Study Protocol

Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Study Protocol of the ALLY Micro-Randomized Trial

J.-N. Kramer[1], MSc; F. Künzler[2], MSc; V. Mishra[3]; B. Presset[4], MSc.; D. Kotz[3], PhD; S. Smith[5], PhD; U. Scholz[6], PhD; T. Kowatsch[1], PhD

[1]Center for Digital Health Interventions, Institute of Technology Management, University of St.Gallen (ITEM-HSG), St.Gallen, Switzerland
[2]Center for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland
[3]Department of Computer Science, Dartmouth College, Hanover, NH, USA
[4]Institute of Sports Studies, University of Lausanne, Lausanne, Switzerland
[5]Institute for Social Research, University of Michigan – Ann Arbor, Ann Arbor, MI, USA & Medical School, University of Michigan – Ann Arbor, Ann Arbor, MI, USA
[6]Department of Psychology, University of Zurich, Zurich, Switzerland

Corresponding Author

Tobias Kowatsch
Center for Digital Health Interventions
Institute of Technology Management
University of St.Gallen
Dufourstrasse 40a
9000 St.Gallen, Switzerland
+41 71 224 72 44
tobias.kowatsch@unisg.ch

Word count: 5070;

Abstract

Background: Smartphones enable the implementation of just-in-time adaptive interventions (JITAIs) that tailor the delivery of health interventions over time to user and time-varying context characteristics. Ideally, JITAIs include effective intervention components and delivery tailoring is based on effective moderators of intervention effects. Using machine learning techniques to infer each user's context from smartphone sensor data is a promising approach to further enhance tailoring.

Objective: The primary objective is to quantify main effects, interactions and moderators of three intervention components of a smartphone-based intervention for physical activity. The secondary objective is the exploration of participants' states of receptivity, i.e. situations in which participants are more likely to react to intervention notifications, through collection of smartphone sensor data.

Methods: In 2017, we developed the **A**ssistant to **L**ift your **L**evel of activit**Y** (Ally), a chatbot-based mHealth intervention for increasing physical activity that utilizes incentives, planning and self-monitoring prompts to help participants meet personalized step goals. We used a micro-randomized trial design to meet the study objectives. Insurees of large Swiss insurance company were invited to use the Ally app over a 12-day baseline and a six-week intervention period. Upon enrolment, participants were randomly allocated to either a financial incentive, a charity incentive or a no incentive condition. Over the course of the intervention period, participants were repeatedly randomized on a daily basis to either receive prompts that support self-monitoring or not, and on a weekly basis to receive one of two planning interventions or no planning. Participants completed a web-based questionnaire at baseline and post-intervention follow-up.

Results: Data collection finished in January 2018. In total, 274 insurees enrolled in the study and installed the Ally app on their smartphones. Main reasons for declining participation were having an incompatible smartphone (37/191; 19.4%) and collection of sensor data (35/191; 18.3%). Step data is available for 227/274 participants (82.8%) and smartphone sensor data is available for 247/274 participants (90.1%).

Conclusions: This study describes the evidence-based development of a JITAI for increasing physical activity. If components prove to be efficacious, they will be included in a revised version of the app that offers scalable promotion of physical activity at low cost.

Trial Registration: ClinicalTrials.gov NCT03384550

Keywords: physical activity; mHealth; JITAI; incentives; self-regulation; planning; state of receptivity

Introduction

Background

Mobile health and sensing technologies (mHealth) recently sparked excitement due to their capability to deliver large-scale personalized behavior change interventions at low cost [1] that can potentially reduce the disease burden associated with health behaviors, such as diet behavior, smoking or physical inactivity [2]. Beyond passive monitoring of health behavior, smartphones and wearables collect a wealth of contextual information (such as time, location, communication logs or physical activities) that can be used to tailor the delivery of interventions to participant states that increase the intervention's likelihood of success. For example, an intervention could only be delivered at points in time when the participant is likely to change her/his behavior (state of opportunity) or is likely to engage with the intervention content (state of receptivity) [3]. Mhealth applications that utilize this kind of dynamic tailoring are referred to as just-in-time adaptive interventions (JITAIs) [3].

During the development process of a JITAI, it is crucial to decide what key intervention components are needed to affect the desired intervention outcome and what information should be used to tailor the delivery of each component to participants over time [4]. The first question involves an empirical evaluation of single candidate intervention components. The second question involves identifying effective time-varying moderators that indicate in which situations the intervention component is or is not effective. Unfortunately, these decisions can hardly be informed by past research because traditional study designs (e.g. randomized controlled trials) rarely evaluate single intervention components or time-varying moderators of intervention effects. To facilitate the development of JITAIs, Klasnja and colleagues therefore proposed the micro-randomized trial (MRT) [5].

The goals of a MRT are to quantify proximal (short term) main effects of single intervention components, to investigate how these effects change over time and to identify which contextual variables moderate the effects of single intervention components. MRTs use repeated randomization of participants to different versions, or presence and absence, of individual intervention components over time, which enables estimation of time-averaged main effects of single intervention components on proximal outcomes, as well as time-varying effects and their contextual moderators. Results of an MRT can consequently inform the researcher, which components to include in an optimized version of the intervention and how to adapt the delivery of each intervention component in order to maximize effectiveness.

In this paper, we describe the rationale and design of a 6-week MRT that evaluates main effects and moderators of three different intervention components of Ally, a chatbot-based mHealth intervention promoting physical activity that is offered to insurees of a large Swiss health insurance company. We also report descriptive statistics from our recruiting process that was carried out exclusively online.

Although MRTs are designed to accommodate contextual moderation, context is likely to be multi-dimensional – e.g. not just time or location but rather the nexus of time and location (or other higher order interactions) define opportune moments for intervention. This limits the approach of investigating single variables as potential tailoring variables within MRTs. A potential way of overcoming this limitation is to train machine-learning models that classify the participant's latent 'states' of intervention receptivity or vulnerability given a vector of high-resolution smartphone sensor data. Research on interruptibility for example demonstrated that similar models successfully predict the quality and quantity of participants' reactions to notifications on their smartphone [6-8]. This approach allows to continuously model each participant's state of receptivity from a vast number of variables. In addition to evaluating candidate intervention components and time-varying moderators, we therefore collect smartphone sensor data to train machine-learning models that predict the participant's state of receptivity. Predictions of these models can in turn be used to inform intervention delivery of a JITAI.

