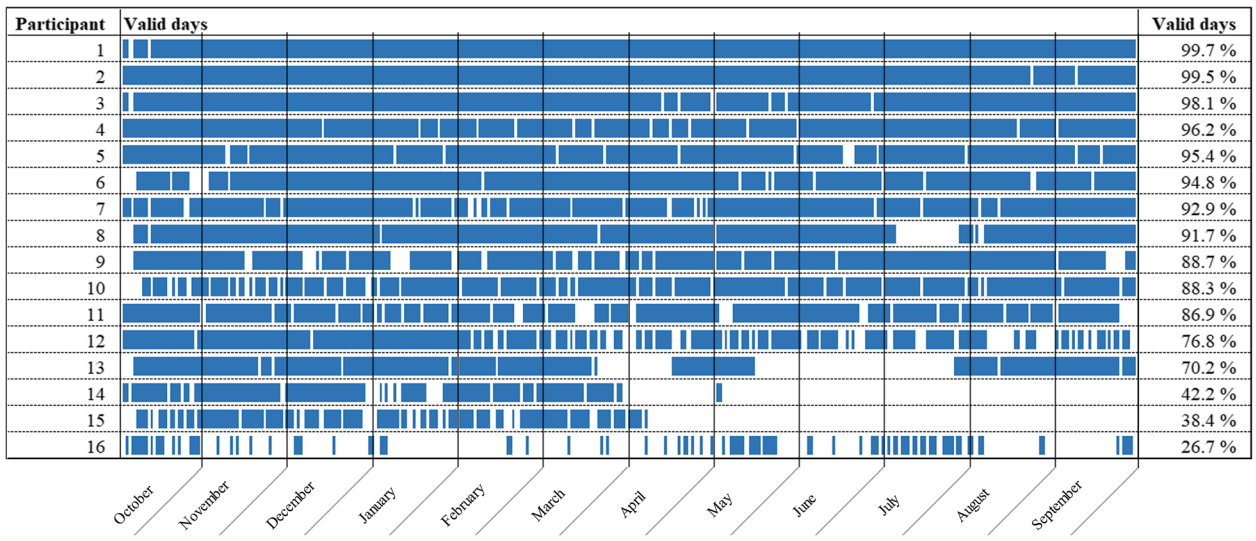


Fig. 1 Activity tracker wear time for 1 year of recording

Table 4 Mean data for Polar M430 and ActiGraph, and correlation, *p*-value, mean absolute percentage error, and Bland-Altman mean difference and limits of agreements (LoA), for steps, moderate-to-vigorous physical activity, and total energy expenditure. Person-days: $n = 203$

Variable	Polar	ActiGraph	Correlation (95% CI)	***P***-value	MAPE	Mean difference	Lower LoA	Upper LoA
Steps	8956 (5106)	5165 (3230)	0.625 (0.44, 0.70)	< 0.001	119.5%	3791	− 4860	12,442
MVPA	143 (97)	44 (32)	0.495 (0.31, 0.53)	< 0.001	373.9%	99.0	−80.4	278.4
TEE	2868 (581)	2967 (458)	0.446 (0.50, 0.69)	< 0.001	10.6%	− 98.7	− 948.5	751.1

Numbers are means (standard deviations). *MVPA* Moderate-to-vigorous physical activity, *TEE* Total energy expenditure, *Correlation* Repeated measurement correlation with 95% confidence interval, *MAPE* Mean absolute percentage error, *LoA* Limits of agreement

activity tracker to get continuous feedback on heart rate during instructor led workouts. In addition, it was also used as a timepiece, and some used it as a tool for measuring sleep and tracking PA, during and after the intervention.

One participant highlighted the usefulness of the activity tracker by saying, "*I had to pay attention to how I performed, so I could increase resistance to get to the [heart rate] level I was supposed to be at" (Participant 1, male 60–70, smartphone).* However, not all who said they used it to track changes during a workout payed much attention to it, as illustrated by one participant who said, "*I didn't put too much into it, but it was fun to keep track [of the activity]*" *(Participant 2, male 50–60, smartphone).*

In addition, many participants used it as a timepiece, and replaced their existing wristwatch with the Polar M430 to accommodate the study. One participant stated that he "*only used it as a watch" (Participant 13, male 60–70, smartphone)* and some simply answered "*Yes" (Participant 6, male 50–60, no smartphone. Participant 16, female 70–80, no smartphone),* when asked in a follow-up question if they simply used the watch as a timepiece.

An often-mentioned useful feature was the ability to track sleep quality and sleep interruptions during the night. For some this was an acknowledgement of what they already knew about their sleep patterns, prompting responses like "*I look at sleep … I am awake a lot*" *(Participant 8, male 60–70, smartphone)* and "*I can see how little sleep I get*" *(Participant 6, male 50–60, no smartphone)*. For others it constituted a source of confusion because the activity tracker was perceived as inaccurate, resulting in quotes like "*tracking sleep … but I don't always think it is accurate*" *(Participant 15, female 60–70, smartphone).*

The reported effects of wearing the activity tracker were different for most participants, and only a few mentioned specific behavioural changes because of the activity tracker. However, one participant said, "*I became more disciplined*" *(Participant 10, female 70–80, smartphone).* Another participant mentioned that he became more conscious about daily activity levels and which types of activity that were effective and stated, "*I am more conscious about moving more while at work …. I take the stairs instead of the escalator*" *(Participant 11, male 60–70, smartphone)*, and "*… more aware of what is effective and what isn't*" (*Participant 11, male 60–70, smartphone).* One participant highlighted this learning effect by saying, "*I learned something from the watch. Things that I thought was [effective] … the watch showed me that it actually wasn't*" *(Participant 4, female 60–70, smartphone).*In addition, during ad-hock discussions with participants when performing technical support on the activity tracker, some participants stated that they compared activity tracker output with each other and found that interesting.

Activity tracker annoyances

This theme summarizes issues that participants found annoying about the activity tracker. Being annoyed with the activity tracker may reduce motivation to wear it. Sources of annoyance should therefore be identified and addressed if possible.

Technical challenges were a major source of annoyance, where participants experienced disconnects between their smartphone and activity tracker, and often found that the activity tracker was difficult to use without assistance. This was repeatedly mentioned during the interviews, prompting responses such as, "*negative about the watch … we got no instructions on how to use it*" *(Participant 1, male 60–70, smartphone),* "*a lot of information at once, considering I hadn't used this [technology] before*" *(Participant 4, female 60–70, smartphone),* and *"I have struggled with the technical aspects" (Participant 13, male 60–70, smartphone).* Several participants mentioned that it could be helpful to have access to the instruction manual, to better understand both the complicated features and the more basic watch features. As stated by two participants: "*It was too complicated … but I didn't spend too much time on it anyway … because we didn't have the instruction manual" (Participant 10, female 70–80, smartphone),* and "*I miss instructions about the watch … unsure how to set time*" *(Participant 7, male 70–80, smartphone).*

Activity tracker inaccuracies was also a major source of annoyance, and sleep feedback was repeatedly mentioned as a source of annoyance because of perceived inaccuracy. Two participants who had contradictory

experiences may best describe this. One participant said, "*The only thing that annoys me … when I feel that I have slept very well … it reports how bad I have slept*" *(Participant 3, male 70–80, smartphone)*, and the other said, "*the sleep thing … I almost got annoyed sometimes … I woke several times per night, and sometimes I am out of bed three-four times ... and it reports that I have slept well*" *(Participant 12, female 70–80, no smartphone).*

One participant also noticed that the pulse sensors was not always accurate, and she got somewhat frustrated about this, stating that, "*I got very caught up in the [low] pulse measurements … resting at 39 [beats per minutes] during the day? I don't get it*" *(Participant 4, female 60–70, smartphone).* Another participant also wondered about the accuracy during exercise sessions, and mention that, "*I wonder if the watch is correct … it's not correct … much lower pulse … not even close*" *(Participant 5, male 70–80, no smartphone).*

Interest in tracking activity was limited for some participants who did not own a smartphone (and we could *not* connect the activity tracker to their phone). When asked about whether they used the watch to track PA, one responded *"No, I didn't, … we could have connected [the watch] to a smartphone, but I didn't have [a smartphone*]" *(Participant 6, male 50–60, no smartphone).* Furthermore, when we asked if they missed any features on the activity tracker, only lack on direct feedback on PA metrics were mentioned by these participants. As stated by two participants without a smartphone: "*Sigh. Yes, steps*" *(Participant 16, female 70–80, no smartphone)* and "*Steps, … I am almost certain it is available on the watch*" *(Participant 14, male 60–70, no smartphone).* In addition, one participant, who owned a smartphone but where the connection between her smartphone and activity tracker was unstable and hampered data transfer, pointed out that this made it more complicated to use the activity tracker and said, *"It was hard to use the watch … I did not see the results as I wanted … I think those who saw their results on their phones got more out of it" (Participant 10, female 70–80, smartphone without successful connection).*

Activity tracker improvements

The final theme captures suggestions that participants reported regarding the choice of activity tracker. Most participants were happy to wear the activity tracker during the intervention, both day and night, and reported no major issues with the day-to-day usage. However, some participants mentioned that the activity tracker could have been more attractive, and some felt it was too large and tight, prompting comments such as, "*It is [for instance] not good looking during the Christmas holiday*" *(Participant 4, female 60–70, smartphone*), "*I take it off when I dress up*" *(Participant 10, female 70–80, smartphone), "I have a more expensive watch I use when I want to look nice" (Participant 1, male 60–70, smartphone),* and "*It is a bit big … also tight*" *(Participant 1, male 60–70, smartphone).* Other participants were not too concerned about the design of the activity tracker, and one even made a point of saying "*I could not be bothered to wear another watch when at parties*" *(Participant 15, female 60–70, smartphone).*

Although the activity tracker had more features than we informed participants about, some pointed out that they knew other people with more advanced activity trackers with more interesting features. One participant said, "*My daughter has a more advanced [watch], with all possible features ... but it is of course more expensive*" *(Participant 13, male 60–70, smartphone).* On the other hand, another participant, who had a daughter with a less complex activity tracker, thought it would be better to use a less complicated activity tracker and commented that "*I liked it better … it was easier to use*" *(Participant 10, female 70–80, smartphone).*

Discussion

Summary of findings

In this feasibility study with 12 months of PA recording, we analysed participant wear time, tested the Polar M430 validity in this sample, and reported participant experiences with long term usage. Wear time was high throughout the study. The Polar M430 over-reports steps (strong correlation) and MVPA (moderate correlation), and under reports TEE (moderate correlation). TEE had borderline acceptable error. Main *motivations* for increased wear time were that they were asked to do it and the ability to track activity progress. Regarding *usefulness*, most participants mainly used the activity tracker as a timepiece, but some also used it to measure heart rate and sleep tracking. In addition, reported positive effects were being more conscious about their day-to-day activity and improving their understanding of the effect of different activity types. Two major sources of *annoyance* were sleep- and -heart rate inaccuracy and limited instruction for use on the activity tracker. Suggestions for *improvement* were that the Polar M430 was big, unattractive, and too complicated to use.

Participant characteristics

We invited 75 participants randomly selected from the seventh wave of the Tromsø Study. Since only 16 accepted the invitation, we included everyone who accepted, resulting in a sex skewed cohort of 70% men. All participants owned a mobile phone, but only 70% owned a smartphone. Smartphone penetration is lower in older age groups [44], which we also saw in this sample, as those owning a smartphone had a lower median

age (64y vs 73y) compared to those who did own a smartphone.

Polar M430

Wear time

In the present study, wear time was high, and most participants wore the activity tracker for the duration of the study. This high wear time is in alignment with a similar study, where Duignan et al. [45] conducted a shorter intervention study (3 months) in a younger sample (mean age: 23.4, SD: 2.8). In this study, 73% of participants still wore an activity tracker after 87 days, with an average wear time of 79 days (90%) among remaining participants. Reasons for loss of participants were mostly technical (e.g. data synchronizing) and loss of activity tracker. However, in a observational study by Hermsen et al. [46] they saw a slow exponential decline in wear time of a hip-worn Fitbit Zip, also mostly due to technical reasons, where only 16% still wore the activity tracker after 320 days. Although anecdotal, this indicates that being part of an intervention with close follow-up of participants increases wear time, as compared to studies where participants are only observed.

Polar M430 validity

In a systematic review of Polar activity trackers [47], we have previously reported that Polar activity trackers show mixed results depending on activity tracker, study setting, and study sample. Furthermore, compared to findings in a previous Polar M430 validation study [29], with a wider range of weight, height, and age, correlations were lower and MAPEs were higher in the present cohort.

The difference in results between the two validation studies shows that it is good practice to perform a separate validation study on participants with similar characteristics as the sample under study, when planning to use a consumer-based activity tracker in clinical research, as suggested by Phillips et al. [23]. The ActiGraph and the Polar M430 are worn of different locations, which may contribute to the large difference in MVPA and steps between devices. Certain activity types, e.g. stationary biking where hands are placed firmly on the bike's handle, will result in more activity on the hip compared to the wrist. TEE is less affected by this difference as resting energy expenditure (energy consumed to maintain body functions at rest) is the main component of TEE and constitutes between 60 and 75% of TEE [48]. In addition, about 10% of TEE is expended from food digestion (dietary induced thermogenesis).

The Polar M430 is not a suitable replacement for the ActiGraph but can be used as a source of additional information for long term monitoring, for some variables.

Participant perspective

Challenges and solutions

In the following, we discuss challenges and potential solutions, drawn from participants' feedback together with experiences from the researcher perspective, and results from the objective data analyses.

Motivation and activity tracker usefulness Most participants were enthusiastic about being invited to participate in the study. This was expressed repeatedly throughout the intervention during ad-hoc encounters. Because of this and because collecting data from the activity tracker was presented as an important part of the intervention, we do not find it surprising that wear time was high during the intervention. This is also in accordance with Duignan et al. [45] who achieved high wear time in a 3-month PA intervention. Wear time during follow-up was higher than expected, as the observational study by Hermsen et al. [46] showed high activity tracker attrition. However, the same study also showed that this attrition was lower in higher age groups, which may be part of the explanation of the high wear time in the present study.

Most users, when buying a new activity tracker, tend to stop using it after a few months, mostly due to loss of motivation [14, 49]. In the present study, only two of 16 stopped using the activity tracker after 6 months (i.e. intervention end). A major reason that participants in the present study wore the activity trackers for a full year, was because they were asked to wear it and they wanted to contribute to the study. This suggests external motivation and, at least for this group, may partly explain why activity tracker usage is not higher in the general population. About 20% of Americans use an activity tracker, with about 10% usage among people aged 55 and above [50]. While some reported annoyance with sleep and heart rate inaccuracies, we believe most participants were not too concerned with activity tracker accuracy, but more concerned about understanding how to use the activity tracker and having access to all collected data.

Similarly, we observed (during ad-hoc interactions with participants during the intervention) that some participants sometimes compared activity tracker output with each other. This may also indicate that having access to activity output for self-monitoring and being able to compare and compete with others was a possible source of motivation for prolonged wear time. This observation supports earlier findings which shows that activity tracker feedback can motivate PA participation in and of itself [14–16]. This effect must thus be considered when planning and analysing results of a PA intervention, to avoid ascribing increased PA participation to the intervention when the activity tracker itself may have

been a major source of motivation. In addition, for participants who found the activity trackers useful during exercise sessions, and those who found it useful for learning which types of activity that were effective, it is likely that these features contributed to the increased wear time. It is apparent that activity tracker output is important for many, and unless there are specific reasons to *not* displaying these outputs, researchers should use an activity tracker that can show output that participants would find relevant to track their own progress (e.g. steps and/or minutes of MVPA).

Activity tracker annoyances We found several sources of annoyance or nuisance among participants, where problems were mostly related to technical problems, activity tracker inaccuracy, and activity tracker complexity.

Technical problems during smartphone and activity tracker setup are likely to occur because of the large variation in participant phone models. It is therefore necessary to schedule enough time available for setup and have technically skilled personnel available who can resolve any issues directly. Too many technical problems may reduce participants motivation to wear an activity tracker. This is also suggested by Hermsen et al. [46] who found that the main reason (57%) for tracker attrition was related to technical problems. In addition, some participants did not bring their smartphone for the setup meeting, and several participants' phones were out of power. Participants should have been reminded to bring a fully charged smartphone, and we should have brought charging equipment to the initial meeting. In addition, some participants lost their charging cable, and one misplaced the activity tracker for a period, showing that replacement equipment should also be available.

We did not specifically ask participants to clean the activity tracker regularly. Because of the long recording period, this caused the optical pulse sensor to become unclean and therefore unreliable. This sensor emits light onto the skin and estimates pulse by analysing changes in light waveform from the reflecting light. The reflecting light is affected by change in blood volume under the skin [51]. Annoyances about heart rate inaccuracy could have been avoided, at least partly, by instructing participants to clean the activity tracker regularly. In addition, the Polar M430 regularly misclassified sleep and non-wear time. Our main reason for selecting the Polar M430 was that it had a very good optical pulse sensor (according to Polar). However, we did not consider that being unable to disable sleep notifications could cause annoyance. We did not perform sleep validation on the Polar M430, which we (in retrospect) should have done to be able to inform participants about the possible inaccuracy of this metric.

Inaccuracy was mentioned as an individual issue and as a source of curiosity when participants compared activity tracker output between themselves and saw different results for the same activity. People are different and activity tracker output will differ between individuals. An additional possible source of variation may be activity tracker firmware, which is routinely updated by vendors. How updates affect activity tracker output are mostly company secrets. We therefore avoided updating the firmware unless we could update all activity trackers simultaneously. However, participants who connected the activity tracker to their smartphone were able to do this update more frequently, which resulted in several weeks where participants had different firmware. Activity tracker inaccuracy has also been identified by e.g. Hardcastle et al. [52] as a source of disappointment and false sense of achievement.

