Using cloud-based physical activity data from Google Fit and Apple HealthKit to expand recording of physical activity data in a population study

André Henriksen[a], Laila Arnesdatter Hopstock[a,b], Gunnar Hartvigsen[c,d], Sameline Grimsgaard[a]

[a] Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway,
[b] Department of Health and Care Sciences, UiT The Arctic University of Tromsø, Tromsø, Norway
[c] Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway,
[c] Norwegian Centre for E-health Research, University Hospital of North Norway (UNN), Tromsø, Norway

Abstract

Large population studies are important sources for medical research. These studies are well planned, well organized, and costly. However, people record health data themselves using different sensors, which are mostly unplanned, unorganized and inexpensive. Nevertheless, self-recorded data might be an important supplement to population studies. The question is how to access and use this data.

In the seventh survey of the Tromsø cohort study, questionnaires and accelerometers were used to collect data on physical activity (PA).

We now plan to collect historical PA data from these participants, using mobile sensor data already stored in the cloud. We will examine the feasibility of this approach and the quality of this data. Objectively measured historical data will provide valuable insights in the potential and limitations of mobile sensors as new data collection tools in medical research.

Keywords

Cohort studies, Human Activity, Fitness Trackers

Introduction

The Tromsø Study is a large ongoing population based study in the municipality of Tromsø, Norway. Researchers have conducted repeated surveys 5-8 years apart for a period of 40 years. The Tromsø Study was initiated in 1974 to understand the reasons for the high prevalence of cardiovascular disease in Norway, as well as to find ways to prevent and intervene [10; 11]. Since then there have been six additional surveys.

In the most recent survey (2015-2016), the seventh Tromsø Study (Tromsø 7), everyone above the age of 40 were invited. More than 21,000 people participated (65% attendance rate) and went through a wide range of examinations, including questionnaires and interviews, physical and clinical examinations and biological sampling. There are more than 50 different research projects in Tromsø 7, several of which uses PA as a predictor, endpoint or adjustment variable.

In all seven surveys, participants completed a questionnaire, including questions about PA. In Tromsø 7 there were six questions related to PA (sedentary behavior, occupational and leisure-time PA, and PA frequency, duration and intensity).

In the sixth Tromsø Study (2007-08), a subsample of 300 participants wore an accelerometer for seven days, to obtain objectively measured data on PA. Although the coinciding questionnaire was previously validated as a tool to collect data of high-intensity leisure activity, it was also shown that moderate-intensity leisure activity was over-reported when compared to the objectively measured data [4].

In Tromsø 7, 6,300 participants wore the ActiGraph wGT3X-BT accelerometer for one week. A subsample of 700 participants wore the CamNtech Actiwave Cardio for 27 hours. With these sensors, researchers were able to objectively measure PA, energy usage, sedentary behavior and sleep. The Actiwave Cardio also measured a single lead ECG giving data on heart rhythm and heart rate.

Since PA is an important life style factor, data collected from accelerometers will be important in several ongoing projects. Combining self-reported PA data with objectively measured PA using accelerometers enhances validity in PA measurements. However, having access to continuous PA data over an extended period can be a valuable addition to the data already collected. The next step is to collect historical PA data for participants in Tromsø 7.

Most people carry a smart phone with them throughout the day [14] and a lot of people also have an activity tracker in form of a smart watch or activity bracelet [15]. These devices have several sensors that can measure PA. Android and iPhone users can opt to upload collected health information to Google Fit or Apple HealthKit, which are cloud-based services that allows device vendors and application developers to store health related information in a common online location. This can be beneficial for users, because they can access this information and see details and summaries for all kinds of health related information in one application. A wide range of health related data types can be stored here, but in this project, we will only look into how to access PA and heart rate data.

