

# Which Components of a Smartphone Walking App Help Users to Reach Personalized Step Goals? Results From an Optimization Trial

Jan-Niklas Kramer, MSc<sup>1,⊗</sup> · Florian Künzler, MSc<sup>2</sup> · Varun Mishra, BTech<sup>3</sup> · Shawna N. Smith, PhD<sup>4,5</sup> · David Kotz, PhD<sup>3,6</sup> · Urte Scholz, PhD<sup>7</sup> · Elgar Fleisch, PhD<sup>1,2</sup> · Tobias Kowatsch, PhD<sup>1</sup>

Published online: 17 March 2020

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society of Behavioral Medicine.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

## Abstract

**Background** The Assistant to Lift your Level of activityY (Ally) app is a smartphone application that combines financial incentives with chatbot-guided interventions to encourage users to reach personalized daily step goals.

**Purpose** To evaluate the effects of incentives, weekly planning, and daily self-monitoring prompts that were used as intervention components as part of the Ally app.

**Methods** We conducted an 8 week optimization trial with  $n = 274$  insureds of a health insurance company in Switzerland. At baseline, participants were randomized to different incentive conditions (cash incentives vs. charity incentives vs. no incentives). Over the course of the study, participants were randomized weekly to different planning conditions (action planning vs. coping

planning vs. no planning) and daily to receiving or not receiving a self-monitoring prompt. Primary outcome was the achievement of personalized daily step goals.

**Results** Study participants were more active and healthier than the general Swiss population. Daily cash incentives increased step-goal achievement by 8.1%, 95% confidence interval (CI): [2.1, 14.1] and, only in the no-incentive control group, action planning increased step-goal achievement by 5.8%, 95% CI: [1.2, 10.4]. Charity incentives, self-monitoring prompts, and coping planning did not affect physical activity. Engagement with planning interventions and self-monitoring prompts was low and 30% of participants stopped using the app over the course of the study.

**Conclusions** Daily cash incentives increased physical activity in the short term. Planning interventions and self-monitoring prompts require revision before they can be included in future versions of the app. Selection effects and engagement can be important challenges for physical-activity apps.

**Clinical Trial Information** This study was registered on ClinicalTrials.gov, NCT03384550.

**Keywords:** Walking · Mobile health · Microrandomized trials · Intervention components · Engagement

There is univocal evidence for a curvilinear dose–response relationship between physical activity and health (1), with the greatest health benefits seen in inactive individuals who become more active. While the current physical-activity guidelines for adults recommend at least 150 min of moderate-intensity aerobic physical

✉ Jan-Niklas Kramer  
[jan-niklas.kramer@unisg.ch](mailto:jan-niklas.kramer@unisg.ch)

<sup>1</sup> Institute of Technology Management (ITEM), University of St. Gallen, St. Gallen, Switzerland

<sup>2</sup> Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

<sup>3</sup> Department of Computer Science, Dartmouth College, Hanover, NH, USA

<sup>4</sup> Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

<sup>5</sup> School of Medicine, University of Michigan, Ann Arbor, MI, USA

<sup>6</sup> Center for Technology and Behavioral Health, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

<sup>7</sup> Department of Psychology, University of Zurich, Zurich, Switzerland

activity per week (2), studies suggest that some of the health benefits of physical activity can be achieved with presumably less intensive activities such as walking (3). Walking is a common and safe type of activity that can be performed by individuals of all ages largely independent of time and location (4). Consequently, walking is a suitable target behavior for many physical-activity interventions.

Today, the widespread adoption of smartphones and activity trackers enables objective monitoring of walking and provides new opportunities for interventions on a large scale. Smartphone apps, for example, facilitate the delivery of health interventions to thousands of individuals at low cost and require minimal human support (5). If effective, these stand-alone interventions could have a substantial impact on public health. Unfortunately, recent meta-analyses of randomized controlled trials reported only small effects for these types of interventions (6, 7). For example, the summary effect of stand-alone physical activity apps was estimated at 477 steps per day, 95% confidence interval (CI): [−230, 1,183] (7).

One approach to developing interventions that are more effective is to conduct so-called optimization trials, that is, evaluating the effects of single-intervention components and their combinations (8), before finalizing and testing the complete intervention in a randomized controlled trial. Results of optimization trials provide valuable data for refinement of interventions and guarantee that every component of an intervention contributes to its overall effectiveness. Klasnja et al. proposed the microrandomized trial (MRT) (9), an optimization trial design for mobile-health interventions. MRTs use repeated randomization (microrandomization) of participants to different versions and/or the presence and absence of individual intervention components over the course of the intervention. This enables the estimation of the time-averaged and time-varying main effects of single-intervention components on proximal (short-term) outcomes, as well as the interaction effects of two (or more) intervention components. In this paper, we present the results from an optimization trial that evaluated the main effects and interactions of three intervention components of a smartphone app to promote walking, using both baseline randomization and microrandomization.

## Theoretical Background

Health-behavior theories can guide the development of interventions by identifying the intervention targets that are most likely to lead to behavior change. Some health-behavior theories, such as the Health Action Process Approach (10), differentiate between *motivational processes* that predict behavioral intentions to change and

*volitional processes* that subsequently translate intentions into actual behavior. Volitional processes are especially important with regard to mobile health (mHealth) apps because users of such apps typically report preexisting intentions to change (11).

Self-regulation of behavior or action control is a central volitional process that ensures alignment of one's own behavior with set behavioral goals. Self-regulation comprises the adoption of a reference value for behavior in the form of a behavioral goal, monitoring goal progress (i.e., evaluating one's ongoing performance relative to the reference value), and reacting according to the evaluation (i.e., increasing effort if one's performance is below the reference value) (12). Previous research demonstrated that physical-activity interventions that support self-regulatory processes have been found to be significantly more effective than other interventions (13), making self-regulation an attractive intervention target. Due to the smartphone's innate capability to passively measure physical activity, smartphone apps can be an excellent tool to monitor goal progress and support self-regulation of physical activity.