Several participants requested an easier way to view activity tracker output. The Polar M430 does not show daily step count automatically. This was annoying to several participants. This is also supported by Hardcastle et al. [52] who identified steps as the most popular feature of an activity tracker. Activity output would likely have been more accessible for participants if we had provided them with the instruction manual, which shows how to access this information. The main reasons for not providing the instruction manual were to prevent participants to change settings (e.g. turn on GPS tracking) or be affected by activity tracker output. However, since wearing an activity tracker is likely to only affect short term behaviour [14] we suggest providing participants with the instruction manual for long-term measurements. The importance of having access to the instruction manual and that lack of instructions are a source of annoyance, is also supported by previous studies on activity tracker use in older adults [53]. In the present study, some participants said that the Polar M430 was too complicated. However, in a study by McMahon et al. [54] on older adults using an activity tracker to increase PA, they showed that although older adults require more time to adopt new technology and needs more technical support [55], they found activity trackers easy to use and useful for PA self-tracking. Although this study used a Fitbit One, a less complex activity tracker compared to the Polar M430, adequate training in the present study would likely empower participants to use the activity tracker as intended.

Activity tracker improvements Participant feedback regarding the activity tracker was mostly related to activity tracker design and available outputs. Hardcastle et al. [52] also found appearance to be important, and the Polar A300 (an earlier Polar model with similar design as the Polar M430) was found to be especially "bulky

and clunky". Similarly, Puri et al. [56] have also shown that aesthetics and comfort are important to increase activity tracker usage. When considering activity trackers in future studies, researchers should therefore consider appearance and usability, and not only price, accuracy, battery life, etc. Many vendors offer multiple versions of the same activity tracker, with different colours, shapes, and materials. Allowing participants to choose between multiple designs may increase wear time. This may be more important in a study with a younger population, as younger participants are more likely to own an activity tracker and may be resistant to replace it or start wearing an additional device. Similarly, because some participants said the activity tracker was too complicated and others said it was too simple, it could be beneficial to have more than one activity tracker available for participants to choose between, at least if the goal is to increase wear time. The drawback is that it is more complicated to compare activity levels between participants using different activity trackers.

Recommendations

From the above discussion, we have extracted the following recommendations that should be considered when planning and performing a study where participants are equipped with an activity tracker over a prolonged period. We have grouped recommendations into three phases: 1) the preparation and planning phase, 2) the setup and training phase, and 3) the recording phase.

Preparation and planning phase

- Budget for a technician who can provide technical support throughout the study and during follow-up.
- Offer activity trackers that can easily display relevant metrics, unless there are specific reasons not to display output.
- Allow participants to choose from multiple activity tracker designs, both in terms of complexity and appearance.
- Validate recent activity trackers in the relevant cohort if no such study exists, to identify acceptable activity trackers.
- Validate all metrics on the selected activity tracker and consider informing participants about untrustworthy metrics.

Activity tracker setup and participant training phase

- Provide adequate time for training and follow-up of participants.
- Remind participants to bring a fully charged smartphone (and bring charging equipment for common phones types) before connecting participants' phones to their activity tracker.
- Instruct participants to clean the activity tracker regularly, to avoid inaccuracy in pulse measurements.
- Provide activity tracker instruction manual to participants, unless there are specific reasons not to.

Recording phase

- Keep close follow-up of participants to increase wear time.
- Have replacement activity trackers and charging equipment available.
- During study or follow-up; update activity tracker firmware simultaneously if possible.

Contribution to the literature

The most important contribution to the literature from this study is the identification of several important success factors that may increase wear time of an activity tracker, when provided to participants in a clinical study for PA tracking over a prolonged period. These factors have been summarized into a list of recommendations for clinical studies where similar methods of PA tracking are used. Following these recommendations may be timesaving for researchers, as well as reduce potential activity tracker annoyance among participants.

Strengths and limitations

The main aim of this paper was to identify factors that contributed to the wear time of the activity tracker. Study participants were recruited from a large ongoing population study, with a well-defined sample in terms of age, lifestyle habits, and health risks. This strength adds to the study's transferability to similar population groups in similar societies [57]. Another strength is the use of a mixed methods approach and the long recording period, which allowed us to identify challenges from multiple perspectives and identify challenges that would not necessarily be detected in a study of shorter duration.

The main limitation is the limited transferability to other populations and age groups. Since participants were part of an intervention, desirability bias may have affected activity tracker wear time. This limits transferability of findings to other study designs. In addition, because only 16 participants were included, the variation in quantitative findings may be due to undetected differences in background characteristics. Participation, although low (21%), is as expected because intervention studies are unavoidably hampered by selection bias because participation demands high motivation and compliance. This challenge is further reinforced in studies that also require considerable efforts from participants, i.e. lifestyle interventions. In addition, older people often decline participation in PA interventions [58]. Acceptance assessment for the underlying feasibility study is

addressed in Deraas et al. [28]. Further, participants were recruited from a population-based health study, and although the attendance was 72% in this age-group [59], this may introduce selection bias.

Conclusions

In this study, long term activity tracker wear time was high. Results indicate that it is feasible to use a consumer-based activity tracker to measure PA over a longer period. Potential success factors for increased wear time includes providing adequate instructions on how to use the activity tracker, allowing participant to choose between different activity tracker designs (appearance and complexity), and offer activity trackers with accurate measurements. Validation studies on recent activity trackers may be needed for the target cohort, to identify such trackers.

Abbreviations

AT: Activity tracker; EE: Energy expenditure; GPS: Global positioning system; MAPE: Mean absolute percentage error; MVPA: Moderate-to-vigorous physical activity; PA: Physical activity; RCT: Randomized controlled trial; TEE: Total energy expenditure

Acknowledgements

The publication charges for this study were granted from the publication fund at UiT The Arctic University of Norway. The Clinical Trial Unit at the University Hospital of North Norway made important contributions to the study by performing participant screening and measurements, as well as facilitating the interviews.

Authors' contributions

AH, LH, SG, ASS, TD, and GH contributed to the design of the study. AH drafted the manuscript with critical review by LH, ASS, SG, GH, and TD. AH (objective) and ASS (interview) collected the data. AH performed the data analysis together with ASS and LH. SG conceived the underlying feasibility study. TD was project lead in the underlying feasibility study. All authors read and approved the final manuscript.

Funding

UiT The Arctic University of Norway's thematic priority grant "Personalized medicine for public health" funded the PhD-grant for the first author. UiT also funded scientific positions for all co-authors. Tromsø County (grant TFK 2016–058) funded data collection and activity trackers in the feasibility study, as well as qualitative data collection and transcription of interviews. The University Hospital North Norway and Stamina Helse contributed to study planning and implementation of the feasibility study. The MILLS fund "1 kr til hjertesaken" contributed funding for the second round of interviews.

Availability of data and materials

The data/transcripts used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was conducted in accordance with the Helsinki declaration. The Regional Committee for Medical and Health Research Ethics in Northern Norway (Reference 1100/2017) approved the study. All participants signed an informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway. [2]Department of Health and Care Sciences, UiT The Arctic University of Norway, Tromsø, Norway. [3]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway.

Received: 28 February 2020 Accepted: 18 August 2020
Published online: 27 August 2020

References

1. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; 2010. Available from: https://www.ncbi.nlm.nih.gov/books/NBK305057.
2. World Health Organization. Fact sheet: physical activity. 2018. https://www.who.int/news-room/fact-sheets/detail/physical-activity.
3. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working G. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet. 2012;380(9838):247–57. https://www.ncbi.nlm.nih.gov/pubmed/22818937.
4. World Health Organization. Global status report on noncommunicable diseases 2014. 2017.
5. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. Lancet Glob Health. 2018;6(10):e1077–e86. https://www.ncbi.nlm.nih.gov/pubmed/30193830.
6. Ding D, Lawson KD, Kolbe-Alexander TL, Finkelstein EA, Katzmarzyk PT, van Mechelen W, Pratt M, Lancet Physical Activity Series 2 Executive C. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet. 2016;388(10051):1311–24. https://www.ncbi.nlm.nih.gov/pubmed/27475266.
7. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, Casanova A, Swaminathan S, Anjana RM, Kumar R, Rosengren A, Wei L, Yang W, Chuangshi W, Huaxing L, Nair S, Diaz R, Swidon H, Gupta R, Mohammadifard N, Lopez-Jaramillo P, Oguz A, Zatonska K, Seron P, Avezum A, Poirier P, Teo K, Yusuf S. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet. 2017;390(10113):2643–54. https://www.ncbi.nlm.nih.gov/pubmed/28943267.
8. Silfee VJ, Haughton CF, Jake-Schoffman DE, Lopez-Cepero A, May CN, Sreedhara M, Rosal MC, Lemon SC. Objective measurement of physical activity outcomes in lifestyle interventions among adults: a systematic review. Prev Med Rep. 2018;11:74–80. https://www.ncbi.nlm.nih.gov/pubmed/29984142.
9. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, Lu X. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform. 2019;93:103153. https://www.ncbi.nlm.nih.gov/pubmed/30910623.
10. Wright SP, Hall Brown TS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol. 2017;312(3):R358–R67. https://www.ncbi.nlm.nih.gov/pubmed/28052867.
11. Bunn JA, Navalta JW, Fountaine CJ, Reece JD. Current state of commercial wearable technology in physical activity monitoring 2015–2017. Int J Exerc Sci. 2018;11(7):503–15. https://www.ncbi.nlm.nih.gov/pubmed/29541338.
12. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, Hamilton CB, Li LC. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth. 2018;6(8):e10527. https://www.ncbi.nlm.nih.gov/pubmed/30093371.
13. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, Gallagher R. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas. 2018;112:85–93. https://www.ncbi.nlm.nih.gov/pubmed/29704922.
14. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth. 2019;7(4):e11819. https://www.ncbi.nlm.nih.gov/pubmed/30977740.
15. de Vries HJ, Kooiman TJ, van Ittersum MW, van Brussel M, de Groot M. Do activity monitors increase physical activity in adults with overweight or obesity? A systematic review and meta-analysis. Obesity (Silver Spring). 2016;24(10):2078–91. https://www.ncbi.nlm.nih.gov/pubmed/27670401.

16. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health. 2015;15(1):585. https://www.ncbi.nlm.nih.gov/pubmed/26104189.
17. Schrager JD, Shayne P, Wolf S, Das S, Patzer RE, White M, Heron S. Assessing the influence of a Fitbit physical activity monitor on the exercise practices of emergency medicine residents: a pilot study. JMIR Mhealth Uhealth. 2017;5(1):e2. https://www.ncbi.nlm.nih.gov/pubmed/28143805.
18. Carmichael H, Overbey DM, Hosokawa P, Goode CM, Jones TS, Barnett CC Jr, Jones EL, Robinson TN. Wearable technology-a pilot study to define "normal" postoperative recovery trajectories. J Surg Res. 2019;244:368–73. https://www.ncbi.nlm.nih.gov/pubmed/31323392.
19. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, Lee JJ, Hilbert V, Valentino M, Taylor DH, Manders ES, Mutalik K, Zhu J, Wang W, Murabito JM. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med. 2017;177(11):1586–93. https://www.ncbi.nlm.nih.gov/pubmed/28973115.
20. Halse RE, Shoneye CL, Pollard CM, Jancey J, Scott JA, Pratt IS, Dhaliwal SS, Norman R, Straker LM, Boushey CJ, Delp EJ, Zhu F, Harray AJ, Szybiak MA, Finch A, McVeigh JA, Mullan B, Collins CE, Mukhtar SA, Edwards KN, Healy JD, Kerr DA. Improving nutrition and activity behaviors using digital technology and tailored feedback: protocol for the LiveLighter Tailored Diet and Activity (ToDAy) randomized controlled trial. JMIR Res Protoc. 2019;8(2):e12782. https://www.ncbi.nlm.nih.gov/pubmed/30801257.
21. Maxwell-Smith C, Cohen PA, Platell C, Tan P, Levitt M, Salama P, Makin GB, Tan J, Salfinger S, Kader Ali Mohan GR, Kane RT, Hince D, Jimenez-Castuera R, Hardcastle SJ. Wearable Activity Technology And Action-Planning (WATA AP) to promote physical activity in cancer survivors: randomised controlled trial protocol. Int J Clin Health Psychol. 2018;18(2):124–32. https://www.ncbi.nlm.nih.gov/pubmed/30487917.
22. Maxwell-Smith C, Hince D, Cohen PA, Bulsara MK, Boyle T, Platell C, Tan P, Levitt M, Salama P, Tan J, Salfinger S, Makin G, Mohan G, Jimenez-Castuera R, Hardcastle SJ. A randomized controlled trial of WATAAP to promote physical activity in colorectal and endometrial cancer survivors. Psychooncology. 2019;28(7):1420–9. https://www.ncbi.nlm.nih.gov/pubmed/30980691.
23. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med. 2018;54(1):144–50. https://www.ncbi.nlm.nih.gov/pubmed/29122356.
24. Creswell JW. Research design. qualitative, quantitative, & mixed method approaches. 4th ed. Los Angeles, London, New Delhi, Singapore, Washington DC: SAGE publications; 2014.
25. Jacobsen BK, Eggen AE, Mathiesen EB, Wilsgaard T, Njolstad I. Cohort profile: the Tromso study. Int J Epidemiol. 2012;41(4):961–7. https://www.ncbi.nlm.nih.gov/pubmed/21422063.
26. Nordic Council of Ministers. Nordic nutrition recommendations 2012: integrating nutrition and physical activity. 2014.
27. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass. 2016;10(9):503–18. https://doi.org/10.1111/spc3.12265.
28. Deraas TS, Hopstock L, Henriksen A, Morseth B, Sand A-S, Njølstad I, Pedersen S, Sagelv E, Johansson J, Grimsgaard S. Complex lifestyle intervention among inactive older adults with elevated cardiovascular disease risk and obesity. 2020. PREPRINT (Version 1) available at Research Square, https://doi.org/10.21203/rs.3.rs-39292/v1.
29. Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L. Validity of the polar M430 activity monitor in free-living conditions: validation study. JMIR Form Res. 2019;3(3):e14438. https://www.ncbi.nlm.nih.gov/pubmed/31420958.
30. Sasaki JE, John D, Freedson PS. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport. 2011;14(5):411–6. https://www.ncbi.nlm.nih.gov/pubmed/21616714.
31. Lee J, Finkelstein J. Consumer sleep tracking devices: a critical review. Stud Health Technol Inform. 2015;210:458–60. https://www.ncbi.nlm.nih.gov/pubmed/25991187.
32. McMinn D, Acharya R, Rowe DA, Gray SR, Allan JL. Measuring activity energy expenditure: accuracy of the GT3X+ and actiheart monitors. Int J Exerc Sci. 2013;6(3):5.
33. Schofield WN. Predicting basal metabolic rate, new standards and review of previous work. Hum Nutr Clin Nutr. 1985;39(Suppl 1):5–41. https://www.ncbi.nlm.nih.gov/pubmed/4044297.
34. Polar Electro. Polar flow. 2018. https://flow.polar.com.
35. Pugliese L, Woodriff M, Crowley O, Lam V, Sohn J, Bradley S. Feasibility of the "bring your own device" model in clinical research: results from a randomized controlled pilot study of a mobile patient engagement tool. Cureus. 2016;8(3):e535. https://www.ncbi.nlm.nih.gov/pubmed/27096135.
36. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being. 2014;9:26152. https://www.ncbi.nlm.nih.gov/pubmed/25326092.
37. Kvale S, Brinkmann S. Interviews. Learning the craft of qualitative research interviewing. 2nd ed. Los Angeles, London, New Delhi, Singapore: SAGE publications; 2009.
38. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008;40(1):181–8. https://www.ncbi.nlm.nih.gov/pubmed/18091006.
39. Bakdash JZ, Marusich LR. Repeated measures correlation. Front Psychol. 2017;8:456. https://www.ncbi.nlm.nih.gov/pubmed/28439244.
40. Evans JD. Straightforward Statistics for the Behavioral Sciences. Pacific Grove: Brooks/Cole Publishing; 1996.
41. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60. https://www.ncbi.nlm.nih.gov/pubmed/10501650.
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.
43. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods. 2016;5(1):80–92. https://journals.sagepub.com/doi/pdf/10.1177/160940690600500107.
44. Berenguer A, Goncalves J, Hosio S, Ferreira D, Anagnostopoulos T, Kostakos V. Are smartphones ubiquitous?: an in-depth survey of smartphone adoption by seniors. IEEE Consum Electron Mag. 2017;6(1):104–10. https://ieeexplore.ieee.org/document/7786986/.
45. Duignan C, Slevin P, Sett N, Caulfield B. Consumer wearable deployments in actigraphy research: evaluation of an observational study. JMIR Mhealth Uhealth. 2019;7(6):e12190. https://www.ncbi.nlm.nih.gov/pubmed/31237237.
46. Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth. 2017;5(10):e164. https://www.ncbi.nlm.nih.gov/pubmed/29084709.
47. Henriksen A, Johansson J, Hartvigsen G, Grimsgaard S, Hopstock L. Measuring physical activity using triaxial wrist worn polar activity trackers: a systematic review. Int J Exerc Sci. 2020;13(4):438–54. https://www.ncbi.nlm.nih.gov/pubmed/32509122.
48. Poehlman ET. A review: exercise and its influence on resting energy metabolism in man. Med Sci Sports Exerc. 1989;21(5):515–25. https://www.ncbi.nlm.nih.gov/pubmed/2691813.
49. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Hum Behav. 2020;102:223–37. https://www.sciencedirect.com/science/article/pii/S0747563219303127.
50. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. 2020. https://pewrsr.ch/37laaN4.
51. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas. 2007;28(3):R1–39. https://www.ncbi.nlm.nih.gov/pubmed/17322588.
52. Hardcastle SJ, Galliott M, Lynch BM, Nguyen NH, Cohen PA, Mohan GR, Johansen NJ, Saunders C. Acceptability and utility of, and preference for wearable activity trackers amongst non-metropolitan cancer survivors. PLoS One. 2018;13(12):e0210039. https://www.ncbi.nlm.nih.gov/pubmed/30596781.
53. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth. 2016;4(1):e7. https://www.ncbi.nlm.nih.gov/pubmed/26818775.
54. McMahon SK, Lewis B, Oakes M, Guan W, Wyman JF, Rothman AJ. Older adults' experiences using a commercially available monitor to self-track their physical activity. JMIR Mhealth Uhealth. 2016;4(2):e35. https://www.ncbi.nlm.nih.gov/pubmed/27076486.
55. Smith A. Older Adults and Technology Use. Pew Research Center. 2014. Available at: http://www.pewinternet.org/2014/04/03/older-adults-and-technology-use/.

56. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth. 2017;5(11):e173. https://www.ncbi.nlm.nih.gov/pubmed/29141837.
57. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004;24(2):105–12. https://www.ncbi.nlm.nih.gov/pubmed/14769454.
58. McMurdo ME, Roberts H, Parker S, Wyatt N, May H, Goodman C, Jackson S, Gladman J, O'Mahony S, Ali K, Dickinson E, Edison P, Dyer C, Age, Ageing Specialty Group NCCRN. Improving recruitment of older people to research through good practice. Age Ageing. 2011;40(6):659–65. https://www.ncbi.nlm.nih.gov/pubmed/21911335.
59. Tromsøundersøkelsen. Tromsøundersøkelsen. 2020. https://uit.no/go/target/367276.

Publisher's Note

Paper IV

Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock L.

Physical activity surveillance during the COVID-19 pandemic: Using consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies.

[Submitted]. doi: 10.2196/preprints.23806.

Physical activity surveillance during the COVID-19 pandemic: Using consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies

André Henriksen[1], Erlend Johannessen[2], Gunnar Hartvigsen[2], Sameline Grimsgaard[1], Laila Hopstock[1]

[1]Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway
[2]Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author
André Henriksen
Department of Community Medicine
UiT The Arctic University of Norway, Tromsø, Norway
Hansine Hansens veg 18
Email: andre.henriksen@uit.no

Abstract

Background: Consumer-based physical activity trackers increase in popularity. The widespread use of these devices and the long-term nature of the recorded data provides a valuable source of physical activity data for epidemiological research. Major challenges include the large number of activity tracker providers and models, and the difference in how and what data are recorded and shared.

Objective: The aim of this study was to develop a system to record data on physical activity from different providers of consumer-based activity trackers, and to examine its usability as a tool for physical activity monitoring in epidemiological research. The longitudinal nature of the data and the concurrent pandemic outbreak allowed us to show how the system can be used for surveillance of physical activity levels before, during, and after a COVID-19 lockdown.

Methods: We developed a system (mSpider) for automatic recording of data on physical activity from participants wearing activity trackers from Apple, Fitbit, Garmin, Oura, Polar, Samsung, and Withings, as well as trackers storing data in Google Fit and Apple Health. To test the system throughout development, we recruited 35 volunteers to wear a provided activity tracker from primo 2019 and onwards. In addition, we recruited 113 participants with privately owned activity trackers worn before, during, and after the COVID-19 lockdown in Norway. We examined monthly change in number of steps, minutes of moderate-to-vigorous physical activity, and activity energy expenditure during 2019-2020 using bar plots and two-sided paired sample t-tests and Wilcoxon signed-rank test.

Results: Compared to March 2019, there was a significant reduction in mean step count and mean activity energy expenditure during the March 2020 lockdown period. The reduction was temporary, and the year to year comparison show a small increase in moderate-to-vigorous physical activity and no change in steps and activity energy expenditure.

Conclusions: mSpider is a working prototype currently able to record physical activity data from providers of consumer-based activity trackers. The system was successfully used to examine change in physical activity levels during the COVID-19 period.

Keywords: Energy expenditure; physical activity; steps, smart watch; fitness tracker; actigraphy; public health; lockdown; Sars-Cov-2; pandemic; COVID-19.

Introduction

Physical activity is an important lifestyle factor [1] associated with a range of health outcomes [2]. Physical activity questionnaires and accelerometers are widely used to measure physical activity in epidemiological studies. The widespread use of advanced consumer-based activity trackers with a

growing list of sensors and capabilities [3] has increased the use of activity trackers for research purposes [4]. New activity trackers are continuously released, and although the validity of most currently used activity trackers is unknow, a recent systematic review showed that inter-device reliability is often very strong [5].

This unique source of longitudinal physical activity recordings can be used to measure change in physical activity over time. It is therefore of interest to develop a system for automatic and continuous recording of physical activity data from available providers. This system can be used in a range of different research projects, including as a tool for physical activity surveillance.

The disease outbreak of COVID-19 (SARS-CoV-2) started in China December 2019, spread rapidly, and became a global pandemic. The first case of COVID-19 in Norway was confirmed February 26th, 2020. March 12th the Norwegian government implemented a lockdown of all schools, kindergartens, universities, high schools, gyms, etc., with additional restrictions in the following days. Although a national curfew was not instigated, people were encouraged to stay at home if possible. The most restrictive measures were gradually lifted from the end of April throughout May 2020. Less intrusive social distancing restrictions were gradually re-introduced throughout the Autumn, but no second lockdown was instigated in 2020.

In addition to the societal cost of the COVID-19 pandemic [6], physical inactivity during lockdown, and failing to revert to normal physical activity routines after the lockdown may cause health harm [7].

The aim of this study was to develop a system for automatic continuous recording of physical activity data from a range of providers of consumer-based activity trackers, and to examine its usability as a tool for physical activity monitoring in epidemiological research. The longitudinal nature of the data, and concurrent pandemic allowed us to examine how this system could be used to monitor change in physical activity before, during, and after the COVID-19 lockdown.

Materials and methods

System architecture

We designed and developed an experimental system, mSpider, intended for automatic and continuously recording of physical activity data using consumer-based activity trackers. The system collects data on physical activity, energy expenditure, pulse, sleep, and related variables over an extended period, and from a range of providers and activity tracker models.

The system consists of three modules (see Figure 1): 1) the *web frontend,* 2) *the server backend*, and 3) the *mobile application*. The *web frontend* is used for managing surveys and to facilitate participant authorization when granting access to their activity tracker data. The *server backend* stores participant authorization access information, handles data transfer between mSpider and the cloud storages of supported providers, and stores downloaded activity tracker data. The *mobile application* further facilitates authorization and data transfer for providers where communication cannot be performed directly between the server backend and the provider cloud storage (e.g. Samsung and Apple activity trackers). For these providers, communication is performed through the provider mobile application and uploaded to the mSpider server backend via the mSpider mobile application.

Figure 1 gives an architectural overview of the mSpider system, which providers are supported, and communications paths between systems. Red dashed lines indicate communication paths for participant authorization. To share data, users of Samsung and Apple activity trackers must install the mSpider mobile application and initiate authorization through this application via the provider mobile application. All other supported providers initiate authorization via the web frontend, using open authorization, and participants are not required to install the mSpider application. Black solid lines between the *server backend* and external systems show providers where the server backend initiates a pull request to fetch data directly from the provider cloud storage, after access is granted

by the participant. Grey dashed lines show providers where data transfer is initiated at the provider side (e.g. Garmin) using a push request to provider-specific interfaces on the server backend. Data collected by the mSpider mobile app are also pushed to the mSpider server backend.

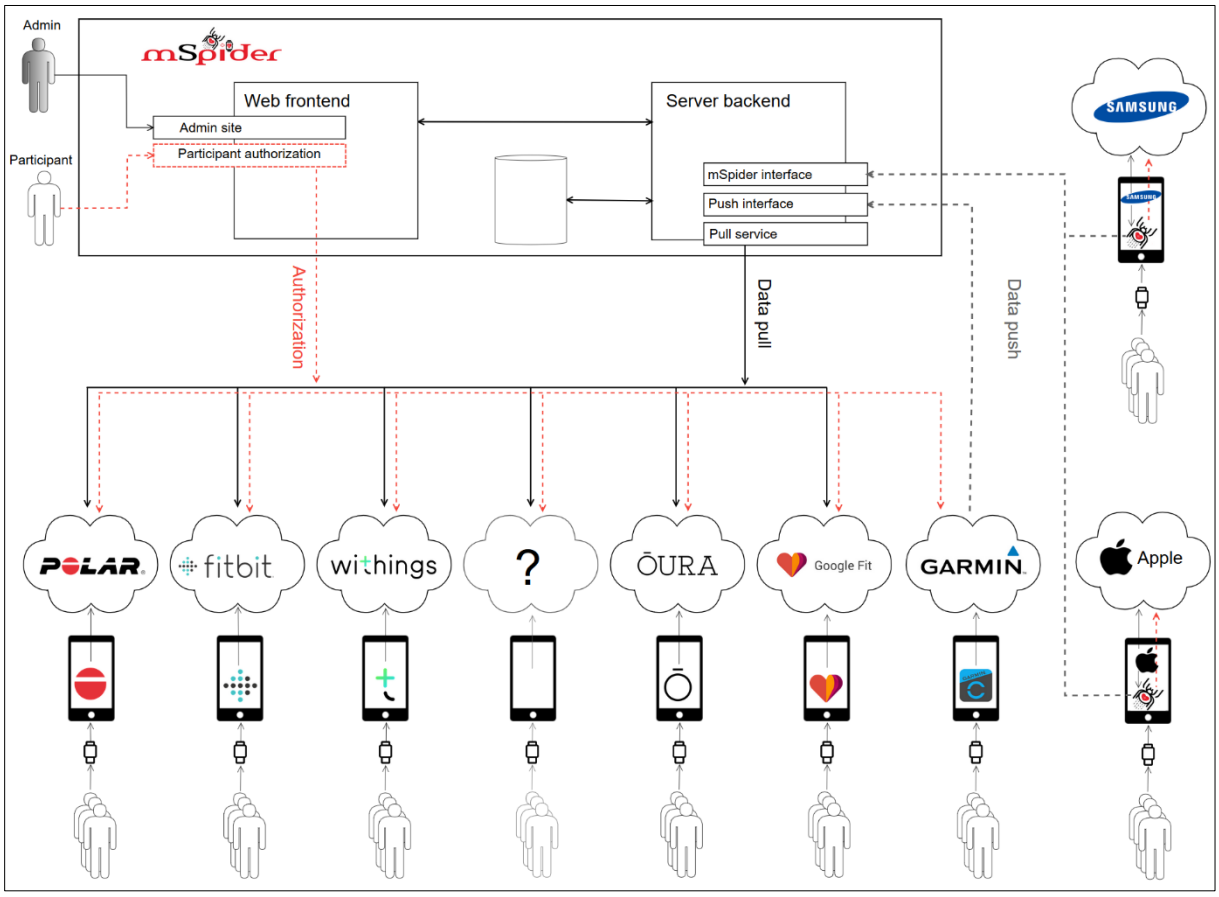


Figure 1. mSpider system architectural overview.

Authorization

Participants authorizes the mSpider system and grants access to their activity tracker data using open authorization (OAuth). Open Authorization is an open protocol for allowing users to securely authorize data sharing between systems, without sharing user logon credentials [8].

Pull requests from the mSpider system to external APIs (e.g. Fitbit Web API) contains a *Client Identifier* and *Client Secret*, identifying mSpider as an authorized application for data retrieval. These credentials are given by the external system (i.e. providers) upon successfully registration of the mSpider app with each provider.

In addition, a *Token Identifier* and *Token Secret* are provided by the external system when an activity tracker user registers to participate in a study. Tokens are used to identify participants in future pull requests to the provider cloud storage (or push request from the provider). No directly identifiable information is transferred between the provider systems and the mSpider system. All communication is encrypted through the secure socket layer (SSL) protocol (i.e. https).

Provider support and available data types

We developed support for activity trackers from Fitbit, Polar, Garmin, Withings, Samsung, Oura, and Apple, as well as providers that store data in Google Fit or Apple Health open health clouds (e.g. Huawei). Except Samsung and Apple, supported providers offers a REST (representational state transfer) API (application programming interface) web service. The REST software architectural style provides a set of constrains for distributed systems [9] and is a style commonly used when developing web services. A RESTful API, i.e. an API using HTTP (Hypertext Transfer Protocol) requests

(e.g. GET, POST), uses a stateless architecture where the necessary information, including participant identification (i.e. tokens), is transferred with the request. To access data from providers not supporting a REST API, the mSpider mobile application was developed using provider specific software development kits (SDK), which gives access to activity tracker data via the provider specific mobile application. Table 1 gives an overview of providers and which API/SDK we used to access data.

Table 1. Provider data access details

Provider documentation	API/SDK	Version
Apple [10]	HealthKit	6.4
Fitbit [11]	Web API	1/1.2
Garmin (Must register to gain access)	Health API	2.9.7
Google [12]	Fit API	1
Oura [13]	Cloud API	1
Polar [14]	AccessLink API	3.36.0
Samsung [15]	Health SDK	1.4.0
Withings [16]	Data API	2.0

API: Application programming interface. SDK: Software development kit.

Each provider offers a different set of data types through their API/SDK. Steps is the only variable supported by all providers. Table 2 gives a list of available variables relevant for the present study for each provider, and how we used these variables to define valid days, i.e. days where activity tracker wear time was sufficient to be included in daily physical activity analysis. A complete list of available variables can be found in the provider documentations (Table 1).

Table 2. Available variables by provider

Provider	Variables	Valid day calculation
Apple	Steps, AEE, REE, sleep	Step>150
Fitbit	Steps, TEE, AEE, LPA, MPA, VPA, sleep	Step>150
Garmin	Steps, TEE, AEE, MPA, VPA	(Sleep + sedentary time + LPA + MPA + VPA) >10 hours
Google Fit	Steps, TEE	Step>150
Oura	Steps, TEE, AEE, sedentary time, LPA, MPA, VPA, non-wear time	Step>150
Polar	Steps, TEE, AEE, sedentary time, LPA, MPA, VPA, sleep	non-wear time<14 hours
Samsung	Steps, AEE, sleep	(Sleep + sedentary time + LPA + MPA + VPA) >10 hours
Withings	Steps, TEE, AEE, LPA, MPA, VPA, sleep	Step>150

TEE: Total energy expenditure, AEE: activity energy expenditure, REE: resting energy expenditure, LPA: light physical activity, MPA: moderate physical activity, VPA: vigorous physical activity.

Recruitment of volunteer and study participants

Volunteers (development phase)

To test the system during development and increase the likelihood of long-term recording, we used convenience sampling to recruit 35 volunteers with the following inclusion criteria; 1) 18 years or older, 2) willing to wear a provided activity tracker for an extended period, and 3) willing to share collected physical activity data. Data from these volunteers were used for system development purposes and were not included in the longitudinal analysis of physical activity.

Volunteers were recruited during the development phase (2019-2020) and equipped with an activity tracker from Apple, Fitbit, Garmin, Huawei, Oura, Polar, Samsung, or Withings. Two participants also

shared mobile phone collected physical activity data stored in Google Fit. One participant withdrew after a few days and two participants withdrew after a few months. We gave no instructions on activity tracker usage, except giving instructions on how to initiate automatic data sharing with the mSpider system. Participants were given written and oral information about the mSpider system, and informed that all collected data would be stored at the activity tracker provider cloud storage. All volunteers signed informed consent.

Study participants (physical activity study)
Through online news media advertisement, we recruited 130 people with privately owned activity trackers, worn before, during, and after the Norwegian COVID-19 lockdown. Inclusion criteria were: 1) owned an activity tracker from Garmin, Fitbit, Withings, or Oura, and 2) willing to share physical activity data. Participants received an e-mail invitation with a letter of information and instructions on how to grant access to the mSpider system. Participants gave informed consent by actively granting access to their data.

Privacy

The 35 volunteers who received an activity tracker were required to register a user account at the activity tracker provider. Although the mSpider system only accessed non-identifiable information, volunteers were informed that by registration of a provider account, all data collected by the activity tracker would be upload to the provider cloud storage, including potential identifiable information (e.g. GPS-data).

The 130 study participants for analysis of activity tracker data already owned an activity tracker and thus already had a provider user account. After downloading the relevant data, we removed user tokens from the mSpider database and thus stored data anonymously.

Data collection

Daily estimates for steps, activity energy expenditure, moderate physical activity, and vigorous physical activity were downloaded from study participants to be included in the physical activity analysis. A variable for moderate-to-vigorous physical activity (MVPA) was created by combining moderate physical activity and vigorous physical activity, for participants where these variables were available. We further downloaded light physical activity, sedentary time, sleep duration, and non-wear time, to be used for activity tracker wear-time estimates. Data download was limited to days between January 1st, 2019 and December 31st, 2020.

Only days where the activity tracker was worn for at least 10 hours were labelled as *valid days* [17]. As this was not possible for all providers (Table 2), days with less than 150 recorded steps were excluded. After data download was completed, we removed the connection between the user's provider and the mSpider tool by deleting user tokens. All data on physical activity was thus stored anonymously. An anonymous online questionnaire was used to collect self-reported data on sex, age, height, and weight.