Compared to the ActiGraph, PA data collected from mobile sensors and connected smart devices are not equally validated for all types of measurements. A systematic review from 2015 indicates that validity of step counting was high, whereas energy expenditure and sleep had a low validity [5]. Validity on PA was inconclusive because of few available studies. In this review, they only considered FitBit and Jawbone, the two leading vendors in the consumer market in 2015. Newer studies seem to indicate the same results [3; 12]. Two studies [13; 19] from 2016 also indicates that modern wrist worn heart rate monitors are accurate and therefore relevant for our study. Moreover, there will be gaps in the data for participants who do not have a wearable fitness tracker, because not everyone carries their smart phone at all times. These limitations are not necessary drawbacks, but they need to be well understood.

In this vision paper, we will present our plans for retrieving historical PA data, by accessing Google Fit and Apple HealthKit cloud services, for users with smart phones and activity wearables. We will only invite people who participated in Tromsø 7, because their data can enhance existing data sets for Tromsø 7 participants, as well as allow us to examine the feasibility and quality of this data compared to subjective and objective PA measurements.

Methods

Cloud based health repositories

On September 17, 2014, Apple released *HealthKit*. Google released *Fit* on October 28, 2014. Users can activate these services to track and store PA, heart rate, diet, sleep patterns and other health metrics. It is possible for users to input information manually, it can be added by third party applications, or it can be automatically collected from internal phone sensors and sensors in connected smart devices (i.e. watches, bracelets, etc.). Users can later view this information or share them with other applications to get a better overview of their health status.

"Google Fit is an open ecosystem that allows developers to upload fitness data to a central repository where users can access their data from different devices and applications in one location" [9]. Google Fit is supported by Android 2.3 and above and consists of four components: 1) the Google Fitness store, 2) the sensor framework, 3) permissions and user controls, and 4) the Google Fit APIs. There are six APIs and the *History API* and the *Sessions API* allows access to historical data and can be used to access relevant information [7; 8]. Users can specify what they want to share with the different applications.

Apple HealthKit was introduced in iOS 8 and provides a "centralized, coordinated, and secure data store for health-related information" [2]. There are three ways to access fitness information through the HealthKit API [1]. The *query* approach is appropriate for extracting snapshots of specified periods within the store. These snapshots contain, among other things, aggregated data like step counts.

It is possible to export health data manually from Google Fit, either from the Google web sites or the Google Fit application. Supported formats includes TCX-files and CSV-files. TCX-files (Training Centre XML) is a specialized XML format for transferring heart rate, training pace, calories, GPS and other PA related information. Apple HealthKit also supports manual data export to XML-files and CSV-files. Exporting data manually from these services is cumbersome and not a straightforward process for all users. It is however, a way to download the data without writing any code, and we can therefore use it as a quick way to access sample data during development and testing.

We do not know how many Tromsø 7 participants who actually use these cloud services, when they started using them, how often and what type of information they add or how accurate this information is. However, among the more than 21,000 participants in Tromsø 7, it is likely that several has used these cloud services for some time. Furthermore, 83% of the population in Norway have a smart phone [14] and it has been estimated that more than one million smart bracelets and watches will be sold in Norway in 2016 [15]. This is in addition to the 720,000 bracelets sold in 2014 and 2015 [15].

Information flow

In the proposed solution, we will create a two-part system. In addition, some enhancements on an existing system is required. The first part is a mobile application that will collect and forward health data to the second part. The second part is a Web Service (WS) that will collect, process and forward data to an existing system called EUTRO. EUTRO is the main storage for research data in the Tromsø Study.

Figure 1 shows a simplified illustration of the information flow in this solution. Every time an authorized health application or device receives data, it can also store (1) this data in the Apple HealthKit (A) and/or Google Fit (B) cloud services. This data can be accessed by other authorized applications and the proposed application (C) will be installed on participant's phones, so that this information can be downloaded (2) from the two cloud services. The data will then be uploaded (3) to EUTRO, via the WS, connecting the data to the correct participant in the Tromsø Study.

