Data collection and smart nudging to promote physical activity and a healthy lifestyle using wearable devices

Seshathiri Dhanasekaran[1], Anders Andersen[1], Randi Karlsen[1], Anne Håkansson[1], and André Henriksen[1]

[1]UiT The arctic university of Norway, Department of Computer Science, Tromsø, Norway, seshathiri.dhanasekaran@uit.no

Abstract

Nudge principles and techniques can motivate and improve personal health through emerging digital devices, such as activity trackers. Tracking people's health and well-being using such devices have earned widespread interest. These devices can continuously capture and analyze health-related data from individuals and communities in their everyday environment. Providing context-aware nudges can help individuals to self-manage and improve their health. In this study, we discuss how a consumer-based activity tracker can be used to track different variables for physical activity (PA) and how it has the potential to be an important source of data for future smart nudging.

Keywords

Digital nudge; physical activity; artificial intelligence; lifestyle; smartwatch

1 INTRODUCTION

A nudge was defined by Thaler and Sunstein in 2008 as "any aspect of the choice architecture that alters people's behaviour predictably without forbidding any options or significantly changing their economic incentives" [1]. Nudges were initially explored in offline decision-making, focusing mainly on personal health or wealth decisions.

The concept of *digital nudging* was introduced in 2016 when the notion of nudging was transferred to digital user interfaces. The term was coined by Weinmass et al. [2], who defined it as the "*use of user-interface design elements to guide people's choices or influence user's inputs in online decision environments.*" Others have later revised this, and Meske and Potthoff [3] further expanded the definition in 2017 to include free decision-making and defined digital nudging as "*a subtle form of using design, information, and interactive elements to guide user behaviour in digital environments, without restricting the individual's freedom of choice.*"

Further, in 2019, Karlsen and Andersen [4] defined *smart nudging* as "*digital nudging, where the guidance of user behaviour is tailored to the current situation of each user.*"

Smart nudging requires a user profile with a broad scope of connected data. The data is investigated before a customized nudge is structured. User acceptance of a customized nudge likely has a higher probability of succeeding (i.e., the user approves the nudge and adheres to the recommendation) than a non-customized nudge [4].

Current evolving digital technologies and devices, such as smartwatches, activity trackers, and smartphones, can provide continuous and long-term collection and analysis of behavioural data. This includes health-related data types like physical activity (PA), pulse, body temperature, stress levels, and sleep, and contextual data (e.g., location by global positioning systems [GPS]) from which other data types can be inferred (e.g., weather). These devices thus provide an opportunity for providing a continuous feedback loop to provide on-time nudges, to allow people to better self-manage their health and make informed choices [5].

From a public health perspective, by collecting this type of data from a large number of the population, these devices can potentially provide insights into different demographic groups, with access to near real-time data collection and evaluation at the population level, without being directly involved with the individual [6].

Recent developments in wearable sensor technology have taken such implementation closer to existence [7, 8]. Adding context-aware nudges to these digital health devices can help individuals identify and self-manage their PA levels to lead a healthy lifestyle.

The World Health Organization recommends that adults perform at least 150-300 minutes of moderate PA each week [9]. Globally, 25% of adults do not reach this goal [9]. Inactive people have a 20%-30% higher risk of death, and it is estimated that five million deaths could be avoided each year if people were more physically active [9]

Understanding an individual's context is complex, and smart device data must be precise to get valuable insights [10]. Natural human behavior consists of multiple simultaneous circumstances [11]. For example, people may be running indoors, outdoors, and on running tracks. The geographical location data can assist in identifying the user's context in this situation.

We propose a system that provides nudges related to the user's context and interests [4]. A system with context-aware data sources such as location data, time, weather, PA, and heart rate data, will help tailor nudges suitable to the user's context. For example, we could present a nudge to a user living close to a ski resort with a previous interest in skiing, suggesting when they can ski in the mountains using weather and location data for context and information about the activity levels for the user the previous days.

To create such a smart nudging system, we must first understand what type of data is available in current digital smart technology and explore how that data can be used.