Statistical analysis

Participant characteristics from the online questionnaire are presented as means, standard deviations, and range. For each participant we used valid days to create monthly and yearly averages for steps per day (steps/day), activity energy expenditure in kilocalories per day (kcal/day), and MVPA in minutes per day (minutes/day) for 2019 and 2020. March 2020 was divided into two periods (up to and after March 12th, i.e. the lockdown date). For each variable we compared the following:

1) 2019 (March-December) with 2020 (March-December)
2) March 2019 with March 1st – 12th 2020
3) March 2019 with March 13th – 31st 2020
4) April 2019 with April 2020, May 2019 with May 2020, etc.

5) March 2020, 1st – 12th with 13th – 31st

We created bar plots to visualize difference between time periods. Normality was checked using histograms. We used two-sided paired sample t-test or two-sided paired Wilcoxon signed-rank test, depending on normality, to test differences in physical activity between time periods. Two-sided p-values <.05 were considered statistically significant. Statistical analyses were performed using R version 4.0.3.

Results

Participant characteristics

Of the 130 recruited study participants, 14 did not respond to the following invitation e-mail and three owned an un-supported activity tracker. A final sample of 113 participants were thus included in the analysis. Of the included participants, 106 completed the online questionnaire and provided their characteristics (Table 3).

Table 3. Participant characteristics (N=106).

Variable	Value	Range
Height, cm	173.5 (8.0)	158 – 194
Weight, kg	76.0 (14.3)	53.5 – 147.0
Body mass index, kg/m^2	25.2 (4.0)	18.3 – 50.3
Age, years	40.6 (10.6)	21 – 69
Females, %	56.2 (59)	NA

Values are means (standard deviations) or percentages (numbers). NA; not applicable.

Altogether 39 participants used Fitbit activity trackers and 74 participants used Garmin activity trackers. No participants owned a Withings-, or Oura activity tracker. Both Fitbit and Garmin provide data on steps, MVPA, and activity energy expenditure. All 113 participants were thus included when generating monthly means for all three variables. Monthly means are calculated from 66.274 measurements (i.e. valid person-days).

Change in physical activity

On average participants walked 797 fewer steps per day in March 13th – 31st, 2020 compared to March 2019 (p=.021). Similarly, participants walked on average 913 fewer steps per day in March 13th – 31st, 2020 (post lockdown) compared to March 1st – 12th, 2020 (pre lockdown) (p<.001). Remaining step comparisons showed no differences.

Mean activity energy expenditure was 74 kcal/day lower in March 13th – 31st, 2020 compared to March 2019 (p=.021). In addition, mean activity energy expenditure was 85 kcal/day lower in March 13th – 31st, 2020 (post lockdown) compared to March 1st – 12th, 2020 (pre lockdown) (p=.001). However, activity energy expenditure was on average 54 kcal/day higher in September 2020 compared to September 2019 (p=.021). Remaining activity energy expenditure comparisons showed no difference.

For MPVA, monthly comparisons showed a significant increase from 2019 to 2020 for May (p=.013) with median difference of eight minutes, September (p=.008) with median difference of three minutes, October (p=.022) with median difference of five minutes, and December (p=.043) with median difference of four minutes, as well as the yearly comparison (p=.026) with median difference of four minutes. Remaining MVPA comparisons showed no difference.

A summary of mean difference per day between periods for steps and activity energy expenditure, with 95% confidence intervals and p-values from each t-test is given in Table 4. The table also gives the median of the difference per day between periods for MVPA, with interquartile ranges and p-values from each Wilcoxon test. Because we used paired tests, analysis only include participants with data in both the pre-period and the post-period, thus is based on data from 76-107 participants.

Figure 2 and Figure 3 gives monthly mean step count and activity energy expenditure from March 2019 thru December 2020. Figure 4 gives median MVPA for the same periods.

Table 4. Difference per day between pre-periods and post-periods.

Monthly comparison 2019-2020	Steps (steps/day)	P-value	AEE (kcal/day)	P-value	MVPA (min/day)	P-value
March-December	349 (-4, 702)	.053	29 (-2, 60)	.066	4 (-6, 4)	.026
March 1st – 12th*	28 (-608, 664)	.930	21 (-40, 82)	.493	-2 (-14, -2)	.566
March 13th – 31st**	-797 (-1468, -126)	.021	-74 (-136, -11)	.021	2 (-11, 2)	.831
April	-123 (-850, 605)	.738	-35 (-105, 34)	.319	-1 (-15, -1)	.810
May	53 (-586, 692)	.869	2 (-59, 64)	.936	8 (-6, 8)	.013
June	301 (-276, 878)	.303	45 (-7, 97)	.092	4 (-10, 4)	.068
July	442 (-232, 1117)	.196	44 (-15, 104)	.141	1 (-14, 1)	.525
August	326 (-271, 922)	.280	24 (-24, 72)	.325	2 (-14, 2)	.529
September	324 (-148, 797)	.176	54 (8, 100)	.021	3 (-7, 3)	.008
October	361 (-290, 1011)	.274	41 (-7, 89)	.096	5 (-6, 5)	.022
November	242 (-442, 927)	.484	42 (-22, 106)	.199	4 (-11, 4)	.338
December	491 (-6, 988)	.053	32 (-21, 84)	.235	4 (-8, 4)	.043

Numbers are mean difference with 95% confidence interval (steps and AEE), or median of the difference with interquartile range (MVPA), followed by p-values from paired sample t-test or paired Wilcoxon signed-rank test.
*Comparing March 2019 with March 1st – 12th 2020.
**Comparing March 2019 with March 13th – 31st.
AEE: activity energy expenditure. MVPA: moderate-to-vigorous physical activity. kcal: kilocalories. min: minutes.

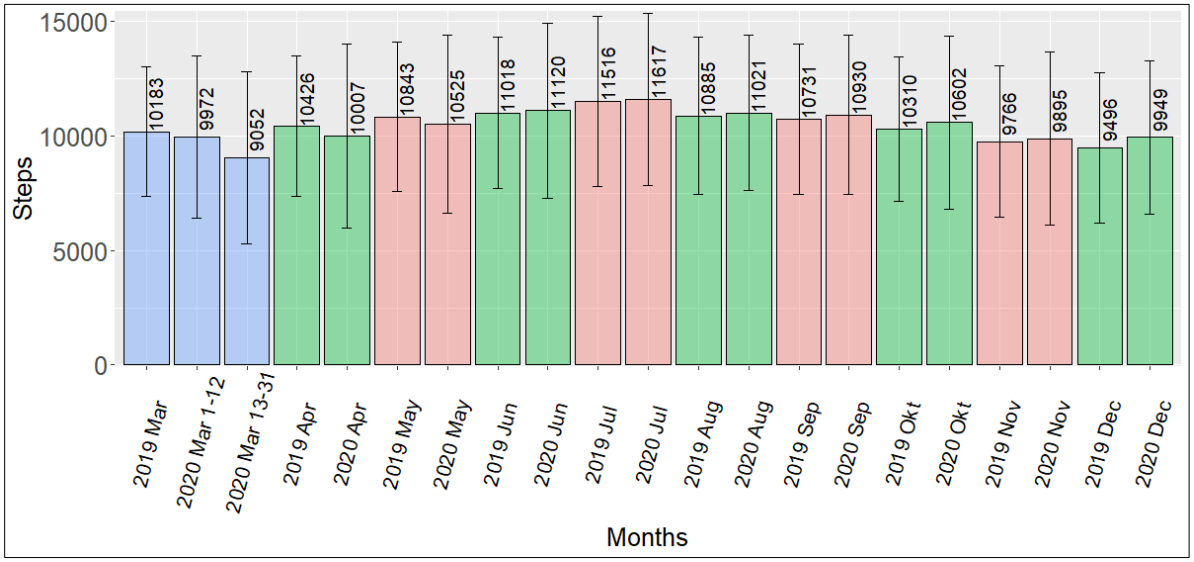


Figure 2. Bar plot of mean step count per day, by month, with standard deviation (error bars).

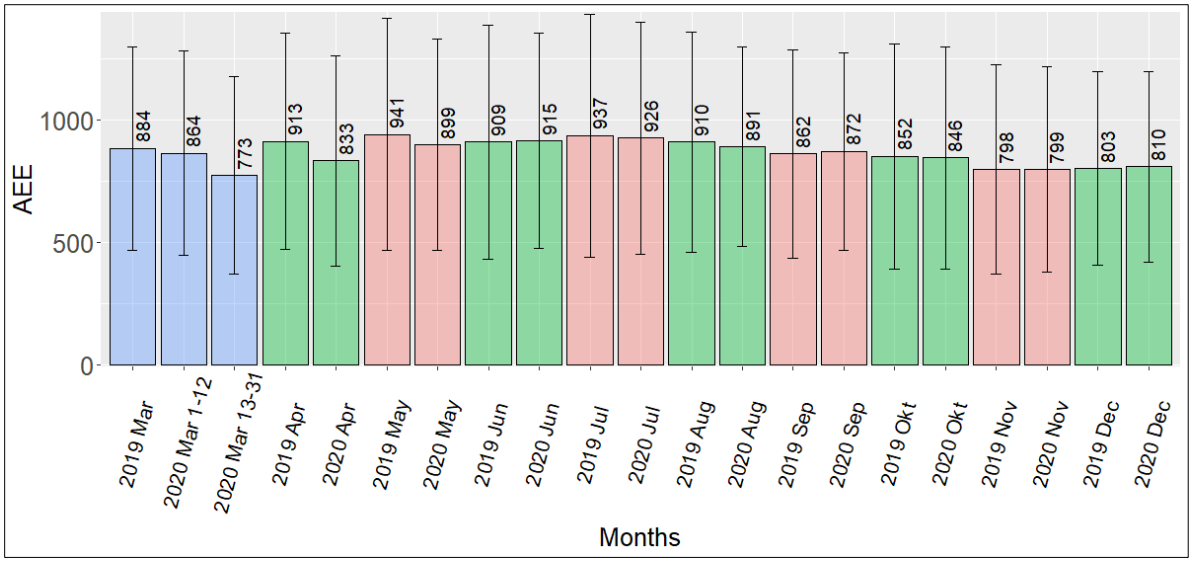


Figure 3. Bar plot of mean activity energy expenditure per day, by month, with standard deviation (error bars).

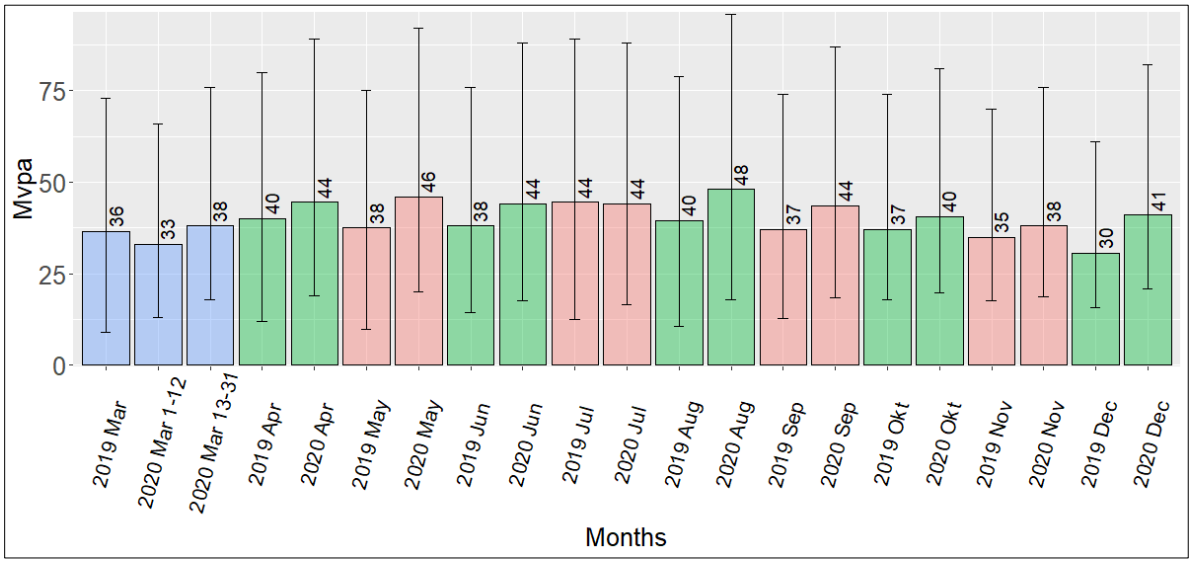


Figure 4. Bar plot of median minutes of moderate-to-vigorous physical activity (MVPA) per day, by month, with inter-quartile range (error bar).

Discussion

Principal findings

In this study, the mSpider-system was successfully used to download historic data on steps, activity energy expenditure, and MVPA from Garmin and Fitbit activity tracker users. The longitudinal data showed changes in physical activity during the COVID-19 pandemic.

Findings indicate a short-term reduction in steps and activity energy expenditure due to the COVID-19 lockdown, but no reduction in MVPA. However, participants increased their level of MVPA the month after the lockdown period (i.e. May) and some months in the autumn of 2020 (i.e. September, October, and December), compared to 2019.

Comparison with previous work

Results in the present study are supported by reports from providers of consumer-based activity trackers. Garmin have released a statement showing that users globally had a distinct decline in step count during the last two weeks of March 2020, and that the reduction in step counts was

compensated by increase in other activities [18]. Withings have reported a temporary decline in step counts among users during national lockdowns [19]. Similarly, a study of UK adults using physical activity data recorded by a smartphone application, showed a significant decrease in physical activity during the March 2020 UK national lockdown [20].

Google trend analysis of community interest in physical activity during the COVID-19 outbreak and lockdown, showed an increase in Google search rates on physical activity topics in Australia, the UK, and the US [21]. A study among German athletes, using activity tracker data, showed that shorter and more vigorous exercise sessions replaced longer sessions [22].

These studies support our finding that although restrictions confined people to their home, they found alternative ways to keep their habitual physical activity level. Conversely, based on online physical activity questionnaires, a study from Thailand did not show any increase in physical activity after the lockdown was lifted [23] and a study from Bangladesh showed high prevalence of inactivity during lockdown [24].

In summary, activity tracker data from several vendors and groups of users including athletes and chronic disease patients, have shown changes in physical activity levels and patterns during the COVID-19 pandemic, but findings vary between countries.

mSpider as a method for collection data on physical activity

The analysis of physical activity changes related to the COVID-19 pandemic period showed that the mSpider system can be a valuable tool for collection of long-term data on physical activity, including historical data, as well as detect changes in physical activity over time.

In the present study, we used the proposed system to access data retrospectively from participants with privately owned activity trackers. Previously, we have successfully used the same technology for long-term prospective physical activity monitoring, among participants in a lifestyle intervention study wearing a provided activity tracker for up to one year [25, 26, 27].

A system similar to mSpider, RADAR-base (Remote Assessment of Disease And Relapses), was used by Sun et al. [28], who observed change in daily steps during national lockdowns, among participants with chronic disease equipped with a Fitbit tracker. RADAR-base is an open source platform for collecting physical activity data from smartphones, Fitbit- and Garmin activity trackers, and some research grade accelerometers [29]. RADAR-base uses similar technology as mSpider, but data collection is limited to only two providers of consumer-based activity trackers.

A study by Radin et al. [30] successfully mapped historic Fitbit data (provided manually by Fitbit) to known influenza outbreaks. This also shows the potential for the proposed system as a tool for disease outbreak surveillance, where clusters of participants with a combination of physical activity reduction and elevated resting heart rate can be used to indicate disease outbreaks in an area. The quality of accelerometer-based physical activity data is dependent on participant wear compliance. Since younger adults tends to be less compliant when wearing accelerometers in research [31], but more likely to own and wear an activity tracker [32], the proposed system also has potential to add to and enrich current methods for physical activity data collected used in epidemiological research.

Summarized, we find the mSpider system to be an interesting supplement to present tools for physical activity monitoring in epidemiological studies. However, major challenges must be kept in mind. First, self-selected users of activity trackers are often more physically active compared to non-users [32, 33]. Second, the accuracy of different activity trackers is highly variable [5, 34, 35]. At the population level, the system may perform better to detect change in physical activity over time, than to estimate the absolute levels of physical activity.

Strength and limitations

The major strength of the present study is the long-term recording, with up to two years of daily

physical activity data per participant. This allowed for month-to-month comparison between 2019 and 2020, thus taking potential seasonal differences in physical activity levels into account.

The study has limitations that can affect the study results. Firstly, the participants were self-selected owners of physical activity trackers, who are likely to be more physically active than the general population. A recent study by Anyan et al [36] investigating physical activity change during the Norwegian lockdown (using questionnaire data), found that 14% of participants reported reduction, 22% reported increase, and 64% reported no change in physical activity level. Therefore, there is a risk of selection bias in this study, i.e. the sample may not be representative of the general population. Nevertheless, the observed changes in physical activity levels in this sample during the study period demonstrates the usefulness of the mSpider system. Further, due to anonymous data collection, we could not link participant characteristics to physical activity data to examine physical activity in strata of sex, age or other characteristics.

Conclusion

mSpider is a working prototype currently able to record physical activity data from several providers of consumer-based activity trackers. The system was successfully used to detect longitudinal changes in physical activity levels before, during, and after the Norwegian COVID-19 lockdown period in 2020. To our knowledge, this is the first study reporting change in physical activity caused by the COVID-19 lockdown in Norway, using two years of objective consumer-based activity tracker data.

Ethical approval

The Regional Committees for Medical and Health Research Ethics North (reference 164780) and the Norwegian Center for Research Data (reference 628485) reviewed the study.

Acknowledgements

The authors would like to thank Martin Mikalsen for his contributions in the early phase of development of the MSpider system. The work was funded by UiT The Arctic University of Norway thematic priority grant “Personalized medicine for public health”. The publication charges for this study were granted from the publication fund at UiT The Arctic University of Norway.