Objectives

To inform the evidence-based development of JITAI for physical activity, the described study has the following objectives:

(1) To quantify main effects and interactions of main effects of three intervention components of Ally, a mHealth intervention for physical activity.

(2) To explore how the effects of intervention components are moderated by contextual factors and change over time.

(3) To collect a wide range of high-resolution smartphone sensor data in order to predict the participants' states of receptivity.

Method

Study Setting

This study is part of a research collaboration between the Center for Digital Health Interventions (CDHI), a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, and the CSS insurance, a large health insurer in Switzerland. Data for this study was collected from October to December 2017 in the German-speaking part of Switzerland. The study is registered on ClinicalTrials.gov (NCT03384550) and was approved by the ethical committee of ETH Zurich.

The Ally Application

Ally is a smartphone-based application that tracks participants' daily step count and provides interventions to help participants reach personalized daily step goals. It contains a dashboard that displays basic information such as the participant's current step count and the step goal of the current day as well as an activity overview of the past seven days (Figure 1). Ally runs on the common operating systems Android and iOS. On Android smartphones, Ally obtains all physical activity related information from GoogleFit, a health-tracking platform developed by Google. On iOS smartphones, the same information is obtained from the HealthKit, an application programming interface (API) for health applications provided by Apple. To obtain smartphone sensor data we used EmotionSense, a framework to support smartphone-based data collection originally developed for experimental social psychology research [9].

Personalized step goals are calculated daily for each participant based on the participant's activity over the past nine days employing the moving-window percentile-rank algorithm described by Adams and colleagues [10]. This adaptive goal-setting algorithm sets the daily step goal to the 60th percentile of the participant's step count distribution of the past nine days meaning that the participant reaches her/his step goal 40% of the times when maintaining her/his recent activity level. Previous studies demonstrated that this adaptive goal setting outperforms static step goals [10, 11]. To facilitate maintenance of behavior change, adaptive step goals are capped at 10,000 steps per day which approximates the WHO recommendations for physical activity [12, 13].

To administer the intervention components evaluated in this study, the Ally app includes a chatbot (Ally) that provides interactive coaching dialogues similar to other messaging Apps such as Apple's iMessage, Facebook's Messenger or WhatsApp. The open source behavioral intervention platform MobileCoach (www.mobile-coach.eu) [14] was used to build the chatbot and deliver the interactive coaching dialogues. In previous studies, MobileCoach-based interventions have successfully reduced problem drinking in adolescents [15] and engaged the majority of participants of a three-month smoking cessation program [16]. Participants interact with Ally by selecting predefined answer options (Figure 1) which trigger a response by the chatbot according to the conversational rules specified in the MobileCoach system.

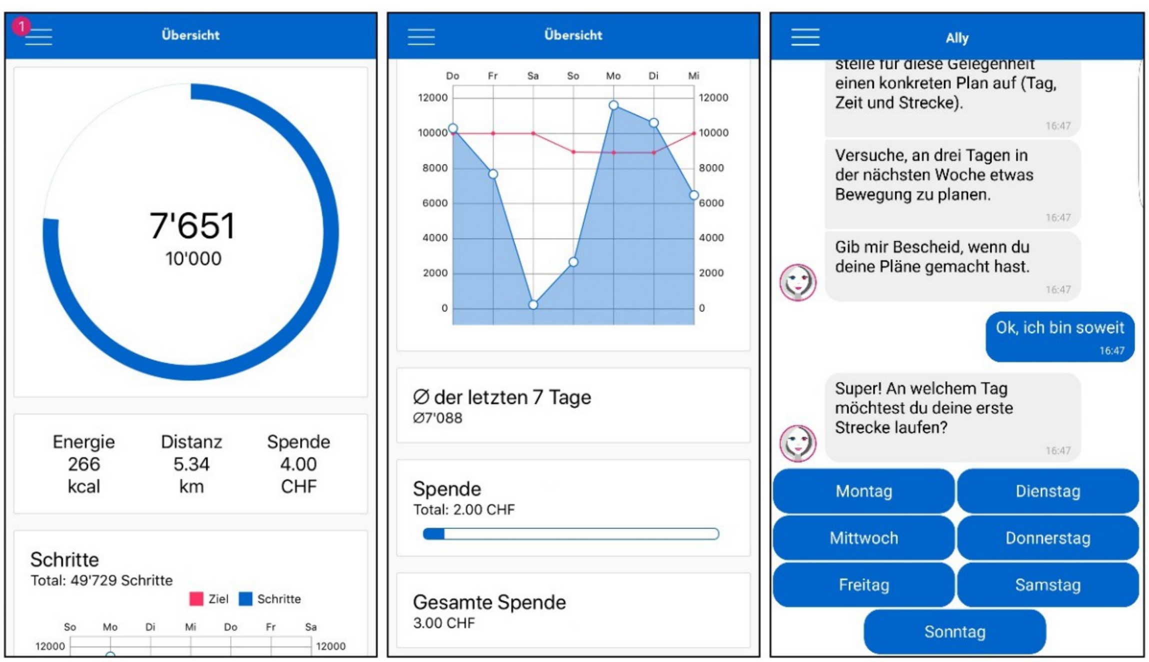

Figure 1. The Ally app: Dashboard with daily (left) and weekly overview (middle) and chat interactions with the Ally chatbot (right)

Beyond specific interventions, the chatbot also communicates the daily step goal in the morning and feedback regarding the goal together with informative facts about physical activity at 8 pm in the evening to all participants.

Study Design

From October to December 2017 insurees of a large Swiss health insurance used the Ally app over a 12-day baseline and a 6-week intervention period. During the baseline period, participants only had access to the dashboard of the app and no interventions were administered. Over the course of the 6-week intervention period, we randomized participants to different versions of three intervention components: daily self-monitoring prompts (two levels, within-subjects), a weekly planning intervention (3 levels, within-subjects) and daily incentives (3 levels, between-subjects). The rationale for these intervention components is described below. To meet study objective three, we aimed to explore if and how participants' reaction to intervention components were dependent on their context. In order to do so, we ideally need to observe reactions to intervention notifications in wide variety of contexts. We therefore sent out intervention and step goal related notifications at random points in time but within pre-specified time windows that guaranteed delivery at appropriate times. For example, daily step goal notifications were delivered at a random point in time between 8am and 10am since users likely expect to be notified about their goal early in the day. Participants completed an online questionnaire at baseline and at post-intervention follow up and received CHF 10 (US$ 10 as of 2017) for the successful completion of both questionnaires. If participants provided consent, they were invited to participate in exit interviews after the end of the study that investigate perceptions of participants and mechanisms of behaviour change.