Planning or formulating implementation intentions is another volitional process that helps to realize intentions to change. Making specific plans about behavior change can support the initiation of goal-directed behavior and enable its execution by insulating it from tempting alternatives or competing goals. Two types of planning are considered important: *action planning*, that is, specifying when, where, and how to perform the target behavior, and *coping planning*, that is, planning how to overcome barriers and deal with setbacks (10). A meta-analysis of 24 randomized and quasirandomized trials has confirmed the positive effect of such “planning interventions” on physical activity (14).

Naturally, the intention to change is a prerequisite for the effectiveness of volitional intervention strategies. However, the large attrition rates observed for many mHealth apps (15) suggest that intentions may diminish over time and volitional intervention strategies need to be supported by intervention strategies that target motivation. Regulatory changes, such as the Affordable Care Act in the USA, have facilitated the use of financial incentives to motivate health behaviors (16). As a consequence, the interest in financial incentives as a behavior change strategy has surged among researchers and practitioners alike and, today, many employers and health insurance companies offer incentive-based health promotion programs (16). Conditional financial incentives (i.e., incentives directly tied to the achievement of behavioral goals) may boost motivation (and subsequent behavior change) by altering the associated outcome expectancies, that is, beliefs about positive and negative consequences of behaviors (10). Indeed, previous randomized studies have

reported significant positive effects of financial incentives on physical activity (17). However, self-determination theory (18) predicts that external rewards reduce intrinsic motivation, if they are perceived as controlling the target behavior. In previous studies, rewards have reduced intrinsic motivation in interesting tasks (e.g., games or puzzles) in samples of children and students (19). If applied to health behavior, such effects would be detrimental because intrinsic motivation has been associated with maintained health-behavior change (20).

## Research Objectives and Hypotheses

We evaluated three intervention components of the Assistant to Lift your Level of activitY (Ally) app: (a) incentives for meeting daily step goals, (b) weekly planning, and (c) daily self-monitoring prompts, that is, short reminders to continue monitoring progress and achievement of daily step goals. Intervention components are described in detail below. We hypothesize that all intervention components encourage participants to walk. Further, we assume an interaction between motivational and volitional intervention components. Specifically, we hypothesize that the effects of planning and self-monitoring prompts will be greater if they are accompanied by incentives. As a secondary objective, we explore the effects of incentives on intrinsic motivation. A further goal of this study was to collect smartphone sensor data to predict participants' states of receptivity, that is, moments in time where participants are more likely to react to intervention push notifications. Results relating to the prediction of participants' states of receptivity will be reported in a separate publication.

## Methods

In this section, we briefly describe the app used in this study, the study design, the randomization of intervention components, and the statistical analyses. More detailed descriptions are available in the study protocol (21). The statistical analysis approach is described in detail in the [Supplementary Material \(Sections I and II\)](#). This study was conducted in collaboration with a health insurer in Switzerland to accompany the development of a large-scale prevention program focused on physical activity and targeted at the general population. We collected data from October to December 2017 in the German-speaking part of Switzerland. We obtained informed consent from all individual participants included in the study. The ethical review board of ETH Zurich approved all study procedures. The study is registered on ClinicalTrials.gov (NCT03384550). The

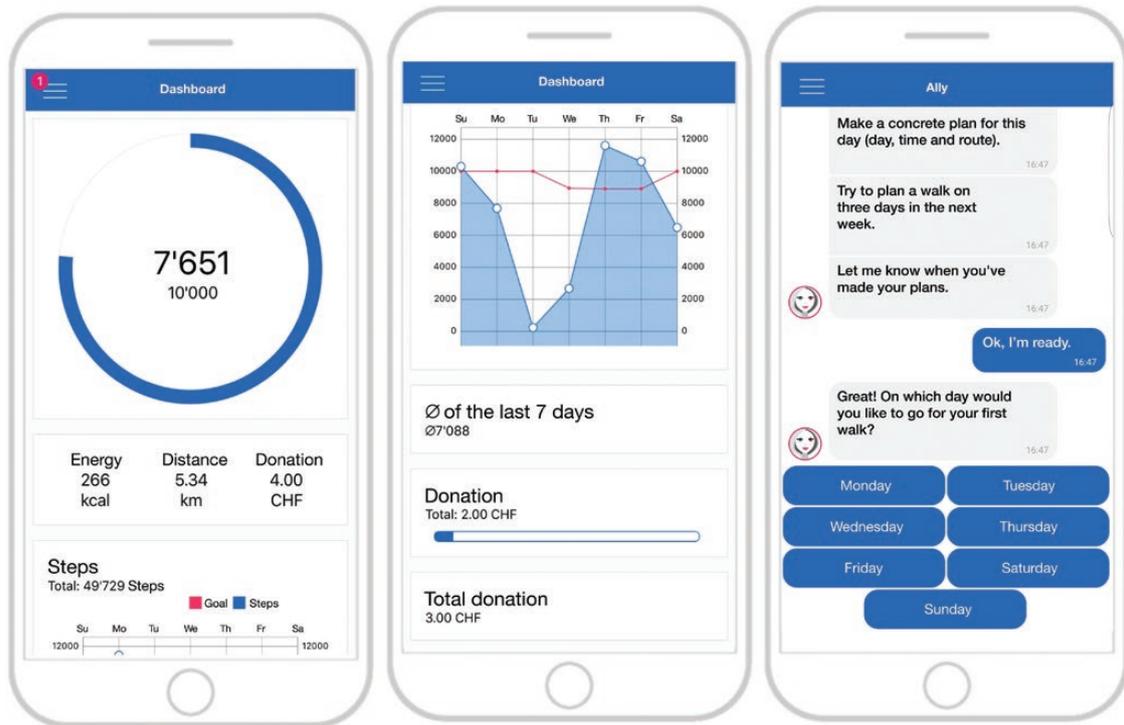
anonymized data set and the analysis script for the primary analyses from this study are available online: [https://osf.io/uc765/?view\\_only=334e8e3df14d4677b9296959d06c2c00](https://osf.io/uc765/?view_only=334e8e3df14d4677b9296959d06c2c00).