Therefore, the objective of this study was to use PA data from consumer smartwatches to understand the context of the consumer and prepare for a future smart nudging solution for leading a healthy lifestyle.

2 RELATED WORKS

A pilot study by Haga et al. [11] was conducted in Canada to promote PA among college students with smartwatches. One hundred seventy-five students completed surveys on stress, diet, PA, and behaviors during weekends. The study did not test the effect of the intervention, so they could not assess the impact of these technologies on health and wellness outcomes [11].

A study by Cherubini et al. [10] aimed to find the correlation between tangible rewards and PA. The study reveals that tangible rewards and motivational messages decrease intrinsic motivation and thus their related PA. One of the primary outcomes of this work is that they learned that tangible rewards do not help establish lasting healthy routines. They observed no significant difference in the number of steps walked during the experiment, whether they offered participants money or not to perform this activity [10].

A study by Hafner et al. [12] suggests that incentivizing PA can lead to increased activity levels. In addition to the overall rise in the amount of activity monitored, the gain also appears to be correlated with higher levels of intensive activity over time, measured in the average number of days of intensive activity per month [12].

Finally, Mozgai et al. [13] proposed a mHealth application that provided optimized and interactive digital content through a mobile application. Novel adaptive logic algorithms used behavioral change techniques like a virtual human coach. The system would be aware of context and personal usage patterns with additional data sources, e.g., user mobility and calendar data.

3 METHOD

Recruitment

We used convenience sampling to recruit six volunteers. Participants had to be 18 years or older, owned and operated an Apple Watch (Apple Inc., CA, USA), and willing to share watch recorded PA data. Participants were recruited among people living in the northern part of Norway, north of the arctic circle.

Equipment

The first Apple Watch was released in 2015. Since then, eight generations have been released, and it is currently one of the most popular smartwatches available. In 2021, the Apple Watch had a 30% market share worldwide [14]. Depending on generation, the watch comes in several sizes (38-45mm), colors, metal finish, memory- and storage capacity, display quality, connectivity capabilities, and sensor support.

The fourth generation Apple Watch was released in 2018. It included a multitude of sensors, including a Global positioning system (GPS), tri-axial accelerometer, gyroscope, altimeter, compass, electrical heart sensors (ECG), and optical pulse sensor (i.e., photoplethysmograph).

Fuller et al. [15] concluded in a 2020 systematic review of smartwatch validation studies that the Apple Watch (with Samsung) had less measurement variability than other brands when estimating step counts and heart rate. The same study also concluded that although energy expenditure on average was overestimated by 58%, it was one of the most accurate brands. A recent paper by Kwon et al. [16] further concluded that the Apple Watch had a mean absolute percentage error (MAPE) of only 1% for moderate-to-vigorous PA and a MAPE of 4% for activity energy expenditure. A MAPE below 10% is generally used as a threshold for an acceptable MAPE when comparing smartwatch data in free-living conditions [17, 18].

Data extraction and variables

We asked participants to export their Apple Watch collected data from the Apple Health web solution to access the data. The exported data contains daily data on PA, workout details, and device information encoded as one Extensible Markup Language (XML) file per participant. The collected data has aggregated the steps count of a person from mobile phones and smartwatches. We excluded mobile phone steps and only analyzed step data retrieved from the smartwatch.

We extracted relevant variables from the source data using *Python* 3.9 [19] with *NumPy* 1.21 [20]. We only extracted data between January 2020 and December 2021. We stored variables in comma-separated value (CSV) files, divided into daily variables for *Active energy, Exercise minutes*, *Steps,* and *Distance*.

Active energy is understood as PA energy expenditure, given as Kcal. *Exercise minutes* are considered minutes of light-, moderate-, and vigorous PA combined. *Distance* is the sum of walking- and running distance, given in kilometers. *Steps* are the sum of daily steps counted on the smartwatch device.

Additional variables exist, including Heart rate, heart rate variability (HRV), oxygen saturation (SpO2), peak oxygen uptake (VO2Max), stand minutes, and sleep data. These additional variables were not used in the current study but can be relevant for data analysis in future smart nudging.