The following subsections first describe details and rationale for each intervention component as well as for potential moderators. Subsequently, we outline how each component was randomized during the intervention period and how we define the proximal outcome to evaluate each component. Table 2 provides a summary of all intervention components.

Intervention components

Self-monitoring prompts

Self-regulatory processes have been identified as a key factor for health behavior change [17, 18]. To support participants' self-regulation, we designed short dialogue-based self-monitoring prompts. Self-monitoring prompts remind the participant of their daily step goal, compare the participant's current step count to their daily goal and provide an estimate of walking minutes necessary to reach the goal together with an actionable tip on how to increase physical activity. These dialogues were designed to support the three sub-processes of the self-regulatory construct action control, self-monitoring, awareness of goals or standards and self-regulatory effort [19, 20]. If a participant had already reached their daily step goal when starting the dialogue, she/he would receive positive and encouraging feedback from the Ally chatbot instead.

Participants were randomized to receive a self-monitoring prompt or no prompt every day during the intervention period except Sunday as this day was reserved for the planning intervention (see below). Self-monitoring prompts were delivered at a random point in time between 10 am and 6 pm.

Self-monitoring preceding the prompt (e.g., how often the participant opened the app) may moderate the effect of the prompt. If a participant is already aware of their current activity level, the information provided by the prompt is likely to be redundant. Additionally, we assume the timing of the self-monitoring prompt to be critical. Research from cognitive psychology demonstrates that people assign more value to performance increases when their current performance is close to a goal rather than further away [21]. We consequently assume that self-monitoring prompts are more effective when they are sent later during the day when participants are more likely to be closer to their step goal. We will therefore explore whether self-monitoring preceding the prompt and time of delivery moderate the effect of the self-monitoring prompts.

Planning

Even if motivation to change exists, previous studies show that on average 47% of people fail to act upon their good intentions [22]. Forming specific plans about when and how to act increases the likelihood of performing the intended behavior [23, 24] and helps to bridge the so-called intention behavior gap. Planning can be further divided into action planning (specifying when, where and how to act) and coping planning (specifying coping responses for barriers and difficult situations) [25]. Plans that are articulated in an if-then format (e.g. "if I am tired at work, I will go for a brief walk to get new energy") are typically referred to as implementation intentions [26].

Every Sunday during the intervention period, participants received either an action planning (AP), a coping planning (CP) or no planning intervention (control; CC). In the action planning condition, Ally asks the participant to plan at least one and up to three walks for the upcoming week. To plan a single walk, the participants need to specify the day of the week, the time and the route that they intend to walk. In order to create flexible plans and thus increase the likelihood of adherence, Ally advises the participant to choose event-related times (e.g. after work) instead of actual times. In the coping planning condition, Ally asks the participant to identify barriers for physical activity by reflecting on the two least active days from the previous week. The participant is then prompted to develop counter-strategies for each barrier using the if-then format of implementation intentions [26]. Ally guides this process using examples for common barriers for physical activity that have been identified in previous studies [27-29], for example: "If I want to go for a walk but I lack motivation, I will think of the benefits of walking for health to motivate myself." Lastly, the participant had the option to anticipate days of the upcoming week where the barrier may arise again. Both action planning and coping planning include reminders for the participant on days when either a walk or a coping reaction was scheduled. To address the third objective of this study, planning interventions were sent out on Sundays at a random point in time between 10 am and 8 pm.

Participants' activity level as well as their location may moderate the effects of action and coping planning. Participants with low activity levels may be more likely to benefit from action planning, which promotes the initiation of action, whereas participants with high activity levels may benefit more from coping planning which prevents routines from distraction [25, 30]. Further, completing the planning intervention can take several minutes and requires a considerable amount of the participant's attention and cognitive capacity. Ideally, the planning intervention should therefore not be delivered in situations where the participant is involved in an attention-consuming activity, such as social activities or work. We will therefore explore whether the participant's geolocation measured by the smartphone (home/work/other) moderates the effect of action and coping planning.

Incentives

Meta-analyses [31, 32] and recent randomized trials [33-35] have demonstrated the ability of financial incentives to increase physical activity. However, financial incentives may reduce intrinsic motivation [36, 37]; thus charity incentives have been proposed as an alternative incentive strategy. Charity incentives, i.e. donations to a charity organization, could foster experiences of autonomy and relatedness, which are known to facilitate rather than impede the build-up of intrinsic motivation [38]. Two recent studies have so far compared financial and charity incentives with mixed results [33, 39].

In this study, participants were randomly allocated to either a financial incentive, a charity incentive or a control condition using an allocation ratio of 1:1:1. In the financial incentive condition, participants received CHF 1 (US$ 1 as of 2017) for each day they met their personalized step goal. In the charity incentive condition, instead of keeping the reward to themselves, participants made a donation of CHF 1 to a charity of their choice. Participants allocated to the control condition received no incentives. Earned rewards (maximum of CHF 42) were paid to participants or donated to charity after completion of the study.

We hypothesize that the presence of incentives moderates the effect of the other intervention components. Both planning and self-monitoring prompts target the participant's self-regulatory processes and thus require the participant to be motivated to reach the provided step goals in order to produce an effect [40]. Since we expect the incentives to increase the motivation of participants, we hypothesize that effects of self-monitoring prompts and planning are more pronounced for participants receiving financial or charity incentives.

Randomization, Allocation Concealment and Blinding

The MobileCoach version used in this study requires the time point of dissemination for all dialogues to be known a priori. Therefore, randomization had to be performed upon enrolment of participants for all intervention components. Each participant was randomized to one out of three incentive conditions using simple randomization and an allocation ratio of 1:1:1. Additionally, participants were randomized to one out of nine planning intervention sequences (S_1-S_9) that determine the order in which the participant received the action planning intervention (AP), the coping planning intervention (CP) or no planning intervention (CC) throughout the intervention period. We used blocked randomization with a block size of nine to achieve balance between the sequences. The resulting intervention schedule (Table 1) is uniform and strongly balanced, which allows controls for time and carry-over effects [41]. To avoid interference of self-monitoring prompts and planning, self-monitoring prompts were not delivered on Sundays. This left 42-6 =

36 available days for delivering self-monitoring prompts. To prevent repetition of content, we created 18 different versions of self-monitoring prompts which we randomly allocated to the 36 days for each participant. Consequently, at each of the 36 days 50% of participants received a self-monitoring prompt and 50% received no prompt on average. All randomizations were performed using random number sequences generated with the shuffle-array package in JavaScript.