Conflicts of interest

None declared

Abbreviations

API: Application programming interface
HTTP: Hypertext transfer protocol
Kcal: Kilocalories
MVPA: Moderate-to-vigorous physical activity
OAuth: Open authorization
SDK: Software development kit

References

1. Lacombe J, Armstrong MEG, Wright FL, Foster C, *The impact of physical activity and an additional behavioural risk factor on cardiovascular disease, cancer and all-cause mortality: a systematic review.* BMC Public Health, 2019. 19(1): p. 900. doi: 10.1186/s12889-019-7030-8.
2. Haskell WL, Blair SN, Hill JO, *Physical activity: health outcomes and importance for public health policy.* Prev Med, 2009. 49(4): p. 280-2. doi: 10.1016/j.ypmed.2009.05.002.
3. Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S, *Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables.* J Med Internet Res, 2018. 20(3): p. e110. doi: 10.2196/jmir.9157.
4. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM, *Wearable Technology and Physical Activity in Chronic Disease: Opportunities and Challenges.* Am J Prev Med, 2018.

54(1): p. 144-50. doi: 10.1016/j.amepre.2017.08.015.
5. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Van Heerden D, Luan H, Cullen K, Slade L, Taylor NGA, *Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review.* JMIR Mhealth Uhealth, 2020. 8(9): p. e18694. doi: 10.2196/18694.
6. Mandel A, Veetil V, *The Economic Cost of COVID Lockdowns: An Out-of-Equilibrium Analysis.* Economics of Disasters and Climate Change, 2020. doi: 10.1007/s41885-020-00066-z.
7. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, Olgin JE, *Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study.* Ann Intern Med, 2020. doi: 10.7326/M20-2665.
8. Perecki A. *oAuth.net*. 2021. [Available from: https://oauth.net/. [Accessed: 2021-01-28].
9. Fielding RT. Architectural Styles and the Design of Network-based Software Architectures [Doctoral dissertation]: University of California, Irvine; 2000.
10. Apple Inc. *Apple HealthKit*. 2020. [Available from: https://developer.apple.com/documentation/healthkit. [Accessed: 2020-10-02].
11. Fitbit Inc. *Web API*. 2020. [Available from: https://dev.fitbit.com/build/reference/web-api. [Accessed: 2020-10-02].
12. Google Inc. *Fit API*. 2020. [Available from: https://developers.google.com/fit/rest. [Accessed: 2020-10-02].
13. Oura Oy. *Cloud API*. 2020. [Available from: https://cloud.ouraring.com/docs. [Accessed: 2020-10-02].
14. Polar Electro Oy. *AccessLink*. 2020. [Available from: https://www.polar.com/accesslink-api. [Accessed: 2020-10-02].
15. Samsung Electronics Co. *Health SDK*. 2020. [Available from: https://developer.samsung.com/health/android. [Accessed: 2020-10-02].
16. Withings. *Data API*. 2020. [Available from: https://developer.withings.com/oauth2. [Accessed: 2020-10-02].
17. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M, *Physical activity in the United States measured by accelerometer.* Med Sci Sports Exerc, 2008. 40(1): p. 181-8. doi: 10.1249/mss.0b013e31815a51b3.
18. Garmin LTD. *The Effect of the Global Pandemic on Active Lifestyles*. 2020. [Available from: https://www.garmin.com/en-US/blog/general/the-effect-of-the-global-pandemic-on-active-lifestyles/. [Accessed: 2020-10-02].
19. Pepin JL, Bruno RM, Yang RY, Vercamer V, Jouhaud P, Escourrou P, Boutouyrie P, *Wearable Activity Trackers for Monitoring Adherence to Home Confinement During the COVID-19 Pandemic Worldwide: Data Aggregation and Analysis.* J Med Internet Res, 2020. 22(6): p. e19787. doi: 10.2196/19787.
20. McCarthy H, Potts HWW, Fisher A, *Physical Activity Behaviour Before, During and After COVID-19 Restrictions: A Longitudinal Smartphone Tracking Study of 5395 UK Adults.* Journal of Medical Internet Research, Preprint. doi: 10.2196/preprints.23701.
21. Ding D, Del Pozo Cruz B, Green MA, Bauman AE, *Is the COVID-19 lockdown nudging people to be more active: a big data analysis.* Br J Sports Med, 2020. doi: 10.1136/bjsports-2020-102575.
22. Zinner C, Matzka M, Leppich R, Kounev S, Holmberg H-C, Sperlich B, *The Impact of the German Strategy for Containment of Coronavirus SARS-CoV-2 on Training Characteristics, Physical Activity and Sleep of Highly Trained Kayakers and Canoeists: A Retrospective Observational Study.* Frontiers in Sports and Active Living, 2020. 2. doi: 10.3389/fspor.2020.579830.
23. Katewongsa P, Widyastari DA, Saonuam P, Haemathulin N, Wongsingha N, *The effects of the COVID-19 pandemic on the physical activity of the Thai population: Evidence from Thailand's Surveillance on Physical Activity 2020.* J Sport Health Sci, 2020. doi: 10.1016/j.jshs.2020.10.001.

24. Rahman ME, Islam MS, Bishwas MS, Moonajilin MS, Gozal D, *Physical inactivity and sedentary behaviors in the Bangladeshi population during the COVID-19 pandemic: An online cross-sectional survey.* Heliyon, 2020. 6(10): p. e05392. doi: 10.1016/j.heliyon.2020.e05392.
25. Deraas TS, Hopstock L, Henriksen A, Morseth B, Sand A-S, Njølstad I, Pedersen S, Sagelv E, Johansson J, Grimsgaard S, *Complex lifestyle intervention among inactive older adults with elevated cardiovascular disease risk and obesity.* BMC Pilot & Feasibility studies, Preprint.
26. Henriksen A, Sand AS, Deraas T, Grimsgaard S, Hartvigsen G, Hopstock L, *Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach.* BMC Public Health, 2020. 20(1): p. 1300. doi: 10.1186/s12889-020-09406-w.
27. Hopstock LA, Deraas T, Henriksen A, Martiny-Huenger T, Grimsgaard S, *Results from a 6-month single-arm complex lifestyle intervention with 6-month follow-up: The RESTART pilot study.* PloS one, Submitted.
28. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, Cummins N, Matcham F, Dalla Costa G, Simblett S, Leocani L, Lamers F, Sorensen PS, Buron M, Zabalza A, Guerrero Perez AI, Penninx BW, Siddi S, Haro JM, Myin-Germeys I, Rintala A, Wykes T, Narayan VA, Comi G, Hotopf M, Dobson RJ, Consortium R-C, *Using Smartphones and Wearable Devices to Monitor Behavioral Changes During COVID-19.* J Med Internet Res, 2020. 22(9): p. e19992. doi: 10.2196/19992.
29. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Boettcher S, Hyve, Dobson R, Folarin A, Consortium R-C, *RADAR-Base: Open Source Mobile Health Platform for Collecting, Monitoring, and Analyzing Data Using Sensors, Wearables, and Mobile Devices.* JMIR Mhealth Uhealth, 2019. 7(8): p. e11734. doi: 10.2196/11734.
30. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR, *Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study.* The Lancet Digital Health, 2020. 2(2): p. e85-e93. doi: 10.1016/s2589-7500(19)30222-5.
31. Roth MA, Mindell JS, *Who provides accelerometry data? Correlates of adherence to wearing an accelerometry motion sensor: the 2008 Health Survey for England.* J Phys Act Health, 2013. 10(1): p. 70-8. doi: 10.1123/jpah.10.1.70.
32. Macridis S, Johnston N, Johnson S, Vallance JK, *Consumer physical activity tracking device ownership and use among a population-based sample of adults.* PloS one, 2018. 13(1): p. e0189298. doi: 10.1371/journal.pone.0189298.
33. Strain T, Wijndaele K, Brage S, *Physical Activity Surveillance Through Smartphone Apps and Wearable Trackers: Examining the UK Potential for Nationally Representative Sampling.* JMIR Mhealth Uhealth, 2019. 7(1): p. e11898. doi: 10.2196/11898.
34. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, Finlayson G, Stubbs J, *How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies.* Br J Sports Med, 2020. 54(6): p. 332-40. doi: 10.1136/bjsports-2018-099643.
35. Evenson KR, Spade CL, *Review of Validity and Reliability of Garmin Activity Trackers.* J Meas Phys Behav, 2020. 3(2): p. 170-85. doi: 10.1123/jmpb.2019-0035.
36. Anyan F, Hjemdal O, Ernstsen L, Havnen A, *Change in Physical Activity During the Coronavirus Disease 2019 Lockdown in Norway: The Buffering Effect of Resilience on Mental Health.* Front Psychol, 2020. 11: p. 598481. doi: 10.3389/fpsyg.2020.598481.

Appendix A

Study protocol, Polar M430 validation study (Paper II).

Protocol: Accelerometer validation study

1. Introduction

Accelerometers measuring physical activity

Physical activity can objectively be measured by accelerometers. Accelerometers are small, light, wearable non-invasive devices that measure acceleration (i.e. movement) in one or more axis, giving an indication of frequency, duration, and intensity of physical activity per time unit [1]. Today, a wide range of accelerometer devices exist, aimed at both research [1] use and the consumer marked [2].

A major limitation of using accelerometers in physical activity research is the lack of standardization of raw data to counts, due to patent protection. Another limitation is the uncertainty of various accelerometer devices to detect various types of physical activity. Thus, validation studies between devices are needed.

Traditionally, research-aimed devices, like the ActiGraph, have been used in population-based studies to collect objective data on physical activity. Even though ActiGraphs are extensively used in research, their accuracy is not agreed upon. When compared with indirect calorimetry (Ultima CPX), previous studies using both ActiGraphs and Actiheart found a high correlation for energy expenditure and steps for the ActiGraph, as well as high accuracy for Actiheart energy expenditure estimation [3].

One study found the ActiGraph to be both valid and reliable for measuring PA when compared with VO2 measurements by the Cosmed K4b wearable metabolic system [4]. Another study concluded ActiGraph did not provide valid estimates for energy expenditure when compared to a portable gas analyser (MetaMax 3B) [5]. Yet another study found ActiGraph to be valid for step counting only at certain walking speeds when compared to manual step counting [6].

Similarly, the ActiWave Cardio (CamNTech) and the Actiheart (CamNTech) has been used in research settings to measure both heart rate and physical activity. Previous validation studies of Actiheart show conflicting results. Some studies indicates that there is a low agreement between the Actiheart compared with Doubly Labelled Water (DLW) [7] , while others report good level of agreement [8]. To our knowledge, no validation studies of ActiWave Cardio have been published.

In a systematic review from 2015, Evenson et al [9] concluded that for some consumer-market brands (here Jawbone and Fitbit), there is an indication that validity of steps is high, but validity for energy expenditure is lower. Previous studies using Polar devices in lab settings, indicates that step and energy expenditure validity is low for the Polar Loop, using various comparison means (Bodymedia accelerometer, manual step counting, indirect calorimetry), but energy expenditure is partly acceptable for the Polar V800 [10-13]. In free-living conditions, there is some evidence that Polar devices may be slightly more valid compared to lab settings, using accelerometers as criterion measure (Bodymedia, ActiGraph, Yamax) [14-16].

The Polar Loop and Polar V800 were released in 2013 and 2014, respectively. Even though Polar has released several devices each year since 2014, no studies have reported on these. Specifically, no validation study on the Polar M430 has been conducted to date.

Accelerometers in population-based studies

Results from the 2007-08 survey of the Tromsø Study (Tromsø 6) shows that, when compared with objective measurements collected using an ActiGraph, only 30 percent of participants achieve the recommended level of moderate to vigorous activity (MVPA), even though 85 percent self-reported meeting these recommendations [17]. Thus, there is a need to collect this data objectively when conducting population studies.

In the last Tromsø Study (Tromsø 7), 6300 participants wore an ActiGraph to collect data on physical activity. In addition, 700 participants wore an Actiwave Cardio to collect data on physical activity and heart rhythm. These data can provide valuable insight in the current health status of the Tromsø population. However, only about 35 percent of participants were measured using these devices, and the duration of the measurements was limited to 8 days for the ActiGraph, and 27 hours for the Actiwave Cardio. The participation rate for Tromsø 7 was 65 percent, which is relatively high, compared to similar studies, but the trend for this and other population studies is that participation rates are dropping. There is a need to find new ways to collect objective data on physical activity, over a longer period of time, with a lower burden for both participants and researchers.

One solution for collecting more data, with more participants, and over a longer period, is by accessing data already being collected through people's smart phones, fitness trackers, and smart watches. The technology used in these devices is, in many cases, very similar to that of research grade accelerometers like the ActiGraph. The validity of these devices is not well known, and validation studies are needed.

Raw data versus processed data

Accelerometers typically give information on number of steps, amount of time spent in various physical activity levels, and an estimation of energy expenditure per time unit. The underlying "counts" per minute (the sum of the acceleration that the accelerometer have registered divided on number of minutes that the accelerometer have been in use) is the commonly used variable for physical activity provided in various accelerometer software. The categorization of levels of intensity (sedentary, light, moderate or vigorous) is based on validated cut-points of counts via algorithms in various software.

With consumer devices, access to raw data is very limited, and most devices do not expose this data. Different devices provides different output variables, where steps and energy expenditure per day is common. In addition, they may output the number of minutes per day in various activity intensity zones; minutes of sleeps per day in various sleep zones, and more. Some devices also measures heart rate using photoplethysmograph, which is an optical technique to estimate heart by monitoring changes in blood volume beneath the skin [18]. The exact algorithm for calculating these variables from the raw data are largely company secrets, and not available to researchers.

Coherence between devices for comparison between studies

In the two population-based studies, the Tromsø 7 Study (Tromsø 7) and the International Project on Cardiovascular Disease in Russia Study (IPCDR), as well as in the pilot intervention study Lasting Lifestyle Change (LLC), four different accelerometer devices were used. In Tromsø 7, a hip-worn ActiGraph wGT3X-BT (ActiGraph Corp) and a chest-worn Actiwave Cardio (CamNtech LT) was used. In IPCDR, a chest-worn Actiheart (CamNtech LT) was used. In LLC, a hip-worn ActiGraph and a wrist-worn Polar M430 watch (Polar Electro) was used. Actiwave Cardio, Actiheart, and Polar M430 additionally provide heart rate measurements. There is a need to perform a validation study to

investigate the correlation of different measurement outputs between these four accelerometers for comparison between studies. Table 1 gives an overview of the devices used in these three studies.

Table 1 Accelerometer overview

	Tromsø 7		IPCDR	LLC	
Device	ActiGraph wGT3X-BT	Actiwave Cardio	Actiheart	ActiGraph	Polar M430
Vendor	ActiGraph Corp	CamNtech LT	CamNtech LT	ActiGraph Corp	Polar Electro
Sensor technology	Accelerometer 3-axial	Accelerometer 3-axial, heart rhythm and heart rate (ECG)	Accelerometer mono-axial, heart rate (ECG)	Accelerometer, 3-axial	Accelerometer 3-axial, heart rate (photoplethysmograph)
Placement	Right hip	Chest (V4/V5)	Chest (high (V1/V2) or low (V4/V5))	Right hip	Non-dominant hand
Fixation	Belt	Electrodes[1]	Electrodes[2]	Belt	Watch
Wear time	8 days and nights	27 hours	5 days and nights	8 days and nights	1 year
Software	ActiLife	Actiwave	Actiheart	ActiLife	Polar Flow
Raw data availability	Yes	Yes	Yes	Yes	No

[2]3M red dot 2570 or Kendall H99SG

[1]Unomedical Unilect Long Term 4060M, Bio Protech ECG electrode E5 Tele815, or 3M red dot 2570

2. Purpose

The purpose of this study is to investigate the difference in measurement output between four different accelerometers with different positioning (Polar M430 versus ActiGraph, Actiwave Cardio, and Actiheart), with and without heart rate monitoring, worn by human adults in a free-living setting.

3. Methods

3.1 Devices

The *Polar M430* sport watch (Polar, Finland) was released in 2017 and is one of Polar's newest watch with integrated GPS, 6 LED wrist-based optical (photoplethysmograph) heart rate sensor, and a 3-axis accelerometer for tracking activity and sleep. It weighs 51 gram, and has up to 20 days of battery life.

The *ActiGraph wGT3X-BT* (ActiGraph, Pensacola, FL, USA) is a research grade devices for measuring acceleration in three axis. Sample rate can be set to 30-100 Hz, i.e. number of samples per second. Its dimensions are 46mm x 33mm x 15mm, and it weighs 19 gram. It can be worn on the wrist, waist, ankle, and thigh, for up to 25 days (maximal battery life).

The *CamNtech Actiwave Cardio* (CamNtech Ltd, Cambridge, UK) is a single-led electrocardiogram (ECG), which can measure ECG waveform for up to 31 hours. ECG sample rate can be specified between 32 and 1024Hz. A built in 3-axis accelerometer can record activity in 25, 32, 50, 64, or 100 Hz. The Actiwave Cardio attaches to the chest using standard ECG electrodes; it weighs 10.3 grams, and has a diameter of 32mm.

The *CamNtech Actiheart* CamNtech Ltd, Cambridge, UK) can record heart rate and activity for up to 21 days. ECG sampling rate is 128 Hz. It has a mono-axial accelerometer, which records using a sample rate of 32 Hz. The Actiheart attaches to the chest using standard ECG electrodes, and it weighs less than 10 grams.

3.2 Placements, setup, and wear time

3.2.1 Placement

The Polar M430 is placed on the wrist of the non-dominant hand. One ActiGraph is placed next (above) to the Polar M430 on the non-dominant hand. One ActiGraph is placed on the right hip. The Actiwave Cardio is placed approximately at the level of the forth intercostal space at the sternum (medial part) and to the left (lateral part). One Actiheart is placed just below the Actiwave Cardio approximately at the fifth intercostal space (medial part) and to the left (lateral part). The other Actiheart is placed approximately at the level of the second intercostal space at the sternum (medial part) and to the left (lateral part). Figure 1 indicates device placement.