Statistical analysis

We used *Python* 3.9 [19] with *Pandas* 1.4.2 [21, 22] and *Matplotlib* 3.5.2 [23] packages to visualize and analyze the PA data for the following variables: *Active energy, Exercise minutes*, *Steps,* and *Distance*.

For each combination of selected variables, we calculated Pearson's correlation coefficients. The strength of the associations is assessed using cutoffs suggested by Evans [24], i.e., very weak, less than 0.2; weak, 0.2-0.4; moderate, 0.4-0.6; strong, 0.6-0.8; and very strong, greater than 0.8. Correlations are presented as scatter plots. Value distributions for each variable are shown as bar plots. Pearson correlations and variable distributions are combined and presented in a *pair* plot. Pearson's correlations are further explained as a heat map.

For Steps and Exercise Minutes, we created *time series* plots for mean daily values to evaluate change in these variables throughout the two years for all the participants.

4 RESULTS

Participant characteristics

The mean age (in 2022) for included participants was 36.3 years (SD=11.7). We collected two years of PA data for each participant, from January 2020 to December 2021. Included participants used the 4th generation Apple Watch or newer. All participants signed informed consent.

Variable correlation and value distribution

Here we present how the four selected variables correlate. (i.e., Active energy [kcal], Exercise minutes [light-, moderate-, and vigorous PA], Steps, and Distance [km]).

Figure 1 gives a *pair* plot showing the Pearson's correlation coefficient between each variable. We used the mean value of every month's activity of all participants as data for the variables. The value distribution for each variable is also provided, shown as natural divided time in months over two years. A corresponding heat map with correlation values is given in Figure 2.
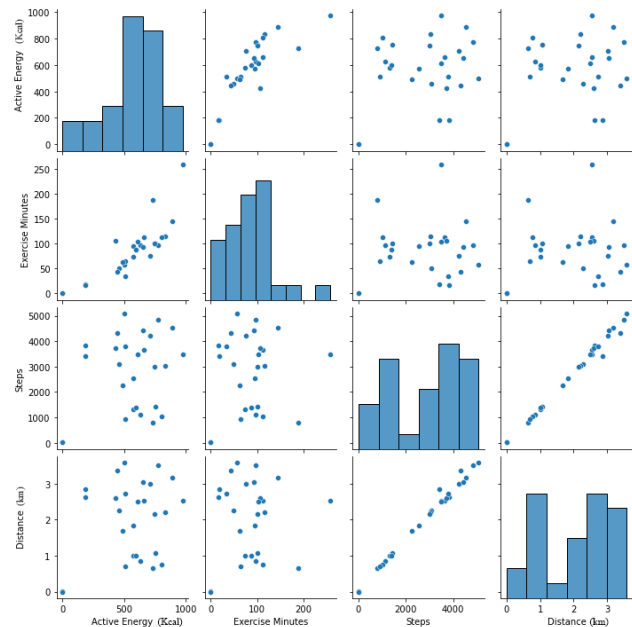
Figure 1. Pair plot with correlations and value distribution for each variable under study.
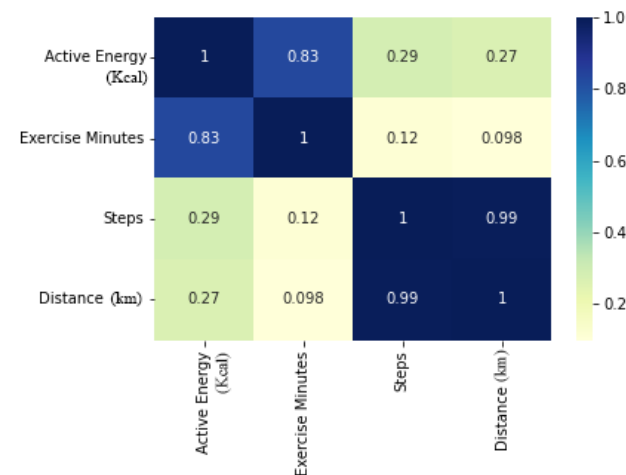

Figure 2. Correlation heat map for each variable under study.