Table 1. Intervention schedule of the planning intervention

	Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**
S_1	AP	AP	CP	CC	CC	CP
S_2	CP	CP	CC	AP	AP	CC
S_3	CC	CC	AP	CP	CP	AP
S_4	AP	CP	CP	AP	CC	CC
S_5	CP	CC	CC	CP	AP	AP
S_6	CC	AP	AP	CC	CP	CP
S_7	AP	CC	CP	CP	CC	AP
S_8	CP	AP	CC	CC	AP	CP
S_9	CC	CP	AP	AP	CP	CC

AP: Action planning
CP: Coping planning
CC: Control condition (no planning)

The fully automated randomization process guarantees allocation concealment for everyone involved in the study. Variables in the dataset indicating intervention allocation are encrypted to blind members of the research team involved in data analysis. A researcher of ETH Zurich who is not involved in data analyses holds the decryption key and is instructed to safely store the key until the analysis script has been finalized. Due to the setting of the study, it is not possible to blind participants to intervention assignments. To reduce the impact of performance and attrition bias, participants were not informed about the details of the intervention components prior to the study.

Measurements

Primary and secondary outcomes

Because the intervention components are randomized on different timescales, we need to define primary and secondary proximal outcomes that correspond to these timescales in order to correctly evaluate the intervention components. The proportion of overall participant-days that step goals are achieved during the intervention period is the primary outcome to evaluate the different incentive conditions. Weekly and daily proportions of participant days that step goals are achieved during the intervention period are the primary outcomes of the planning and self-monitoring prompts respectively. On the same timescales, differences in steps per day measured with the smartphone are investigated as a secondary outcome.

For financial and charity incentives, post-intervention differences in intrinsic and extrinsic motivation and differences in app engagement and non-usage attrition during the intervention period are evaluated as additional secondary outcomes. Dimensions of intrinsic and extrinsic motivation are measured using the Behavioral Regulation for Exercise Questionnaire (BREQ-2) [42]. Because the external regulation subscale in the BREQ-2 exclusively relates to external regulation by other people, it is substituted by the more generally worded external regulation subscale of the Situational Motivation Scale [43]. Subscales of both instruments have shown good reliability (Cronbach's α = .73 - .86 (BREQ-2) [42] and Cronbach's α = .86 (SIMS external regulation subscale [43])). Validity has been confirmed by factor analysis (BREQ-2) [42] and correlational analysis (SIMS) [43]. We measure engagement with the Ally app using the number and length of app launch sessions per day. An app launch session is defined as any interaction of the participant with the Ally app, separated by five minutes between events. If a participant left the app open and did not take action for five minutes or more, then the next interaction with the app counts as a new session. We coded a participant as "non-usage attrition observed" when she/he stopped using the Ally app at least seven days before the end of the study.

Table 2. Overview of intervention components of the Ally app

Component	Intervention options	Randomization	Mode of delivery	Time of delivery	BCTs [44][a]	Proximal outcome
Self-monitoring prompts	Prompt	Daily except Sunday; allocation ratio 1:1	Chat	randomly between 10 am and 6 pm	1.6; 2.2; 4.1	Daily proportion of participant days that step goals were achieved
	Control (no prompt)		-	-	-	
Planning	Action planning	Sundays; allocation ratio 1:1:1	Chat	randomly between 10 am and 6 pm	1.4	Weekly proportion of participant days that step goals were achieved
	Coping planning		Chat		1.2	
	Control (no planning)		-	-	-	
Incentives	Cash incentives	Upon enrolment; allocation ratio 1:1:1	Dashboard/ Chat	Daily	10.2	Overall proportion of participant-days that step goals were achieved
	Charity incentives		Dashboard/ Chat	Daily	10.3	
	Control (no incentives)		-	-	-	

[a]: 1.2 problem solving; 1.4 action planning; 1.6 discrepancy between current behavior and goal; 2.2 feedback on behavior; 4.1 Instruction on how to perform a behavior; 10.2 material reward (behavior); 10.3 non-specific reward

Other outcomes

As a preliminary pre-post evaluation of the Ally app, self-reported health outcomes and targeted mediators of behavior change were assessed at baseline and at post intervention follow up. In addition, we assessed participant's perceptions of the Ally app, the intervention components, the chatbot and predictors of technology acceptance at post intervention follow-up. An overview of all measured variables is available in Multimedia Appendix 1.

Sensor data

Drawing on previous literature on context-aware mobile notification management systems [45], we identified smartphone sensors that may aid with predicting the participants' state of receptivity. Sensor data were obtained from participants during the intervention period. Table 3 provides a summary of these sensors, their collected data and their sensing frequency. In line with these studies, we operationalize state of receptivity by using the response rate (i.e. whether a participant responds to a notification or not) and the response time (i.e. time between notification and response) to notifications of the Ally app.

Table 3. Summary of collected sensor data

Sensor	Variable	Data type	Frequency[a]
GPS	Location	3D Float	every 10 min
Accelerometer	Physical activity	Categorical	continuous
Time	Time	Integer	continuous
Proximity	Proximity of the phone	Binary (near, far)	continuous
Wi-Fi	Wi-Fi connection	Categorical / String	every 10 min
Bluetooth	Bluetooth connection	Categorical / String	every 10 min
Ambient light	Ambient light	Float	continuous
Battery status	Battery status	Float (charged in %)	continuous
Screen events	Screen on / off	Binary (on/off)	continuous

[a]Estimated frequencies only. Actual frequencies may vary depending on device and operating system.

Sample Size

We used a simulation-based approach to estimate the power of our study design and determine the necessary sample size. Because interaction effects require a greater number of participants to be detected with adequate power [46], we focused the power analysis on the two-way interaction effect of the between-subject factor incentives and the within-subject factor planning. We systematically varied the probability of reaching the step goal p(SG) in the incentive control group when no planning intervention is provided (0.30, 0.40, 0.50). These values seem reasonable given the fact the probability of step goal achievement according to the goal setting algorithm is 0.40. We further varied the increase in probability due to incentive and planning main effects (0.05, 0.10, 0.15) and the interaction effect (0.05, 0.10, 0.15) for sample sizes ranging from $n = 20$ to $n = 400$. These effect sizes were based on previous studies on the use of incentives to promote physical activity [34, 35]. One hundred simulations were generated for each scenario. *P*-values of interaction effects were obtained by fitting generalized estimating equations (GEE) models to the simulated data and power was calculated as the proportion of *p*-values below the significance level of $\alpha = .05$. Figure 2 displays simplified results of this simulation with constant main effects of .15 and different values for p(SG) and the interaction effect. The black horizontal line indicates the recommended level of power of $1\text{-}\beta = .80$.