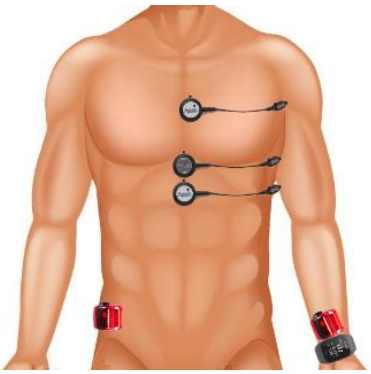


Figure 1 Device placement

3.2.2 Setup and export

Participant height, weight, and age, is specified on each device during setup. When exporting data, epoch length for the ActiGraph, Actiwave Cardio, and Actiheart, is set to 10, 10, and 15 seconds, respectively.

ActiGraph setup:

- Select start time: 00:00 the following day (“tomorrow”)
- Use stop time: Tick box, and add 1 day to “todays” date, so that a full day of recording is specified.
- Sample rate: 100 Hz
- Idle sleep mode: Disabled

Actiheart setup:

- Long Term recording
- Start: 00:00 the following day (“Tomorrow”)
- Recording control: 15 seconds epoch
- Disable HRV: *not* ticked

ActiWave Cardio setup:

- Start on: 00:00 the following day (“tomorrow”)
- ECG Sample at: 128 Hz
- ECG Resolution: 9 bits
- Channel 1: Enable ECG
- Accelerometer: Tick all three boxes (x, y, z)
- Accelerometer: Sample at 32 Hz

3.2.3 Wear time

The Polar M430, the ActiGraphs, and the Actihearts are set to record for one full day. The Actiwave Cardio is set to record for the maximum 27 hours. All recordings starts at 00:00 the day after the devices are placed on the participant. Table 2 gives and overview of sample rates, recording duration, and export epochs.

Table 2 Alternative device setup

Device	Accelerometer sample rate (Hz)	ECG sample rate (Hz)	Epoch (second)	Recording duration
Polar M430	-	-	-	1 day
Actiwave Cardio	32	128	10	27 hours
Actiheart	32	128	15	1 day
ActiGraph	100	-	10	1 day

3.3 Sample

Because of the high number of devices and the high requirement for compliance, we will use convenience sampling and recruit people we believe will adhere to the protocol. We aim to include 20 adult participants, aged above 20 years and below 70 years.

3.4 Analysis

Table 3 gives an overview of available output variables from all devices, including the frequency of those variables. The Polar M430 uses proprietary unknown algorithms to calculate all output variables. For Polar M430, raw data are not available. The Actiwave, Actiheart, and Actiwave Cardio export raw data, and/or aggregated in data in a predefined number of seconds (epochs). Epoch data are used to calculate steps, calories, activity intensity, non-wear time, sleep, and heart rate/pulse in order to compare with Polar variables.

Table 3 Output variables: Polar M430, ActiGraph, Actiwave Cardio, and Actiheart

Output variables	Polar M430	ActiGraph wGT3X-BT	Actiheart	Actiwave Cardio
Steps per day	X			
Calories per day	X			
Sedentary minutes per day	X			
Light physical activity minutes per day	X			
Moderate physical activity minutes per day	X			
Vigorous physical activity minutes per day	X			
Non-wear time minutes per day	X			
Restless sleep minutes per day	X			
Restful sleep minutes per day	X			
Heart beats per minutes (BPM) frequency	12/hour	-	4/minute	1/second
ECG	-	-	4/minute	128/second
Accelerometer frequency	-	100/second	4/minute	32/second

Epoch (export)	-	10 seconds	15, 30, 60 seconds	10 seconds
Number of accelerometer axis	3	3	1	3

References

1. Butte, N.F., U. Ekelund, and K.R. Westerterp, *Assessing physical activity using wearable monitors: measures of physical activity.* Med Sci Sports Exerc, 2012. **44**(1 Suppl 1): p. S5-12.
2. Henriksen, A., et al., *Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables.* J Med Internet Res, 2018. **20**(3): p. e110.
3. McMinn, D., Acharya, R., Rowe, D. A., Gray, S. R. and Allan, J. L., *Measuring activity energy expenditure: accuracy of the GT3X+ and actiheart monitors* International Journal of Exercise Science, 2013.
4. O'Neil, M.E., et al., *Measuring reliability and validity of the ActiGraph GT3X accelerometer for children with cerebral palsy: a feasibility study.* J Pediatr Rehabil Med, 2014. **7**(3): p. 233-40.
5. Gastin, P.B., et al., *Validity of the ActiGraph GT3X+ and BodyMedia SenseWear Armband to estimate energy expenditure during physical activity and sport.* Journal of Science and Medicine in Sport, 2018. **21**(3): p. 291-295.
6. Riel, H., et al., *Comparison between Mother, ActiGraph wGT3X-BT, and a hand tally for measuring steps at various walking speeds under controlled conditions.* PeerJ, 2016. **4**: p. e2799.
7. Campbell, N., et al., *The Actiheart in Adolescents: A Doubly Labelled Water Validation.* Pediatric Exercise Science, 2012. **24**(4): p. 589-602.
8. Villars, C., et al., *Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men.* J Appl Physiol (1985), 2012. **113**(11): p. 1763-71.
9. Evenson, K.R., M.M. Goto, and R.D. Furberg, *Systematic review of the validity and reliability of consumer-wearable activity trackers.* The International Journal of Behavioral Nutrition and Physical Activity, 2015. **12**: p. 159.
10. Wahl, Y., et al., *Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions.* Front Physiol, 2017. **8**: p. 725.
11. Wang, L., et al., *Evaluation on Step Counting Performance of Wristband Activity Monitors in Daily Living Environment.* IEEE Access, 2017. **5**: p. 13020-13027.
12. Fokkema, T., et al., *Reliability and Validity of Ten Consumer Activity Trackers Depend on Walking Speed.* Medicine & Science in Sports & Exercise, 2017. **49**(4): p. 793-800.
13. Roos, L., et al., *Validity of sports watches when estimating energy expenditure during running.* BMC Sports Sci Med Rehabil, 2017. **9**: p. 22.
14. Brooke, S.M., et al., *Concurrent Validity of Wearable Activity Trackers Under Free-Living Conditions.* Journal of Strength & Conditioning Research, 2017. **31**(4): p. 1097-1106.
15. Hernandez-Vicente, A., et al., *Validation study of Polar V800 accelerometer.* Annals of Translational Medicine, 2016. **4**(15): p. 278.
16. Simunek, A., et al., *Validity of Garmin Vivofit and Polar Loop for measuring daily step counts in free-living conditions in adults.* Acta Gymnica, 2016.
17. Emaus, A., et al., *Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study.* Scand J Public Health, 2010. **38**(5 Suppl): p. 105-18.
18. Allen, J., *Photoplethysmography and its application in clinical physiological measurement - IOPscience.* 2007.

Appendix B

Letter of information and consent form, Polar M430 validation study (Paper II).

Participant: _____

Information about validation project: Activity measurement

The Purpose of this project is to test the agreement between different devices capable of measuring movement and heart rate.

During the measurement periode:

- All devices are worn at the same time
- All devices should be constantly worn, during the whole periode (day and night)
- All devices are water resistant
- The ActiGraphs worn on the hip and waist can be removed when showering, to avoid wet belts.
- If a device is removed, remember to re-attach it in the correct place and with the correct orientation.
- If an electrode loosens, replace it with a new one, and re-attach the device.
- Chest worn devices can be removed and re-attached by pressing two buttons. One button on each part of the device.
- Do your activities as you normally would.

Stored information is limited to movement and heart rate, to compare these readings from different devices. When initializing each device, age, gender, weight, and height must be entered, in order to be able to calculate energy expenditure.

Your devices can be removed when you wake up on _____day _____ _____.

Polar M430 watch
Measures movement and pulse.
Worn on the wrist (non-dominant hand).
Must have skin contact.

ActiGraph
Measures movement.
One device is worn on the wrist (non-dominant hand) above the Polar watch.
One devices is worn on the right hip.
The black “button” should be pointing upwards.

Actiwave Cardio
Measures movement and heart rate.
Worn on the chest between two Actihearts, attached with electrodes.

Actiheart
Measures movement and heart rate.
Worn on the chest, attached with electrodes. One Actiheart is worn above the ActiWave Cardio and one below the ActiWave Cardio.

Participant: _____

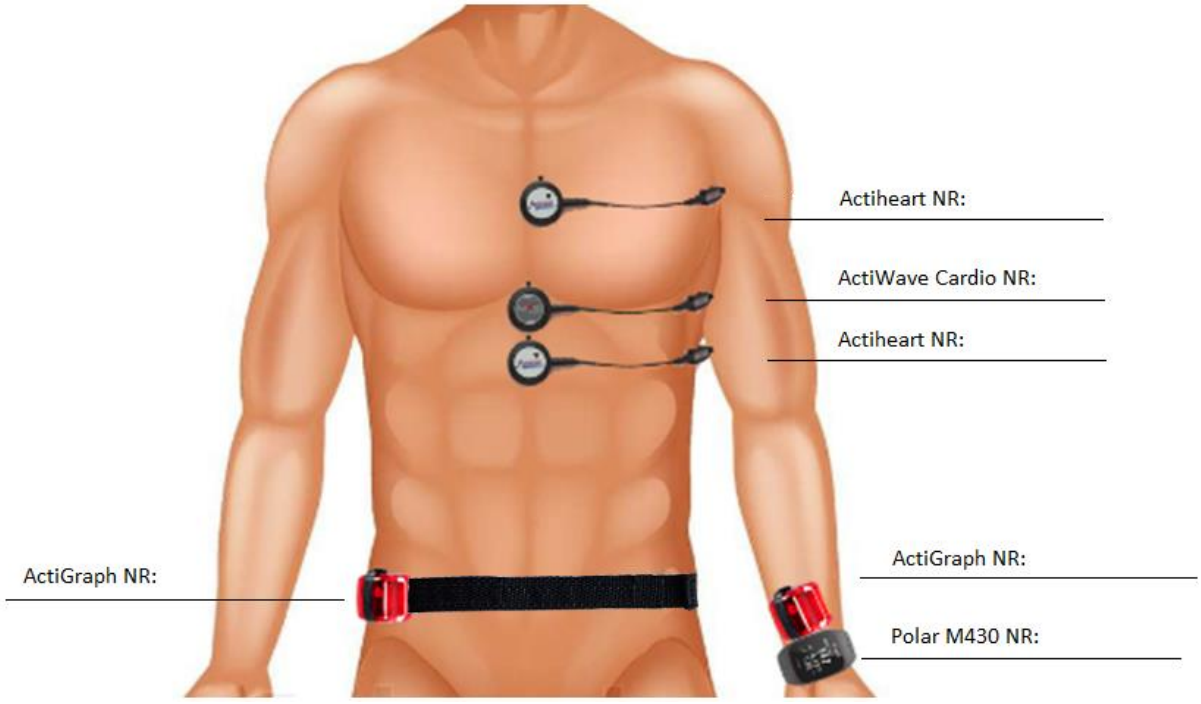
Device placement

Devices must be worn as illustrated below:

- Polar M430 Watch: On the wrist (non-dominant hand)
- One ActiGraph: On the wrist (non-dominant hand), above the Polar M430 watch.
- One ActiGraph: On the right hip.
- ActiWave Cardio: Middle two electrodes on the chest
- One Actiheart: Top two electrodes on the chest.
- One Actiheart: Bottom two electrodes on the chest.

Non-dominant hand means that if you are right handed, the devices should be placed on your left wrist. If you are left-handed, the devices should be placed on your right wrist.

The ActiGraphs on the wrist and hip should be worn in such a way that the black “button” points upward when you stand with your arms hanging down.



UiT
NORGES
ARKTISKE
UNIVERSITET

Consent for participation in accelerometer (physical activity sensor) validation study

I have received written (separate information sheet) and oral information about the purpose of this study and its contents, and I had the opportunity to ask questions about the study before consent.

I hereby consent to participate in this study

Place and date

Participant's signature

Participant's name in capital letters

I hereby confirm that I have provided information about the study

---

Place and date

Signature

Appendix C

Letter of invitation (Norwegian), RESTART pilot and feasibility study (Paper III).

UiT
NORGES
ARKTISKE
UNIVERSITET

Tromsø-
undersøkelsen

OLA NORDMANN
VEGEN XX
90XX TROMSØ

Tromsø, xx.xx.xx

Invitasjon til å delta i forskningsprosjekt om varige livsstilsendringer

I Tromsøundersøkelsen ble det gjort en rekke målinger og registreringer av faktorer som kan ha betydning for den enkeltes risiko for hjerte- og kar sykdommer. Vi gjennomfører nå et forskningsprosjekt for å måle langtidseffekten av et program for å øke fysisk aktivitet og endre kosthold blant personer som har forhøyet risiko for hjerte- og kar sykdommer. I første fase skal det gjennomføres en studie som varer i 6 måneder for å samle erfaringer om studieopplegget.
Ved å delta får du tilbud om å trene to dager i uka på Stamina, kostholdsveiledning og rådgivning til å endre livsstil. Du finner mer detaljert informasjon om dette i vedlagte «Forespørsel om deltakelse i forskningsprosjektet *Varige livsstilsendringer*».

Prosjektstart er siste uke av september i år, og vil foregå ved UNN og Stamina i Tromsø sentrum. Prosjektet avsluttes før påske 2018.

Dersom du kan tenke deg å delta, eller ønsker mer informasjon om prosjektet, kan du kontakte **Forskningsposten v/studiesykepleier Elin Hanssen, tlf. 776 26026/26909, sende en e-post til forskningsposten@unn.no**, eller returnere svarslippen som du finner nederst i brevet i vedlagte frankerte konvolutt innen 13.09.2017.

Med hilsen

For Tromsøundersøkelsen

Heidi Johansen
Prosjektleder Tromsøundersøkelsen
Institutt for samfunnsmedisin
UiT, Norges arktiske universitet

Prosjektleder «Varige livsstilsendringer»

Trygve S. Deraas
Forsker, Institutt for samfunnsmedisin
UiT, Norges arktiske universitet

✂……
Prosjekt «Varige livsstilsendringer»

☐ **Ja, jeg vil ha nærmere informasjon om studien**

Navn: …… Adresse: ……

Telefonnummer: …… E-postadresse: ……

Kontaktinformasjon:
Tromsøundersøkelsen Telefon: 776 207 00
Forskningsposten UNN Telefon: 776 26 909

UNIVERSITETSSYKEHUSET NORD-NORGE
DAVVI-NORGGA UNIVERSITEHTABUOHCCEVIESSU

Appendix D

Letter of information and consent form (Norwegian), RESTART pilot and feasibility study (Paper III).

FORESPØRSEL OM DELTAKELSE I FORSKNINGSPROSJEKTET

VARIGE LIVSSTILSENDRINGER

Dette er et spørsmål til deg om å delta i et forskningsprosjekt som har til hensikt å undersøke langtidseffekten av et program for å øke fysisk aktivitet og endre kosthold blant personer som har deltatt i den syvende Tromsøundersøkelsen (Tromsø 7), hvor de har fått påvist forhøyet risiko for hjerte- og karsykdom. I første fase av studien (pilotstudien) ønsker vi å undersøke gjennomførbarhet og samle erfaringer med studieopplegget som senere skal testes i en stor studie (hovedstudien).

Fysisk inaktivitet, overvekt og fedme øker risikoen for hjerte- og karsykdom. Flere studier tyder på at fysisk aktivitet bidrar til å redusere risiko for hjerte- og karsykdom og diabetes. Det er imidlertid få studier som har vist langvarig effekt av fysisk aktivitet og kostholdstiltak hos personer med økt risiko for hjerte- karsykdom. Likeledes er effekten av smartklokker med løpende tilbakemelding lite undersøkt. Denne pilotstudien samler erfaringer og tester utforming og logistikk av et studieopplegg, som vi har til hensikt å undersøke videre i en større studie. Pilotstudien gjennomføres i samarbeid med Stamina Helse og Stamina Trening, som har lang erfaring med oppfølging av personer med helseutfordringer. Det gjennomføres to fellesaktiviteter på Stamina hver uke, tirsdager og torsdager fra klokken 14.15 til 16.00. I tillegg vil det avsettes tid til kostholdsveiledning og motiverende gruppesamtaler.

HVA INNEBÆRER PROSJEKTET?

For å delta i undersøkelsen må du ha deltatt i den syvende Tromsøundersøkelsen og ha fått påvist forhøyet risiko for å få hjerte- og karsykdom, overvekt og at du har lavt nivå av fysisk aktivitet. Du må være mellom 5574 år, ha kroppsmasseindeks (BMI) på 30 eller mer og økt risiko for hjerte-karsykdom. Du må også være motivert og samtykke til å delta. Du kan ikke ha en alvorlig sykdom eller sykdom som begrenser din deltakelse i prosjektaktiviteter.

Personer som ønsker å være med i studien vil etter et telefonintervju bli invitert til å møte på Forskningsposten på UNN for å undersøke om de fyller kravene for deltakelse. I så fall blir de bedt om å signere samtykke til deltakelse. Studiedeltakere vil deretter fylle ut spørreskjema og undersøkes på Forskningsposten (høyde/vekt, hofte/midjemål, blodtrykk, hjerterytme, EKG, kroppssammensetning, blodprøver). De vil også gjennomgå en kondisjonstest for å undersøke fysisk form. Deretter følger de et trenings- og kostholdsprogram på Stamina i en periode på totalt 26 uker der de følges opp av teamet på Stamina. Midt i studieperioden og ved studiens slutt blir de innkalt til de samme undersøkelser som over på Forskningsposten på UNN.