In the pair plot (Figure 1), we checked the correlation between the variables with monthly mean value as data. Each scatter point denotes a monthly mean data of the variables.

Steps have a very strong correlation with Distance (0.99), a very weak correlation with Exercise Minutes (0.12), and a weak correlation with Active Energy (0.29). Exercise Minutes have a very weak correlation with Distance (0.098) and a very strong correlation with Active Energy (0.83). Active Energy has a weak correlation with Distance (0.27).

Change in step counts

Figure 3 gives a time series visualization plot for *Step* data for 2020 and 2021, using mean value of all participants.
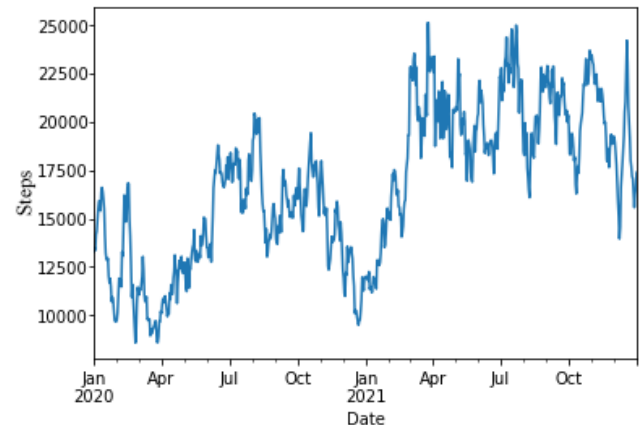

Figure 3. Change in daily mean steps between January 2020 and December 2021

The average step count data from 2020 and 2021 shows a higher step count between June and September compared to the remaining months. The lowest average daily Step count was found in March 2021 (2890 steps). The highest average daily Step count was in April 2022 (13750 steps).

The Step count data for 2021 shows a lower level of recorded steps during the year's initial months and an increased number of steps from April to December, especially between April and August.

Change in Excercise Minutes

Figure 4 gives a time series visualization plot for the *Exercise Minutes* data for 2020 and 2021, using mean value of all participants.
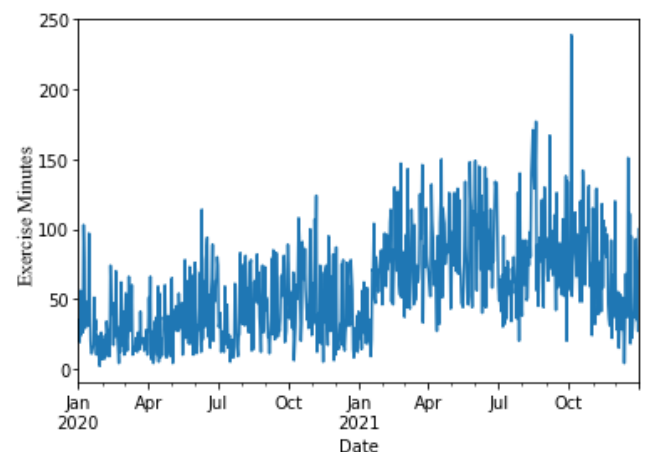
Figure 4. Mean Exercise Minutes for the two years

The mean *Exercise Minute*s was lowest from March 12, 2020, to April 10, 2020; this time period was also the Norwegian COVID-19 lockdown period. The Exercise Minutes did not diverge much from the yearly mean during the COVID-19 lockdown period. From December 2020 to January 2021, participants' Exercise Minutes were reduced.

5 DISCUSSION

In this study we collected and analyzed PA data for 2020 and 2021 from six participants wearing an Apple Watch, recruited in the Arctic region. We analyzed different variables for PA collected by the watch.

The correlation plots show that only Distance and Step count, *Exercise Minutes* and *Activity Energy* are linear dependent. Remaining combination of variables does not have a clear linear dependency.

We observed a decline in S*teps* and *Exercise Minutes* during the COVID-19 locktown period (from March 2020). After June 2020, people seamed to find ways to be active. A potential reason for the reduction of performed physical activity in January and February 2021, may be that these are the two coldest months of the winter season. Furthermore, although not a lockdown, there were COVID-19 restriction during this period.