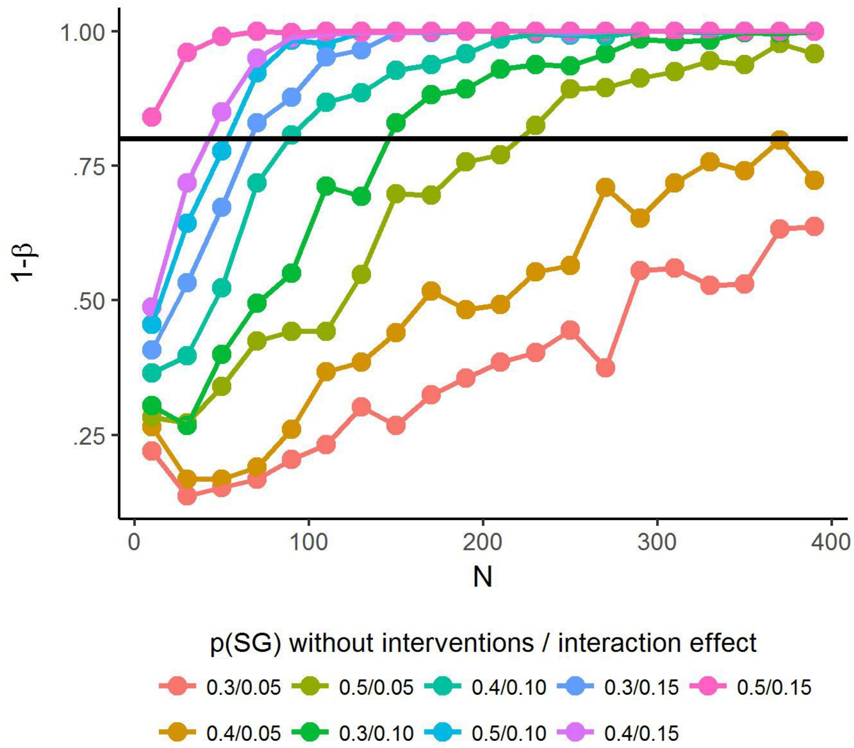

Figure 2. Results of the simulation-based power analysis.

Simulations indicate that a sample size of roughly N = 220 is sufficient to detect an interaction effect of .05 with a power of 1-β = .80 and α = .05 for p(SG) = .50. Sample sizes to detect an interaction effect .05 considerably increase for smaller values of p(SG) and smaller main effects (not shown). We therefore considered a sample size of N = 220 to be most feasible and accounting for drop out we set the target sample size for our study to N = 300.

Recruitment & Eligibility

We invited insurees via email to participate in our study. Based on a previous study in the same population and with a similar recruiting process [47], we expected a participation rate of approximately 3%. We initially sent the invitation email to 10,000 insurees. However, because initial participation was lower than expected, an additional 20,000 insurees were invited to meet the required sample size.

The invitation email contained brief information about the study, eligibility criteria and emphasized the benefits of participation. No details about the different intervention conditions were disclosed to the insurees. By following a link in the invitation mail, interested insurees were forwarded to an online survey platform, where they were screened for eligibility. Eligibility criteria were:

- German-speaking
- >= 18 years old

- enrolled in a complementary insurance program
- being free of any medical condition which prohibits increased levels of physical activity
- not actively using an activity tracker or a comparable smartphone app
- not working nightshifts

Since meeting the first three eligibility criteria could be determined from the insurance company's database, only insurees meeting these criteria were invited to participate. Due to legal regulations in Switzerland, the Ally app could be offered to insurees with complementary health insurance plans only. Note however, that in Switzerland 75% of people are enrolled in complementary insurance plans [48]. We excluded insurees working nightshifts because interventions were sent out on pre-specified times during the day only. Eligible insurees could subsequently obtain detailed information about the goals and procedure of the study, provide consent to participate and enroll in the study. After enrolment, participants completed the first online-questionnaire and subsequently received a six-digit code together with instructions on how to download and install the Ally application. Participants had to enter the six-digit code once upon first opening the Ally app to connect survey data and app data and to ensure that only study participants were using the app.

Statistical Analyses

To reach the study's objectives, we pre-specified the following analyses:

Primary Analyses

To evaluate main effects and interactions of intervention components we will use the centered and weighted generalized estimating equation (GEE) approach described in Boruvka et al. [49]. This approach guarantees unbiased effect estimates when treatment and moderator variables are time-varying. We will fit a series of models that separately evaluate each main effect and interaction of intervention components of interest on the components appropriate proximal outcome. For all main effects and interactions that include comparisons of multiple conditions, the main comparisons of interest are between the respective intervention and control conditions. Moderations of main effects by time and contextual factors are investigated by adding a term for the interaction between the main effect and the respective moderator to the model. However, these analyses are exploratory and may investigate different forms of relationships (e.g. linear or quadratic) and different forms of operationalizing the pre-specified moderators. Missing data on covariates and on the dependent variable will be imputed using multiple imputation provided the missing at random assumption is justified.

We will perform sensitivity analyses to assess the robustness of the results of the primary analyses. These analyses include a per-protocol analysis and an adjusted analysis, in which effect estimates are adjusted for a linear trend of time, main effects of the remaining intervention components, baseline step count and covariates of physical activity. For all tests, we use 2-sided p-values with $\alpha < .05$ level of significance.

Secondary Analyses

Secondary analyses focus on the analysis of intervention components on participants' step counts, and the effects of incentives on app engagement, non-usage attrition and motivation. Steps per day are analysed using the same modelling approach as described above. Again, if missing data can assumed to be missing at random, we plan to impute missing step counts using multiple imputation. Because evidence suggests that participant days with less than 1000 steps are unlikely to represent accurate activity data [50, 51], those days will be set to missing before imputation.

Generalized linear models will be used to analyse the effect of incentives on engagement and non-usage attrition. One-way analysis of variance (ANOVA) is performed for each subscale of the BREQ-2 to analyse the effect of incentives on the different forms of intrinsic and extrinsic motivation. P-values will be adjusted according to the Holm-Bonferroni method [52]. If the omnibus test of the ANOVA is significant, we will investigate contrasts between the three incentive groups. Again, the main comparison of interest is between the intervention groups and the control group. An overview of all planned statistical analysis is available in Multimedia Appendix 2.