## The Ally App

The Ally app is a research app developed in 2017 to help people increase and maintain their physical activity. Ally aims at supporting physical-activity behavior change by means of a dashboard that visualizes steps per day in daily and weekly overviews as depicted in [Fig. 1](#) (middle and left) and by providing additional interventions via a chatbot interface as depicted in [Fig. 1](#) (right). We designed Ally as a chatbot because evidence suggests that computers framed as humans are perceived as social actors, which results in fundamentally social relationships (22), and that a therapeutic alliance, or the physician–patient relationship, is linked to treatment success in medical studies (23) and psychotherapy (24).

After the mobile app has been installed, Ally welcomes each participant personally, tracks physical activity using the smartphone's built-in accelerometer, and guides participants through physical-activity interventions using chat-based interactions. Every time Ally initiates a conversation, a push notification appears in the status bar and on the lock screen of the participant's smartphone, prompting the participant to chat with Ally. Participants can respond to Ally by selecting one of several predefined answer options ([Fig. 1](#), right), which, in return, triggers a response by Ally according to specified conversational rules. Every day, the Ally app sets a personalized activity goal based on each participant's past activity data. Using the adaptive approach described by Adams et al. (25), daily step goals correspond to the 60th percentile of the step-count distribution of the past 9 days. Consequently, step goals are set slightly above the participant's current average activity level. To facilitate the maintenance of physical activity, step goals are capped at 10,000 steps per day, which approximates the World Health Organization's physical activity recommendations (26).

Ally was developed using the MobileCoach platform ([www.mobile-coach.eu](http://www.mobile-coach.eu)), an open-source server-client software for the design of ecological momentary assessments and digital health interventions (27). We supported both common mobile platforms, that is, Google's Android and Apple's iOS, to reach a market share of 99.3% in Switzerland (28). On Android, Ally obtains all physical activity-related information from GoogleFit, a health-tracking platform developed by Google. On iOS, the same information is obtained from the HealthKit, an application programming interface for health apps provided by Apple. The Ally app requires Android version 4.0 or higher or an iPhone model 5 or newer.



**Fig. 1.** The Ally app: dashboard with daily overview (left), weekly overview (middle), and chat interactions with the Ally chatbot (right).

## Study Design

The study consisted of a 2 week run-in and baseline period and a 6 week intervention period. Data was collected from participants' smartphones via the Ally app and from two online questionnaires at the beginning and at the end of the study. Participants received 10 Swiss Francs (CHF; approximately equal to US\$10) for participation in the study and completing both questionnaires.

We invited 30,000 insureds of the health insurance company to participate in the study via an email invitation. Interested insureds could click on a link in the invitation email to be forwarded to an online survey platform where they were screened for eligibility. Eligibility criteria were: (a) German speaking, (b) aged 18 years or older, (c) enrolled in a complementary insurance program, (d) being free of any medical condition that prohibits increased levels of physical activity, (e) not actively using an activity tracker or a comparable smartphone app, and (f) not working night shifts. Eligible insureds could subsequently obtain detailed information about the study goals and procedures, provide consent to participate, and enroll in the study. After enrollment, participants completed the first online questionnaire and received a six-digit code, together with instructions on how to download and install the Ally app. Participants had to enter the 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. The baseline period started once participants had installed the app. During this period, Ally counted and displayed steps per day and sent occasional messages that were unrelated to physical activity to foster participants' interest in the study. However, the app's dashboard did not display any information related to financial incentives and the app sent no intervention-related messages. Two weeks after sending out the invitation emails, the baseline period ended and the 6 week intervention period started for all participants. During the intervention period, the Ally app set daily step goals and delivered interventions to support step-goal achievement.

## Intervention Components and Randomization

Because the MobileCoach version used in this study requires dissemination time points for chatbot dialogs to be known a priori, randomization for all intervention components (including sequences of microrandomized component delivery) was performed in advance upon enrollment of participants in the study. The Ally chatbot delivered intervention prompts at random times during the day within prespecified time windows. The random timing allows to observe participants' reaction to intervention notifications in a wide variety of contexts, thereby facilitating the prediction of participants' states of receptivity from smartphone sensor data, a secondary

goal of this study (see above). A detailed overview of the intervention components, including their behavior change techniques (29), randomization, and delivery, is available in the [Supplementary Material \(Section IV\)](#).

### *Incentives*

At the beginning of the study, we randomized participants to receiving either one of two types of financial incentives (cash incentives or charity incentives) or to a no-incentive control group for the duration of the study, with a randomization probability of .33 for each group. Participants in the cash-incentive group received CHF 1 for each day they reached their personalized step goal. Participants in the charity-incentive group earned the same amount, which was donated automatically to a charity organization of the participant's choice. We chose the amount of CHF 1 per day because the lowest incentive value that has produced significant changes in physical activity in previous studies is around US\$ 1 per day ( $\approx$  CHF 1 per day) (17).

### *Planning*

Every Sunday, participants received an action-planning intervention (specifying date, time, and location of up to three brisk walks for the upcoming week), a coping-planning intervention (anticipating up to three barriers for physical activity and planning counter-strategies for the upcoming week), or no planning intervention. Ally sent brief reminders to participants on days when a brisk walk was planned or if a barrier for physical activity was anticipated. Ally sent planning interventions according to a uniform and strongly balanced intervention schedule that controlled for time and carryover effects during the 6 week intervention period. At the beginning of the study, we randomized participants to one of the nine different sequences of the intervention schedule that determined the order of planning and control conditions during the study. To guarantee balance between the sequences, we used blocked randomization with a block size of 9 and a randomization probability of 0.11. Ally delivered planning interventions on Sundays at a random time between 10 AM and 6 PM.

