MyBehavior: Automatic Personalized Health Feedback from User Behaviors and Preferences using Smartphones

Mashfiqui Rabbi[1], **Min Hane Aung**[1], **Mi Zhang**[2], **Tanzeem Choudhury**[1]
[1]*Cornell University,* [2]*Michigan State University*
ms2749@cornell.edu, msa242@cornell.edu, mizhang@egr.msu.edu, tanzeem.choudhury@cornell.edu

ABSTRACT
Mobile sensing systems have made significant advances in tracking human behavior. However, the development of personalized mobile health feedback systems is still in its infancy. This paper introduces MyBehavior, a smartphone application that takes a novel approach to generate deeply personalized health feedback. It combines state-of-the-art behavior tracking with algorithms that are used in recommendation systems. MyBehavior automatically learns a user's physical activity and dietary behavior and strategically suggests changes to those behaviors for a healthier lifestyle. The system uses a sequential decision making algorithm, *Multi-armed Bandit*, to generate suggestions that maximize calorie loss and are easy for the user to adopt. In addition, the system takes into account user's preferences to encourage adoption using the *pareto-frontier* algorithm. In a 14-week study, results show statistically significant increases in physical activity and decreases in food calorie when using MyBehavior compared to a control condition.

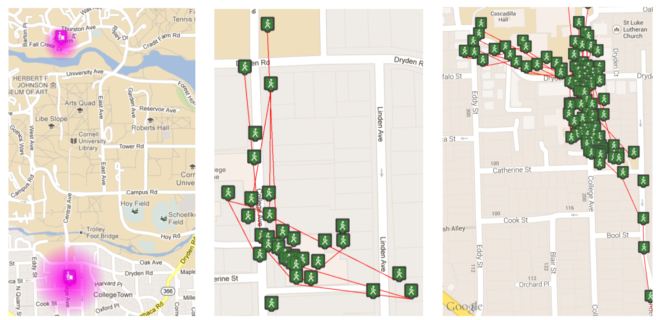
Author Keywords
Mobile Phone Sensing, Machine learning, Mobile Health, Health Feedback

ACM Classification Keywords
H.1.2 User/Machine Systems; I.5 Pattern Recognition

General Terms
Systems, Design, Experimentation, Scalibility, Performance

INTRODUCTION
In 2007, the World Health Organization (WHO) declared obesity as a global epidemic [9]. Over one-third of the US adult population is classified as obese [39]. Obesity is now a public health issue and addressing obesity-related problems is beyond the capacity of the healthcare industry [42]. Therefore, scalable solutions that can promote healthier lifestyles outside of clinical settings are desirable.

One way to tackle obesity is to create a calorie deficit by way of decreased food intake and increased physical activity. There is a growing trend in the development of tracking devices and applications to monitor and regulate physical activity and food intake [38][28]. However, existing work makes little use of the tracked data to provide personalized feedback that fit well into a users's routine. Feedback is often limited to either overall statistics [11][25], visualization of entire self-tracked data [33][15] or generic suggestions [24][46] that are not personalized to a user's behaviors and lifestyle. However, we can go beyond these paradigms and take advantage of more fine-grained information contained in the data. With a deeper analysis of the self-tracked data, patterns that characterize both healthy and unhealthy behavior can be revealed. These patterns then can be leveraged to generate personalized and actionable suggestions that relate to a user's behaviors.



(a) (b) (c)

Figure 1: Visualization of user behaviors over a week (a) Heatmap of places a user stayed stationary (b) Location traces of frequent walks for the same user (c) Location traces of frequent walks for another user.

To this end, we created a mobile application called MyBehavior. The novel functionality is an intelligent engine that provides personalized suggestions by learning a user's physical activity and dietary behaviors. For example, Figure 1(a-b) show learnt behaviors of one user's stationary locations and the routes of frequently taken short walks over a week. Then suggestions are issued to take small walks near the stationary locations (Figure 1a) or continue with the existing walk (Figure 1b). Similar contextualization of suggestions can be done on a per person basis. For instance, Figure 1c shows one walking behavior over a week for a different user which can be used for personalized suggestions.

MyBehavior's intelligent suggestion engine is built upon two well-known decision theory models. The first is the multi-armed bandit (MAB) [45] which dynamically learns and influences user behaviors by suggesting actions that maximize the chances of achieving calorie loss goals. Maximization is achieved by strategically suggestion a combination of frequent and infrequent healthy behaviors. The frequent vs. infrequent behavior suggestions map to the "$exploit$ vs. $explore$" principle that often underpins MABs. For example, if the user makes a 20 minute walk to work 4-5 days a week and goes to the gym once or twice a month then MyBehavior would more often suggest that the user walk to work and would occasionally suggest to increase gym visits. The assumption is that walking to work is more regular and will be *lower-effort* to adopt while also yielding more aggregate calorie loss compared to going to the gym. Prioritizing frequent behaviors also means that these behaviors are practiced and therefore the user is likely to be good at those actions (i.e., users have *mastery* or *self-efficacy*). Such low-effort change and self-efficacy are well-grounded in persuasion [16] and behavior change theories [4]. A further function of the system is to keep users in the loop by giving users control to prioritize suggestions that they prefer to follow. User preferences are then balanced with the machine generated suggestions using the second decision theory model, the pareto-frontier [48].

The blending of algorithms with behavioral theories into a usable and deployable mobile application required several iterations of research and development. We previously published early ideas of health feedback automation along with a feasibility pilot study [47]. The previous version used MAB to generate suggestions without considering user preference. This paper presents MyBehavior 2.0 which builds and improves on our earlier work and conducts more extensive testing and evaluation. Specifically the contributions include:

1. The design of an improved system to create actionable suggestions that takes into account both users behaviors and preferences. MyBehavior interfaces allowed users to easily input their preferences. User preferences and behaviors are utilized to generate a set of suggestions using Multi-armed bandit and pareto-frontier. Both of these models operationalizes the principles of behavior change theories.
2. A energy efficient, deployable android application that provides automated feedback based on real-time activity tracking, food logging and user preferences,
3. A 14-week study with 16 participants to demonstrate MyBehavior's efficacy quantitatively. Participants using MyBehavior followed more suggestions with more calorie loss (increased activity or decreased calorie intake) compared to a control condition with prescribed recommendations from health experts. These improvements lasted beyond the initial novelty period and continued over 5-9 weeks.

MOTIVATING DESIGN OF MYBEHAVIOR

In this section, we discuss the motivation and vision that led to the development of MyBehavior.

Low effort and self-efficacy: Social cognitive theory [4] suggests that in order to voluntarily initiate an action, a person needs a sense of self-efficacy or confidence to perform the action. The more frequently a person performs an action in a certain context the more self-efficacy increases and the less effortful the action is perceived to be. The Fogg behavior model applies this theoretical principle to technology design by creating tools to prompt low effort actions that can be triggered even when motivation is low [16]. MyBehavior leverages the principles of low effort and self-efficacy to create suggestions that focus on repeated actions in distinct contexts.

Personalization of suggestions: Proponents of small data [14] and N-of-1 interventions [53][17] argue that each individual is unique and heterogeneous. This uniqueness means personalized intervention should perform better than one-size-fits-all interventions that may fail to satisfy a person's specific requirements. To our knowledge, so far such personalization is provided only through human health coaches. MyBehavior aims to build an automated suggestion generation system that personalizes without a human health coach.

Mobile recommender system for health feedback: Over the last decade, search engines (e.g., Google, Bing) have transformed the way we acquire information. Similarly, movie [37] or news [27] recommendation systems influence our media consumption. However, no automated health recommendation system has yet utilized the vast amount of personal data collected using mobile or wearable devices. To our knowledge, MyBehavior is a first step in filling this gap and is the first adaptive health suggestion generator. MyBehavior also tackles the practical challenges of usability, privacy and battery life.

MYBEHAVIOR APPLICATION DEVELOPMENT

The development of automated health suggestion generation from logged data has little precedence. We had to overcome both technical and user-centered challenges in order to create a system that can promote change for real world users. To this end, MyBehavior was honed using an iterative development process that spanned nearly 2.5 years. During this period, several MyBehavior prototypes were created and deployed. These deployments revealed a core set of requirements that an automated health feedback application needs to satisfy. Here we describe the two significant versions of MyBehavior from this iterative process.

MyBehavior 1.0: An Automated Health Suggestion Engine with Multi-armed Bandit Algorithm

MyBehavior 1.0 solves the important hurdle of transforming raw log data into personalized health suggestions. This system is comprised of two modules: (1) a logging and behavior mining module to track and mine user behaviors, (2) an automated suggestion generating module that utilizes the behavior data to suggest small changes that maximize chances of calorie loss. Details of MyBehavior 1.0 along with results from a 3-week pilot deployment to determine feasibility of automated feedback and usability concerns can be found in [47]. This paper significantly extends the previous work by developing the MyBehavior 2.0 system and by quantitatively evaluating its effectiveness through a 14-week user study. To provide context, we briefly describe key aspects and findings of MyBehavior 1.0 with more technical details.

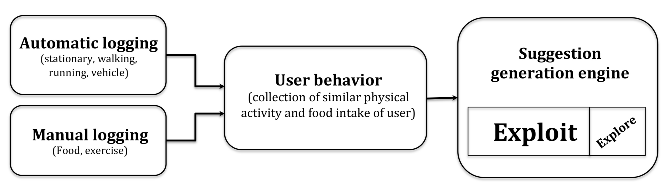
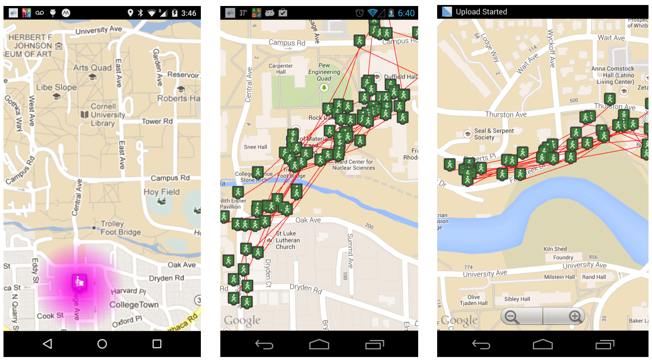


Figure 2: MyBehavior 1.0 processing pipeline

Logging and Behavior mining module

MyBehavior 1.0 uses a combination of automatic sensing and manual logging to record user's food intake and physical activity. Stationary, walking, running, and driving activities are automatically inferred using the technique described in [28].

The inferred activities are also tagged with corresponding locations. Physical activities that are not automatically tracked can be manually logged from a database of 800 activities [1]. Foods can be manually logged using the USDA database which contains more than 8000 food items [19].



(a) **(b)** **(c)**

Figure 3: A few clusters representing different user behaviors (a) a stationary cluster (b) a walking cluster (c) another walking cluster

Unsupervised clustering algorithms are then used to find different user behaviors from log data. Food categories are clustered based on common ingredients. For example, different types of burgers would be clustered together if they share a common bread or meat type. Similarly, manually logged exercises are also grouped by type. For instance, yoga or other types of gym workout would be clustered together. The automatically classified physical activities, namely stationary, running and walking, are also sub-categorized into similar behavior classes using clustering. For stationary events, GPS distance is used to separate different stationary episodes, for example being stationary at home would be a different cluster from being stationary at work. For walking and running activities, discretized Fréchet distance [52] is used to match similar walking or running trajectories in a computationally efficient manner. A dataset of activities collected from 20 users is utilized to derive appropriate thresholds for clustering. The BIRCH online clustering algorithm [55] is used to efficiently cluster activities. Figure 3 shows a few generated clusters representing different user behaviors.

Suggestion generation module

Once MyBehavior learns user behaviors, it determines a set of high calorie loss suggestions that involve small changes to the user's existing behaviors. In the following, we describe how we model this goal as an algorithmic optimization problem.