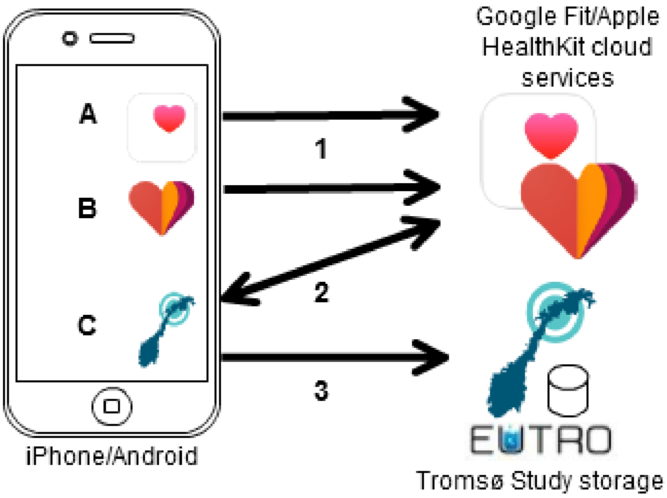

Figure 1 - Information flow

System architecture

The proposed system can collect data from potentially thousands of participants, process this data into usable information, and forward it to EUTRO.

Figure 2 illustrates the architecture for this system.

Internal phone sensors and *smart device sensors* generates data which, in addition to manually entered health information, can be shared with *third party health applications* and/or stored in the *Apple HealthKit/Google Fit application* running on the user's phone. The *Apple HealthKit/Google Fit application* will forward this information to the appropriate cloud service. *Third party health applications* can also forward this data to other devices or application specific cloud services.

Participants must install the *Tromsø Study application* on their phones. This application will, upon user authorization, access relevant health data from *Apple HealthKit* and *Google Fit cloud services*. After authorization, the participants must enter a unique number, identifying them as participants in Tromsø 7. This allows linkage of new data to existing data for study participants. Participants will also be able to specify which health data the application can access.

In addition to health data, it is important to have access to sensor and device meta-data. This is necessary in order to examine the data quality, and to examine whether this data collection strategy is a feasible method for collecting objective PA data. After the data transfer to the *Web Service* in the *Tromsø Study backend* is complete, participants can uninstall the *Tromsø Study application*. We are only focusing on historical data in this study and will only download data once. Future PA data is not part of this solution.

The *Web Service* receives the data and processes it into information that can be stored in *EUTRO*. Depending on what data participants have stored in the cloud services, it is possible to store a range of different information. Examples include daily step count, calories burned, heart rate, workouts, sleep patterns and more. The *Web Service* will forward the processed information to *EUTRO*, where it is stored and made available for active projects in the Tromsø Study.