State of Receptivity

We will compare several different methodological approaches to predict the participants' state of receptivity. First, we plan to evaluate the performance of supervised learning algorithms in predicting response rate and response time. These algorithms have produced predictions of acceptable accuracy in previous studies on interruptibility [45]. Second, we plan to frame the problem at hand as a classification problem. A classifier will be trained to learn to differentiate between contexts in which the notification is sent (and are assumed to represent non-receptive contexts) and contexts in which the participant interacts with the app (and in turn are assumed to represent receptive contexts). To this end, we aim to use generalized linear models as a starting point before exploring online learning algorithms that can learn and adapt to each participant's preferences, and any change thereof. This analysis strategy, however, is preliminary at the time of writing, as the final analysis will consider additional factors such as the quality and distribution of collected data.

Results

Recruitment

Of all 30,000 invited insurees, 749 (2.50%) clicked the link in the invitation mail and were subsequently screened for eligibility. Of those, 694 (92.7%) were eligible and 382 (51.0%) provided informed consent to participate. Of all insurees who provided informed consent, 274 (71.7%) successfully completed the baseline survey and installed the Ally app on their smartphone (Figure 3). Invited insurees were given the opportunity to select reasons why they declined participation from a list of predefined answer options using a separate survey (n = 191). A link to this survey was included in the invitation mail and placed on the informed consent screen. Possession of an incompatible smartphone (37/191, 19.4%) and unwillingness to share smartphone sensor data (35/191, 18.3%) were the most frequently stated reasons to decline participation.

Thirty-two out of 274 participants (11.7%) did not receive any interventions, because they stopped using the app before the start of the intervention period. Due to technical errors, six participants did not receive the interventions they were randomized to (for example, a self-monitoring prompt was sent out on a day where the participant was randomized to not receiving a prompt). For the six participants these errors affected between 1 and 25 out of 42 participant days. Steps per day measured with the smartphone are available for 227/274 participants (82.8%) and smartphone sensor data are available for 247/274 participants (90.1%). After completing the six-week intervention period, 181/274 (66.1%) participants filled out the web-based follow-up survey. Data collection finished in January 2018.

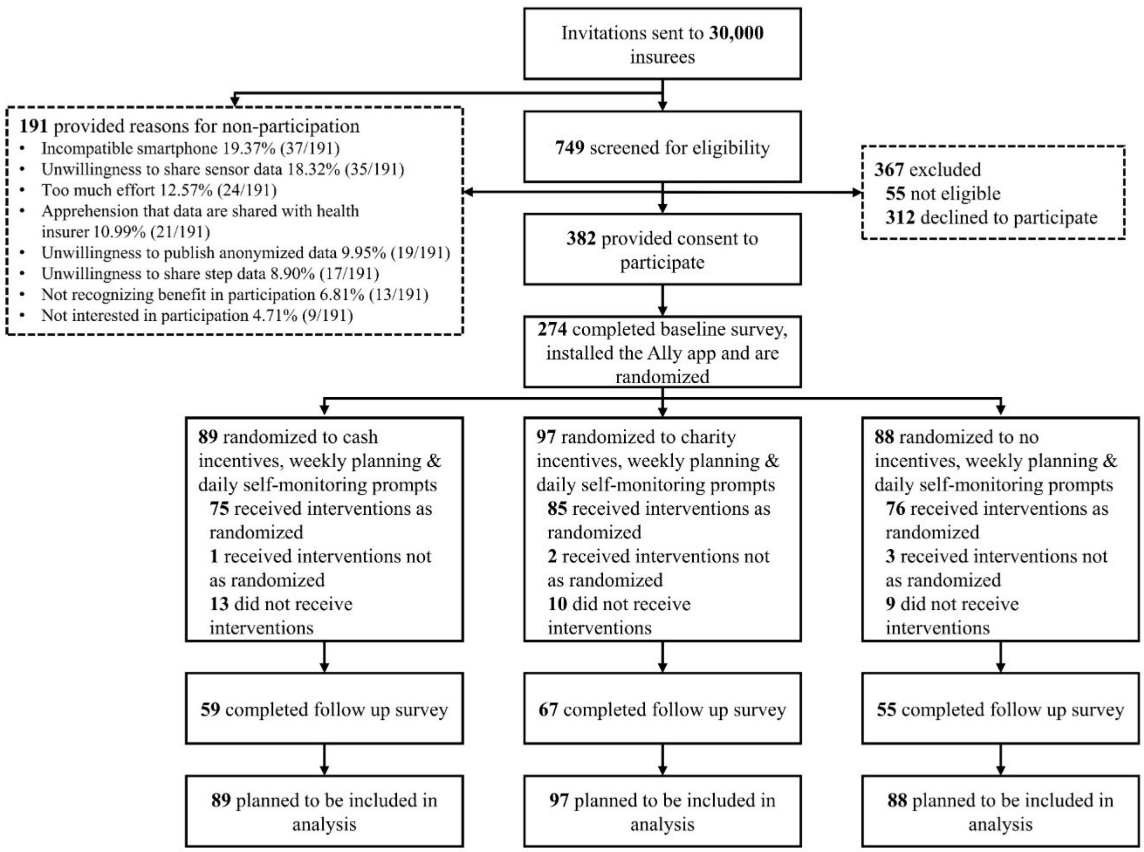

Figure 3. Participant flow

Expected Results and Dissemination

We will start data analyses after publication of this study protocol. We anticipate submitting results to a peer-reviewed journal in 2019. Preliminary results of the study may be presented at conferences, workshops, symposia etc. Results of the analysis of sensor data to predict the participants' state of receptivity will be published separately in a peer-reviewed journal or conference proceedings.

Discussion

This study protocol describes the design of a micro-randomized trial that investigates the effectiveness of three intervention components as well as associated moderators to guide the design of a JITAI for physical activity. This study is among the first to generate data for the evidence-based development of a JITAI for physical activity. In addition, a data collection strategy is described that enables the parallel collection of sensor data needed to build predictive models that, when implemented into a JITAI, allow real-time prediction of the state of receptivity to better inform adaptive intervention delivery by highlighting the best times for mobile intervention delivery. Insights from this study are of value for anyone involved in the development of mobile health interventions and support important decisions such as which components to include in a mHealth intervention or how to tailor intervention delivery to participants over time.