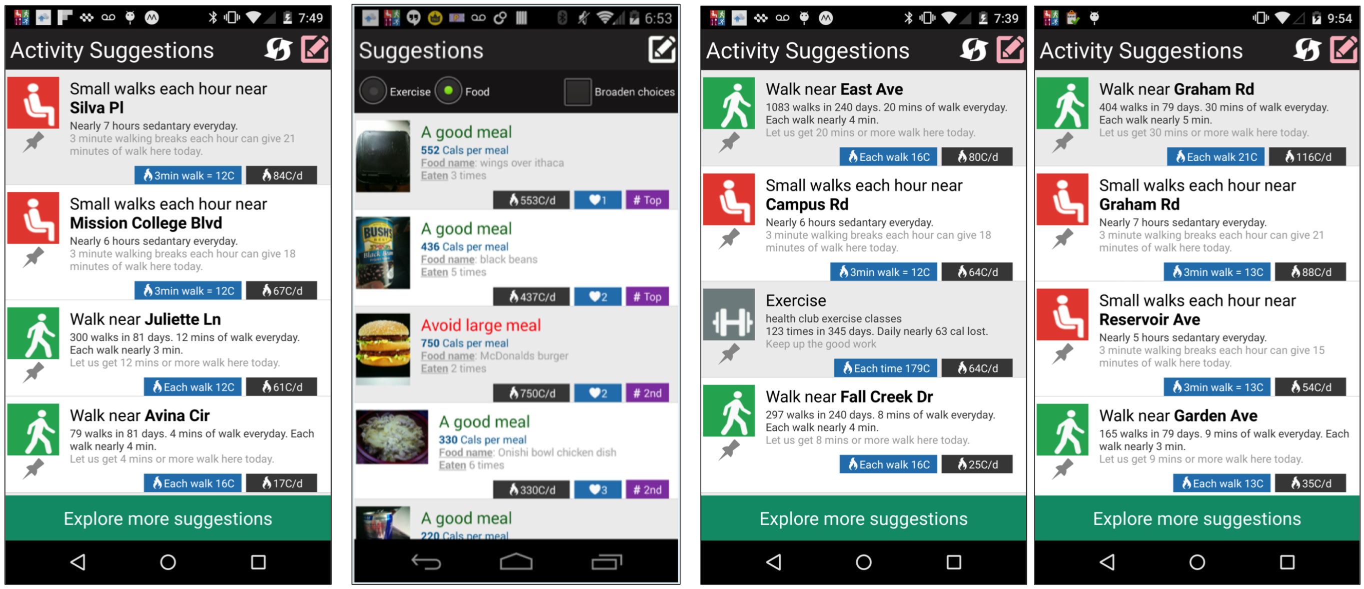
Any optimization algorithm requires an appropriate objective function. In MyBehavior, this function is grounded in principles of persuasion and behavior change theories. Users often take actions that are easy to do [16]; from a psychological perspective, an action is easy if it relates to a user's lifestyle [1] and has been frequently done before [4]. Given this insight, MyBehavior sets up the objective function as the multiplication of frequency of a user behavior and average calorie benefit when that behavior is performed. Frequency of a behavior is simply the size of the behavior's corresponding cluster. For dietary behaviors, the average calorie count is the mean of all foods in its corresponding cluster. For activity clusters, we use the metabolic equivalent of task (MET) scale [26] to get calorie count for each activity instance in the clusters. Finally, the mean calorie loss is calculated over the whole cluster.

A simple suggestion making strategy could be a list of suggestions ranked according to their frequency and average caloric benefit. However such a simple strategy would not take into account a person's lifestyle changes over time (e.g., seasonal changes or major life events). Moreover, MyBehavior's influence in itself may also cause changes. Formally, this would mean that past frequent behavior would not be entirely comprehensive and future proof. One approach to protect against such future scenarios is to *exploit* most frequent past calorie loss behaviors, while including a few suggestions that *explore* past infrequent behaviors to see whether the user starts doing them frequently. Finally, personal data is limited and scarce. Thus the algorithm should not be highly parameterized nor should it require a lot of data in order to generate useful suggestions.

The multi-armed bandit (MAB) algorithm can address the aforementioned issues in its core optimization process. To illustrate this process, we briefly introduce the classic MAB problem. Consider a scenario where a gambler needs to sequentially choose from a set of slot machines with different reward distributions initially unknown to the gambler. Each time an arm is selected (pulled), a reward is drawn from that arm's reward distribution. The goal is to maximize the long term cumulative rewards obtained from the slot machines. Stated this way, the tradeoffs between explore and exploit is straightforward. Clearly, the long term reward would be maximized by pulling the arm whose mean payoff is the highest (exploitation). Finding this arm however, entails exploration - each arm pull provides incremental information about the payoff distribution for that particular arm. MyBehavior's suggestion making algorithm faces the same problem as the gambler. Initially, a user's most frequent calorie burning behaviors are unknown to the system. Over time these behaviors are revealed once food and physical activities are logged and clustered. However, potential changes in future behaviors cannot be known. Therefore the system also explores like the gambler by suggesting non-frequent behaviors to see if the user will frequently adopt them in the future.

Exploit-explore is common to all bandit algorithms but there are different strategies that are used [7]. In MyBehavior we use a EXP3 strategy [7]. In EXP3, most beneficial actions are frequently exploited with seldom exploration of less beneficial ones. In MyBehavior, 90% of the suggestions are exploited from frequent behaviors associated with high-calorie values and 10% of the time non-frequent behaviors are explored. One of the features of EXP3 is that it can adapt to changes in underlying payoff functions. This means if the user starts following new suggestions or the user's lifestyle changes (e.g., moving to a new location) then underlying caloric benefits of certain behavior will change. EXP3 strategy would tune to those changed circumstances quickly.

MyBehavior generates 10 food and 10 activity suggestions separately. The suggestion engine does not mix foods with exercise as the joint space of possibilities would make it a



(a) (b) (c) (d)

Figure 4: MyBehavior app screenshots (a) a set of activity suggestions for a user (b) a set of food suggestions for the same user (c) a set of suggestions at a different time for the same user (d) a set of activity suggestions for a different user

combinatorially hard problem and lead to more complicated suggestions. For activity suggestions, changing stationary behavior is added to the mix of walking, running or manually input exercise suggestions. MyBehavior suggests users to change every hour of stationary event in a specific location with 3 minutes of walking. Such a mix often results into non-trivial changes in suggestions: for instance, Fig. 4(c) shows a ranking of MyBehavior suggestions where simply changing regular stationary episodes with 3 minutes of walk for every stationary hour can yield more calorie expenditure compared to the user's gym exercises. Regarding food suggestions, a separate bandit generates food suggestions that take into account intake frequency and calorie. MyBehavior makes a distinction between suggestions for meals and those for snacks, as the number of calories consumed can be different for these two food clusters.

Figure 4 shows different suggestions generated by MyBehavior. As seen in the screenshots, semantically meaningful messages are added with every suggestion. For suggestions generated by exploiting, MyBehavior asks the user to either continue positive activities (i.e., good calorie foods, walking, or exercise), make small changes in some situations (i.e., stationary activities), or avoid negative activities (i.e., frequent large meals). On the other hand, suggestions generated during exploration phase, the system asks the users to consider trying out the suggestions. All MyBehavior suggestions change overtime and are different for different users. Figure 4(a) and (c) are physical activity suggestions from the same user on different days. Figure 4(d) shows suggestions generated for a different user demonstrating the scalability of the system.

Finally, modeling MyBehavior as a MAB also has additional advantages. MAB is an online algorithm and incrementally makes decisions in a computationally efficient manner. This means MyBehavior can compute all suggestions inside the phone which is an added privacy feature. MABs also have fewer parameters and are easy to learn.

Deployment and lessons learned

MyBehavior 1.0 was deployed in a 3-week pilot with 9 users (4 female). At the end of this study, we conducted semi-structured interviews with the participants about their experiences with MyBehavior. We also asked the participants to indicate whether they would be willing and able to follow each suggestion on an average day on a scale of 1-5 (5=Strongly Agrees that s/he can follow the suggestions; 1=Strongly Disagrees). Each participant rated 15 suggestions.

In the interviews, users reported MyBehavior suggestions to be actionable. In the suggestion rating survey, MyBehavior received an average of 3.4 out of 5 ($\mu = 3.4; \sigma = 1.4$). However, several areas of improvement are also identified. They are as follows:

1. Difficulty in manual logging: Users reported the manual food logging process to be self-reflective. However, they also found the searching and adding appropriate food items to be long and cumbersome. Furthermore, many manually added exercises were repeatedly done (e.g., gym or fitness classes). Users wanted quick ways to add repeated exercise rather than searching them every time.
2. Lack of human control: Although MyBehavior can dynamically adapt to lifestyle changes, on occasion MyBehavior was slow to adapt. For example, a user regularly played soccer with his friend but when his friend moved to a new location he could no longer do that activity. The user was frustrated that he could not remove the soccer suggestion. On the other hand, users are sometimes highly motivated about certain activities that they did not repeat much in the past. For example, several users wanted, "going to the gym" as a top suggestion even though they did not frequently go to gym in the past.

The initial 3-week pilot study was not long enough to show statistically significant changes in behavior. But, the study confirmed that automatically generated suggestions are indeed actionable and provided important usability feedback.

MyBehavior 2.0: Easier logging and human-in-the-loop

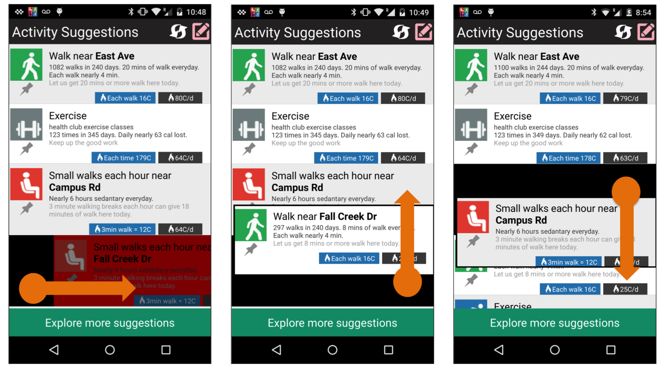
Based on insights and user feedback from the first pilot, we develop MyBehavior 2.0. We focus on easier ways to log food and exercise. In addition, we include the provision for user customization on MyBehavior generated suggestions. Below we describe these changes in detail.

Easier logging

To reduce the burden of food logging, MyBehavior 2.0 contains a crowd-sourcing functionality that returns calorie information using photographs of the food taken by the user. This is similar to the implementation of the PlateMate framework by Noronha et al. [38] which coordinates Amazon Mechanical Turk [3] (AMT) labelers to ascertain nutritional information. Results from PlateMate deployment showed the nutritional information derived using the AMT was comparable to trained dieticians. In a pilot test, we found that AMT labelers often made mistakes if a very large database was used. This is overcome by using a smaller list of the 40 most frequently selected foods by nearly 50 million MyFitnessPal [33] mobile app users. For each image, 5 AMT labelers are asked for calorie information and the median is used as the final calorie number. In addition, we ask the AMT labelers to rate similarity of a food image with 16 other previous food images from the user. This similarity information is used for clustering similar foods. In a test image set of 50 images, we found an average of only ± 70 calories between ground truth values and the AMT labeled calorie estimate. In addition to this reduced burden in food logging, MyBehavior 2.0 also allows users to select from list of their past exercises for easier manual exercise logging.

Incorporating Human Customization in Suggestions

The second major modification in MyBehavior 2.0 is giving the user control to customize suggestion set. This is achieved by allowing the user to remove the suggestions the user does not want or is unable to follow due to a change in circumstances. In terms of interaction, users can swipe from left to right and remove suggestions (Figure 5(a)); a removed suggestion is never considered in the future. In addition, MyBehavior 2.0 allows the users to re-sort the suggestions in order of their preference. The user can long-press a suggestion and move the suggestion above or below another suggestion (Figure 5(b-c)). For instance, if a user prefers to go to the gym even though s/he did not do it often before, the user can simply move the gym suggestion to the top. This re-sorting creates a new ranking based on the user's preference in addition to the system generated suggestions. With this dual information, a final ranking is determined that considers both factors according to B.J. Fogg's [16] behavior model; where both preference and ability (i.e. perceived effort level) are important factors in how actionable a suggestion would be. We illustrate what Fogg's behavior model would suggest with an example. Let us assume that there are three suggestions for a user: walking near the office, walking near the



(a) (b) (c)

Figure 5: Keeping human in the loop (a) dismissing a suggestion by removal (b) Moving a suggestion above (c) Moving a suggestions below

home, and going to gym. The user frequently walks near the office and prefers doing this. User also has a high preference for going to the gym, but is not good at gym work and goes infrequently. In addition, the user frequently walks near her house but is not keen on this activity. In this scenario, Fogg's behavior model would suggest that walking near the office is the most actionable. However, choosing between walking near home and going to gym would be a tie since one is easier to do while the other is more preferred.

