Original Paper

Smartphone pervasive sensing of physical activity of overweight adults in a long-running randomized controlled trial

Simon Kamronn, Lars Kai Hansen, Jakob Eg Larsen
Cognitive Systems, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark

Corresponding Author:
Simon Kamronn
Cognitive Systems
Department of Applied Mathematics and Computer Science
Technical University of Denmark
Kgs. Lyngby
Denmark
Phone: +45 40929792
Email: simon@kamronn.com

Abstract

Background: Clinical trials are expensive why it should be a priority to acquire as much data as possible during the trial. The burden on participants and staff is often the limiting factor on the amount of data feasibly acquired, which is why trials may benefit from incorporating the readily available small sensor-packed ubiquitous device; the smartphone.

Objective: The aim of this study was to assess whether a smartphone can assist or replace existing practices in evaluating a physical activity intervention study in overweight sedentary adults.

Methods: We introduce the smartphone as an additional sensing device in a physical activity intervention study that investigates the effects of active commuting and leisure-time exercise on a range of biological measures.

Results: We find the smartphone is able to measure multiple modalities ubiquitously over a long duration and a hierarchical Bayesian analysis reveal estimates that are well in line with an independent analysis of the biological measures.

Conclusions: The smartphone has in this study shown that it, while not being without limitations, is able to augment current research methodologies and add value in historically infeasible ways. We can now ask questions that factorizes temporally on a minute-scale resolution and conceptually over the domains of everyday life.

Trial Registration: Clinicaltrials.gov NCT01962259

Keywords: mobile health, rct, smartphone, physical activity

Introduction

With more than half of the western population being overweight and nearly a quarter obese [1], the obesity pandemic and the associated non-communicable diseases [2] has become one of modern times biggest challenges and is only getting worse. Behavioural aspects of managing body weight by being physically active are largely not understood [3], and sometimes even counter-intuitive [4], why more research using the latest available technology is needed [5].

Before trying to manage body weight we must first monitor and understand the underlying dynamics. In this study we ask what effect active commuting (bicycle) and leisure-time activities have on physical activity in everyday life, when not exercising. Will the extra bouts of exercise be absorbed and decrease activity in other domains, maintaining status quo, or will they inspire and motivate more activity?

Conducting a randomized controlled trial (RCT) is typically expensive but also necessary if we want to understand systems in which causality is important. It is expensive primarily due to the excessive time spent on compliance and
acquiring data (clinical sampling) and even though the sampling is time consuming, it is often only possible to shed light on the big picture questions. Pervasive monitoring can take the analysis to more detailed levels and explore novel directions [6]. Instead of viewing a three months period as one sample, activity tracking using smartphones can break that period into samples of one minute or less. Quantifying energy expenditure in terms of domains, places, and type of activity is simply not feasible in a traditional long-running study but the sensors of smartphones allow us to query these different directions and may provide insights otherwise missed.

ActiGraph (ActiGraph GT3x, ActiGraph Corp, Florida, USA), the de-facto standard of monitoring physical activity, has established its usefulness by being widely adopted and validated, but at least three aspects are in its disfavor. It uses an obsolete and unknown algorithm based on how mechanical activity trackers worked once and by being static is not able to adapt to different activities, e.g. walking or running, or different groups of people, e.g. children, adults, and handicapped, which would enable much more accurate estimates [7]. Secondly, conducting a large scale, long-term study quickly becomes cumbersome as the price of acquisition and the personal burden on participants of carrying and maintaining an extra device can be limiting factors in studies. Lastly, and this is for many monitoring methods: the psychological effects of being in a study, and in particular being reminded continually from a device on the hip, can have significant impact on the behaviour that is being studied [5]. Simply by carrying a pedometer, it was found in a systematic review [8], people increase their physical activity by 26.9%.

Given the dominance of ActiGraph and the many previously established health outcomes based on the device [9,10], [11] investigated to what extent Android smartphone based activity sensing correlated with ActiGraph in both laboratory and free-living conditions. They found that physical activity estimates from the smartphones were highly correlated with estimates from ActiGraph and with one another, and that it mattered little in what position the smartphone was carried. Using a smartphone has several advantages, beyond avoiding to carry an additional device. The vast majority already owns a smartphone which makes it a cheap and easy way to deploy scalable, real time, and truly pervasive studies, if it reliably can estimate energy expenditure and activity. It further has the ability to provide feedback to participants [4,12,13], which may be an important component in fighting obesity after tracking and understanding it, hereby closing the loop and hopefully guiding us towards more healthy choices.