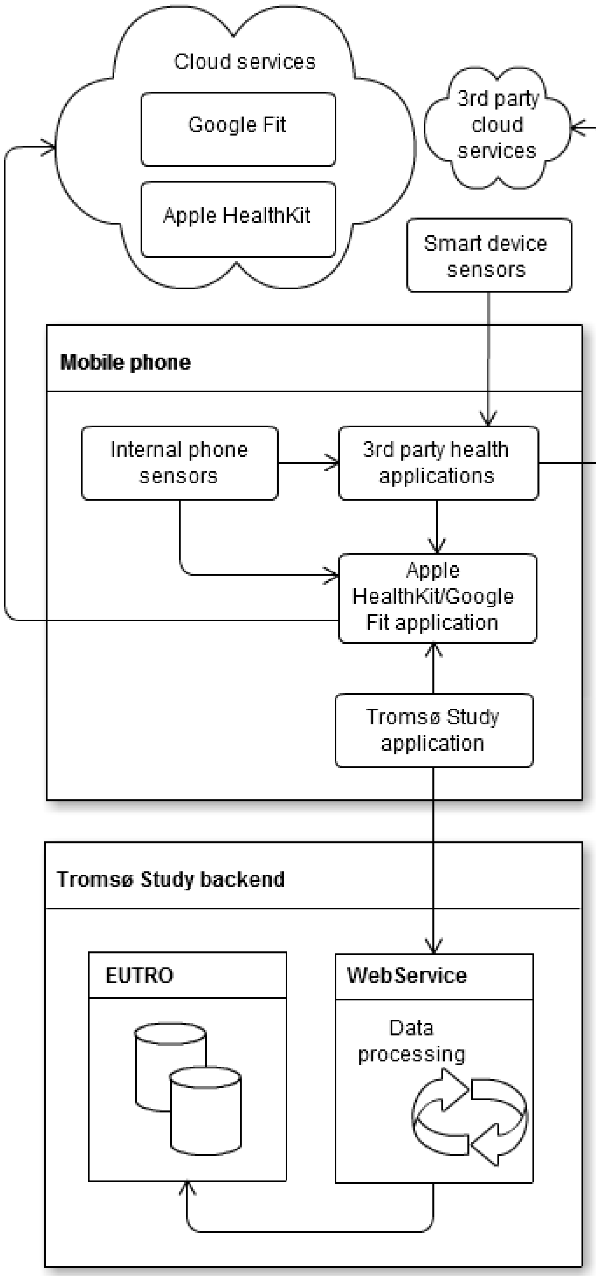

Figure 2 - System architecture

We can potentially invite more than 21,000 participants, so it is important that the system scales. Not everyone will connect at the same time, but the amount of available data can be extensive, depending on how much data participants have stored in the cloud and how much they are willing to share.

Mobile application

Our goal is to develop Android and iPhone applications that can read PA and heart rate data from these cloud services. To avoid having to create multiple code bases, we will write the application once and compile it for two platforms. Several solutions allow this. One promising alternative is *React Native* by Facebook [6], which makes it possible to program the application in JavaScript and compile two versions, one for Android and one for iPhone. React Native also supports coding natively, which we will need in order to access the two cloud services.

Through the application, participants can specify which data they agree to transfer, as well as from which period they want to share data. It is not possible to run one query to return everything. Each type of data must be accessed and transferred individually. It will therefore take some time to collect and transfer the data. Ideally, participants will be willing to share all available data, but because some participants may be reluctant to share everything, they will get the option to specify what they share. Although the amount of data and the level of details for each participant will vary because of this, we hope this will result in a higher participation rate. Figure 3 shows a mock-up of the proposed mobile application, where the user can select what to share and from which period they want to share data.