### *Self-monitoring prompts*

Ally supports participants' self-monitoring with brief reminder messages that include the daily step goal, the difference between the user's current step count and the goal, and an estimation of walking minutes needed to reach the goal. Additionally, the reminder messages included an actionable tip on how to increase daily step counts (e.g., "you can add a few steps simply by walking during everyday activities, for example when making a phone call, listening to music or brushing

your teeth"). For each day from Monday through Saturday, we randomly assigned participants to either receive or not receive a self-monitoring prompt (probability 0.50). Participants selected to receive a self-monitoring prompt received a randomly selected prompt from a pool of 18 different prompt conversations. Ally delivered self-monitoring prompts at a random time between 10 AM and 6 PM.

### **Outcomes**

The primary outcome of this study was the proportion of participant days that daily step goals were achieved. We investigated steps per day, obtained from participants' smartphones, as a secondary outcome. Smartphones are capable of accurately measuring physical activity in controlled and uncontrolled settings (30), but smartphones may underestimate absolute step counts in free-living conditions because they are not always continuously carried (31). We measured participants' intrinsic and extrinsic motivation using the Behavioral Regulation for Exercise Questionnaire-2 (32).

### **Statistical Analyses**

To evaluate the effect of each intervention component and the interactions of interest, we aggregated outcomes to the time scale of randomization of the respective intervention component. That is, to estimate the effect of incentives, randomized once at baseline, we compared participants' total proportion of step goals achieved (calculated over the complete intervention period) between the incentive groups using a linear regression model with incentive group membership represented by dummy-coded variables. For planning interventions, randomized weekly, we compared participants' weekly proportion of step goals achieved (calculated separately for each of the 6 weeks of the intervention period) for the various planning conditions. For self-monitoring prompts, randomized daily, we compared the binary daily indicator of step-goal achievement for days when self-monitoring prompts were either present or absent. To estimate the treatment effects of planning interventions and self-monitoring prompts, we followed the analysis approach by Boruvka et al. (33) for data from microrandomized trials. This method produces unbiased causal treatment effects in situations where treatments are repeatedly randomized and covariates are time varying. In our case, this method simplifies to an analysis using generalized estimating equations (GEE) (34) with an independent covariance structure. As in multilevel modeling, GEE models account for the nested structure of longitudinal data. We estimated intervention effects on participants' step counts using the same analysis approach.

We estimated all effects in a complete case analysis using available data only. We conducted sensitivity analyses for missing data (intention-to-treat analysis) and for adjustment of covariates of physical activity (adjusted intention-to-treat analysis). Covariates included in all adjusted models were age, gender, baseline step count, smartphone operating system, and employment. In addition, we adjusted longitudinal models for linear time trends and a binary indicator for weekend days. We investigated time-varying effects of microrandomized intervention components by adding the interaction between interventions and day in study (for self-monitoring prompts) and week in study (for planning exercises) to the model. To account for missing data, we assumed missing data to be “missing at random” (MAR) and used multiple imputation to create 10 complete data sets. We then fitted models to each complete data set separately and pooled the results over all data sets using Rubin’s rules (35). All analyses were prespecified (21).

We conducted a priori power analyses using a simulation-based approach that assumed a proportion of step goals achieved of 50% without interventions, intervention main effects on step-goal achievement of 15%, and interaction effects of 5%. Based on these assumptions, we determined that we needed a sample size of 220 to detect interaction effects with a power of  $1-\beta = .80$ , assuming a type-1 error rate of 5%. We conducted all analyses in R, version 3.5.1 (36).

## Results

### Sample

Of all 30,000 invited insureds, 749 were screened for eligibility and 382 were classified as eligible and provided consent to participate. Of those,  $n = 274$  completed the baseline survey, installed the Ally app, and were randomly assigned to a group and to interventions as described above (Fig. 2). Due to technical errors, six participants did not always receive the interventions they were assigned. Like all participants, we analyzed these six participants according to their randomized intervention schedules. Comparisons of participants’ baseline step counts (Table 1) with large-scale step count data from physical activity app users in Switzerland (37), and of SF-12 component summary scores with the German 12-item Short Form norm sample (38), indicate that (on average) participants in our study were healthier and more active than the general population. On average, participants walked 6,336 steps per day (standard deviation [ $SD$ ] = 2,701) during the baseline period and significantly increased their average daily step count by 438 steps, 95% CI: [134, 742], during the intervention period. However, a graphical illustration of participants’

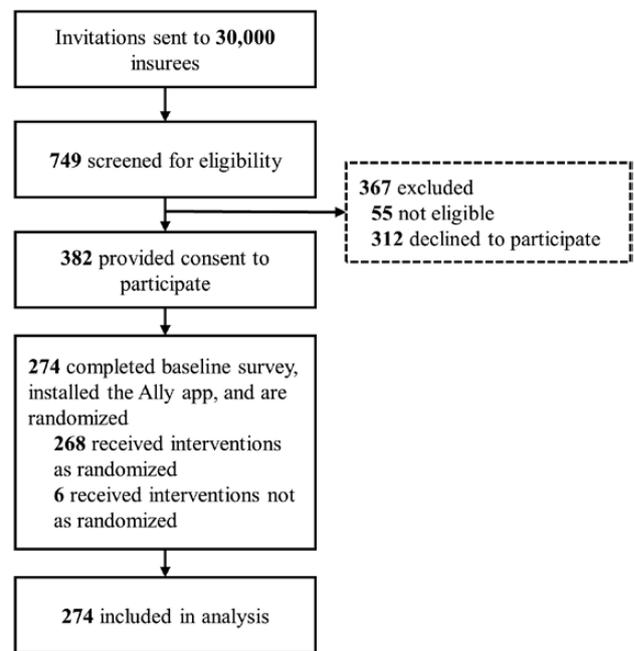


Fig. 2. Participant flow.