Our study illustrates potential and challenges associated with mHealth studies. The study's digital recruitment and data collection process allowed recruiting more than 270 participant in less than a week and the collection of a unique and powerful high-resolution dataset containing real-world behavioural and contextual data. In line with other mHealth studies [53], we observed a larger drop in app usage at the beginning of the study, potentially complicating interpretation of our findings. Likewise, step and sensor data was missing for some participants. Explanations for missing data include never reacting to a message of the Ally chatbot, which was required to request step counts from GoogleFit or the HealthKit, or denying app permissions to collect sensor data. Sending invitations via email and to insurees of one insurer only, the restricted range of compatible smartphones and the requirement to share sensitive data (e.g. GPS sensor data) are likely contributing to a self-selection of participants in our study. This limits the generalizability of our findings and conclusions.

If intervention components prove to be effective, we plan to include them in a revised version of the Ally app that provides just-in-time adaptive support depending on identified moderators and predicted states of receptivity. We plan to evaluate this revised version in a randomized controlled trial.

Authors' contributions

Jan-Niklas Kramer, Tobias Kowatsch and Urte Scholz developed the concept for intervention components and for the Ally app. Florian Künzler, Varun Mishra and Tobias Kowatsch were responsible for app design and implementation. Jan-Niklas Kramer, Florian Künzler, Tobias Kowatsch and Shawna Smith developed the study design described in this protocol. Jan-Niklas Kramer and Shawna Smith developed the methodological approach for the analyses of the different intervention components and Florian Künzler, Varun Mishra, David Kotz and Tobias Kowatsch developed the methodological approach for the analyses of smartphone sensor data. Bastien Presset developed concept and methodology for the qualitative exit interviews. Jan-Niklas Kramer wrote the manuscript incorporating critical reviews from all authors. All authors reviewed and approved the manuscript before submission.

Acknowledgements

This study is funded by CSS insurance, Switzerland. The CSS insurance supported the recruitment of participants but had no role in app development, study design, data collection, data analysis and interpretation, writing the manuscript or in reviewing and approving the manuscript for publication.

Conflict of Interest

Jan-Niklas Kramer, Florian Künzler and Tobias Kowatsch are affiliated with the Center for Digital Health Interventions (CDHI, www.c4dhi.org), a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. Tobias Kowatsch is also co-founder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways and has used the open source MobileCoach platform for that purpose, too. However, Pathmate Technologies is not involved in the intervention described in this paper. No other conflicts of interests are declared.

Multimedia Appendix

Multimedia Appendix 1: Study timeline including intervention components and assessment of outcomes

Multimedia Appendix 2: Overview of variables, measures and methods of analysis

References

1. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med. 2015;7(283):283rv3-rv3. PMID: 25877894.
2. World Health Organization (WHO). World health statistics 2017: monitoring health for the SDGs. Geneva: 2017.
3. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med. 2016:1-17. PMID: 27663578.
4. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med. 2016;51(5):843 - 51. PMID: 27745684.
5. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol. 2015;34(S):1220-28. PMID: 26651463.
6. Pielot M, Dingler T, Pedro JS, Oliver N, editors. When attention is not scarce-detecting boredom from mobile phone usage. Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing; 2015, Sep 07-11; Osaka, Japan. New York: ACM; 2015.

7. Fischer JE, Greenhalgh C, Benford S, editors. Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications. Proceedings of the 13th international conference on human computer interaction with mobile devices and services; 2011, Aug 30 - Sep 02; Stockholm, Sweden. New York: ACM; 2011.
8. Pejovic V, Musolesi M, editors. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2014, Sep 13-17; Seattle, USA. New York: ACM; 2014.
9. Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A, editors. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. Proceedings of the 12th ACM international conference on Ubiquitous computing; 2010: ACM.
10. Adams MA, Hurley JC, Todd M, Bhuiyan N, Jarrett CL, Tucker WJ, et al. Adaptive goal setting and financial incentives: a 2×2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health. 2017;17(1):286. PMID: 28356097.
11. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One. 2013;8(12):e82901. PMID: 24349392.
12. Tudor-Locke C, Bassett DRJ. How many steps/day are enough? Preliminary pedometer indices for public health. Sports Med. 2004;34:1-8. PMID: 14715035.
13. Tudor-Locke C, Hatano Y, Pangrazi RP, Kang M. Revisiting" how many steps are enough?". Med Sci Sports Exerc. 2008;40(7):S537-S43. PMID: 18562971.
14. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A Novel Open Source Platform for the Design of Evidence-based, Scalable and Low-Cost Behavioral Health Interventions - Overview and Preliminary Evaluation in the Public Health Context. 14th annual Wireless Telecommunications Symposium (WTS 2015); New York, USA: IEEE; 2015.
15. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web-and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol. 2017;85(2):147-59. PMID: 27606700.
16. Castro RP, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement Within a Mobile Phone–Based Smoking Cessation Intervention for Adolescents and its Association With Participant Characteristics and Outcomes. J Med Internet Res. 2017;19(11). PMID: 29092811.
17. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol. 2009;28(6):690-701. PMID: 19916637.
18. Mann T, De Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol. 2013;32(5):487. PMID: 23646832.
19. Carver CS, Scheier MF. Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. Psychol Bull. 1982;92(1):111-35. PMID: 7134324.

20. Sniehotta FF, Nagy G, Scholz U, Schwarzer R. The role of action control in implementing intentions during the first weeks of behaviour change. Br J Soc Psychol. 2006;45(1):87-106. PMID: 16573874.
21. Heath C, Larrick RP, Wu G. Goals as reference points. Cogn Psychol. 1999;38(1):79-109. doi: 10.1006/cogp.1998.0708.
22. Sheeran P. Intention—Behavior Relations: A Conceptual and Empirical Review. European Review of Social Psychology. 2002;12(1):1-36. doi: 10.1080/14792772143000003.
23. Gollwitzer PM, Sheeran P. Implementation intentions and goal achievement: A meta-analysis of effects and processes. Adv Exp Soc Psychol. 2006;38(1):69-119. doi: 10.1016/S0065-2601(06)38002-1.
24. Bélanger-Gravel A, Godin G, Amireault S. A meta-analytic review of the effect of implementation intentions on physical activity. Health Psychol Rev. 2013;7(1):23-54. PMID: 10.1080/17437199.2011.560095.
25. Sniehotta FF, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur J Soc Psychol. 2005;35(4):565-76. doi: 10.1002/ejsp.258.
26. Gollwitzer PM. Implementation intentions: Strong effects of simple plans. Am Psychol. 1999;54(7):493-503. doi: 10.1037/0003-066X.54.7.493.
27. Reichert FF, Barros AJ, Domingues MR, Hallal PC. The role of perceived personal barriers to engagement in leisure-time physical activity. Am J Public Health. 2007;97(3):515-9. PMID: 17267731.
28. Cerin E, Leslie E, Sugiyama T, Owen N. Perceived barriers to leisure-time physical activity in adults: an ecological perspective. J Phys Act Health. 2010;7(4):451-9. PMID: 20683086.
29. Zunft H-JF, Friebe D, Seppelt B, Widhalm K, de Winter A-MR, de Almeida MDV, et al. Perceived benefits and barriers to physical activity in a nationally representative sample in the European Union. Public Health Nutr. 1999;2(1a):153-60. PMID: 10933635.
30. Ziegelmann JP, Lippke S, Schwarzer R. Adoption and maintenance of physical activity: Planning interventions in young, middle-aged, and older adults. Psychol Health. 2006;21(2):145-63. PMID: 21985115.
31. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med. 2013;45(5):658-67. PMID: 24139781.
32. Barte JC, Wendel-Vos GW. A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. Behav Med. 2017;43(2):79-90. PMID: 26431076.
33. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol. 2016;4(12):983-95. PMID: 27717766.