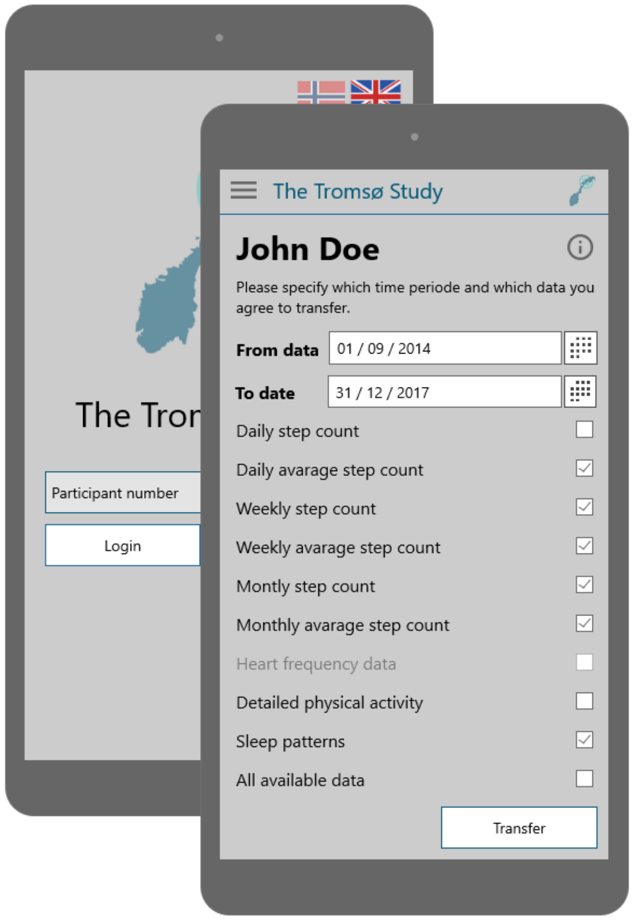

Figure 3 - Mobile application mockup

EUTRO

Data collected in the Tromsø Study is stored in EUTRO, a system for managing research projects, biological material, health data and meta data [16; 17]. Historical PA data must also be stored here. EUTRO does not currently have an interface for receiving information from mobile applications or other external systems. This means that EUTRO must be extended with an interface to receive data from the new WS, which will serve as an intermediate server for collecting and processing data from the mobile application, as well as a prototype for the changes needed in EUTRO. Until EUTRO is updated, collected data can be manually imported into EUTRO using existing features.

Web Service

The mobile application's only function is to transfer health data from the two mentioned cloud services. The planned WS will be the most complex part of the system. This is because the transferred data will be of different types, have different detail levels and be from different periods. This data must be processed into usable and comparable information.

The data that is most likely to be available is step count data, which might be aggregated hourly, daily, weekly or monthly, or averaged over a longer period.

The following types of health data will be collected, processed and forwarded, if they are available:

- Daily, weekly and monthly step counts
- Daily, weekly and monthly average step counts
- Daily, weekly and monthly average heart rate
- Sedentary behavior
- Sleep patterns
- Exercise bouts, including detailed heart rate, intensity, average speed and inclination
- Estimated precision on each data set
- Sensor meta data

Security is important when working with health data. Information transferred between systems will be encrypted using SSL. Authentication is achieved by using a unique identification number included in the invitation letter. When the mobile application has completed the data transfer, it will call a method to disable the identification number. This will further enhance security. The WS will be REST-full, that is, stateless and self-contained, encouraging a simple, lightweight and fast interaction between server and client. This is important to support the high number of potential participants.

Challenges

There are several challenges in this project. Motivating participants to install the application in order to share their information, is probably going to be the main challenge.

Wearing an ActiGraph, a standalone physical device, not connected to anything is one thing. With the ActiGraph, it was expected to be challenging to motivate participants to wear the device for several days, because of the inconvenience it might cause. However, compliance among participants in Tromsø 7 was high, with a 94% acceptance rate. Installing an application with the purpose of sharing personal health related data is another thing altogether. There are several challenges with this approach where trust and motivation probably are the two most important.

Installing any application on your personal phone, where you have so much personal data stored, requires a certain level of trust. There are many fraudulent applications which only function is to hijack your phone and access your personal information. As a participant, you must trust that the application only does what it claims it will do. Specifically, that the application will only access data the participant has agreed to share, and that the system is secure and data interception is impossible during transfer.

We need to find ways to motivate participant to share their data. The Tromsø 7 survey contained questions, tests and measurements, divided into two phases. Those attending both phases have already spent about 4 hours contributing data to Tromsø 7. Participants in the PA project, wearing the ActiGraph for a week, have contributed even more. Asking participants to use more time can potentially be a challenge. However, because the effort needed to share data is very low, once the application is installed, we are confident that many will be willing to participate in this extension to the PA project. After installing the application, participants only have to input an identification number and select which data to share. The rest is automatic and requires no additional input.

However, installing the application and using it properly can be a challenge for some participants. In a future survey, it is possible to set up a station where people can get assistance with the setup. Because the data collection phase of Tromsø 7 is completed, there is no infrastructure to provide such assistance now. This will be resolved by making a detailed instruction that must be included in the invitation letter. In addition, making an instruction video might result in some additional participants.