Given this insight from Fogg's behavior model, we repurpose a principled technique in decision theory called *pareto-frontier* to balance between preference and low-effort [48]. Before moving into the details of the algorithm, we introduce some notations. We denote the set of suggestions as X where an element $x_j \in X$ is a suggestion. For a suggestion x_j, ν_j refers to its rank from MAB algorithm whereas p_j refers to its rank after users finishes reordering the suggestions (Figure 5(a-c)). Thus a higher rank or value of ν_j or p_j means the suggestion is more low effort or more preferred respectively. With this notation, the pareto algorithm works as follows. Let us assume that for two suggestions x_i, x_j, preferences and low-effort ranks are p_i, p_j and ν_i, ν_j respectively. If x_i's both preference and low-effort ranks are higher than x_j then x_i ranks higher (or is more actionable) than x_j and we say that x_i *pareto-dominates* x_j. If x_i's preference is higher than x_j while the low-effort rank is lower than x_j (i.e., $p_i > p_j$ and $\nu_i < \nu_j$) or the other way around (i.e., $p_i < p_j$ and $\nu_i > \nu_j$) then x_i and x_j receive the same rank and the more actionable suggestions can not be decided. Note here, that pareto-frontier makes no assumption about scale of p or ν and can still balance between them. Finally, the ranking process works iteratively as shown in Algorithm 1. It starts with a set of all available suggestions X. At every iteration, a set of suggestions X_i are selected that pareto-dominates rest of the suggestions. X_i are then ranked higher than the rest and are removed from the set of X. The process then repeats.

Finally a specific case that needs special attention in the pareto ranking is when a *new* suggestion x arrives with low-effort rank ν and unknown preference p since the user never ranked it. In this case, a fair policy is adopted that acts as follows: If x_1 and x_2 are two other suggestions such that $\nu_1 > \nu > \nu_2$ and $p_1 > p_2$ then no matter what the unknown

input : A set of suggestions X annotated with user preference and caloric benefit used in MAB

Initialize an index value $i = 1$;

while X *is non-empty* **do**
- find subset X_i in X that pareto dominates $X - X_i$;
- rank suggestion(s) in X_i with i;
- increment i by one and remove X_i from X;

end

Algorithm 1: Ranking suggestion with pareto-frontier

value of x's preference is, x would not be pareto dominated by x_2 since x has a higher low-effort rank than x_2. Since the value of p is unknown, it is fairly assumed that this unknown value to be less than a known value p_2. This would assign x the same rank as x_2 which is lower than x_1.

Pilot deployments with MyBehavior 2.0

We conducted a 3-week pilot to study the improvement of MyBehavior 2.0. We recruited 7 users (3 female) for 21 days. Users were interested to see the crowd-source based calorie content of the foods and were satisfied with the accuracy of food labeling and clustering. Specifically we counted the number of foods logged per day over 3 weeks. We have found that the number of foods recorded per day per user with crowd-based approach ($\mu = 4.2, \sigma = 2.5, q_{25} = 2.2, q_{50} = 4.1, q_{75} = 6.0$) is higher than the manual logging approach from MyBehavior 1.0 ($\mu = 2.4, \sigma = 1.7, q_{25} = 1.4, q_{50} = 2.0, q_{75} = 2.9$). This increase is also statistically significant (Wilcox ranksum test, $z = 2.5, p = 0.013$).

To measure the benefit of incorporating human preference, we showed MyBehavior 1.0 generated suggestions to the users after the study. They were asked to rate 8 food and 8 activity suggestions between a scale of 1 to 5. This rating represents whether users liked the suggestion and would act on it on an average day (1 = disagree and 5 = agree). After users finished rating the default set of suggestions (from 1.0), they were instructed on the use of the remove and reorder functions to incorporate their preferences. Users on the average changed 3.5 suggestions out of 16 suggestions. When users finished providing their preferences, we ran the pareto-frontier algorithm and then showed users the revised suggestions. We asked the users to rate again. Ratings without incorporating the human preference are similar to results from the MyBehavior 1.0 deployment ($\mu = 3.5, \sigma = 1.2$). However, after incorporating human preference using pareto-frontier algorithm, there is a statistically significant increase of almost 19% ($\mu = 4.2, \sigma = 1.1$).

EVALUATION

After two iterations of improvements, we conducted a 14-week deployment and evaluation study of MyBehavior 2.0. The purpose of the evaluation is two fold: (1) to test whether MyBehavior has better efficacy compared to control condition, and (2) to assess if MyBehavior can enable change beyond the initial novelty period. In the rest of this section, we detail the user study design and report results.

Study design considerations

Several pilot studies have been done to test early adoption and to make design improvements to MyBehavior, as have been argued by Klasanja et al. [23]. In order to quantitatively demonstrate early efficacy, Dallery et al. [12] and Onghena et al. [40] argue for small scale within subject trials, sometimes referred as "single case experiments". These experiments achieve sufficient statistical power with large number of repeated samples from a single individual. MyBehavior suits this requirement since enough repeated samples can be collected with automated sensing or daily manual logging [12].

In our study, we follow a single case experiment paradigm called *multiple baseline* design [12]. In a multiple baseline design, subjects are initially exposed to the control condition, which is followed by the experiment condition. However, the duration of the control condition before the experiment condition varies for different users. Such a variation is made as a *replication* strategy to show that the desired dependent variable consistently changes in the desired direction after the experiment condition starts. In our study, participants are exposed to either 2, 3 or 4 weeks of control condition before using MyBehavior as part of multiple baseline design. We also run the experiment condition for longer (7-9 weeks) than control condition. We do so to investigate MyBehavior's influence beyond initial novelty periods.

Study procedure and participants

We sent an invitation for participating in MyBehavior's user study through Cornell University's Wellness Center's email list. Interested individuals emailed back an investigator and were requested to fill out a prescreening survey. The survey asked for age, gender, experience in using smartphones etc. We also asked *readiness* to act on healthier behavior as defined by the Trans-theoretical model (TTM) [46][1]. We only included participants with (i) sound proficiency in using smartphones (ii) are either in ready or acting stages of TTM since in these stages people are willing or acting towards changing their behaviors [47]. The study investigators met with eligible participants and installed MyBehavior on their phones. In these meetings, we provided basic instructions to use MyBehavior. Participants also entered their gender, weight, height and weekly weight loss goals. Then the Harris-Benedict equation [18] is used to translate weight loss goals to daily calorie intake and expenditure goals.

The day after the face-to-face meeting, participants started the *baseline phase* of the study. In this phase, calorie goals were displayed in an on-screen widget in the phone's home screen. This widget also incorporated realtime updates of user's daily calorie intake and expenditure. We also added a daily chronological summary of physical activities and food intake. No suggestions were provided in this baseline phase. Note here

[1]TTM defines several stages to readiness: "Precontemplation" represents a stage of not feeling the need to change while in "Ready" stage there is intention to start eating well or doing exercise in near future but not taking actions. "Acting" stage on the other hand represents already taking actions but still need to strengthen commitment, or fight urges to slip. Finally, "Maintaining" stage means a lifestyle with regular health eating and exercise.

Variable	n(%)
Gender	
Male	7(43.7)
Female	9(56.3)
Age	
18 - 29	4(25.0)
30 - 39	6(37.5)
40 - 49	3(18.7)
> 50	3(18.7)
Stage of behavior change before the study	
Ready	7(43.7)
Acting	9(56.3)
Previous experience with self-tracking	
Maintained food diary	13(81.3)
Maintained exercise diary	11(68.7)

Table 1: User demographics in the long term study

that such widgets and daily logs are common for many modern health and fitness applications [15][33]. We ran the baseline phase for 3 weeks, since starting to use a health application often makes users more active temporarily even though no intervention is used. Such an effect is often referred to as "novelty effect" [49]. After the baseline phase, participants were exposed to the control condition of the study. Participants received generic prescriptive recommendations generated from a pool of 42 suggestions for healthy living, such as "walk for 30 minutes" and "eat fish for dinner". A certified fitness professional created these generic suggestions after following National Institute of Health resources [36][35]. An external nutrition counselor also reviewed the suggestions to ensure that they were both healthy and achievable. We followed the multiple baseline design as described before and continue the control condition for different durations for different participants. The control condition ranged between 2-4 weeks depending on participants. Each day of the control phase, 8 physical activity and 8 food suggestions were randomly selected from the 42 prescriptive suggestions. These suggestions are shown in a list similar to MyBehavior suggestions are shown in Figure 4[2]. After the control phase, participants received MyBehavior suggestions for 7-9 weeks. Total participation period did not exceed 14 weeks for any participant. Participants were compensated $120 for their regular participation in the study.

We recruited 16 participants. Table 1 shows the participant demographics. Our sample size was determined by following the literature of single case experiment design [12]. The literature argues that $n \geq 4$ is sufficient for statistical power if enough repeated samples are collected per participant.

[2]Screenshots of suggestions during control phase and the list of 42 suggestions are added as supporting material of this paper.

Daily phone survey
1. How many suggestions were you able to follow today?
2. How many suggestions did you want to follow?
3. How well did the suggestions relate to your life. • likert scale 1-7 • 1- doesn't relate to your life • 7- relates to your life perfectly
4. Did you encounter any barrier to follow the suggestions today (e.g., weather or deadline)? • Yes/No
5. Rate your emotional state today • photographic affect meter (PAM) scale [44]

Table 2: Users answered the above 5 questions in a daily phone survey

Outcome measures of the study

We utilize the food and exercise log data to measure changes in food calorie intake and calorie loss in exercise. During the study, we also used an in-phone survey that users filled out daily. The survey asks 5 questions as listed in Table 2. For the number of suggestions followed, we use self-report since it is hard to objectively judge whether an activity is done as part of regular actions or as a result of the suggestion. We ask how many suggestions users *wanted* to follow to measure user intentions or attitude [2]. Past literature shows that attitudes or intentions often indicate 19%-39% of future behavior [5]. A higher score in the 3rd question means the suggestions relate to a user's life and are potentially easy to implement. We ask the 4th and 5th questions because we want to investigate how MyBehavior suggestions perform against negative life circumstances as barriers and negative emotions have been shown to reduce chances of change [20].

Although weight loss is MyBehavior's main long term goal, calorie loss or user intentions to follow suggestions are important mediators to achieve weight loss. Recent work on adaptive interventions in clinical psychology (e.g., Behavior Intervention Technology [31]) and just-in-time adaptive interventions [34] argue that calorie loss or positive activities are essential subaims and are valid outcome measures for weight reduction applications.

Analysis plan

We analyze the efficacy of MyBehavior against control condition by modeling our outcome measures (e.g., caloric loss or number of suggestions followed) as continuous variables using mixed effect models against time. We use mixed effect models [41] since they can handle imbalanced control vs. experiment conditions [10][43] and correlated data points from the same user [13]. In the models, we use intercept and time as random effects to respectively allow for inter-subject variations in initial starting points and growths over time [8][32]. Including these random effects significantly increased likelihood over fixed-effect-only models in likelihood ratio tests [50]. Such an increase in model fit (i.e., likelihood) means inter-subject variability exists in our dataset and including random effect is necessary to properly isolate inter-subject variability from actual trends in fixed effects. As fixed effects, we use time and intervention type (i.e., control vs experiment). Intervention types are coded 0 for control and 1 for experiment phase. For time, the first week of the control phase is coded as 0 and incremented by 1 after each subsequent week. We observed non-linear changes in outcome measures over time, so we use non-linear time effect up to cubic polynomials [50]. In general, considering such polynomial time effects shows significant improvements in likelihood ratio tests compared to models without such polynomial time effects. On exception is for number of minutes walked where time or its polynomial forms as fixed effects did not improve the likelihood significantly. This approach of centering [50] time and intervention adjusts for time related effects (e.g., weather effect, or changes due to logging for longer periods) and isolates the change with MyBehavior over control as the co-efficient of intervention fixed effect (β_i). In other words, β_is reflect changes (e.g., number of minutes walked more) at the points of introducing MyBehavior. Finally, for the survey response of number of suggestions followed, we additionally include emotional state, barrier and their interaction with intervention types as fixed effects. We add these extra terms to explore interplay between MyBehavior and emotional states/barriers. Emotional states are coded as 0,1,2,3 respectively for negative high, negative low, positive low and positive high. Barriers are coded as 1 for presence of barrier and 0 for absence. Both barriers and emotional states are considered as categorical in the mixed model. The analyses are run using Matlab's statistical analysis toolbox with maximum likelihood.