A sharp increase in people's *Exercise Minutes* after the COVID-19 restricitons were lifted, i.e., after September 25, 2021, was also observed. However, three weeks later, *Exercise Minutes* had returned to normal levels (i.e., before September 25).

Physical activity findings from the present study is in accordance with a previous study by Henriksen et al. [6], where two years of PA data were collected using consumer-based activity trackers. Results showed that change is PA due to the Norwegian COVID-19 lockdown was distinct and clearly detected.

The Tromsø study, the longest-running population study in Norway [25] was initialized in 1974 with the overarching goal of combating the cardiovascular disease epidemic in the northern par part of Norway. The seventh survey of The Tromsø study saw a steady decrease in cardiovascular risk variables while obesity continued to rise [26]. In this study, and similar population studies, data on physical activity are traditionally collected using questionnaires and research-based accelerometers. The present study suggests future population studies could investigate new technologies and tactics to improve data gathering efficiency and reduce participant burden. Smartwatches and activity trackers could be investigated as a potential source of PA data, adding to already existing methods for PA data collection.

The advantage of the data collection method in the present study, is that it shows that Exercise Minutes (and other variables) can be recorded with a smart watch, and that daily levels of PA can be measured. However, evaluating PA using *Steps* and *Activity Energy* alone is not sufficient.

Depending on which activity tracker a participant wears, a multitude of other variables can be extracted. For the Apple Watch, in addition to PA, various sleep and heart rate variables exists, including: pulse, heart rate variability (HRV), oxygen saturation (SpO2), peak oxygen uptake (VO2Max), Electrocardiograms (ECG), time in bed, average sleep minutes, and more.

For participants already owning an activity tracker, data can be collected without adding much burden to participants.

Conclusions

Understanding the context of the user is critical for smart nudging. But, we faced the toughest pandemic and lockdown in the form of COVID- 19 effect on our data [26].

People found new ways of performing workouts and physical activity during COVID-19, and people are interested in following the same even after removing the restrictions due to their constraints and comfort.

We are working on data analysis techniques to predict the activity type with the impact of context to assess how context-aware nudges can work better for them. We also plan to evaluate the effect of weather conditions and calendar data on physical activity in the future to tailor smart nudges for the users. Input from smartwatch data is an essential source for future smart nudging in this context. We will use this study and methods to analyze the context impact on physical activity and increase physical activity with context-awareness.

6 SUMMARY

To understand users' context and PAs in the Arctic regions, we collected and analyzed smartwatch activity for 2020 and 2021. We enrolled six participants. Our finding indicates that PA patterns changed due to COVID-19, albeit temporarily. Despite governmental restrictions, participants found other means of performing PA, even though they could not leave their house as much as usual. This study focuses on understanding the context of the users to assess and prepare nudges for healthy lifestyle patterns.

7 REFERENCES

[1] Thaler, R.H. and Sunstein, C.R. (2009), *Nudge: Improving decisions about health, wealth, and happiness* (Penguin Books).

[2] Weinmann, M., Schneider, C., and vom Brocke, J., "Digital Nudging", *SSRN Electronic Journal*, 2015

[3] Meske, C. and Potthoff, T. (2017), 'The DINU-Model - a process model for the design of nudges', *25th European conference on information systems (ECIS)* (Guimarães, Portugal: Research-in-Progress Papers), 2587-97.

[4] Karlsen, R. and Andersen, A., "Recommendations with a Nudge", *Technologies,* 7 (2), 2019

[5] Shaw, R.J., Bonnet, J.P., Modarai, F., George, A., and Shahsahebi, M., "Mobile health technology for personalized primary care medicine", *Am J Med,* 128 (6), 555-7, 2015

[6] Henriksen, A., Johannessen, E., Hartvigsen, G., Grimsgaard, S., and Hopstock, L.A., "Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic: Development and Usability Study", *JMIR Public Health Surveill,* 7 (4), e23806, 2021

[7] Meske, C., Amojo, I., Poncette, A.-S., and Balzer, F. (2019), 'The Potential Role of Digital Nudging in the Digital Transformation of the Healthcare Industry',

Design, User Experience, and Usability. Application Domains (Springer International Publishing).