We apply the smartphone as a monitoring device in a physical activity intervention study spanning six months with the primary aim of investigating the effects of active commuting and leisure-time activities on physically inactive obese participants. Will we be able to measure the effects of interventions based on smartphone sensing data and will they be in accordance with the clinical results?

Methods

Participants

Caucasian, overweight, and physically inactive healthy adults were recruited in the Copenhagen area from November 2013 to October 2015. Inclusion criteria included age of 20 to 45, body mass index (BMI) from 25 to 35 kg/m^2, and a self-reported physically inactive lifestyle. The full list of of criteria are available in [14].

Study design

The study design is described in full elsewhere [14] but most importantly it was designed as a randomized controlled trial to test the effectiveness of three interventions on physical activity and one control group. They were prescribed a daily exercise energy expenditure of 320 kcal for women and 420 kcal for men. The four groups are:

**BIKE**: bicycle commuting to work  
**MOD**: leisure time moderate activity (50% of VO\(_2\)max)  
**VIG**: leisure time vigorous activity (70% of VO\(_2\)max)  
**CON**: The control group was asked to maintain the same lifestyle as prior to the study.

Clinical sampling

Participants were asked to attend three different test days at baseline, 3 months, and 6 months where a range of physiological tests were conducted and biological samples acquired. Immediately after the test-days a free-living assessment period (7-14 days) followed in which they were monitored on physical activity (ActiGraph GT3x, 7 days), sleep, diet, and daily energy expenditure through doubly labelled water [15]. They were furthermore required to wear a Polar heart rate monitor (RC3 GPS, Polar Electro Oy, Kempele, Finland) every time they exercised and weigh...
themselves every day of the study using a wireless internet-connected bodyweight (WiThings Body Composition WiFi scale, WiThings Europe, Issy-les-Moulineaux, France).

**Pervasive sampling**

A mobile sensing application developed for a previous smartphone sensing study [16] was adopted and configured for use in the present study. The application was configured to acquire acceleration (50Hz sampling frequency), GPS-based location (6-minute sampling interval), recognized activity through the Google Activity Recognition API, step count, and screen on/off status.

**Implementation details**

To accommodate the high sampling rate of the accelerometer without reducing the battery-life of the smartphone, samples are 'batched', i.e., queued in hardware before read and saved to the storage. At the time of the study only two smartphones supported batching, LG Nexus 5 and Samsung S5. The former was chosen as it was successfully used previously [16] but during the enrollment period it was unfortunately decommissioned and Samsung S5 was used hereafter. A total of 19 LG and 11 Samsung smartphones are used in the cohort that completed the experiment.

**Missing data**

In an experiment with limited control some level of missing data must be expected. In the present study we also experienced significant missing data. The observed patterns of missingness are not always "random" or explainable. For example, during night-time the smartphones may reduce the sampling of the accelerometer, such loss of data may increase variance but not introduce a bias, as the effects can be modeled. In addition data loss occurred during the day with no apparent pattern. In some cases all the modalities are lost, indicating a general failure to record, send, or receive the data, and at other times it is a single modality indicating a failure on the smartphone. If such losses occur without dependence of the outcome variables it can again lead to increased variance but should not introduce a bias.

**Pre-processing**

**Activity level calculation**

The primary outcome is the level of physical activity but since that is not a well-defined measure, especially using smartphone sensors, *activity counts* are used to make the conclusions comparable with ActiGraph results. Activity counts are a measure of physical activity within a time span, or more precisely, a motion summary count calculated from the area under the accelerometer curve [7]. In practice this means that the signal is bandpass filtered with a range that preserves as much human motion related acceleration as possible while discarding noise, which is around 0.25 - 5Hz. The absolute values are then summed in bins of 1 second. This leaves activity counts for each dimension but can without loss of information be reduced to the vector magnitude count (VMC) which is the euclidean distance spanned by the count vector.