Ethics

Actively installing the application and selecting which data to share requires informed consent from all study participants. We will apply for recommendation from The Regional Committee for Medical and Health Research Ethics, and for approval from the Norwegian Centre for Research Data on the processing of personal data.

Expected results

Scenarios

The most likely scenario is a person who only have a mobile phone and no wearable smart device. For this person, only a limited set of data will be available, and only in periods the person has the phone with him/her. For several hours of the day, as well as during the night, the internal mobile sensors will not collect any activity data. In this scenario, we expect that the system will get access to daily step count, which is unlikely to be accurate because the phone is not carried around all day.

A more data rich scenario will be a person that has an activity bracelet or watch that also can collect heart rate. For this person, it will be possible to collect heart rate data for specific days, as well as to compare this data with different levels of PA. If this person brings a GPS-enabled phone during workouts, or has GPS built into the bracelet/watch, it will be possible to access average speed, max speed, altitude differences, average heart rate, peak heart rate and more. As an example, measuring how the heart rate recovers after exercising is an indicator of cardiovascular health.

Data usage

The information collected from participants in Tromsø 7 is extensive. All study participants completed comprehensive questionnaires, sharing information about their lifestyle, general health, medical history, drug usage, diet, alcohol consumption, smoking, education, social status, PA and other areas. Measurements of height, weight, hip-waist, blood pressure, heart rate, SpO_2 and pain sensitivity were also collected. Biological samples of blood, saliva, urine, feces and samples from the nose/throat were also taken. A subsample underwent dental examinations, ECG-recording, cognitive tests, physical function tests, carotid ultrasound, eye tests, lung function tests, measurements of bone densitometry and body composition (by DEXA), echocardiography, and heart-, lung- and carotid auscultation. Many of the participants also attended earlier surveys in the Tromsø Study, which gives the possibility to analyze repeated measurements, enriching the data further.

Because the validity of the data collected by wearable devices is unknown or low for most data types related to PA, step count data will be collected first. Heart rate data for participants using heart rate monitors will also be collected since these sensors are considered accurate. Participants who already wore accelerometers in Tromsø 7 will be invited first. With these participants, existing accelerometer data can be compared with participant’s phone and wearable sensors collected in the same period. This will help determine how accurate step count and heart rate data collected from these sensors are. If accurate, all remaining participants in Tromsø 7 will be invited. This will potentially result in objective measurements for step counts for a proportion of the 21,000 participants in Tromsø 7.

Discussion

PA is an important lifestyle factor and is relevant in multiple projects in Tromsø 7, both as a predictor, endpoint and adjustment variable [18]. Data from accelerometers include heart rate, PA, sleep patterns, energy usage and sedentary behavior. However, only one week of PA data was collected using accelerometers. To get a broader picture of PA in a population, it is of interest to include historical PA data stored in Google and Apple cloud services. Data from accelerometers used in Tromsø 7 is considered accurate. This level of accuracy is, for the most part, not achievable through mobile and smart device sensors.

Participants will have a wide range of mobile devices, resulting in data from a wide range of different sensors. For those wearing a smart watch or bracelet with activity and pulse sensors, there will also be a lot of variety in how accurate these sensors are. Until data collection starts, the accuracy of the data is unknown. In addition, from participants that does not have a smart watch or activity bracelet, activity data will only be available for parts of the day. All these limitations are okay, if we know about them. The data should not be used alone, but rather as supplements to the data already collected in Tromsø 7.

Population based studies are valuable sources for new knowledge and for monitoring the status and development in population health and for disease risk factors. However, these types of studies are costly, and epidemiologists are concerned that participation rates are declining both in Norway and internationally. Hence, we need to motivate study participation and develop new tools for data collection. Smart phones come forward as promising data collection tools. They are prevalent, it is easy for study participants to donate their data, and large amounts of data can be collected from many study participants at a low cost.