Given significant intervention effects are achieved with mixed effect models, we explore the real-world end effect of MyBehavior in post-hoc analysis. We compare 2-4 weeks of using control condition to last 3 weeks of using MyBehavior. We consider the last 3 weeks of MyBehavior to measure change beyond initial novelty periods. Specifically, we describe the mean and standard deviations for these two conditions. We then use student t-tests and Cohen-d to measure the statistical significance and effect size. Similar pre-post analysis to measure real world end effect has been done in [8].

Results

Comparison with the control condition

Table 3 shows the results from the mixed model analyses for different outcome measures. Due to space limitations, we only include the relevant statistics. In 2nd column, we report the coefficient of intervention fixed effect (β_i) and its significance. In third column, we also report the standard model fit statistics that underpin the values of β_is. We include standard model fit statistics namely deviance, AIC and BIC scores [50]. We add significance of the fitted models (LR) against unconditional mean models (i.e., a baseline mixed model with only intercept as both fixed and random effects) [50] using a likelihood ratio test. From table 3, we observe that all the fitted mixed models for different outcome measures are significant improvements over the unconditional mean model. Furthermore, use of MyBehavior compared to control condition results in increased number of suggestions followed ($\beta_i = 1.2, p < 0.0005$), walking minutes ($\beta_i = 10.1, p < 0.005$) and calories burnt in non-walking exercises ($\beta_i = 42.1, p < 0.05$) per day. Calorie consumption also decreased per meal ($\beta_i = -56.1, p < 0.05$).

Figure 6 shows different outcome measures (i.e., number of suggestions followed, minutes walked, calories burnt in exercise, calorie intake in meals) over time as commonly reported in multiple baseline designs [12][6]. All these values are predicted from the mixed models. For each outcome measure, we create three groups representing 2, 3, 4 weeks of using control conditions before exposing to MyBehavior. A dotted line shows the start time of using MyBehavior. Improvements in all outcome measures can be seen to occur in Figure 5 after the introduction of the MyBehavior phase irrespective of the start times. However, patterns over time differ for different outcome measures. Minutes walked did not change much over time. On the other hand, food calories consumption generally decreased over time although introduction of MyBehavior had some effect. Non-walking exercises generally decreased in control over time, but were sustained during MyBehavior usage.

Subjective responses namely number of suggestions participants wanted to follow ($\beta_i = 2.9, p < 0.0005$) and relatedness of suggestions to life ($\beta_i = 0.5, p < 0.0005$) were also higher for MyBehavior compared to control (Table 3). Including emotional state, barriers and their interactions with interventions significantly improved likelihood of predicting number of suggestions followed compared to excluding them in likelihood ratio tests ($p = 0.05$). This means that there are significant interactions of MyBehavior vs. control with emotional state and barriers. Figure 7 visualizes these interactions as distributions of number of suggestions followed for different emotional states and barrier conditions.

Pre-post real-world effect analysis

Pre-post analysis is summarized in Table 4. For all the outcome measures, values of Cohen-d indicate medium to large effects of MyBehavior. Although not shown in the table, all these changes are also statistically significant ($p < 0.05$) in

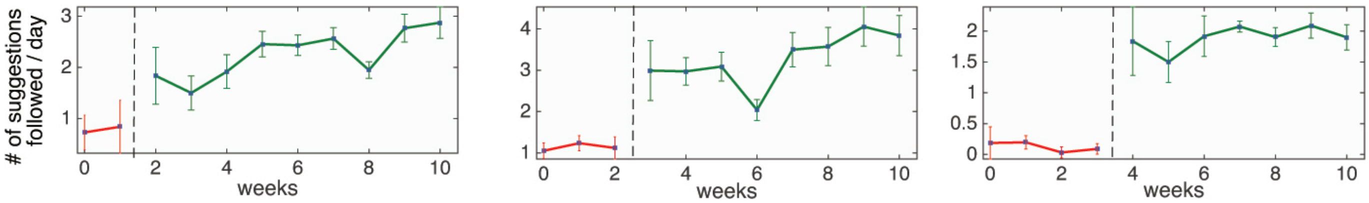
Outcome measure	β_i	$-2logL$	AIC	BIC	LR
# of sug. followed	1.2***	2491	2517	2576	***
# of sug. wanted	2.9***	2496	2518	2568	***
relatedness	0.5***	1551	1573	1623	**
walking/day (min)‡	10.1**	4795	4809	4839	***
exercise/day (cal)^a	42.1*	10959	10973	11006	**
each meal (cal)	−56.1*	16151	16165	16200	***

$^{***}p < 0.0005$; $^{**}p < 0.005$; $^{*}p < 0.05$; $\sim p > 0.1$

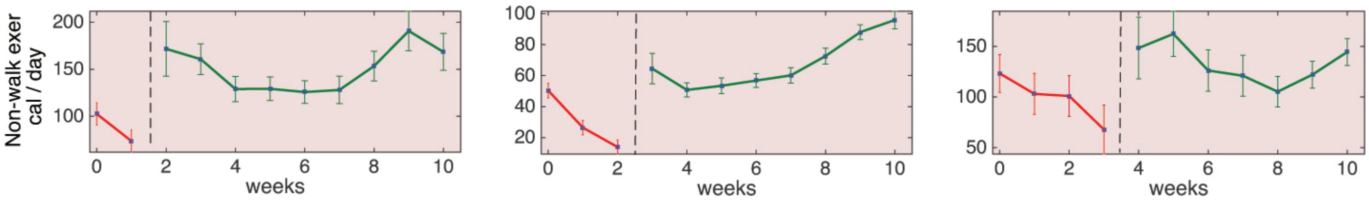
^{a} non-walking exercises combined

‡ without time as fixed-effect

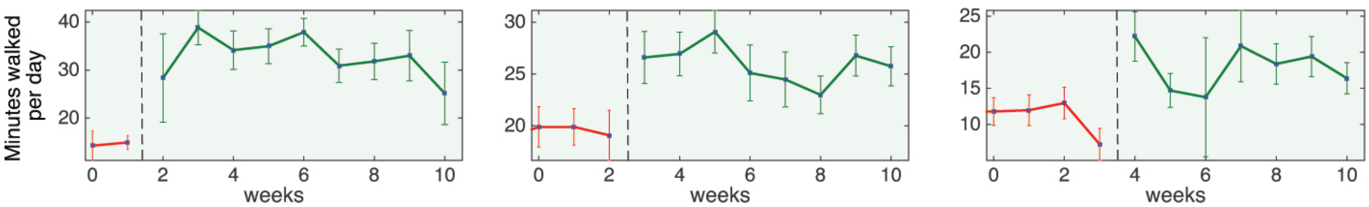
Table 3: Summary of statistical differences between control and MyBehavior as collected from survey, physical activity and dietary logs



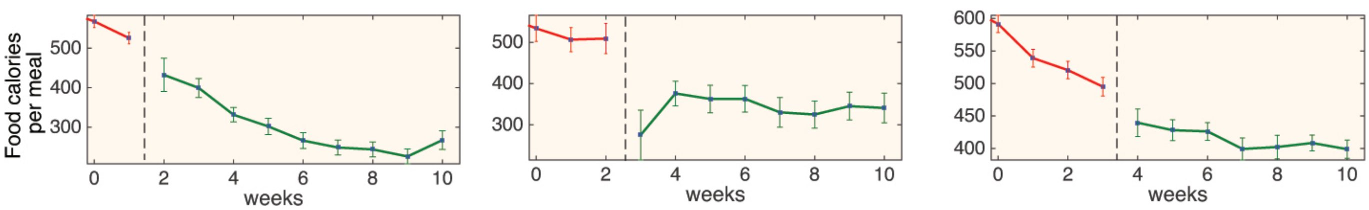
(a) Number of suggestions followed over weeks of the study.



(b) Calories lost in non-walking exercises per day across the study



(c) Minutes walked per day during the study



(d) Calories consumed in per meal

Figure 6: Changes in user behavior as predicted by the mixed model for multiple baseline design. The dotted lines represent the start of the intervention of MyBehavior. Left, middle, and right figures respectively show results from participants where intervention were started after 2, 3 and 4 weeks of using the control. Red color represents control phase where as green represents periods of using MyBehavior.

student t-tests. An additional result we point to is the changes in number of suggestions followed for barriers and emotional states. Users followed more MyBehavior suggestions where there was no barrier ($p < 0.001, d = 0.84$) such as bad weather. Similar significant increase is also found for positive emotion ($p < 0.001, d = 0.82$). Furthermore, MyBehavior suggestions were still followed more than control suggestions even when there were barriers ($p < 0.001, d = 0.44$) or when the user experienced negative emotion ($p < 0.001$, $d = 0.55$). However, effect sizes are smaller for barrier and negative emotions.

DISCUSSION AND RELATED WORK

Primary findings

To the best of our knowledge, MyBehavior is the first recommendation system to automatically generate health feedback from physical activity and food log data. It utilizes concepts of low-effort [16] and self-efficacy [4] from behavior change theory literature and operationalized them in machine learning optimization functions. Through several deployments, we created a usable MyBehavior app that utilize the benefits of algorithmic computation in usable form.

In a 14-week study, participants subjectively reported MyBehavior suggestions to be more related to their life and they wanted to follow the suggestions in higher numbers. We believe such higher actionability and relatedness result from MyBehavior's prioritization of low effort suggestions. The higher actionability and relatedness also translated to actual behavior with increased walking, exercise and decreased food calorie intake. These favorable results are replicated as part of multiple baseline design as shown in Figure 6. This adoption may result from low-effort suggestions that should enable actual adoption according several behavior change theories [16][21][4][20][2]. Finally, in the pre-post real-world effect analysis, MyBehavior suggestions were followed more during no-barrier or positive emotions states compared to barriers or negative emotional states. We believe this happens because low effort suggestions similar to MyBehavior are adopted in higher numbers during high motivation states like no-barrier or positive emotions [16]. Nonetheless, some MyBehavior suggestions were followed during barrier or negative emotional states. According to Fogg [16], low-effort suggestions similar to MyBehavior may still stay actionable

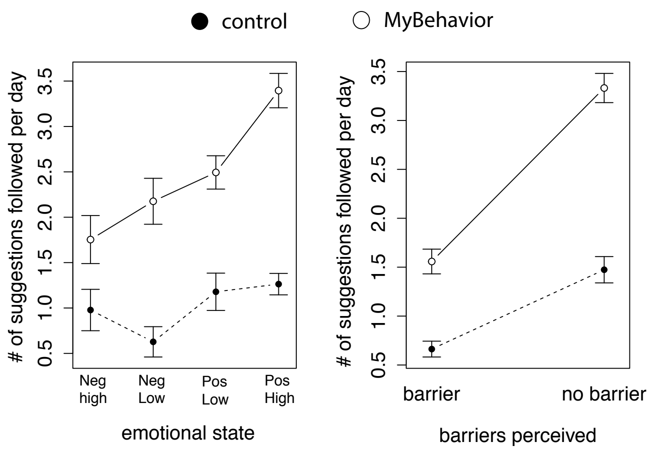


Figure 7: Number of suggestions followed for control and experiment conditions with respect to barriers and emotional states

Outcome measure	Control	MyBehavior	Cohen-d
# of sug. followed	1.1 (1.1)	3.1 (2.7)	0.76
# of sug. wanted	2.1 (1.2)	4.4 (2.4)	1.07
relatedness	3.8 (1.1)	4.5 (1.2)	0.54
walking/day (min)	14.5 (5.9)	24.9 (7.4)	1.41
exercise/day (cal)[a]	83.5 (33.1)	126.7 (35.3)	1.23
each meal (cal)	540 (137.2)	362 (134.1)	1.30
# of sug. followed[b]	1.3 (2.2)	3.4 (2.8)	0.84
# of sug. followed[c]	0.6 (2.1)	1.6 (2.5)	0.44
# of sug. followed[d]	1.2 (1.9)	3.2 (2.6)	0.82
# of sug. followed[e]	0.7 (1.5)	1.9 (2.1)	0.55

[a] non-walking exercises combined
[b] for no barrier, [c] with barrier
[d] for positive emotion, [e] for negative emotion

Table 4: Pre-post analysis for the control condition and last 3 weeks of experiment condition. Means and standard deviations (within bracket) are shown along with effect size measures.

in low motivation states like with barriers and negative emotions.