**Domain inference**

Spatial location data is complex and noisy so to reduce complexity we use a simple behavioral cluster representation [17]. We identify 4 different *domains* for each participant, namely: Home, Leisure, Travelling, and Unknown. Home is manually identified by having the most samples associated and typically active during the evening/night. Travelling is identified as being en route between two locations with a sufficiently high average speed. Leisure is identified as samples that are not Home, Work or Travelling. We also wanted to identify when participants were at work but it turned out to be impossible for more than half of the participants due to either too much missing data, unemployment, or a non-stationary work such as postman. For this reason the locations identified as work has been merged into the Unknown category.

**Aggregation**

The accelerometer samples at 50Hz and therefore gathers a very large amount of data in six months. To enable statistical analysis all modalities are summarized in 5 minute bins. Continuous data such as the activity count is summed and with categorical data the most frequent category within a bin is used.
Statistical Analysis

A long running experiment conducted in the real world in which the participants are left to their own devices will naturally include many confounding factors which are mostly unobservable. Due to the randomized treatment assignment we can theoretically ignore all covariates but may be able to detect a weaker signal by including them [18]. Including all covariates and interactions will provide the most flexible model [19] but also a model that is often over-parameterized and non-identifiable why covariates should be chosen carefully [20]. Especially when the number of participants is low or covariates are correlated with treatment assignment. In model-based analysis of randomized controlled trials the response variable can be approximated with a hierarchical Bayesian model

\[ y_i = \mu + \mu_{s_i}^{pret} + \tau_{t_i} + \alpha_{g_i} + \beta X_i + \epsilon_i, \]

where \( y_i \) is the response, \( \mu \) is the intercept, \( \mu_{s_i}^{pret} \) is the subject-specific pre-treatment mean, \( \tau \) is the treatment assignment, \( \alpha \) are random covariates, \( \beta \) are regression coefficients, \( X \) are the linear regressors, and \( \epsilon_i \sim N(0, \sigma^2) \).

The \( g[\cdot] \) notation specifies which group, of a particular covariate, that sample \( i \) is from. \( s[\cdot] \) denotes participants and \( t[\cdot] \) denotes treatment. \( \mu_{s_i}^{pret} \) and the covariates are sum-to-zero contrasted which ensures the model is identifiable and the posterior distribution of \( \tau \) is thus a measure of the treatment effect. This is a hierarchical Bayesian model when proper prior distributions are placed on the parameters [21] and otherwise known as a linear mixed model.

Varying effects

It may be overreaching to assume the treatment effects are constant over the duration of the experiment, and certainly after the experiment is finished, so the treatment effects have been allowed to vary with the natural periods of the study. The intervention is divided into 4 periods; pre-treatment, 0 to 3 months, 3 to 6 months, and after 6 months. The pre-treatment period is already modelled in the pre-treatment mean \( \mu_{s_i}^{pret} \) and thus not modelled in the treatment effect. It is further assumed that the effect varies with domain (Home, Leisure, Travelling, Unknown). We thus have a treatment effect partitioned into 4 groups, 3 periods, and 4 domains.

Controlling covariates

To control for some of the variability caused by observed baseline variables they are included in the model in the term \( \alpha_{g_i} \), also called varying intercepts. Seasonal, weekly, and daily variation are modelled by including the month, weekday, and hour of the day as covariates, i.e. for month a 12-dimensional varying intercept. Each of these have a normally distributed zero-centered prior with standard deviation of 1. We further control for gender (male, female), relationship status (single, cohabiting), education (no college, college, grad school), and employment status (employed, student). In the analysis of the smartphone data the type of smartphone is also added as varying intercept. These have a normally distributed zero-centered prior with standard deviation of 0.01.

Linear regressors

In the linear term \( \beta \) we include age, body mass index (BMI), maximal oxygen uptake (VO\textsubscript{2}\text{max}), temperature and rainfall in the vicinity and within the hour of a sample, and the number of sick/off-days they have from the study in a given period. All the coefficients have a normally distributed zero-centered prior with standard deviation of 0.005.

Missing data

Because the activity level data is originally on the scale of a second and then summed over a fixed bin size \( N \), if any samples in a bin is missing we get that

\[ y_i = \sum_{n=1}^{N_{obs}} y_{i,n}^{obs} + \sum_{n=1}^{N_{mis}} y_{i,n}^{mis}. \]

One way of dealing with the missing samples is treating them as latent variables, or model parameters, but if the bin size is small enough or data is missing at random, we can assume equal observations within the bin, i.e. \( y_i = N y_i^{obs} \) for any \( n \in [1, \ldots, N_{obs}] \), and thus assume...
\[ y_i^{\text{obs}} = \frac{N^{\text{obs}}}{N} y_i, \]

which scales the regression so we calculate an approximately correct likelihood. To relax this assumption the scaling is itself being scaled with a parameter \( \lambda_{s[i]} \) for each participant

\[ y_i^{\text{obs}} = \lambda_{s[i]} \frac{N^{\text{obs}}}{N} y_i, \]

which enables the model to learn if a particular participant has a behaviour that warrants an even lower scaling.