step counts over time suggests a curvilinear increase of step counts that starts already during the baseline period and stabilizes around 2 weeks into the intervention period (Supplementary Material, Section IV). The proportion of participants using the Ally app declined over the course of the study. At the end of the study, 83 of 274 participants (30.3%) had stopped using the Ally app. Participants who stopped using the app differed significantly from participants who continuously used the app with regard to steps per day recorded during the intervention period (4,441 steps [ $SD = 2653$ ] vs. 6,979 steps [ $SD = 2,909$ ]) but not with regard to other characteristics, including steps per day at baseline (5,916 steps [ $SD = 2,544$ ] vs. 6,408 steps [ $SD = 2,727$ ]; see the Supplementary Material, Section IV).

### Incentives

Averaged over all study days in the intervention period, participants in the cash-incentive group had an 8.1% greater probability of reaching their daily step goals, 95% CI: [2.1, 14.1], than control-group participants. Charity incentives were associated with a 6.9% greater probability of goal attainment, 95% CI: [1.0, 12.8], but this effect was no longer statistically significant after adjusting the analysis for missing data. We observed that incentives had similar but statistically not significant effects on participants’ step counts. Participants in the cash-incentive group walked on average 783 steps more, 95% CI: [−135, 1,701], than participants in the control group, and participants in the charity-incentive group walked 602 steps more, 95% CI: [−305, 1,509], than participants in the

**Table 1.** Baseline characteristics of participants ( $n = 274$ )

Variable	Value <sup>a</sup>
Age	41.73 (13.54)
Gender	
Female	158 (57.66)
Male	111 (40.51)
Missing	5 (1.82)
Education	
No university degree	100 (36.50)
University degree	164 (59.85)
Missing	10 (3.65)
Employment	
Full-time	152 (55.47)
Part-time	76 (27.74)
Not working	38 (13.87)
Missing	8 (2.92)
Smartphone	
iOS	186 (67.88)
Android	88 (32.12)
Intention to increase physical activity	
Yes	223 (81.39)
No	48 (17.52)
Missing	3 (1.09)
Baseline step count	6,336 (2,701)
Self-reported intrinsic motivation <sup>b</sup>	3.96 (0.88)
Self-reported extrinsic motivation <sup>b</sup>	2.93 (0.75)
Sitting (hr/day) <sup>c</sup>	7.00 [4.00, 9.00]
MVPA (hr/day) <sup>c</sup>	1.75 [1.17, 3.00]
BMI	24.44 (4.15)
SF-12 physical component summary	53.32 (4.58)
SF-12 mental component summary	51.17 (8.11)

*BMI* body mass index; *MVPA* moderate-to-vigorous physical activity.

<sup>a</sup>Reported numbers are mean (standard deviation) for continuous variables and  $n$  (%) for categorical variables unless indicated otherwise.

<sup>b</sup>Items were answered on a five-point Likert scale.

<sup>c</sup>Reported numbers are median [interquartile range] due to nonnormality.

control group. At postintervention follow-up, incentive group differences with regard to extrinsic and intrinsic types of motivation were small and statistically not significant (Supplementary Material, Section IV).

## Planning

Out of three possible plans, participants articulated 0.6 action plans and 0.4 coping plans per week on average. Neither action-planning nor coping-planning interventions significantly affected weekly proportions of step goals achieved or weekly average steps

per day when averaged over incentive conditions (Table 2). Adding the interaction between planning exercises and incentive types revealed larger effects from both planning exercises in the no-incentive control group and decreased effects in the cash- and charity-incentive groups, although neither main effects nor interactions were statistically significant. However, when adjusting the effect estimates for covariates and missing data, there was a statistically significant main effect of action planning on step-goal achievement in the no-incentive control group (5.8%, 95% CI: [1.2, 10.4]) that decreased significantly in the cash-incentive group (interaction term:  $-7.1\%$ , 95% CI:  $[-14.0, -0.1]$ ) and not significantly in the charity-incentive group (interaction term:  $-4.9\%$ , 95% CI:  $[-11.8, 2.0]$ ). The interactions between planning exercises and week in study were statistically not significant. Models of steps per day revealed no meaningful or statistically significant main effects and interactions (Supplementary Material, Section II).

## Self-monitoring prompts

As with the planning interventions, the effect of self-monitoring prompts was statistically not significant for daily step-goal achievement or steps per day (Table 2) when averaged over incentive conditions. When adding the interaction effect between incentive conditions and self-monitoring prompts to the model, the main effect of self-monitoring prompts in the incentive control group and the interaction terms were statistically not significant. Results were similar in models of steps per day (see Supplementary Material, Section II). In the complete case analysis, we found a statistically significant linear change in the effect of self-monitoring prompts on step-goal achievement over time, leading to a statistically significant positive effect of self-monitoring prompts around 4 weeks into the study. However, this time-varying effect was not robust to sensitivity analyses.