Related work and MyBehavior

MyBehavior's automated personalization scheme differs from prior pervasive health literature. Ubifit [11] or BeWell [25] relied on overall statistics (e.g., total amount of activity) for providing feedback without personalized actionable suggestions. On the other hand, previous literature on life-logging [56] relied on visualizing the whole personal data that users have to interpret and find actionable information by themselves. MyBehavior, in comparison, breaks down each user's behavior and finds personalized actionable suggestions. Such an approach is only remotely similar to tailored health communication [24] approaches where suggestions are tailored for groups of users with similar age, gender or stages of behavior change. This also means MyBehavior personalization approach is the first automated N-of-1 [17][53] or small data [14] system that treates each user differently in creating its suggestions.

In addition to personalization, MyBehavior also explored the space of generating low-effort suggestions automatically for the first time. According to B.J. Fogg's behavior model, low-effort is similarly important as motivation [16]. However, earlier literature on gamification [30], goal seeking [54] or self-regulation [51] relied on increasing user motivation. To the best our knowledge, easiness or low-effort suggestions were only provided through health coaches in earlier work. Nonetheless the work on increasing user motivation is orthogonal to our work on low-effort. Higher user motivation can increase probability of executing low-effort suggestions [16].

Using Multi-armed Bandit (MAB) also solves a few practical issues of generating suggestions. MAB is an online learning algorithm that learns, adapts and decides simultaneously. All of these learning and adaption are done with relatively less data since MABs are not heavily parameterized. This is crucial at the early stages when less data is available from users. MAB's online nature also means model update needs only processing the latest data with less computation. A competing technique to MAB is the Markov Decision Processes (MDP) [45], the most used reinforcement learning algorithm for decision making. In comparison to MABs, MDPs are highly parameterized and often require large amount of data to train. Because of MAB's low computational requirement, MyBehavior can generate all suggestions inside the phone without significantly lowering the battery. This also means location and activity traces do not need to leave the phone, which is an added privacy feature [22]. Finally, MABs are used before for personalized recommendation in other domains. For example, Yahoo on their front page uses MABs to suggest personalized news articles [27]. Google also uses MABs to dynamically serve their advertisements [29].

CONCLUSION

In this paper, we present the design, implementation, and evaluation of the MyBehavior smartphone app that provides personalized health suggestions automatically. We build the underlying automatic suggestion generation system using two different decision theory techniques, namely, multi-arm bandit and pareto-frontier algorithm. The combination of these techniques provides a novel way to tailor feedback without requiring expensive and difficult-to-scale interventions from health coaches. We present the results from a 14-week study that shows significant improvement over an appropriately chosen control condition that lasted beyond the initial novelty phase. As more and more people use automated technologies to track their health, we believe MyBehaviors ability to auto-personalize suggestions holds great promise for providing actionable feedback at scale.

ACKNOWLEDGMENT

This work has been partially supported by Intel Science and Technology Center for Pervasive Computing. We thank Julie Kientz for her feedback on early manuscript and for suggesting the multiple baseline design.

REFERENCES

1. B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon. 2011 compendium of physical activities: a second update of codes and met values. *Medicine and science in sports and exercise*, 43(8):1575–1581, 2011.
2. I. Ajzen. Theory of planned behavior. *Handb Theor Soc Psychol Vol One*, 1:438, 2011.
3. Amazon Mechanical Turk. http://www.mturk.com/, 2013. [Online; accessed 19 March 2013].
4. A. Bandura and D. C. McClelland. Social learning theory. 1977.
5. Behavioral Intentions. http://chirr.nlm.nih.gov/behavioral-intention.php, 2013. [Online; accessed 26 February 2015].
6. A. Biglan, D. Ary, and A. C. Wagenaar. The value of interrupted time-series experiments for community intervention research. *Prevention Science*, 1(1):31–49, 2000.
7. S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012.
8. M. N. Burns, M. Begale, J. Duffecy, D. Gergle, C. J. Karr, E. Giangrande, and D. C. Mohr. Harnessing context sensing to develop a mobile intervention for depression. *Journal of medical Internet research*, 13(3), 2011.
9. B. Caballero. The global epidemic of obesity: an overview. *Epidemiologic reviews*, 29(1):1–5, 2007.
10. A. Cnaan, N. Laird, and P. Slasor. Tutorial in biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med*, 16:2349–2380, 1997.
11. S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 1797–1806. ACM, 2008.
12. J. Dallery, R. N. Cassidy, and B. R. Raiff. Single-case experimental designs to evaluate novel technology-based health interventions. *Journal of medical Internet research*, 15(2), 2013.
13. P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger. *Analysis of longitudinal data*. Oxford University Press, 2002.
14. D. Estrin. Small data, where n = me. *Commun. ACM*, 57(4):32–34, Apr. 2014.
15. Fitbit, Inc. http://www.fitbit.com/, 2013. [Online; accessed 19 March 2013].
16. B. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 40. ACM, 2009.
17. W. M. Grove. Thinking clearly about psychology.
18. J. Harris and F. Benedict. Biometric studies of basal metabolism. *Washington, DC: Carnegie Institution*, 1919.
19. D. Haytowitz, L. Lemar, P. Pehrsson, J. Exler, K. Patterson, R. Thomas, M. Nickle, J. Williams, B. Showell, M. Khan, et al. Usda national nutrient database for standard reference, release 24, 2011.
20. G. Hochbaum, I. Rosenstock, and S. Kegels. Health belief model. *United States Public Health Service*, 1952.
21. D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
22. P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. Exploring privacy concerns about personal sensing. In *Pervasive Computing*, pages 176–183. Springer, 2009.
23. P. Klasnja, S. Consolvo, and W. Pratt. How to evaluate technologies for health behavior change in hci research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3063–3072. ACM, 2011.
24. R. Kukafka. Tailored health communication. *Consumer Health Informatics: Informing Consumers and Improving Health Care*, pages 22–33, 2005.
25. N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. T. Campbell. Bewell: A smartphone application to monitor, model and promote wellbeing. In *5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth2011)*, 2011.
26. J. Lester, C. Hartung, L. Pina, R. Libby, G. Borriello, and G. Duncan. Validated caloric expenditure estimation using a single body-worn sensor. In *Proceedings of the 11th international conference on Ubiquitous computing*, Ubicomp '09, pages 225–234, New York, NY, USA, 2009. ACM.
27. L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.
28. H. Lu, J. Yang, Z. Liu, N. D. Lane, T. Choudhury, and A. T. Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 71–84. ACM, 2010.
29. T. Lu, D. Pl, and M. Pl. Showing relevant ads via lipschitz context multi-armed bandits. In *Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010.
30. S. McCallum. Gamification and serious games for personalized health. *Stud Health Technol Inform*, 177:85–96, 2012.

31. C. D. Mohr, M. S. Schueller, E. Montague, N. M. Burns, and P. Rashidi. The behavioral intervention technology model: An integrated conceptual and technological framework for ehealth and mhealth interventions. *J Med Internet Res*, 16(6):e146, Jun 2014.
32. L. G. Morrison, C. Hargood, S. X. Lin, L. Dennison, J. Joseph, S. Hughes, D. T. Michaelides, D. Johnston, M. Johnston, S. Michie, et al. Understanding usage of a hybrid website and smartphone app for weight management: A mixed-methods study. *Journal of medical Internet research*, 16(10), 2014.
33. MyFitnessPal, LLC. http://www.myfitnesspal.com/, 2013. [Online; accessed 19 March 2014].
34. I. Nahum-Shani, S. N. Smith, A. Tewari, K. Witkiewitz, L. M. Collins, B. Spring, and S. Murphy. Just in time adaptive interventions (jitais): An organizing framework for ongoing health behavior support. *Methodology Center technical report*, (14-126), 2014.
35. National Heart, Lung and Blood Institute, National Institutes of Health. Getting started and staying active. http://www.nhlbi.nih.gov/health/public/heart/obesity/lose_wt/calories.htm, 2011. [Online; accessed 19 March 2013.
36. National Heart, Lung and Blood Institute, National Institutes of Health. Healthy eating plan. http://www.nhlbi.nih.gov/health/public/heart/obesity/lose_wt/calories.htm, 2013. [Online; accessed 19 March 2013].
37. Netflix. http://www.netflix.com/, 2013. [Online; accessed 19 March 2014].
38. J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos. Platemate: crowdsourcing nutritional analysis from food photographs. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 1–12. ACM, 2011.
39. C. L. Ogden, M. D. Carroll, B. K. Kit, and K. M. Flegal. Prevalence of obesity in the united states, 2009–2010, 2012.
40. P. Onghena and E. S. Edgington. Customization of pain treatments: Single-case design and analysis. *The Clinical journal of pain*, 21(1):56–68, 2005.
41. R. L. Ott and M. T. Longnecker. *An Introduction to Statistical Methods and Data analysis, 4th*. New York: Duxbury Press, 1993.
42. T. Philipson and R. Posner. Is the obesity epidemic a public health problem? a decade of research on the economics of obesity. Technical report, National Bureau of Economic Research, 2008.
43. J. Pinheiro and D. Bates. Nlme: Software for mixed-effects models, 2000.
44. J. P. Pollak, P. Adams, and G. Gay. Pam: a photographic affect meter for frequent, in situ measurement of affect. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 725–734. ACM, 2011.
45. W. B. Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.
46. J. O. Prochaska and W. F. Velicer. The transtheoretical model of health behavior change. *American journal of health promotion*, 12(1):38–48, 1997.
47. M. Rabbi, A. Pfammatter, M. Zhang, B. Spring, and T. Choudhury. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: A randomized controlled trial on adults. *JMIR mHealth uHealth*, 3(2):e42, May 2015.
48. M. C. Roberts, A. S. Dizier, and J. Vaughan. Multiobjective optimization: Portfolio optimization based on goal programming methods.
49. E. M. Rogers. *Diffusion of innovations*. Simon and Schuster, 2010.
50. J. D. Singer and J. B. Willett. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford university press, 2003.
51. L. Springvloet, L. Lechner, H. de Vries, M. J. Candel, and A. Oenema. Short-and medium-term efficacy of a web-based computer-tailored nutrition education intervention for adults including cognitive and environmental feedback: Randomized controlled trial. *Journal of medical Internet research*, 17(1):e23, 2015.
52. R. Sriraghavendra, K. Karthik, and C. Bhattacharyya. Fréchet distance based approach for searching online handwritten documents. In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, volume 1, pages 461–465. IEEE, 2007.
53. A. Tashakkori and C. Teddlie. *Mixed methodology: Combining qualitative and quantitative approaches*, volume 46. SAGE Publications, Incorporated, 1998.
54. B. E. Wilde, C. L. Sidman, and C. B. Corbin. A 10,000-step count as a physical activity target for sedentary women. *Research Quarterly for Exercise and Sport*, 72(4):411–414, 2001.
55. T. Zhang, R. Ramakrishnan, and M. Livny. Birch: an efficient data clustering method for very large databases. In *ACM SIGMOD Record*, volume 25, pages 103–114. ACM, 1996.
56. L. M. Zhou and C. Gurrin. A survey on life logging data capturing. *SenseCam 2012*, 2012.