[8] Reeder, B. and David, A., "Health at hand: A systematic review of smart watch uses for health and wellness", *J Biomed Inform,* 63, 269-76, 2016

[9] World Health Organizaiton 'Physical activity', <https://www.who.int/news-room/fact-sheets/detail/physical-activity>, accessed 2022-01-20.

[10] Cherubini, M., Villalobos-Zuñiga, G., Boldi, M.-O., and Bonazzi, R., "The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk", *ACM Transactions on Computer-Human Interaction,* 27 (2), 1-44, 2020

[11] Haga, S.B., Shaw, R., Kneifel, C., Bond, S.J., and Ginsburg, G.S., "Promoting Wellness Through Mobile Health Technology in a College Student Population: Protocol Development and Pilot Study", *JMIR Res Protoc,* 9 (4), e16474, 2020

[12] Hafner, M., Pollard, J., and Van Stolk, C., "Incentives and Physical Activity: An Assessment of the Association Between Vitality's Active Rewards with Apple Watch Benefit and Sustained Physical Activity Improvements", *Rand Health Q,* 9 (1), 4, 2020

[13] Mozgai, S., Hartholt, A., and Rizzo, A., "An Adaptive Agent-Based Interface for Personalized Health Interventions", *IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, 118-19, 2020

[14] GmbH, S. (2022), 'Smartwatch market share worldwide from 2020 to 2021, by vendor', <https://www.statista.com/statistics/1296818/smartwatch-market-share/>, accessed 05-23.

[15] Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M.A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L., and Taylor, N.G.A., "Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review", *JMIR Mhealth Uhealth,* 8 (9), e18694, 2020

[16] Kwon, S., Kim, Y., Bai, Y., Burns, R.D., Brusseau, T.A., and Byun, W., "Validation of the Apple Watch for Estimating Moderate-to-Vigorous Physical Activity and Activity Energy Expenditure in School-Aged Children", *Sensors (Basel),* 21 (19), 2021

[17] Nelson, M.B., Kaminsky, L.A., Dickin, D.C., and Montoye, A.H., "Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types", *Medicine & Science in Sports & Exercise,* 48 (8), 1619-28, 2016

[18] Roos, L., Taube, W., Beeler, N., and Wyss, T., "Validity of sports watches when estimating energy expenditure during running", *BMC Sports Sci Med Rehabil,* 9, 22, 2017

[19] Python Software Foundation (2022), 'Python', (3.9 edn.: Guido van Rossum).

[20] The NumPy development team (2022), 'NumPy', (1.21.0 edn.: Sebastian Berg, Ralf Gommers, Charles Harris, Stephan Hoyer, Melissa Weber Mendonça, Inessa Pawson, Matti Picus, Stéfan van der Walt, Eric Wieser).

[21] McKinney, W. (2010), 'Data structures for statistical computing in Python', in Stefan van der Walt and Jarrod Millman (ed.), *Proceedings of the 9th Python in Science Conference*, 56-61.

[22] The pandas development team (2020), 'pandas-dev/pandas: Pandas', (1.4.2 edn.: Zenodo).

[23] The matplotlib development team (2022), 'Matplotlib', (3.5.2 edn.: John Hunter, Darren Dale, Eric Firing, Michael Droettboom).

[24] Evans, J.D. (1996), *Straightforward statistics for the behavioral sciences.* (Brooks/Cole Publishing).

[25] Jacobsen, B.K., Eggen, A.E., Mathiesen, E.B., Wilsgaard, T., and Njolstad, I., "Cohort profile: the Tromso Study", *Int J Epidemiol,* 41 (4), 961-7, 2012

[26] Hopstock, L.A., Grimsgaard, S., Johansen, H., Kanstad, K., Wilsgaard, T., and Eggen, A.E., "The seventh survey of the Tromso Study (Tromso7) 2015-2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey", *Scand J Public Health*, 14034948221092294, 2022

8 ACKNOWLEDGEMENT