Gjennom studieperioden blir du som deltaker bedt om å fylle ut spørreskjema og bruke aktivitetsmålere som bæres på hofta hele døgnet i en uke og deretter returneres til studieledelsen.
Deltakerne vil også få utlevert pulsklokke (Polar M430) som bør brukes mesteparten av døgnet i hele studieperioden. Denne er forhåndsinnstilt med en fiktiv identitet som er koblet til deltakernes nummer i prosjektet for å ivareta deltakernes personvern og sikre at vi får samlet inn data som kan brukes til å beregne søvn, stillesitting, puls og fysisk aktivitet til studiens egen datalagringsserver på UiT. Hvis deltakerne endrer innstillingene på klokka vil det kunne forhindre at vi får samlet inn de data vi ønsker, og vi anmoder derfor om at deltakerne ikke bruker pulsklokka på annen måte enn avtalt med prosjektleder eller prosjektmedarbeidere.

Deltakere kan bli forespurt om muntlige intervjuer i løpet at studieperioden på 26 uker. Samtalene vil bli tatt opp på bånd og deretter skrevet ut og analysert. Båndet slettes når intervjuene er skrevet ut. Alle opplysninger behandles konfidensielt og all presentasjon av resultater vil skje i anonymisert form. Når alle intervjuene er analysert vil alle intervjudata bli slettet.

Kondisjonstest

Dette er en test for å måle maksimalt oksygenopptak (VO_2max) ved fysisk anstrengelse. Testen innledes med 20 minutters rolig oppvarming på tredemølle eller på ergometersykkel. Selve testen tar ca. 8 minutter og foregår ved at deltaker går/løper på en tredemølle. Under testen måles hjertefrekvens, oksygenopptak (O_2) og utlufting av karbondioksid (CO_2) ved at deltaker puster gjennom en slange med et munnstykke. Testen er ikke farlig, men kan oppleves litt ubehagelig dersom deltaker ikke er vant til å presse seg. Testdeltakere oppfordres til å yte maksimalt, men kan avslutte testen på et hvilket som helst tidspunkt dersom de ikke klarer mer.

Treningsprogram

Aktiviteten omfatter trening i gruppe to ganger ukentlig med vekt på styrke, utholdenhet og bevegelighet. Treningen foregår både utendørs og inne i sal på Stamina. Deltakere oppfordres og veiledes til egenaktivitet i tillegg til den organiserte aktiviteten.

Blodprøver

Det vil bli tatt blodprøver for å undersøke blodprosent, nyre og leverfunksjon, muskelenzymer, langtidsblodsukker (HbA1c) og nivået av fettstoffer i blodet (kolesterol og triglyserider) både ved studiestart, etter 13 uker og ved studiens avslutning. Det tas også urinprøve for å sjekke nyrefunksjonen. Stoffskifteprøver tas ved studiestart.

Tidsskjema for studien

Aktivitet	sep.17	okt.17	nov.17	des.17	jan.18	feb.18	mar.18
Rekruttering							
Intervju deltakere							
Basismålinger							
Oppstart intervensjonen							
Intervensjonens periode							
Registrering							

MULIGE FORDELER OG ULEMPER

Fysisk aktivitet antas å ha mange positive effekter på helsa. Det er vist at fysisk aktivitet forebygger hjerte- og karsykdom. Det å være fysisk aktiv gir økt velvære, trivsel og mental helse og dermed økt livskvalitet.

Når personer som har vært lite fysisk aktive øker sin aktivitet er det en viss fare for å få belastningsskader. For å unngå dette vil vi i prosjektet ta utgangspunkt i den enkelte deltakers utholdenhet, styrke og bevegelighet. Deltakere kan oppleve stølhet og smerter i muskulatur etter trening. Dette er ufarlig, men kan være litt ubehagelig. Ved oppstart av trening kan enkelte oppleve leddsmerter som følge av uvant belastning. Det er da viktig å opplyse om dette slik at belastningsmengden kan tilpasses.

FRIVILLIG DELTAKELSE OG MULIGHET FOR Å TREKKE SITT SAMTYKKE

Det er frivillig å delta i studien. Du kan når som helst og uten å oppgi noen grunn trekke ditt samtykke til å delta i studien. Dersom du ønsker å delta, undertegner du samtykkeerklæringen på siste side.

Dersom du senere ønsker å trekke deg eller har spørsmål til prosjektet, kan du kontakte prosjektleder: Lege, ph.d. Trygve S. Deraas, tlf. 934 40 708, eller ledende forskningstekniker Anna-Kirsti Kvitnes tlf 77 64 48 19, e-post: anna.kirsti.kvitnes@uit.no

HVA SKJER MED INFORMASJONEN OM DEG?

Informasjonen som registreres om deg fra spørreskjema, kondisjonstester og blodprøver skal kun brukes for å studere gjennomførbarhet og effekten av et program for å øke fysisk aktivitet. Alle opplysninger og prøver vil bli behandlet uten navn og fødselsnummer. Det vil ikke være andre gjenkjennende opplysninger om deg i analysematerialet. En kode knytter deg til dine opplysninger og prøver gjennom en navneliste, og denne kodelisten vil bli oppbevart elektronisk og utilgjengelig for uvedkommende. Det vil ikke være mulig å identifisere deg i resultatene av studien når disse publiseres. Når analysene er avsluttet vil opplysningene om deg føres tilbake til Tromsøundersøkelsens helseregister.

Hvis du samtykker til å delta i studien har du rett til få innsyn i hvilke opplysninger som er registrert om deg. Du har videre rett til å få korrigert eventuelle feil i de opplysningene vi har registrert. Dersom du trekker deg fra studien vil det ikke samles inn flere opplysninger om deg. Opplysninger som allerede er innsamlet vil ikke bli slettet.

HVA SKJER MED PRØVER SOM BLIR TATT AV DEG?

Blodprøvene som blir tatt analyseres fortløpende og etter analyser blir det resterende blodet destruert. Det opprettes ikke noen biobank for studien.

FORSIKRING

Studiedeltakere der dekket av Pasientskadeerstatningen ved besøk på Forskningsposten på UNN, og de er dekket av Stamina Helse sin forsikring ved aktivitet i regi av Stamina.

ØKONOMI

Studien er finansiert gjennom forskningsmidler fra næringsrettede midler til regional utvikling (RDA-midler) og av Det helsevitenskapelige fakultet, UiT- Norges arktiske universitet.

Det er gratis å delta i studien og prosjektet dekker utgiftene til trening på Stamina Helse og Stamina Trening i 26 uker. Deltakere må selv dekke utgiftene til transport til Stamina. Bruk av offentlig transportmiddel til UNN i forbindelse med besøk på Forskningsposten dekkes av studien i form av et Sentrumsgavekort på kr 400 og deles ut ved start av studien. Deltakerne får utdelt en pulsklokke (Polar M430) til bruk i prosjektet.

GODKJENNING

Prosjektet er godkjent av Regional komite for medisinsk og helsefaglig forskningsetikk, REKnr. 2017/1100.

SAMTYKKE TIL DELTAKELSE I PROSJEKTET «VARIGE LIVSSTILSENDRINGER» 2017-2018

JEG ER VILLIG TIL Å DELTA I PROSJEKTET

Sted og dato	Deltakers signatur
	Deltakers navn med trykte bokstaver

Appendix E

Standard operating procedure Polar M430/H10 (Norwegian), RESTART pilot and feasibility study (Paper III).

Generell prosedyre UiT

Standardprosedyre:
Utlevering av pulsklokke og pulsbelte

Polar M430
Polar H10

Gjeldende versjon: 1.0

Opprettet av:
André Henriksen

Endret av:

Dato:
22.09.2017

Dato:

Standardprosedyren består av 3 dokumenter:

1. Standardprosedyre 1 av 2: Initialisert av pulsklokke
2. Standardprosedyre 2 av 2: Utlevering av pulsklokke
3. Informasjonsskriv pulsklokke

Introduksjon

Polar M430 Pulsklokke og Polar H10 pulsmåler m/belte

Alle deltakere skal få utlevert en Polar M430 pulsklokke som skal bæres på venstre eller høyre håndledet, hele dagen og hele natten, i hele studiens varighet. De skal i tillegg bruke en Polar H10 pulsmåler med pulsbelte på fellestreninger.

Mottak av deltaker

Velkommen

I dette prosjektet måler vi fysisk aktivitet, stillesitting, søvn og puls med en pulsklokke og et pulsbelte.

Her ønsker vi å samle inn pulsdata gjennom hele studien for å kunne forske mer på hvordan pulsen endres seg over tid. Pulsklokken måler også fysisk aktivitet, stillestilling og søvn, på samme måte som aktivitetsmåleren på hoften som skal bæres en uke. Vi ønsker å samle inn denne dataen for å kunne sammeligne data fra aktivitetsmåleren på hoften og pulsklokken.

Pulsklokken *(vise pulsklokken)* plasseres på høyre eller venstre håndledd og bæres i seks måneder, dag og natt. Bortsett fra når det lades.
Pulsbeltet (*vise pulsbeltet*) tas på før trening og tas av etterpå.

Pulsklokken og pulsbeltet leveres inn etter seks måneder, dvs siste samling på UNN

Takk til deltager.

Utstyr: Pulsklokke M430 og Pulsmåler H10

- Polar M430 Pulsklokke
- Ladekabel til pulsklokke m/strøm
- Polar H10 pulsmåler m/pulsbelte
- Informasjonsskriv

Initialisering av pulsklokke

1. Be deltaker feste pulsklokken på venstre eller høyre arm
2. Sjekk med deltaker at dette er den armen han/hun vanligvis vil ha klokken på, og noter ned svaret.
3. Be deltaker ta fram sin telefon og slå på Wifi og Bluetooth
4. Be deltaker laste ned «Polar Flow» app
5. Neste fire punkter er ulik for Android/iPhone

Android

a) Åpne «Play Butikk» og søk etter «Polar flow». Installer
b) Be deltaker starte «Polar Flow» app, samt godkjenne «polar electro End-User software agreement» (kan hende det står på norks)
c) Appen vil be om tilgang til enhetens plassering. Velg «Tillatt». **NB:** *Informer deltaker om at vi ikke skal samle inn GPS-data, men siden dette er en GPS-klokke trenger den tilgang til GPS (enhetens plassering)*
d) De trenger ikke aktivere GPS/possisjonstjenesten selv om appen ber om det.

iPhone

a) Åpne «App store» og søk etter «Polar Flow». Hent og installer.
b) Be deltaker starte appen
c) De må trykk «ok» når de får beskjed om at «polar flow ønsker å gjøre data tilgjengelig for bluetooth-enheter i nærheten selv om appen ikke er i bruk»
d) Trykk på «Kom i gang»

Resten er felles

6. Be deltaker legge inn e-postadressen på formen: varigendringXX@helsefak.uit.no, der XX erstattes med deltakerens deltakernummer (01, 02, .., 18,19,20). **NB***: Husk ledende 0 på tall under 10.*
7. Passordet legges inn av prosjektmedarbeider. **NB:** *Denne skal ikke deles med deltaker.*
8. For de med Android: Deltaker kan legge inn kjønn, fødselsdato, høyde og vekt hvis de vil. Trykk «Fortsett».
9. Klokken kobles til telefonen ved å:

   a. Holde inn knappen nederst til venstre på klokken i **3 sekunder**, inntil klokken viser «Koble til enhet»
   b. Holde klokken inntil telefonen, mens Polar Flow-appen er aktiv
   c. Klokken viser en pin kode som må skrives inn på telefonen. Legg inn denne og trykk «sammenkoble»/ «ok»
   d. Synkroniseringen kan ta litt tid.
   e. Hvis synkroniseringen ikke blir vellykket, prøv på nytt.
10. Polar Flow-appen kan nå lukkes
11. Klokken er nå koblet til deltakers telefon. Pulsmåler er allerede koblet til klokken.

Utlevering av Pulsklokke

Forklaring til deltakerne

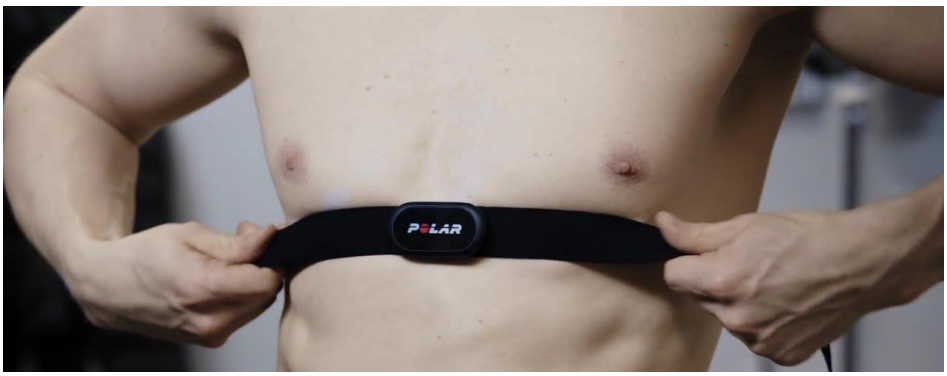
1. Deltakeren skal utføre sine aktiviteter som vanlig.
2. Pulsklokken brukes kontinuerlig til stoppdato, gjerne også under søvn.
3. Pulsbeltet brukes ved fellestreninger
4. Informer om punktene i infoskrivet.
5. Del ut informasjonsskriv, pulsklokke og pulsbelte

Feste pulsklokken

- Pulsklokken festes på høyre eller venstre arm. Som en vanlig klokke.

Feste pulsmåler/pulsbeltet

- Før trening klikkes pulsmåleren fast i pulsbeltet (vis hvordan dette gjøres)
- Baksiden av pulsbeltet må fuktes med vann.
- Pulsbeltet settes rundt livet, slik at pulsmåleren er ca midt under brystmusklene
- Etter trening klikkes pulsmåleren av pulsbeltet, og begge deler skylles i vann



Appendix F

Letter of information Polar M430/H10 (Norwegian), RESTART pilot and feasibility study (Paper III).

Informasjon om pulsklokken Polar M430 og pulsmåler H10

Polars pulsklokke M430

Pulsklokken måler bevegelse og puls. Data fra pulsklokken vil brukes for å måle aktivitetsnivået.

Daglig bruk

Klokken bæres hele dagen og natten så langt det er mulig. Klokken tåler vann og kan brukes i basseng og ved dusjing. Bluetooth skal være slått på hele tiden.

Lading

Klokken må lades hver søndag med medfølgende ladekabel.

1) Sett i ladekabel på klokken
2) Slå på Bluetooth og Wifi på telefonen
3) Hold inne knappen nederst til venstre på klokken i 2 sekunder

Dataoverføring

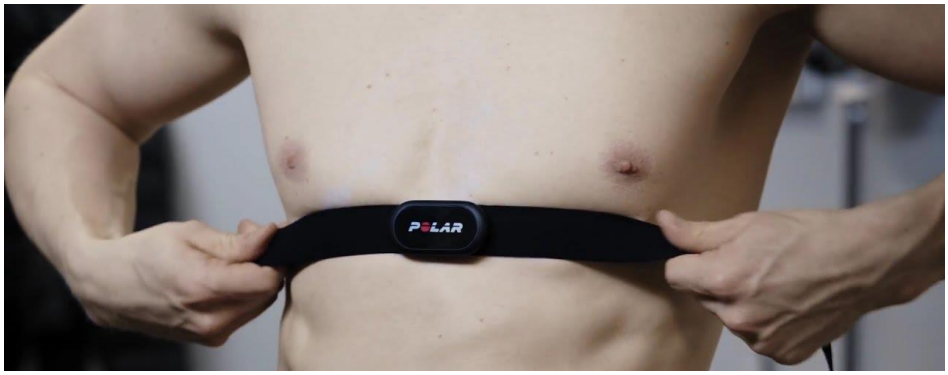
Data overføres fra klokken til en app (PolarFlow) kontinuerlig (Bluetooth må være på) når klokken er i nærheten av din mobiltelefon. PolarFlow-appen må ikke fjernes/avinstalleres fra telefonen. Dataoverføring krever at telefonen er koblet til et nettverk (WiFi). Sørg for at Bluetooth og WiFi og er aktivert på telefon og den er koblet til et nettverk når klokken lades.

Polars bulsmåler H10 m/belte

Aktivitetsmåleren festet til et elastisk bånd når den er i bruk og kobles av igjen når treningsøkten er over.