Protocol

Toward Increasing Engagement in Substance Use Data Collection: Development of the Substance Abuse Research Assistant App and Protocol for a Microrandomized Trial Using Adolescents and Emerging Adults

Mashfiqui Rabbi[1], PhD; Meredith Philyaw Kotov[2], MS; Rebecca Cunningham[3,4,5], MD; Erin E Bonar[2,4], PhD; Inbal Nahum-Shani[6], PhD; Predrag Klasnja[7], PhD; Maureen Walton[2,4], PhD, MPH; Susan Murphy[1], PhD

[1]Department of Statistics, Harvard University, Cambridge, MA, United States
[2]Department of Psychiatry and Addiction Center, University of Michigan, Ann Arbor, MI, United States
[3]Department of Emergency Medicine, University of Michigan Medical School, Ann Arbor, MI, United States
[4]Injury Prevention Center, University of Michigan, Ann Arbor, MI, United States
[5]School of Public Health, University of Michigan, Ann Arbor, MI, United States
[6]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States
[7]School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Mashfiqui Rabbi, PhD
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA, 02138
United States
Phone: 1 6036671797
Email: mrabbi@fas.harvard.edu

Abstract

Background: Substance use is an alarming public health issue associated with significant morbidity and mortality. Adolescents and emerging adults are at particularly high risk because substance use typically initiates and peaks during this developmental period. Mobile health apps are a promising data collection and intervention delivery tool for substance-using youth as most teens and young adults own a mobile phone. However, engagement with data collection for most mobile health applications is low, and often, large fractions of users stop providing data after a week of use.

Objective: Substance Abuse Research Assistant (SARA) is a mobile application to increase or sustain engagement of substance data collection overtime. SARA provides a variety of engagement strategies to incentivize data collection: a virtual aquarium in the app grows with fish and aquatic resources; occasionally, funny or inspirational contents (eg, memes or text messages) are provided to generate positive emotions. We plan to assess the efficacy of SARA's engagement strategies over time by conducting a micro-randomized trial, where the engagement strategies will be sequentially manipulated.

Methods: We aim to recruit participants (aged 14-24 years), who report any binge drinking or marijuana use in the past month. Participants are instructed to use SARA for 1 month. During this period, participants are asked to complete one survey and two active tasks every day between 6 pm and midnight. Through the survey, we assess participants' daily mood, stress levels, loneliness, and hopefulness, while through the active tasks, we measure reaction time and spatial memory. To incentivize and support the data collection, a variety of engagement strategies are used. First, predata collection strategies include the following: (1) at 4 pm, a push notification may be issued with an inspirational message from a contemporary celebrity; or (2) at 6 pm, a push notification may be issued reminding about data collection and incentives. Second, postdata collection strategies include various rewards such as points which can be used to grow a virtual aquarium with fishes and other treasures and modest monetary rewards (up to US $12; US $1 for each 3-day streak); also, participants may receive funny or inspirational content as memes or gifs or visualizations of prior data. During the study, the participants will be randomized every day to receive different engagement strategies. In the primary analysis, we will assess whether issuing 4 pm push-notifications or memes or gifs, respectively, increases self-reporting on the current or the following day.

Results: The microrandomized trial started on August 21, 2017 and the trial ended on February 28, 2018. Seventy-three participants were recruited. Data analysis is currently underway.

Conclusions: To the best of our knowledge, SARA is the first mobile phone app that systematically manipulates engagement strategies in order to identify the best sequence of strategies that keep participants engaged in data collection. Once the optimal strategies to collect data are identified, future versions of SARA will use this data to provide just-in-time adaptive interventions to reduce substance use among youth.

Trial Registration: ClinicalTrials.gov NCT03255317; https://clinicaltrials.gov/show/NCT03255317 (Archived by WebCite at http://www.webcitation.org/70raGWV0e)

Registered Report Identifier: RR1-10.2196/9850

(JMIR Res Protoc 2018;7(7):e166) doi:10.2196/resprot.9850

KEYWORDS

engagement; microrandomized trial; just-in-time adaptive intervention

Introduction

Substance use remains a major public health issue [1,2] due to its associations with risky behaviors (eg, intoxicated driving, violence, risky sexual behaviors, etc) and short-term (eg, injury) and long-term health consequences (eg, development of substance use disorders) [3-6]. Substance use typically begins during early adolescence (typical age, 12-17 years), with emerging adults (typical age, 18-25 years) having the highest prevalence [5]. According to the National Survey on Drug Use and Health, 11.5% of adolescents reported alcohol use and 6.1% reported binge drinking (5 or more drinks) in the past month; 59.6% of emerging adults reported drinking and 37.7% reported binge drinking in the past month [5]. Marijuana is the most commonly used illicit drug in the United States, with 7.4% of adolescents and 19.6% of emerging adults reporting its use in the past month [5]. In addition, the legalization of medical and recreational use of marijuana in several states has paralleled the decreasing perceptions of risk [7-9]. This trend is concerning because marijuana use may affect the neuro-maturational development of the brain among youth [10-12], potentially compromising decision-making and inhibitory control functioning [10-13]. Finally, prescription pain medications are the next most commonly misused substance, with 7.4% of adolescents and 2.8% of emerging adults reporting use in the past month [5]. Importantly, the risk of overdose increases when binge drinking is combined with prescription opioids and sedatives [14].

Longitudinal panel studies show that, as youth develop, evolving interactions between individual, social, and community risk and protective factors can decrease or accelerate substance use trajectories [15-18]. However, there is a critical knowledge gap about the temporal processes that underlie substance use among youth (eg, youth have greater rates of binge drinking). Knowledge about these temporal processes can enable better understanding of the real-time antecedents and sequelae of substance use and can inform the development of just-in-time adaptive interventions (JITAIs) (eg, for craving, high-risk activity spaces, negative affect, and stress) [19-21]. Mobile phones provide a promising platform in this regard. Mobile phones are with us most of the time, which creates opportunities for frequent in situ data collection about the temporal processes underlying substance use. Among tech-savvy adolescents and emerging adults, mobile phones are also pervasive; 92% of teens (age, 13-17 years) go online daily, and 73% of them have a mobile phone, which is fairly balanced across racial groups, with 85% of African Americans, 71% of Caucasians, and 71% of Hispanics having a mobile phone [22].

However, sustaining self-reported data collection is challenging in mHealth [23,24]. Self-reporting rate is generally low for mHealth data collection apps [25], and the same is true for substance use apps. A recent review of mHealth apps for substance use (including short message service [SMS] text messaging and apps) concluded that engagement is a critical limitation, with most use declining quickly over brief periods of time (2 weeks to 3 months) [26], However, only few studies included adolescents or emerging adults [27]. The limited research on behavioral health apps among youth also found engagement challenging, potentially because youth can become habituated to apps [24] or due to competing demands from the frequent use of other apps, such as social media [22] or entertainment apps [28].

To date, there has been little work on increasing self-reporting rates of substance use data [29-31]. Sensor technologies can partially mitigate low self-reporting rates because sensors require no additional effort other than carrying the device [32,33]. However, detecting substance use by sensors is new and requires further validation. Bae et al [34] used mobile phone data to detect drinking episodes; SCRAM [35,36] and BACTrack Skyn [37] are wearable sensors that can continuously measure blood alcohol levels. However, such sensors do not exist for other substances, and important correlates of substance use such as stress level and mood cannot yet be reliably detected using sensors [38]. Therefore, self-reporting remains a valuable method to obtain substance use-related data. Previously, financial incentives were used, often along with frequent staff contact, to increase the rates of self-reporting [39-41]; however, both these strategies are prohibitively expensive for larger studies over longer periods of time. Other approaches are needed for engaging participants to self-report.

In this paper, we describe a mobile phone app, Substance Abuse Research Assistant (SARA), intended to enhance participant engagement in self-reporting. We also describe a micro-randomized trial (MRT) [42,43] design to rigorously test several *engagement strategies* we built into SARA. To our knowledge, the SARA study will be the first [44] to examine how the effect of engagement strategies may vary with both time and context (eg, negative affect, stress, loneliness). Our vision for SARA is an "engagement first" approach, where engagement strategies are employed to increase or sustain self-reporting over extended periods of time. An important goal of SARA is to reduce staff time and financial incentives using nonfinancial engagement strategies that are grounded in behavioral science theories (such as operant conditioning [45-48] and reciprocity [49-51]). In the future, we plan to use the collected data to trigger interventions aimed at reducing substance use. This paper provides a detailed description of SARA, including the theoretical foundation of SARA's different engagement strategies, and describes the study design we used to test the efficacy of SARA's engagement strategies over time.

Methods

Substance Abuse Research Assistant

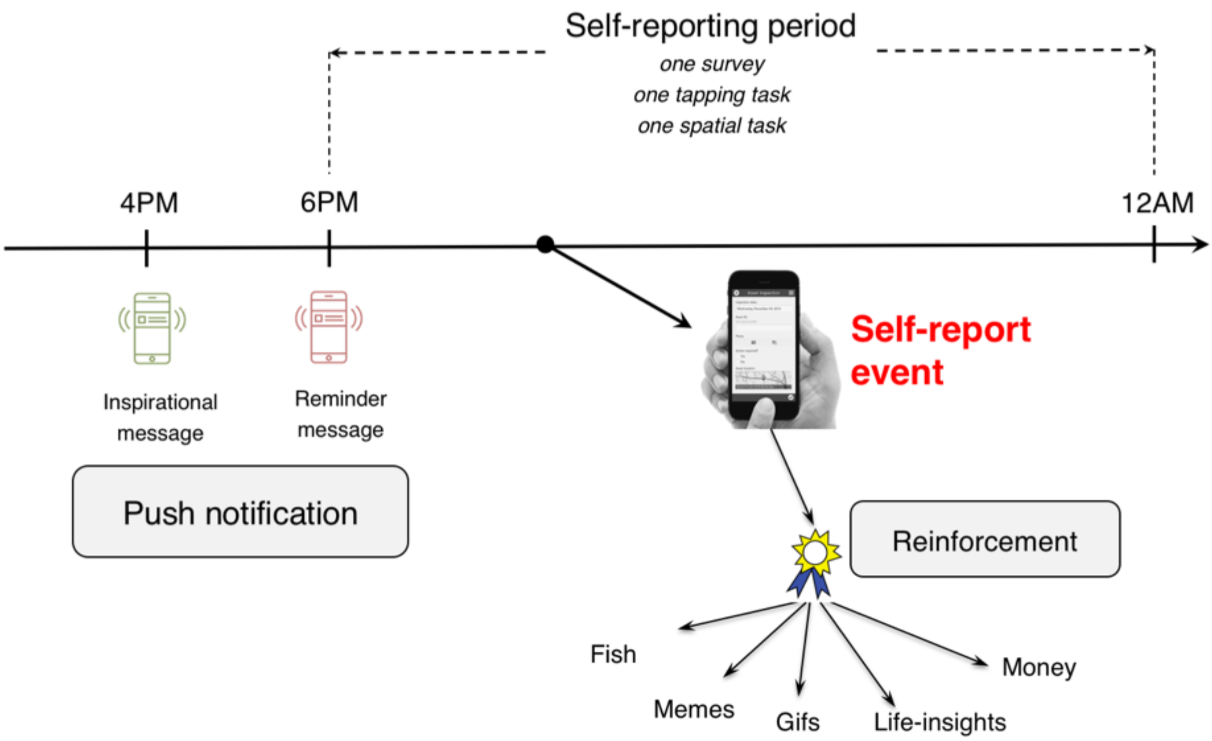
SARA is a mobile phone app aimed at increasing self-reporting of substance use data from adolescents and emerging adults. The app runs on both Android and iOS platforms. To collect data about correlates of substance use, every day, participants are prompted to complete a survey and 2 active tasks. SARA's key innovation is the variety of engagement strategies it incorporates to incentivize and support this data collection. The base engagement strategy is a virtual aquarium, which starts empty, but as a participant provides more data, the fish population grows and treasures accumulate. In addition to the aquarium, other strategies such as push notifications with inspirational messages, memes, and informative visualizations of self-report data are used to enhance engagement with self-reporting. SARA consists of 2 modules: the data collection module and the engagement module (Figure 1).

Data Collection Module

SARA's data collection module deals with procuring self-reported data. SARA currently supports 2 types of self-reported data collection. The first is provided by active tasks. Active tasks, which were first introduced in the Apple Research Kit [52], constitute intuitive user interactions that allow researchers to *objectively* measure reaction time, spatial memory, gait, problem-solving skills, etc. SARA's first active task serves as a measure of spatial memory, a random sequence of 5 seashells lights up in a 2-dimensional grid of 9 seashells. Participants are then asked to repeat the sequence. The second active task is a tapping task in which participants tap 2 buttons alternately for 10 seconds. The number of completed taps gives a measure of reaction time. The reason for choosing the tapping and spatial tasks is that spatial memory and reaction time may vary based on substance use-related intoxication [53,54].