**Completely missing data**

Often we have no data within the summarized bin and so have to treat the sample as missing to ensure a proper posterior distribution. In these cases we treat the missing sample as a latent variable so that \( y_i^{\text{mis}} \sim N(0, \sigma_{\text{mis}}^2) \). If the amount of observed samples in a bin is less than 10% it is treated as unobserved as well.

**Non-compliance**

Not carrying the smartphone, especially while exercising, is a situation of non-compliance and unfortunately often observed in the data. We are able to estimate the compliance at few intervals due to multiple data sources but often we are not. As we observe many instances of non-compliance during the prescribed exercises, which we know due to Polar HR data, all periods containing exercises have been removed from the data and the samples are treated as unobserved, following standard practice [22].

**Inference**

The model is implemented in the probabilistic programming framework Edward [23] because it enables fast inference even when scaling to millions of samples. Relying on variational inference may be less accurate than sampling methods such as Markov Chain Monte Carlo (MCMC). To test this MCMC inference using Stan [24] was carried out for a subset of the data, which showed very similar results to Edward for the same data. To select model architecture, i.e. which covariates to include and how to manage missing data, the model with the highest log-likelihood was selected.

**Results**

Of the 47 recruits receiving a smartphone, 30 participants completed the study. The distribution of participants among the interventions are: 6 CON, 7 BIKE, 12 MOD, and 5 VIG. The mean age of participants was 36 years (range: 22–45), with 19 females (63%). Most were cohabiting (n=22, 73%), in employment (n=27, 90%), had no college degree (n=17, 57%), and only 2 had gone to grad school (7%). After summarization we have a total of 2,311,151 samples, or 8024 days. For our outcome variable we have an observation rate of 30.5% and the other modalities are adaptively sampled, i.e. without a fixed sampling rate. The total number of days for each participant in the study is capped at 300 from the first day, including baseline, and the average number of days in the study is 275 with a standard deviation of 27.
Posterior estimate of each treatment relative to CON in period 1 (baseline to 3 months), 2 (3–6 months), and 3 (after 6 months) of the study. The whiskers indicate 95% interval. For smartphone data we see a positive effect for all of the interventions in the first period and then a decline in the second. For the ActiGraph data we see a more varied result with positive effects in the second period as well.

Effects of interventions

Treatment effects measured by non-exercise activity of each intervention contrasted to the control group for each of the two periods in the study measured by the two devices are shown in Figure 1. As a participant-specific pre-treatment intercept is also modelled, the effects are in contrast to the estimated baselines as well. In the effects measured by the smartphone we see a pattern in which it appears that the participants initially are motivated to be more active but after 3 months the activity declines, not unlike similar studies [4]. However, even though the effects for MOD and VIG in the second period are not significantly different from CON, they are neither significantly different from the effects in period one, so while we do observe a relative decrease in activity with respect to CON, we can not be sure of an absolute decrease of each group. BIKE is the group with the most significant drop in activity with both period two and three being less active than CON and we see the same for the third period of MOD. The third period for VIG, however, is the most active of them all, a result currently with no obvious explanation as it is an uncontrolled period without data from any other modality.

In the effects measured by ActiGraph the effects for the first period at 3 months is positive but becomes negative in the second period for BIKE, following the trend observed from the smartphone. MOD and VIG, however, becomes more positive and stays the same, respectively, which is almost the opposite of what we see in the smartphone data.