## Exploratory Analysis

Effects of planning interventions and self-monitoring prompts likely depend on whether participants engage with the respective intervention content. The MobileCoach platform used for intervention development allowed us to monitor whether participants responded to the initial intervention-related message from the Ally chatbot. In our study, response rates to intervention conversations varied between 40.6% for coping-planning exercises and 55.4% for self-monitoring prompts. The low response rate may provide a possible explanation for the negligible effects of planning interventions and self-monitoring prompts in our prespecified analysis because this analysis averages

**Table 2.** Primary and secondary outcomes

Outcome		Mean ( <i>SD</i> )	Difference to the control condition	
			Main model <sup>a</sup>	Adjusted model <sup>b</sup>
<b>Incentives</b>				
Cash incentives	% of total step-goal achievement	61.4 (17.6)	<b>8.1 [2.1; 14.1]</b>	<b>8.9 [3.0; 14.8]</b>
	Total mean steps/day	7,382 (2,851)	783 [−135; 1,701]	887 [−50; 1,832]
Charity incentives	% of total step-goal achievement	60.2 (19.9)	<b>6.9 [1.0; 12.8]</b>	5.0 [−0.9; 10.9]
	Total mean steps/day	7,201 (2,824)	602 [−305; 1,509]	635 [−154; 1,424]
No incentives	% of total step-goal achievement	53.3 (16.3)	–	–
	Total mean steps/day	6,599 (2,599)	–	–
<b>Planning</b>				
Action planning	% of weekly goal achievement	59.3 (26.7)	1.1 [−2.0; 4.1]	1.8 [−1.1; 4.7]
	Weekly mean steps/day	7,386 (3,194)	101 [−163; 366]	148 [−67; 363]
Coping planning	% of weekly goal achievement	57.9 (25.6)	−0.3 [−3.1; 2.5]	−0.6 [−3.6; 2.3]
	Weekly mean steps/day	7,172 (3,066)	−113 [−351; 125]	−44 [−243; 154]
No planning	% of weekly goal achievement	58.2 (24.9)	–	–
	Weekly mean steps/day	7,285 (2,964)	–	–
<b>Self-monitoring prompts</b>				
Prompt	% of daily goal achievement	60.5 (48.9)	1.1 [−1.1; 3.2]	1.6 [−0.5; 3.6]
	Steps/day	7,507 (4,137)	43 [−114; 200]	32 [−106; 169]
No prompt	% of daily goal achievement	59.5 (49.1)	–	–
	Steps/day	7,464 (4,036)	–	–

Differences to control conditions correspond to point estimates with 95% confidence intervals in brackets. Boldface indicates statistical significance at  $\alpha \leq .05$ . *SD* standard deviation.

<sup>a</sup>Model based on complete cases without adjustment for covariates.

<sup>b</sup>Model adjusted for missing data and covariates of physical activity (age, gender, baseline step count, smartphone operating system, and employment). Effects for planning interventions and self-monitoring prompts were further adjusted for linear time trends and repeated measures.

effects over all participant days irrespective of participants' engagement with the intervention. We, therefore, decided to conduct an exploratory analysis of the intervention components' main effects with recoded treatment indicators that differentiate whether a participant engaged with the intervention content or not (see [Supplementary Material Section III](#) for details). This analysis was not prespecified. Here, we define “engagement with the intervention” as responding to the first message of the chatbot in an intervention-related conversation. To adjust for possible confounding in this analysis, we added known covariates of physical activity to the model.

Compared to not receiving a self-monitoring prompt, engaging with self-monitoring prompt conversations was associated with a significantly increased likelihood of step-goal achievement (5.5%, 95% CI: [2.7, 8.2]) and significantly higher step counts (405 steps, 95% CI: [189, 621]). Conversely, not engaging with self-monitoring prompt conversations was associated with

a significantly lower probability of reaching daily step goals (−9.3%, 95% CI: [−13.0, −5.6]) and lower step counts (−754 steps, 95% CI: [−1,077, −431]). Likewise, compared to not receiving a weekly planning exercise, engaging with action-planning and coping-planning conversations was associated with a higher proportion of step-goal achievement—which was statistically significant for action planning (3.9%, 95% CI: [0.2, 7.6]) but not for coping planning (3.5%, 95% CI: [−0.03, 7.0])—and associated with significantly higher step counts for both planning exercises (action planning: 421 steps, 95% CI: [127, 715]; coping planning: 475 steps, 95% CI: [128, 822]). Not engaging with planning conversations was associated with a lower proportion of step goals met, but this difference was not statistically significant (action planning: −2.0%, 95% CI: [−6.5, 2.5]; coping planning: −3.4%, 95% CI: [−7.3, 0.5]). Not engaging with planning conversations was also associated with lower step counts, which was statistically significant for coping planning (−579 steps, 95% CI:

[−942, −216]) but not for action planning (−250 steps, 95% CI: [−701, 201]).

## Discussion

This optimization trial quantified the effects of three different physical-activity interventions of a smartphone application to promote walking. Daily cash incentives—but not charity incentives—promoted the achievement of personalized step goals during the 6 week intervention period. This result is in line with previous research that has reported positive effects of cash incentives on physical activity (17) and contributes to the existing research on charity incentives, which has so far reported mixed results (39, 40). We found that cash and charity incentives had no effect on participants' self-reported levels of intrinsic motivation, although activity levels and intrinsic motivation at baseline were high. Previously, low baseline levels of behavior and low levels of intrinsic motivation have been used to explain missing effects of incentives on intrinsic motivation in studies of health behaviors (41). The results from this study, however, suggest that incentives can affect behavior without undermining intrinsic motivation even if baseline levels of intrinsic motivation and behavior are high. Presumably, the incentives in our study were too small to be perceived as “controlling” by participants and, consequently, did not reduce intrinsic motivation.

We found no meaningful and statistically significant effects for the remaining interventions: self-monitoring prompts, action planning, and coping planning. This result contrasts with previous research that reported positive effects of similar intervention prompts (42) and planning interventions (14) on physical activity. However, action planning appeared to be an effective intervention for participants who did not receive financial incentives but had no effect for participants in either of the incentive groups. These interaction effects can possibly be explained by the high baseline activity levels of study participants. Specifically, study participants may have had only limited room for increasing their daily physical activity in response to both interventions. Thus, in addition to cash incentives, action planning can be a powerful intervention, although effects of both components may not fully unfold if offered simultaneously.