The second type of data collection is a daily survey in which participants report their daily feelings and activities (ie, stress level, mood, loneliness, hopefulness, amount of free time, and excitement) [55-58]. On Sundays, an extra set of questions assesses the past week's frequency of substance use (alcohol, cannabis, and tobacco), motives and riskiness of use (for alcohol and marijuana), impulsivity, and intentions to avoid alcohol and marijuana use in the upcoming week [8,59-63]. Multimedia Appendix 1 contains the survey questions and the screenshots of the active tasks.

Figure 1. Daily timeline for engagement strategies in Substance Abuse Research Assistant (SARA). Push notifications are sent prior to data collection. Reinforcements are provided after data collection.



In the current version, participants self-report 1 survey and 2 active tasks (1 tapping task and 1 spatial task) every day between 6 pm and midnight. We selected this time window for the following reasons: (i) 6 pm to midnight provides a large enough time period to self-report, (ii) to capture a summary of most of the day using the survey in the evening, and (iii) to capture intoxication via active tasks since substances are typically used by youth in the evening [64].

Engagement Module

The engagement module of SARA contains several engagement strategies to increase self-report completion. These engagement strategies can be grouped as follows: (1) postdata collection rewards, (2) predata collection incentives, and (3) reminders to self-report. In addition, SARA contains other enhancements to support engagement, for example, occasional human support by SMS text messages or phone calls when participants temporarily disengage; a small amount of money, and an overall coherent user experience so that the app is easy to use, and the engagement strategies appear to be part of one app. Since SARA focuses on reducing financial incentives, we took several measures so that the engagement strategies are indeed effective for the target population.

First, we grounded the strategies in behavioral theories on influencing the intended action. Second, we followed a user-centered design process, where we conducted 3 focus groups (N=21, mean age 19.9 years) and a pilot study (N=17, mean age 21.2 years); a majority of the participants in these studies (67% in focus groups and 100% in the pilot study) reported binge drinking (5+ drinks at 1 occasion) or marijuana use in the past 3 months. We describe below the engagement strategies that are supported by theories and the user-centered design process.

Postdata Collection Rewards

Providing rewards after the successful completion of an intended behavior (eg, self-reporting in SARA) is a well-established method to shape future behavior [45,46]. Over the last few decades, the operant conditioning literature has extensively investigated how consequences shape behavior. According to the operant conditioning theory, the following are the 2 key ideas to influence the effectiveness of rewards: (1) Immediate contingent reward: rewards are more efficacious if they are given immediately and only after the intended behavior happens. Delaying rewards after the intended behavior or providing rewards after nonintended behaviors will make the reward less efficacious [47,48]. (2) Value of the reward: the rewards need to be valuable enough to trigger the intended behavior. One method to provide high enough value is to provide a variety of rewards; thus, even if one reward is less effective in a particular context, there is a higher likelihood that another reward can substitute for the less effective reward [65]. SARA employs these 2 ideas from operant conditioning as follows. First, ensuring immediacy and contingency was trivial; we provide the rewards immediately and only after self-reports. To ensure value, we have 2 types of rewards, (1) a growing virtual aquarium and (2) memes and life insights. We present below their design along with supporting evidence from prior work and user-centered design to show that the strategies can indeed engender reward value.

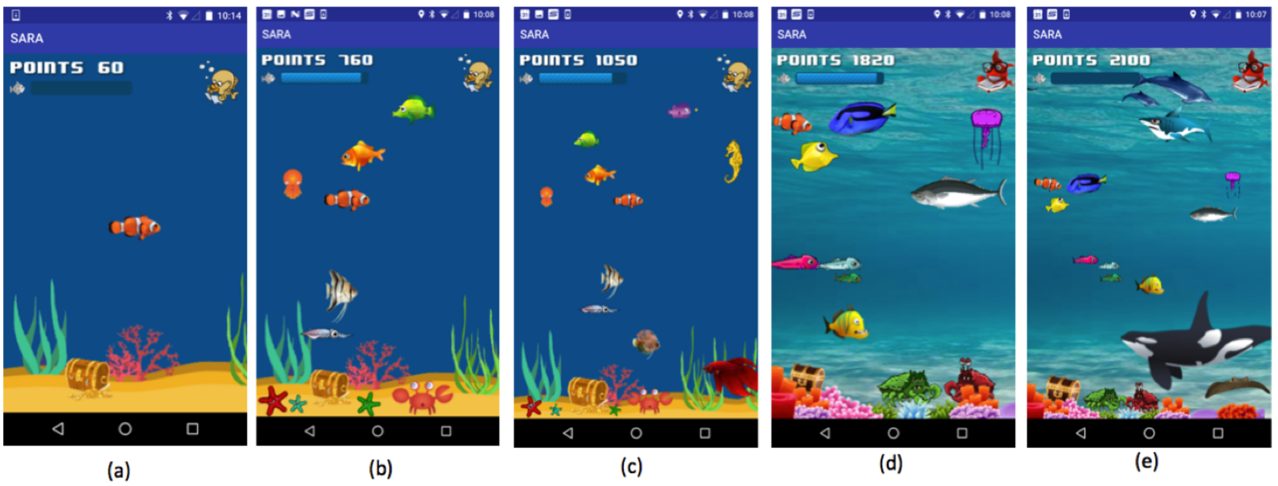
A Growing Virtual Aquarium

In SARA, a virtual aquarium environment grows richer as more self-reports are completed (Figure 2). Every time participants finish either the survey or the 2 active tasks, they earn 30 points toward the aquarium. For the longer survey on Sundays, 50 extra points are rewarded. New fish are unlocked as specific numbers of points are accumulated. Multimedia Appendix 2 lists the fish in the SARA aquarium and the corresponding numbers of points that unlock them. SARA is set up so that 1 fish can be unlocked almost every day if both the survey and active tasks are completed. Every time a fish is unlocked, a fun fact about the fish is also given; for example, when a goldfish is unlocked, participants see the message "Do you know goldfish can recognize faces?" An exception to the 1-fish-a-day rule is made for the first 2 days of the study, when SARA provides 2 fish per day. Initially, these extra fish are given to quickly condition the participant to the fact that interesting fish are unlocked if they self-report [48]. SARA makes the aquarium environment more game-like by introducing levels; after 15 days of self-reporting, participants graduate from a fishbowl environment to a sea environment. Levels help prevent cluttering as more fish are unlocked, while increasing the participants' interest. In addition, for streaks of self-reporting, participants can earn treasures such as pearls and gemstones. Multimedia Appendix 3 lists the different pearls and gemstones available in SARA and the corresponding self-reporting streaks that can unlock them.

The growing aquarium generates reward value based on the conceptualization of engagement as a "subjective experience" [44,47,48]. The aquarium is intended to generate positive subjective experience by creating enjoyment through collection of fish in a game-like environment. Rewarding self-reporting with points or fish also intends to promote positive subjective experiences by linking self-reporting behaviors with positive emotions (eg, joy and pride). Furthermore, once participants are initially engaged, the aquarium extends the positive experience by adding complexity through the aforementioned fun fish fact, levels, and treasures. Adding such complexity makes participants feel that their efforts are being reciprocated by the designers who have invested additional effort in creating new features and challenges; this sense of reciprocation may motivate participants to engage further [66].

Finally, an aquarium representation was chosen because aquariums have been used successfully to represent rewards in wellness apps in the past—notably in Fish n' steps [67] and BeWell [68]. A recent commercial game known as "Abyssrium," where a user has to grow an aquarium over time, has been downloaded more than 30 million times and has received the game of the year award in 2016 [69]. Participants in the user-centered design process also found the aquarium metaphor appealing (eg, focus group participants rated the SARA aquarium 3.9 stars out of 5 for use in a research study).

Figure 2. The evolution of Substance Abuse Research Assistant (SARA)'s aquarium environment as data is collected. Panels a, b, c, d, and e respectively show the state of aquarium if a participant self-reports for 1, 7, 14, 24, and 30 days.



Memes, Gifs, and Life Insights

Although the aquarium is expected to promote engagement, it may lose its novelty over time. Therefore, SARA includes other post-data collection rewards in the form of memes or gifs and life insights. Once participants complete the daily survey part of the self-input, they may receive a meme or an animated gif. The meme or gif is intended to be either funny or inspirational. Memes and gifs are chosen because they can generate reward value by positive emotions and encouragement [44,47,48,70]. Furthermore, the nonjudgmental nature of included memes or gifs is consistent with other substance use interventions [70]. The memes and gifs in SARA were generated using Amazon's Mechanical Turk and reviewed by undergraduate research assistants (RAs) who were of the same age as the target population. Participants in the user-centered design process for SARA also found the memes or gifs as acceptable forms of rewards (eg, on a scale of 1=strongly dislike to 5=strongly like memes or gifs as rewards for self-report, the average rating among a pilot study sample was 3.85).

In addition to memes or gifs, the participants may receive a life insight after they complete the active tasks portion of the data collection. Life insights are visualizations of self-reported data from the past. SARA's life insights are trends of the various data collected using daily survey and active tasks over the past 7 days. SARA contains a life insight for each of the following data types: (1) daily stress, (2) amount of free time in the day, (3) degree of loneliness in the day, (4) level of fun on the day, (5) how new and exciting were the days, (6) tapping speed, and (7) seconds taken to finish the spatial task. Note that 1-5 are gathered from the daily survey and 6-7 are gathered from the active tasks (see Multimedia Appendix 4). Life insights can generate reward value because individuals strive to understand themselves and gain self-relevant knowledge [71-73]. People are frequently unclear about their personal abilities and they learn about themselves by attending to and seeking self-relevant information [71,72,74,75]. Consistent with this notion, previous work has demonstrated that people are interested in receiving feedback about their past self-reported experiences [76]; in fact, most health apps and wearables (eg, fitbit) use visualizations of past data to provide feedback to their users. Participants in the user-centered design process for SARA were also quite interested in seeing their data on life insights (eg, on a scale of 1=strongly dislike to 5=strongly like life insights as rewards for self-report, the average rating among a pilot study sample was 3.92).

Predata Collection Incentive

Sociopsychological perspectives [49-51] suggest that reciprocity, that is, returning a favor, is an innate human tendency. Drawing on these perspectives, SARA sometimes provides incentives before (ie, not conditional on) self-reporting to facilitate participant reciprocation via subsequent self-reporting. SARA may issue a youth-focused inspirational message as a push notification at 4 pm, 2 hour before the data collection period starts. We selected 4 pm because adolescents or emerging adults are likely to be out of school at that time and hence are likely to notice the notification. This time is also close enough to data collection time (6 pm) so that providing an incentive may facilitate participant reciprocation via survey or active task completion. To facilitate participant reciprocation, we provide inspirational messages. From the user-centered design process, we found inspirational quotes in the form of song lyrics and celebrity quotes, which might be appealing to youth. Please refer to Multimedia Appendix 5 for the list of quotes used in SARA. Once again, this repository of messages was assembled and filtered by the undergraduate RAs who were of the same age as our target population. In the pilot study, we asked participants how much they liked the inspirational quotes as an incentive for self-report (1=strongly dislike, 5=strongly like); the average rating was 3.3.

Reminder Notifications

Past research has demonstrated that reminders can increase engagement [44]. SARA thus provides a message at 6 pm to remind participants to report data. The reminder message is sometimes appended with additional content, such as "you are close to unlocking a new fish," "you are close to finishing a streak and earning some money," or "it only takes a minute to collect data in SARA." The additional content tries to increase adherence by altering the perception of the value of self-reporting by reminding the participants of rewards that follow or that self-reporting does not require a lot of effort [77].

We selected 6 pm to send the reminder notification because it was the start time for the daily 6 pm to midnight self-reporting period discussed previously.