Physical inactivity is an emerging and important disease risk factor in western populations. We aim to develop solutions to access data on PA from smartphones using internal and connected sensors, and to examine the strengths, possibilities and the potential of smart phones as data collection tools for future medical research on PA.

Future work

Collecting historical data is only the first step in including PA and heart rate data from mobile sensors in the Tromsø Study. The next step will be to collect the same kind of data for future activity. That is, to collect all PA and heart rate data for several months using mobile sensors and sensors in smart watches, bracelets and other smart devices. The experience from accessing historical data will be valuable going forward. After implementing the system described in this paper, we will expand the system design into a solution for collecting future data.

Conclusion

In this paper, we described our plans for collecting up to two years of historical PA and heart rate data from participants in the Tromsø 7 survey. Accessing this data is possible through Google Fit and Apple HealthKit cloud services. Access is restricted, and participants must consent to donate their data to the Tromsø Study database. The quality of the data is unknown and we therefore wish to examine it. The data will be a supplement to existing PA data from accelerometers and questionnaires in Tromsø 7. In order to validate these sensors as potential new tools for objectively measure PA, we will compare data collected from mobile sensors with existing accelerometer data.

References

[1] Apple Inc., HealthKit API Reference, in, 2016.
[2] Apple Inc., Introduction to HealthKit, in, 2016.
[3] E.E. Dooley, N.M. Golaszewski, and J.B. Bartholomew, Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices, *JMIR Mhealth Uhealth* **5** (2017), e34.
[4] A. Emaus, J. Degerstrøm, T. Wilsgaard, B.H. Hansen, C.M. Dieli-Conwright, A.-S. Furberg, S.A. Pettersen, L.B. Andersen, A.E. Eggen, L. Bernstein, and I. Thune, Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromsø study, *Scandinavian Journal of Public Health* **38** (2010), 105-118.
[5] K.R. Evenson, 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* **12** (2015), 159.
[6] Facebook Inc., React Native, in, 2016.
[7] Google Inc., Fit APIs, in, 2016.
[8] Google Inc., Fit History API, in, 2016.
[9] Google Inc., Fit SDK overview, in, 2016.
[10] B.K. Jacobsen, A.E. Eggen, E.B. Mathiesen, T. Wilsgaard, and I. Njølstad, Cohort profile: The Tromsø Study, *International Journal of Epidemiology* **41** (2012), 961-967.
[11] I. Njølstad, E.B. Mathiesen, H. Schirmer, and D.S. Thelle, The Tromsø study 1974–2016: 40 years of cardiovascular research, *Scandinavian Cardiovascular Journal* (2016), 1-6.
[12] R.E.R. Reid, J.A. Insogna, T.E. Carver, A.M. Comptour, N.A. Bewski, C. Sciortino, and R.E. Andersen, Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment, *Journal of Science and Medicine in Sport* (2016).
[13] S.E. Stahl, H.S. An, D.M. Dinkel, J.M. Noble, and J.M. Lee, How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough?, *BMJ Open Sport Exerc Med* **2** (2016), e000106.
[14] Statista, Percentage of people who use a smartphone in Norway from 2012 to 2016 in, 2016.
[15] Tek.no, Hele 800 000 nordmenn forventes å kjøpe et smartbånd i år, in: *Tek.no*, 2016.
[16] UiT - Department of Community Medicine, Eutro, in, 2011.
[17] UiT - Department of Community Medicine, Eutro, in, 2015.
[18] UiT - Department of Community Medicine, Hva forsker vi på?, in, 2016.
[19] M.P. Wallen, S.R. Gomersall, S.E. Keating, U. Wisloff, and J.S. Coombes, Accuracy of Heart Rate Watches: Implications for Weight Management, *PLoS One* **11** (2016), e0154420.

Address for correspondence

André Henriksen, Department of Community Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway.
E-mail: andre.henriksen@uit.no