This study also revealed important limitations of both planning exercises and self-monitoring prompts that, together with absent main effects, suggest that these components should be revised if they are to be included in future versions of the app. Most importantly, participants responded only to roughly half of the self-monitoring prompt and planning conversations that were initiated by the Ally chatbot, ultimately limiting the potential of both interventions to affect behavior. Indeed, engaging

with self-monitoring prompt and planning conversation was associated with increased walking in our exploratory analysis. Although this finding is in line with a possible true intervention effect, it may also reflect the effect of unobserved confounding variables. For example, the time that participants carried their smartphone could affect both recorded step counts and responding to intervention conversations. In fact, the significantly lower step counts, which were associated with ignoring self-monitoring prompt and coping planning conversations, can most plausibly be explained by the presence of confounding variables. Thus, while the exploratory analyses remain inconclusive regarding effects of coping planning and self-monitoring prompts, they illustrate that reaching and engaging participants can be a major challenge for smartphone-based physical activity interventions.

A further limitation of planning exercises and self-monitoring prompts in the present study was that they were delivered to participants at random times, which may have contributed to the limited engagement with the interventions. One possible strategy to increase engagement with the interventions is to use intelligent notification management algorithms that use smartphone sensor data to predict opportune moments for intervention delivery, that is, moments when participants will most likely react to a smartphone push notification. In fact, research in the field of interruptibility has revealed that the application of such algorithms increases the response rate to smartphone notifications (43). As a consequence, tailoring the delivery of self-monitoring prompts and planning exercises to those opportune moments could increase the intervention's likelihood of success.

In addition, the low number of action and coping plans completed by participants indicate that planning interventions may have been too great a burden for participants. Indeed, planning up to three brisk walks per week or planning coping strategies for potential barriers requires a considerable amount of participants' time and cognitive resources. In our follow-up survey, around 35% of participants found the planning exercises difficult to complete, supporting the conclusion that the burden of planning interventions has prevented at least some participants from completing their weekly plans. Furthermore, 50% of participants indicated that they would prefer to plan their activity or coping strategies on a daily basis rather than a weekly basis. Thus, one way to reduce the burden of planning interventions could involve prompting participants to plan one activity or coping strategy daily, for example, prior to days with typically lower step counts.

Because the Ally app, like most mHealth apps, collects and visualizes data and provides feedback and an easy way to monitor goal progress, it holds inherent value for a health-conscious and motivated subgroup of the population (11). Indeed, baseline characteristics indicate that our

study attracted individuals who were more active, educated, and healthy than the general Swiss population. In addition, a considerable proportion of participants stopped using the app over the course of the 8 week study, although attrition appeared to be somewhat lower compared to similar studies of mHealth apps (42, 44). Interestingly, participant attrition was similar in the different incentive groups, suggesting that incentives tied to behavior do not necessarily promote engagement with the app. Unfortunately, selection effects and lack of sustained participant engagement can limit the health impact of physical activity apps (and other health apps) considerably. Future research, therefore, needs to identify effective strategies for reaching more vulnerable populations and engaging health-app users in the long term.

There are some limitations to the present research. To begin with, the selective sample in our study may limit the generalizability of the reported intervention effects. For example, there may be no interaction effect between action planning and incentives in less-active populations. Furthermore, we were not able to separate increases in step counts from increases in the time that participants were carrying their smartphone. It is possible that the reported intervention effects reflect, at least partly, increases in the time the smartphone was carried. Lastly, although all participants indicated upon enrollment that they were not using any comparable apps or devices for tracking physical activity, we cannot exclude that such apps or devices were used or that participants primarily used the Apple Health or GoogleFit applications that were required for the Ally app to obtain step counts. Use of such additional apps or devices could potentially affect the use of the Ally app and the effectiveness of interventions.

## Conclusion

Daily financial incentives seem to be a suitable intervention as part of the Ally app. Notification-dependent interventions, that is, planning exercises and self-monitoring prompts, are limited by low engagement of participants with the interventions and high intervention burden and, therefore, require revision. Selection effects and attrition appear to be important challenges that can restrict the health impact of physical-activity apps.

## Supplementary Material

Supplementary material is available at *Annals of Behavioral Medicine* online.

**Acknowledgments:** This study is funded in part by CSS Insurance, Switzerland. CSS Insurance supported the recruitment of study

participants but had no role in app development, study design, data collection, data analysis and interpretation, writing the manuscript, or reviewing and approving the manuscript for publication.

## Compliance with Ethical Standards

**Authors' Statement of Conflict of Interest and Adherence to Ethical Standards** Elgar Fleisch co-chairs the Center for Digital Health Interventions (CDHI), a joint initiative between 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 scientific director and Jan-Niklas Kramer and Florian Künzler are doctoral researchers at the CDHI. Elgar Fleisch and Tobias Kowatsch are also co-founders of Pathmate Technologies, a university spin-off company that delivers digital clinical pathways and has used the open-source platform MobileCoach for that purpose too. However, Pathmate Technologies is not involved in the intervention described in this paper. Varun Mishra, Shawna N. Smith, David Kotz, and Urte Scholz report no conflicts of interest.

**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, David Kotz, and Tobias Kowatsch were responsible for app design and implementation. Jan-Niklas Kramer, Florian Künzler, Tobias Kowatsch, Shawna Smith, David Kotz, and Elgar Fleisch developed the study design and methodology. Jan-Niklas Kramer and Shawna Smith analyzed the data. Jan-Niklas Kramer wrote the manuscript incorporating critical reviews from all authors.