A comparison to loss of fat

In [25] the fat mass of the same participants as in the present study was measured with dual-energy X-ray absorptiometry. Here a significant loss of fat mass was found in all three groups compared to CON in the first period (baseline to 3 months), and a sustained fat mass in the second period (from 3 to 6 months), for BIKE and MOD but further decreased for VIG. If we compare the loss of fat mass to the physical activity effect (Figure 1) measured by the smartphone, those results are highly correlated and make sense if we assume that an increase in non-exercise activity leads to loss of fat mass. If we were to include activity measured during prescribed training exercises, the measured effect would be even greater and probably better correlated to loss of fat mass. This suggests that smartphones are able to predict the loss of fat mass.
Figure 2: Effect of domains between groups

Treatment effect summed over the periods. The posterior estimates are not contrasted to CON (as in Figure 1) and therefore not directly interpretable as a treatment effect why no y-axis is shown but they can be used for comparison between domains and interventions. For the smartphone we see e.g. MOD and VIG are more active during Leisure, and that Travelling and Unknown generally elicit higher activity.

Figure 3: Effect of domains over time

Treatment effect summed over the interventions. In the smartphone data we see a general increase of activity in the second period followed by a decrease after the end of the study in the third period. The ActiGraph data shows a change in which domain the activity is high between the periods.

Effects of domain and time

In Figure 2 we see significant differences in the amount of physical activity spent in the different domains even though the samples are distributed relatively evenly with about a third in the Unknown domain and an equal share in the three others. The Unknown domain is unfortunately also the most active across all interventions. Secondary is Travelling which suggests the participants regularly walk between destinations, e.g. to the grocery store, as motorized transportation typically elicit a low accelerometer response. The groups CON and BIKE are more active at Home than away whereas MOD and VIG are inactive at Home and VIG is very active in Leisure. The exercises are excluded from the data and can therefore not explain why VIG is so active during Leisure but some of the explanation might be that it has less Unknown samples than e.g. MOD. In the ActiGraph data the effect of domains are less pronounced, perhaps
because VIG in Leisure is so dominating, but apart from that the trend follows the smartphone data rather well. We see a bit more activity at Home in MOD and VIG which might be due to participants not carrying their phones.

If Figure 3 the effect of domains is shown for each period where the third period is after the 6 months when the study is completed but participants have continued collecting data on their smartphones for a time up to 300 days post commencement. It appears the total level of physical activity increases from the first to the second period which naively contradicts the decrease of treatment effect (Figure 1). The explanation is simply that while the treatment effect is in contrast to CON, the domain effect is a sum over all the groups and can therefore not be used to evaluate the effect of interventions. It is, however, interesting, and perhaps expected, that the level of activity decreases significantly in all domains after the study is completed. The results from the ActiGraph (Figure 3) show a quite different pattern in which it seems the Travelling domain has absorbed activity from both Leisure and Unknown from period 1 to 2.

**Why do the results differ between devices?**

Data from the ActiGraph is acquired at three times; at baseline, 3 months, and 6 months, and for a maximum of 7 days each time. This leads to less data and therefore less stable inference even though the posterior estimates are significant. In a free-living study there will be unobservable confounding factors and one might be the psychological effect of carrying an accelerometer on the hip. The ubiquitous nature of the smartphone leads in this case to more reliable estimates.

![Figure 4: Linear regressors](image)

Coefficients $\beta$ of the linear regressors with whiskers indicating 95% interval. A positive value indicates a positive correlation with, and therefore increase in, physical activity. Age is slightly negatively correlated, BMI, VO$_2$ max, temperature, rain and sick days have no effect and the number of days off from the study is also negatively correlated.
The varying intercepts with whiskers indicating 95% interval. All groups (e.g. gender or education) are sum-to-zero contrasted. Being a cohabiting male without a graduate school degree you have a higher probability of being physically active. It is uncertain if being employed or a student is associated with higher levels of activity.

**Characteristics**

External factors and personal characteristics may influence a measured response to a degree in which the effect of interest is either lost in noise or incorrectly estimated due to e.g. confounders. For this reason all relevant information of the participants and their surroundings are included in the model either as a linear predictor or a varying intercept.

**Linear predictors**

In Figure 4 estimates of the regression coefficients are illustrated in a horizontal boxplot with whiskers indicating 95% intervals. Most of the coefficients are either very small or insignificant but age is interestingly a predictor of less physical activity, which is a common finding [26,27]. The number of days off the study is the strongest predictor of less activity indicating that for whatever reason participation was hindered also limited their non-exercise activity. In the ActiGraph data VO\textsubscript{2}max is a small but positive contributor and we see a moderately strong contribution from the number of sick days a participant has had during a given period in the study. It is unsure if sick days leads to more exercise, which would be unexpected, or the activity during the days of wearing the ActiGraph happen to correlate with the number of sick days.