34. Patel MS, Asch DA, Roy Rosin M, Small DS, Bellamy SL, Heuer J, et al. Framing Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. Ann Intern Med. 2016;164:385-94. PMID: 26881417.
35. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Eberbach K, et al. Individual Versus Team-Based Financial Incentives to Increase Physical Activity: A Randomized, Controlled Trial. J Gen Intern Med. 2016:1-9. PMID: 26976287.
36. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull. 1999;125(6):627-68. PMID: 10589297.
37. Cerasoli CP, Nicklin JM, Ford MT. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. Psychol Bull. 2014;140(4):980. PMID: 24491020.
38. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, USA: Plenum; 1985. ISBN: 0470479213.
39. Harkins KA, Kullgren JT, Bellamy SL, Karlawish J, Glanz K. A Trial of Financial and Social Incentives to Increase Older Adults' Walking. Am J Prev Med. 2017;52(5):e123-e30. PMID: 28062271.
40. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors: The health action process approach. Eur Psychol. 2008;13(2):141-51. doi: 10.1027/1016-9040.13.2.141.
41. Hills M, Armitage P. The two-period cross-over clinical trial. Br J Clin Pharmacol. 1979;8(1):7-20. doi: 10.1111/j.1365-2125.1979.tb05903.x.
42. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. Journal of Sport and Exercise Psychology. 2004;26(2):191-6. doi: 10.1123/jsep.26.2.191.
43. Guay F, Vallerand RJ, Blanchard C. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). Motivation and emotion. 2000;24(3):175-213. doi: 10.1023/A:1005614228250.
44. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med. 2013;46(1):81-95. PMID: 23512568.
45. Künzler F, Kramer J-N, Kowatsch T, editors. Efficacy of mobile context-aware notification management systems: A systematic literature review and meta-analysis. 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob); 2017, Oct 09-11; Rome, Italy: IEEE; 2017.
46. Montgomery AA, Peters TJ, Little P. Design, analysis and presentation of factorial randomised controlled trials. BMC Med Res Methodol. 2003;3(1):26. PMID: 14633287.
47. Kowatsch T, Kramer J-N, Kehr F, Wahle F, Elser N, Fleisch E. Effects of Charitable Versus Monetary Incentives on the Acceptance of and Adherence to a Pedometer-Based Health Intervention: Study Protocol and Baseline Characteristics of a Cluster-Randomized Controlled Trial. JMIR Res Protoc. 2016;5(3):e181. PMID: 27624645.

48. Eisler R, Lüber A. 2006. Wie wichtig ist den Schweizern eine Spitalzusatzversicherung? https://www.comparis.ch/~/media/files/mediencorner/studies/2006/krankenkassen/spitalzusatzversicherungen_studie.pdf. Archived at: http://wwwwebcitationorg/6yvDHGMkm.
49. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. Journal of the American Statistical Association. 2017. doi: 10.1080/01621459.2017.1305274.
50. Sequeira MM, Rickenbach M, Wietlisbach V, Tullen B, Schutz Y. Physical activity assessment using a pedometer and its comparison with a questionnaire in a large population survey. Am J Epidemiol. 1995;142(9):989-99. PMID: 7572981.
51. Bassett Jr DR, Wyatt HR, Thompson H, Peters JC, Hill JO. Pedometer-measured physical activity and health behaviors in United States adults. Med Sci Sports Exerc. 2010;42(10):1819. PMID: 20305579.
52. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics. 1979:65-70.
53. Modena BD, Bellahsen O, Nikzad N, Chieh A, Parikh N, Dufek DM, et al. Advanced and Accurate Mobile Health Tracking Devices Record New Cardiac Vital Signs. Hypertension. 2018:HYPERTENSIONAHA. 118.11177. PMID: 29967036.
54. Ware Jr JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220-33. PMID: 8628042.
55. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Pratt M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35:1381-95. PMID: 12900694.
56. Lippke S, Schwarzer R, Ziegelmann JP, Scholz U, Schüz B. Testing stage-specific effects of a stage-matched intervention: A randomized controlled trial targeting physical exercise and its predictors. Health Educ Behav. 2010;37(4):533-46. PMID: 20547760.
57. Scholz U, Keller R, Perren S. Predicting behavioral intentions and physical exercise: a test of the health action process approach at the intrapersonal level. Health Psychol. 2009;28(6):702. PMID: 19916638.
58. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol. 2011;56(3):161. PMID: 21767036.
59. Venkatesh V, Morris MG, Davis GB, Davis FD. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly. 2003;27(3):425-78. doi: 10.2307/30036540.
60. Venkatesh V, Thong JYL, Xu X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly. 2012;36(1):157-78.
61. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, O'Donnell J. Response to a relational agent by hospital patients with depressive symptoms. Interacting with Computers. 2010;22(4):289-98.

62. Spittaels H, Verloigne M, Gidlow C, Gloanec J, Titze S, Foster C, et al. Measuring physical activity-related environmental factors: reliability and predictive validity of the European environmental questionnaire ALPHA. Int J Behav Nutr Phys Act. 2010;7(1):48. PMID: 20504339.
63. Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of research in Personality. 2007;41(1):203-12. doi: 10.1016/j.jrp.2006.02.001.