**Ethical Approval** The study was performed in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments and was approved by the ethical review board of ETH Zurich.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## References

1. Arem H, Moore SC, Patel A, et al. Leisure time physical activity and mortality: A detailed pooled analysis of the dose-response relationship. *JAMA Intern Med.* 2015;175:959–967.
2. World Health Organization. *Global Recommendations on Physical Activity for Health.* Geneva, Switzerland: World Health Organization; 2010.
3. Murphy MH, Nevill AM, Murtagh EM, Holder RL. The effect of walking on fitness, fatness and resting blood pressure: A meta-analysis of randomised, controlled trials. *Prev Med.* 2007;44:377–385.
4. Morris JN, Hardman AE. Walking to health. *Sports Med.* 1997;23:306–332.
5. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. *Nat Biotechnol.* 2016;34:239–246.
6. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: Behavior change techniques, systematic review and meta-analysis of randomized controlled trials. *Ann Behav Med.* 2016;51:226–239.
7. Romeo A, Edney S, Plotnikoff R, et al. Can smartphone apps increase physical activity? systematic review and meta-analysis. *J Med Internet Res.* 2019;21:e12053.

8. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: Making decisions based on the results of a factorial screening experiment. *Transl Behav Med.* 2014;4:238–251.
9. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychol.* 2015;34S:1220–1228.
10. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors: The health action process approach. *Eur Psychol.* 2008;13:141–151.
11. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *J Med Internet Res.* 2017;19:e125.
12. Carver CS, Scheier MF. Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychol Bull.* 1982;92:111–135.
13. 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;28:690–701.
14. Bélanger-Gravel A, Godin G, Amireault S. A meta-analytic review of the effect of implementation intentions on physical activity. *Health Psychol Rev.* 2013;7:23–54.
15. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *J Med Internet Res.* 2019;21:e14567.
16. Madison K, Schmidt H, Volpp KG. Smoking, obesity, health insurance, and health incentives in the Affordable Care Act. *JAMA.* 2013;310:143–144.
17. Mitchell MS, Orstad SL, Biswas A, et al. Financial incentives for physical activity in adults: Systematic review and meta-analysis [published online ahead of print May 15, 2019]. *Br J Sports Med.* 2019. doi: [10.1136/bjsports-2019-100633](https://doi.org/10.1136/bjsports-2019-100633).
18. Deci EL, Ryan RM. *Intrinsic Motivation and Self-Determination in Human Behavior.* New York, NY: Plenum, 1985.
19. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull.* 1999;125:627–668; discussion 692.
20. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: A systematic review. *Int J Behav Nutr Phys Act.* 2012;9:78.
21. Kramer JN, Künzler F, Mishra V, et al. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: Protocol of a microrandomized trial. *JMIR Res Protoc.* 2019;8:e11540.
22. Nass C, Steuer J, Tauber ER. Computers are social actors. Presented at: Conference on Human Factors in Computing Systems; April 24–28, 1994. Boston, MA; 1994.
23. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: A systematic review. *Lancet.* 2001;357:757–762.
24. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chic).* 2018;55:316–340.
25. Adams MA, Hurley JC, Todd M, et al. Adaptive goal setting and financial incentives: A 2 × 2 factorial randomized controlled trial to increase adults' physical activity. *BMC Public Health.* 2017;17:286.
26. Tudor-Locke C, Hatano Y, Pangrazi RP, Kang M. Revisiting “how many steps are enough?”. *Med Sci Sports Exerc.* 2008;40:S537–S543.
27. Filler A, Kowatsch T, Haug S, et al. 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. Presented at: 14th Annual Wireless Telecommunications Symposium (WTS 2015), April 15–17 2015; New York, NY. New York, NY: IEEE; 2015: 1–6.
28. Statcounter GlobalStats. Mobile Operating System Market Share in Switzerland, February 2019. Available at <http://gs.statcounter.com/os-market-share/mobile/switzerland/>. Accessibility verified March 21, 2019.
29. Michie S, Richardson M, Johnston M, 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;46:81–95.
30. Hekler EB, Buman MP, Grieco L, et al. Validation of physical activity tracking via android smartphones compared to actigraph accelerometer: Laboratory-based and free-living validation studies. *JMIR Mhealth Uhealth.* 2015;3:e36.
31. Duncan MJ, Wunderlich K, Zhao Y, Faulkner G. Walk this way: Validity evidence of iphone health application step count in laboratory and free-living conditions. *J Sports Sci.* 2018;36:1695–1704.
32. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *J Sport Exerc Psychol.* 2004;26:191–196.
33. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. *J Am Stat Assoc.* 2018;113:1112–1121.
34. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. *Biometrics.* 1988;44:1049–1060.
35. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* Hoboken, NJ: John Wiley & Sons; 2004.
36. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2014.
37. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. *Nature.* 2017;547:336–339.
38. Morfeld M, Kirchberger I, Bullinger M. *SF-36 Fragebogen Zum Gesundheitszustand: Deutsche Version des Short Form-36 Health Survey.* Göttingen, Germany: Hogrefe; 2011.
39. Finkelstein EA, Haaland BA, Bilger M, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): A randomised controlled trial. *Lancet Diabetes Endocrinol.* 2016;4:983–995.
40. 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:e123–e130.
41. Promberger M, Marteau TM. When do financial incentives reduce intrinsic motivation? Comparing behaviors studied in psychological and economic literatures. *Health Psychol.* 2013;32:950–957.
42. Shcherbina A, Hershman SG, Lazzeroni L, et al. The effect of digital physical activity interventions on daily step count: A randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. *Lancet Digit Health.* 2019;1:e344–e352.
43. Künzler F, Kramer J-N, Kowatsch T. Efficacy of mobile context-aware notification management systems: A systematic literature review and meta-analysis. Presented at: Fifth International Workshop on Pervasive and Context-Aware Middleware (PerCAM'17); October 9. Rome, Italy; 2017.
44. Dorsey ER, Yvonne Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The use of smartphones for health research. *Acad Med.* 2017;92:157–160.