**Varying intercepts**

We observe large significant effects of most covariates when modelling the smartphone data with the one exception that it is uncertain whether having a job or being a student is important (Figure 5). Notably it seems men are more active than women, which has been previously observed [26], highly educated people are less active than lower educated, and participants with a partner are more active, perhaps due to motivational support. Covariate estimates from the ActiGraph data are quite similar with the exception of having a reversed relationship between employed and student participants but as the ratio between these groups are very skewed, this result should not be interpreted strongly.

Time-dependent intercepts show the expected behaviour of high activity during the day (Figure 6), more activity in the summer months (Figure 8), and less activity in the weekends compared to the weekdays (Figure 7).
Discussion

Principal findings

This is the first long-term randomized controlled trial to investigate the effects of active commuting and leisure-time exercise on physical activity in sedentary obese and overweight people using smartphones and ActiGraph. We find the smartphone is a resourceful platform for research but also that it has both strengths and limitations that needs to be better quantified.

Strengths

As a strength the smartphone may lower the bar to including multidimensional, longitudinal data streams in any study regardless of participant technological proficiency and furthermore have the potential to not only provide easier enrollment but also recruitment through platforms such as the ResearchKit (Apple, Cupertino, CA, USA), increasing reach and decreasing cost of marketing [28]. Selection bias may be lessened by using smartphones as the primary sensor since the vast majority owns a smartphone already and time-demanding activities such as spending multiple days on physiological examinations may no longer be needed, reducing the burden on participants. Reducing examinations can in turn allow researchers to spend more time on activities that improve retention. Clinicians will initially, and rightly, argue that high-variance data as we acquire from most smartphone sensors, are no substitution for low-variance momentary clinical measurements and depending on the endpoint that may be correct, but the sheer volume and variety of data made available through smartphones not only enables low-variance estimates, they allow researchers to explore new directions.

Limitations

In terms of limitations, a too short time-span was available to tailor the existing mobile sensing application before enrolling the first participant which also meant that insufficient time was available to conduct a pilot study to sufficiently test the application. In hindsight it would have been prudent to improve sampling consistency and server infrastructure to reduce the amount of missing data. Newer developments in this area with ResearchKit and ResearchStack [29] will undoubtedly improve much upon this issue and general adaptation. Having access to raw data is always preferred but sampling multidimensional streams at a high frequency introduces a multitude of potential issues with battery-life, storage, data-transfer, etc., that can be mitigated by edge-computing; carefully choosing some pre-processing to take place on the smartphone.

Previous research find strong correlations between activity estimates from smartphones and ActiGraph [11] why the observed discrepancies in the present study are likely due to the difference in the time-span covered by the observations and non-observable confounders. The relatively small sample size and psychological effects of carrying a physical activity tracker could also cause enough bias to throw off the analysis [5]. An explanation that is supported by the agreement of linear regressor and covariate estimates.

Conclusions

Pervasive sensing technologies show great potential in clinical research as either a substitution, enhancement, or addition to existing methodology. We find that the smartphone provides reliable estimates in correspondence with independently measured loss of fat mass. In the results we see a decrease in physical activity with time but we also see a shift in the distribution of activity over the domains with activity moving from Home to Leisure and that VIG in general has a much greater tendency to be active in Leisure than at Home. By combining multi-modal data it is possible to decompose measured effects into time, domain, and many other directions which may contain previously hidden insights to research questions.

Acknowledgements

The authors would like to thank the team behind GO-ACTIWE; Bente Stallknecht, Mads Rosenkilde, Jonas Salling Quist, Martin Bæk Petersen, and Anne Sofie Gram, for conducting the randomized controlled trial and collecting ActiGraph data. We would also like to thank TrygFonden for funding the project.
Conflict of Interest
None declared

Abbreviations
BMI: body mass index
RCT: randomized controlled trial

References


Time-dependent varying intercepts

**Figure 6:** Hourly variation
The effect of hour of the day on the activity level. The effects between devices are quite similar and follow a trend that is agreeable with previous results [27].

**Figure 7:** Daily variation
The effect of day of the week on the activity level. Somewhat similar variation between devices where most days are not significantly different from each other and the weekend seems to be for resting.

**Figure 8:** Monthly variation
The effect of month of the year on the activity level. We see little correspondence between the devices and note that the variations observed in the smartphone data are closer to what would be expected with better weather yielding higher activity.