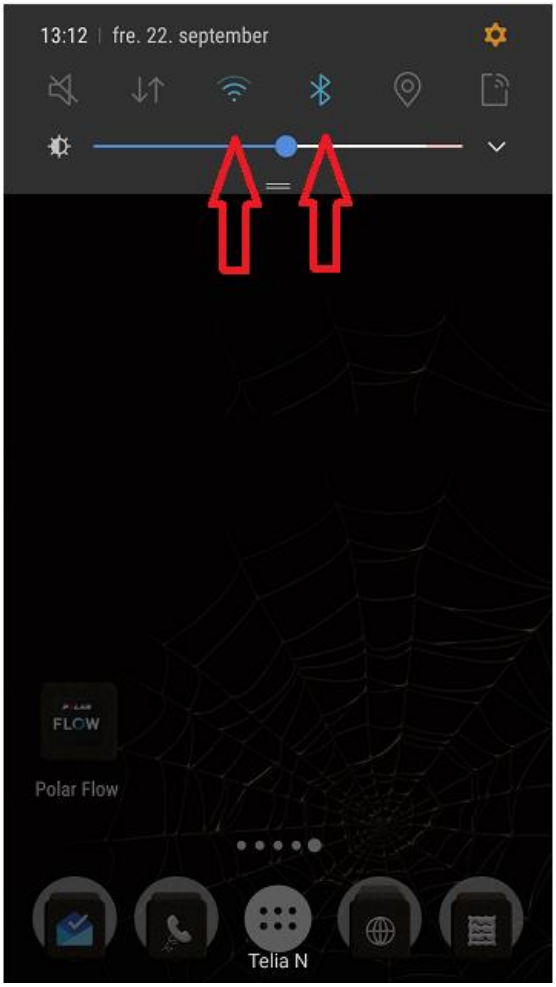
Vær oppmerksom på følgende:

- Utfør alle dine aktiviteter som vanlig
- Bruk klokken hele døgnet, også om natten
- Bruk pulsmåler m/belte ved fellestreninger
- Pulsmåler må bæres direkte mot hud, som i bildet under



Hvordan slå på Bluetooth/Wifi: Android

Dra med en finger ned fra toppen av telefonen og trykk på Bluetooth-symbolet for å slå på Bluetooth, og Wifi-symbolet for å slå på Wifi.



Hvordan slå på Bluetooth/Wifi: iPhone

Dra med en finger opp fra bunnen av telefonen og trykk på Bluetooth-symbolet for å slå på Bluetooth, og Wifi-symbolet for å slå på Wifi.

Appendix G

Letter of information (Norwegian), mSpider volunteers (Paper IV)

UiT
NORGES
ARKTISKE
UNIVERSITET

Deltaker: _____

Informasjon om studie for innsamling av aktivitetsdata via smartenheter

Hensikten med studien er todelt, der DEL 1 er å teste ut datauthenting fra ulike aktivitetsmålere og smartklokker (smartenheter), og DEL 2 er å undersøke samsvar mellom målinger av bevegelse og hjertefrekvens mellom ulike aktivitetsmålere.

Datainnsamling i DEL 1 gjøres via utlevert smartenhet. Denne bæres fra du får den utlevert og så lenge du ønsker å gå med den. Det er ønskelig at du bærer enheten hele døgnet over flere måneder, gjerne et helt år, men du kan selv bestemme når du tar den av og på. Når du ikke lenger ønsker å delta skal enheten returneres.

Datainnsamling i DEL 2 gjøres på et senere tidspunkt, og innebærer å gå med flere målere i én eller flere begrensede perioder. Eget informasjonsskriv og eget samtykkeskjema for denne delen blir utlevert før denne innsamlingen starter.

Informasjon som lagres i DEL 1 hentes automatisk fra smartenheten, datalagringsløsningen til smartenhetens leverandør eller dens samarbeidspartner (f.eks. Google/Apple). Dette inkluderer informasjon som steg, tid i ulike aktivitetssoner, og pulsdata. I DEL 2, der vi måler aktivitet med akselerometer/ hjertefrekvensmåler, samler vi i tillegg inn informasjon om aktivitetsmønstre og hjertefrekvens, samt grunnleggende informasjon om alder, kjønn, vekt og høyde. Denne grunnleggende informasjonen er nødvendig for å kunne beregne energiforbruk.

Du har fått utlevert følgende enhet (leverandør/modell): _____ /_____

Følgende mobil app må installeres på din mobiltelefon: _____

Ved oppstart

- Smartenheten plasseres på håndledd (ikke dominant hånd). Ikke-dominant hånd vil si at dersom du er høyrehendt skal målerne stå på venstre håndledd.
- Du må installere tilhørende mobil app på din mobiltelefon (Polar, Fitbit, Garmin, etc).
- Du må opprette en privat brukerkonto i appen og legge inn riktig informasjon om høyre, vekt, alder, etc. **NB: All data du legger inn, samt registrert aktivitetsdata, vil også lagres hos leverandøren av smartenheten.**
- Du må koble smartenheten til din mobiltelefon via tilhørende mobil app.
- Du vil få tilsendt en link per e-post, der du godkjenner at datainnsamlingsstudien kan hente ut din aktivitetsdata fra smartenhetens lagringsløsning.

I måleperioden

- Smartenheten må lades med jevne mellomrom (ca. en gang i uken).
- Smartenheten må være koblet til din telefon via tilhørende app.
- Smartenheten bæres helst hele døgnet
- Smartenheten bæres over så mange dager og måneder som mulig

UiT
NORGES
ARKTISKE
UNIVERSITET

Samtykke til deltakelse i studie for datainnsamling fra smartklokker

Jeg har mottatt skriftlig (eget informasjonsskriv) og muntlig informasjon om studiens formål og innhold, og jeg har hatt mulighet til å stille spørsmål om studien før samtykke.

Jeg samtykker herved i å delta i studien.

Sted og dato

Deltakers signatur

Deltakers navn med trykte bokstaver

Jeg bekrefter å ha gitt informasjon om prosjektet

Sted og dato

Signatur

Appendix H

Letter of information (Norwegian), mSpider COVID-19 physical activity study (Paper IV)

Kontinuerlig innsamling av aktivitetsdata fra forbrukerbaserte aktivitetsmålere

Dette er en invitasjon om å delta i et forskningsprosjekt hvor formålet er å samle inn data fra ulike aktivitetsmålere/smartklokker. I dette skrivet gir vi deg informasjon om formålet med prosjektet og hva deltakelse innebærer.

Data fra aktivitetsmålere samles inn ved hjelp av mSpider, et datainnsamlingssystem utviklet ved UiT Norges arktiske universitet. Aktivitetsmålere samler inn data basert på bevegelse (fysisk aktivitet, hvile, søvn) samt puls.

Formålet med forskningsprosjektet er todelt. Vi vil teste ut datauthenting fra ulike aktivitetsmålere. Vi vil også studere ulike typer data fra aktivitetsmålere over tid, herunder se på hvordan aktivitetsnivået blant deltakerne har endret seg pga. COVID-19.

Datainnsamlingen gjøres enten via utlevert aktivitetsmåler eller via en aktivitetsmåler du allerede har.

For deg som har din egen aktivitetsmåler

Du bærer din måler som vanlig. Vi vil samle inn data både framover og bakover i tid.

For deg som får utlevert en aktivitetsmåler

Utlevert måler bæres fra du får den utlevert og så lenge du ønsker å gå med den. Det er ønskelig at du bærer måleren hele døgnet, over tid. Når du ikke lenger ønsker å delta skal måleren returneres.

Data samles automatisk inn til datalagringsløsningen til aktivitetsmålerens leverandør (Fitbit, Polar, Garmin, Withings, Oura, Apple, Samsung, Google), via tilhørende app for mobiltelefon. Basert på innsamlet data fra aktivitetsmåleren (bevegelse og puls) beregnes ulike variabler f.eks. antall steg og energiforbruk. Noen aktivitetsmålere kan registrere GPS posisjoner. Registrering av GPS kan forhindres ved å avslå tilgang til GPS i tilhørende mobil app. I tillegg vil annen data du selv registrerer i appen (f.eks. kjønn, alder, høyde, vekt) samles inn. I dette forskningsprosjektet samler vi ikke inn slik identifiserende data fra aktivitetsmålerens leverandør (f.eks. GPS, kjønn, alder, høyde og vekt). Mer informasjon om personvernerklæringene til disse leverandørene er samlet her: https://mspider.org/tos.

Hva innebærer det for deg å delta?

Ved oppstart

- Aktivitetsmåleren plasseres på ikke-dominant hånd håndledd (dvs. dersom du er høyrehendt; venstre håndledd, dersom du er venstrehendt; høyre håndledd).
- Tilhørende mobil app må installeres på din mobiltelefon (f.eks. Polar, Fitbit, Garmin).

- Privat brukerkonto i appen må opprettes, og riktig informasjon om kjønn, alder, høyde, og vekt må legges inn. **NB: Data du legger inn, samt bevegelsesdata, vil også lagres hos leverandøren av aktivitetsmåleren.**
- Aktivitetsmåleren må kobles til tilhørende app på din mobiltelefon.
- Datainnsamling for dette forskningsprosjektet må godkjennes via tilsendt nettlenke.

I måleperioden
- Aktivitetsmåleren må lades med jevne mellomrom.
- Aktivitetsmåleren må være koblet til din telefon via tilhørende app.
- Aktivitetsmåleren bæres helst hele døgnet
- Aktivitetsmåleren bæres over så mange dager og måneder som mulig

Generell informasjon

Hvem er ansvarlig for forskningsprosjektet?
UiT Norges arktiske universitet er ansvarlig for prosjektet.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Datainnsamlingen for dette forskningsprosjektet vil da opphøre.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger
Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.
Kontaktinformasjon oppbevares så lenge du er deltaker. Kun prosjektansvarlig har tilgang til din kontaktinformasjon. Navn og kontaktinformasjon lagres ikke sammen med øvrig data. Innsamlet data er avidentifisert.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?
Opplysningene anonymiseres når prosjektet avsluttes og/eller måleutstyret leveres tilbake.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Disse rettighetene omfatter de opplysninger forskningsprosjektet har kontroll over. For innsyn, sletting og retting av øvrig informasjon som lagres hos leverandøren av aktivitetsmåleren, må det tas direkte kontakt med relevant leverandør. Vi oppfordrer deltakere til å lese personvernerklæringen til den relevante leverandøren (https://mspider.org/tos).

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.
På oppdrag fra UiT Norges arktiske universitet har NSD – Norsk senter for forskningsdata vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- UiT Norges arktiske universitet ved André Henriksen
  andre.henriksen@uit.no, +47 91 36 83 82
- UiTs personvernombud Joakim Bakkevold
  personvernombud@uit.no, +47 77 64 63 22/+47 97 69 15 78

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, ta kontakt med:

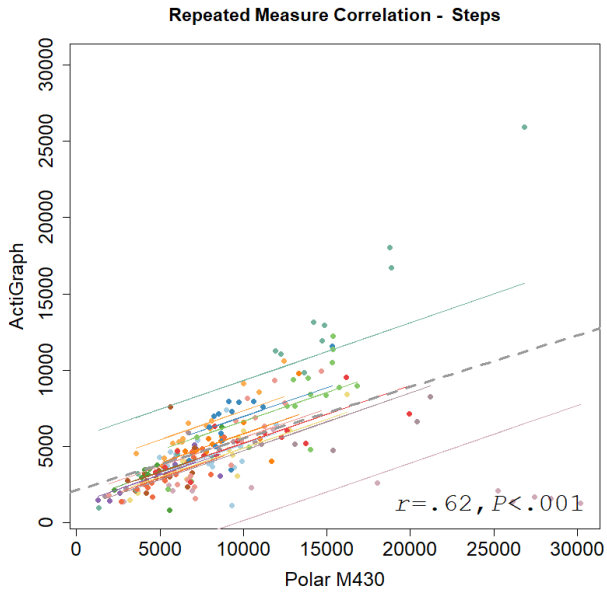
- NSD – Norsk senter for forskningsdata
  personverntjenester@nsd.no, +47 55 58 21 17

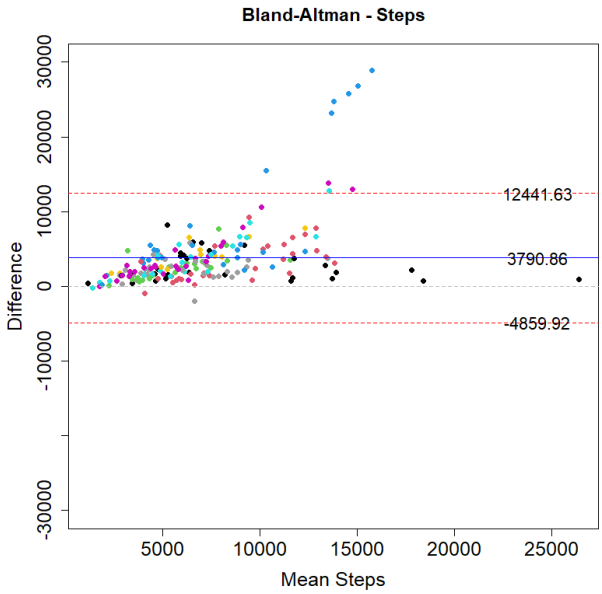
Med vennlig hilsen

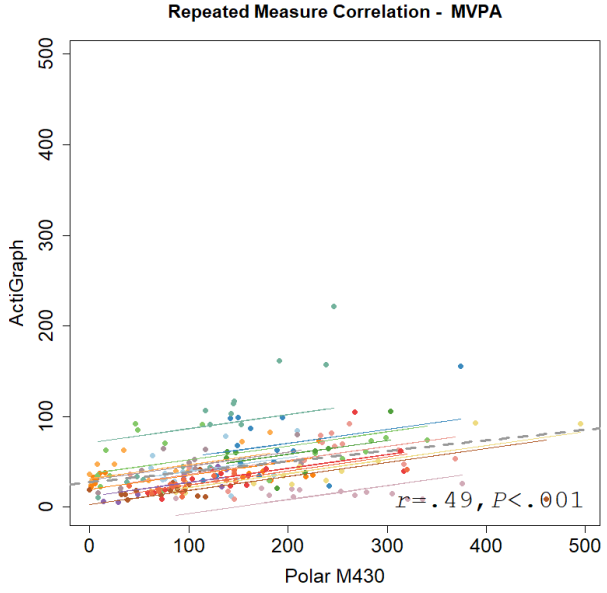
André Henriksen
Prosjektansvarlig

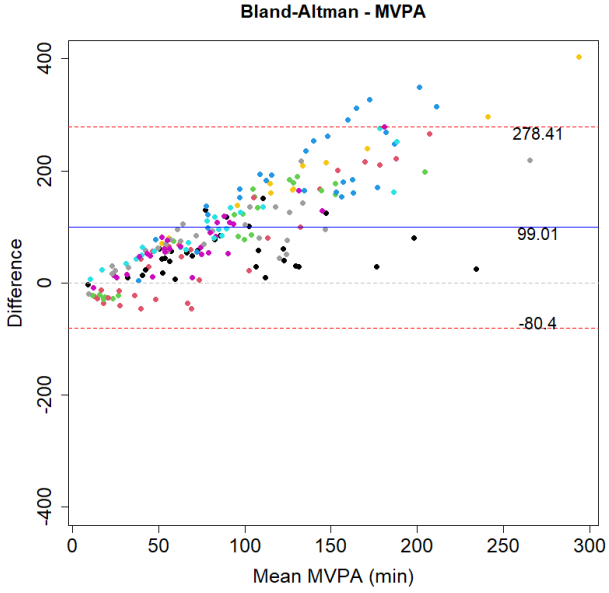
Appendix I

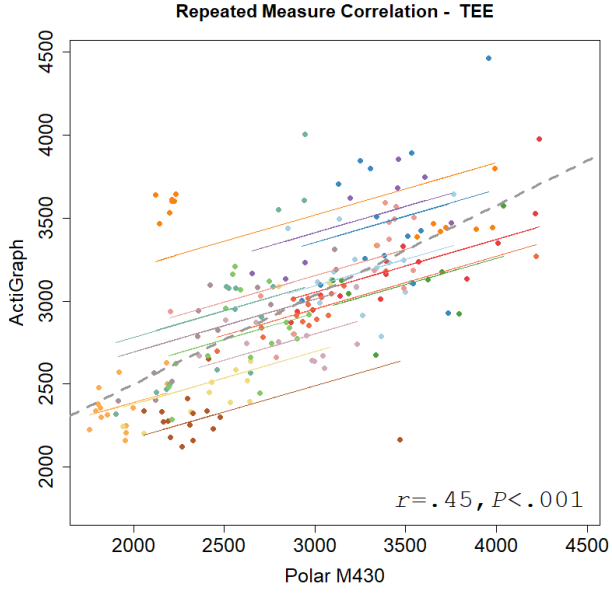
Correlations and Bland-Altman plots, with separate colours for each participant, for steps, MVPA, and total energy expenditure (Paper III)

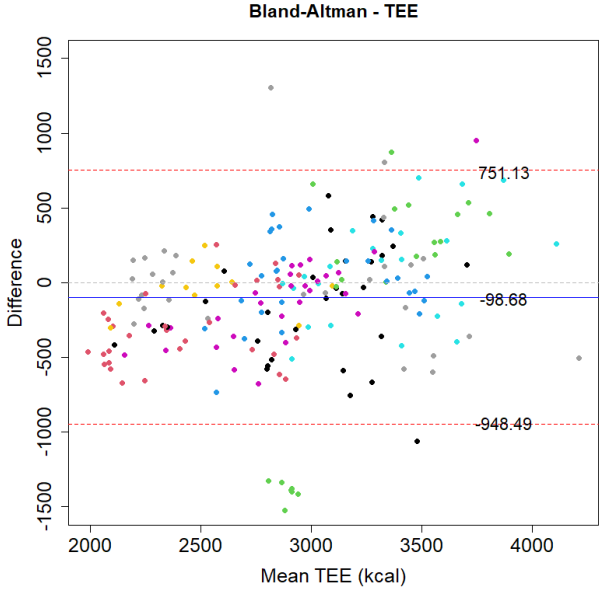

Repeated Measure Correlation - Steps
ActiGraph
Polar M430
r=.62, P<.001


Bland-Altman - Steps
Difference
Mean Steps
12441.63
3790.86
-4859.92


Repeated Measure Correlation - MVPA
ActiGraph
Polar M430
r=.49, P<.001


Bland-Altman - MVPA
Difference
Mean MVPA (min)
278.41
99.01
-80.4


Repeated Measure Correlation - TEE
ActiGraph
Polar M430
r=.45, P<.001


Bland-Altman - TEE
Difference
Mean TEE (kcal)
751.13
-98.68
-948.49