Other Enhancements to Support Engagement

SARA includes a few additional enhancements to support the abovementioned engagement strategies:

Financial Incentives and Human Support

One potential issue with the aquarium, memes, and life insights is that they are new rewards and participants may need time to perceive their full value. For example, unlocking fish and growing a virtual aquarium will be new to participants at the start of the study and they need to receive rewards several times before understanding what to expect. Therefore, participants may need additional sources of reinforcement that are rewarding right from the start [65]. Earlier work has demonstrated the utility of financial incentives [39,40] and human support [41] in promoting engagement. Participants in the user-centered design process for SARA also were very interested in financial incentives (eg, 95% of focus group participants reported that money bonuses would very much increase self-reporting in SARA). Hence, SARA includes relatively minimal financial incentives and human support to supplement its core engagement strategies; for every 3-day streak of self-reporting, that is, completing the survey and active tasks each day, participants can earn 1 dollar. For completing the longer weekly survey on Sunday, an extra 50 cents can be earned. For a 90% self-report completion rate, most participants can earn US $12 or less in a 30-day study (US $13 if self-report completion rate is 100%). Note that this is a fraction of what daily substance use studies normally pay for self-reporting (eg, US $1-4 dollars per day) [21,78].

In addition, if participants do not self-report, they receive SMS text messages from a study phone number. The first SMS text message is sent after 2 days of no self-reporting. If participants still do not self-report, a SMS text message is sent after 3 additional days of no self-reporting. The SMS text message follows a prespecified template (see Multimedia Appendix 6), which can be automated in future versions of SARA. After 7 days of no self-reporting, participants receive a phone call from a study team member. SMS text messaging and phone calls stop if participants neither respond nor self-report for 3 weeks.

User Experience

In SARA, we maintain a coherent user experience using an aquarium theme throughout the app. For example, in the spatial task, we use seashells to match the aquarium theme instead of flowers originally used by the Apple Research Kit [52,79]; after data collection, rather than only providing memes or life insights, we use animations consistent with the aquarium theme; divers swim into the aquarium and inform the participants that they earned a reward (eg, meme or life insight). Further attention-to-detail is provided to improve user experience; for example, Sunday's survey contains 2 parts where participants answer a few questions about their day, followed by a few questions about the past week. Since Sunday's survey is longer than other days, we include a fun question right after the daily questions to entertain and energize participants before asking them additional questions about the past week. This fun question is randomly selected from a set of 5 questions in Multimedia Appendix 7.

With this, we conclude the description of different engagement strategies in SARA. Figure 3 provides a summary of the engagement strategies and how these strategies affect various theoretical constructs to influence self-report completion in SARA.

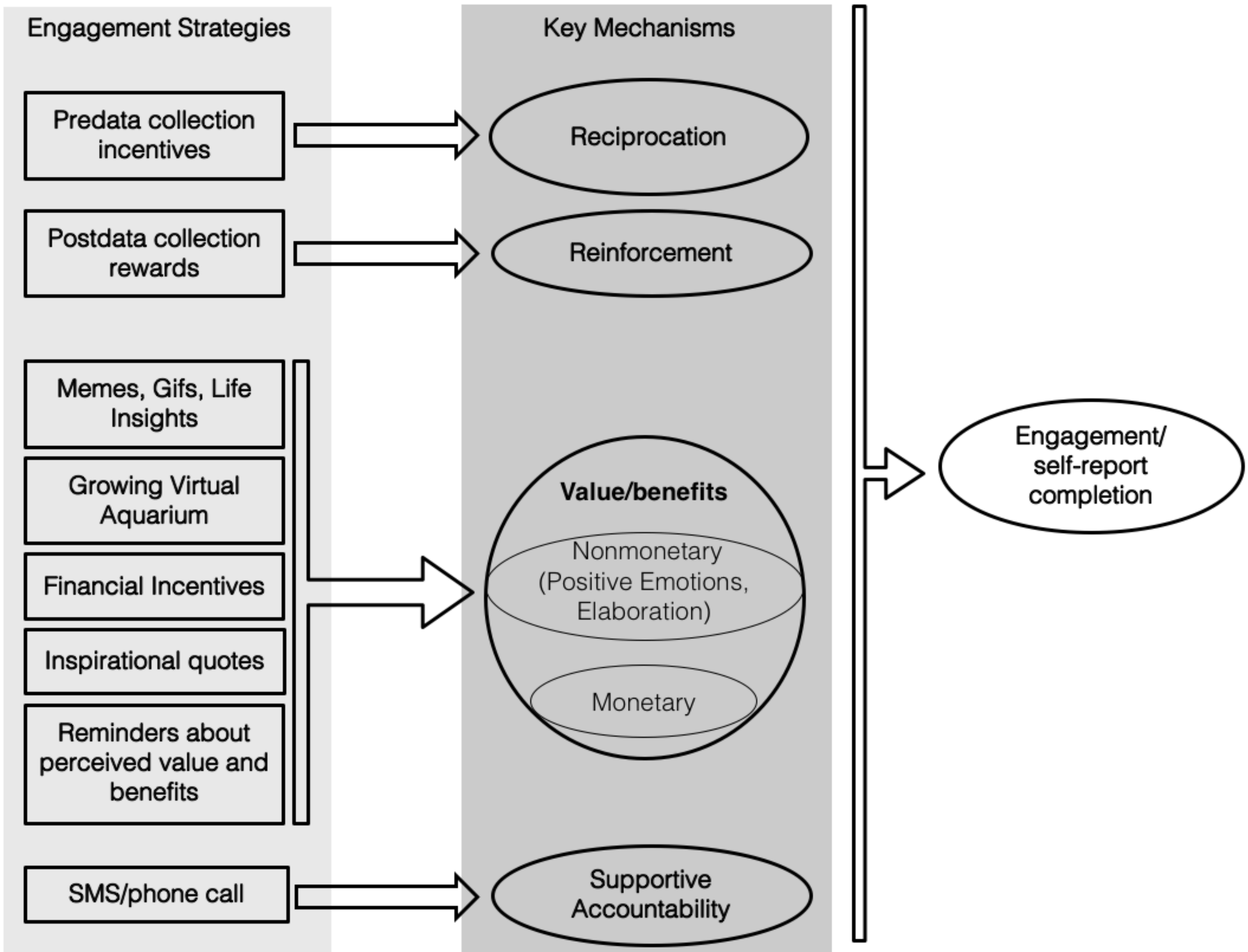
Implementation Details

We used the cross-platform JavaScript framework Ionic to build the Android and iPhone versions of SARA. The aquarium part uses the Phaser 2D game library for the animations. All the self-reported data in SARA are encrypted to comply with the Health Insurance Portability and Accountability Act (HIPAA). All data are stored in Amazon S3, a HIPAA-compliant data storage service, and the communication between Amazon S3 and the mobile phone app is encrypted with RSA 2048 and AES-256 [80].

Objective

In the previous section, we described several of SARA's novel engagement strategies. Although SARA engagement strategies were designed to influence self-report completion among youth who use substances, the effectiveness of these strategies is unknown. Important questions include whether different SARA engagement strategies lead to higher self-report completion and how their effectiveness is moderated by context (eg, negative affect, stress, loneliness, etc). Ineffective engagement strategies may aggravate participants, and participants may habituate even to an initially effective engagement strategy [81,82], resulting in a decrease, over time, in the effectiveness of the strategy. Similar drawbacks might occur if an engagement strategy is used in a context in which it is ineffective; indeed, the current context of each participant may influence the effectiveness of an engagement strategy [77]. One way to gain insight into these questions is by experimentally manipulating the engagement strategies over time. In other words, the manipulation can inform the development of a policy for adaptive engagement strategy delivery to keep participants engaged in self-reporting. Recently, MRTs have been proposed as a method to develop JITAIs for mHealth [42,43]. In SARA, the intervention is composed of the engagement strategies, such as giving a meme or issuing an inspirational message. In MRTs, study participants are randomized sequentially, often multiple times a day, to receive different intervention components. Several key elements in MRTs are decision points and proximal outcomes. *Decision points* refer to the time points when participants are randomized. In SARA, there is a 4 pm decision point at which a participant is randomized to receive an inspirational message via push notification or receive nothing. Once a participant is randomized at a decision point, the *outcome* of using an engagement strategy can be measured *proximally*. A proximal measure for the 4 pm inspirational message is whether or not the participant self-reports later on the same day. For more details on MRTs, we refer the readers to [42,43,83]. In the following, we describe the MRT protocol for SARA.

Figure 3. A conceptual diagram of how the different engagement strategies should affect self-report completion.



Study Protocol

We plan to run a 30-day MRT. Prior to the study, potential participants are screened, and eligible participants complete an in-person intake session with a study recruiter. Within the 30-day study, each participant is randomized at each of 4 decision points per day. After the 30 days, each participant takes part in a follow-up phone interview and provides feedback on using SARA for a month. The details of the study are as follows:

Study Setting and Eligibility

The 30-day MRT is conducted at the University of Michigan. Study participants are recruited from the University of Michigan Hospital Pediatric and Adult Emergency Departments. The study is approved by the Institutional Review Board of the University of Michigan Health System (HUM00121553) and is registered at ClinicalTrials.gov (NCT03255317). Patients are eligible for screening if they are aged between 14 and 24 years, understand English, are medically stable, are able to provide informed consent or assent (eg, not cognitively impaired or intoxicated), and are accompanied by a parent or guardian (for patients aged between 14 and 17 years). Individuals are eligible if they 1) have an Android or an iPhone mobile phone on which the app can be downloaded and 2) screen positive for past-month binge drinking (≥4 drinks for females and ≥5 drinks for males, on at least 1 occasion) or past-month cannabis use without a medical marijuana card.

Baseline Procedure

At the University of Michigan Emergency Department, recruiters monitor incoming admissions and identify patients in the target age range who do not meet the screening exclusion criteria (eg, presentation for sexual assault, droplet precautions, active vomiting, roomed in critical care). After completing an online consent or assent, participants self-administer a screening survey on a tablet, which contains questions regarding their demographics [84,85], health behaviors such as alcohol and marijuana use [8] and sleep habits [86], cell phone capabilities, and social media use. The screening survey takes approximately 8 min to complete, and participants receive a small gift valued at US $1.00 (eg, headphones, water bottle, mobile phone armband) for completing it. Those who report past-month binge drinking or marijuana use and have access to their mobile phone while in the Emergency Department provide written consent or assent and self-administer an approximately 20-min baseline survey with items about (1) substance use frequency, consequences, overdose, and driving under the influence [86-88]; (2) violence involvement, injury, and risky sex behaviors [6,89-92]; (3) coping and mindfulness [93,94]; (4) social influences [95-97]; and (5) motivation or self-efficacy to reduce their alcohol and marijuana use [98,99].

Then, the recruiters collect contact information from the participants. Recruiters ensure that participants install SARA on their phone during the intake since in the pilot study prior to this trial, we have found that many potential participants did not install the app after they left the intake session. Basic instructions on how to use the app are provided during intake, followed by motivational statements regarding participant's stated barriers to using SARA regularly. Participants receive US $20 cash for completing their intake visit.

The 30-Day Microrandomized Trial

Following the intake session, participants start the 30-day MRT with SARA. Each day, participants are requested to complete 1 survey and 2 active tasks between 6 pm and midnight. Participants are randomized at the following 4 decision points each day:

1. *4 pm reciprocity notification*: Every day at 4 pm, with a probability of .5, participants are randomized to either receive a push notification with an inspirational message—a song lyric or a quote from a contemporary celebrity—or to not receive anything.
2. *6 pm push reminder notification*: Every day at 6 pm, a reminder notification is issued. The notification may be one of the following 2 types, randomized at .5 probability:
   a. A simple reminder to complete a survey and 2 active tasks
   b. A reminder to complete a survey and 2 active tasks, along with an additional persuasive message. The additional persuasive message may be one of the following 3 messages:
      - "Do you know it only takes a minute to fill out the survey and active tasks?"
      - "Do you know you can earn money if you complete a 3-day streak?"
      - "You are close to unlocking the next fish for the aquarium."
3. *Reinforcement after survey completion*: If the survey is completed, participants are randomized at .5 probability to receive or not receive either a meme or a gif as a positive reinforcement. If a reinforcement is delivered, there is a .5 probability to receive a meme and a .5 probability to receive a gif.
4. *Reinforcement after active task completion*: If active tasks are completed, participants are randomized at .5 probability to receive or not receive a life insight as a positive reinforcement.

Note that not all the engagement strategies in SARA are randomized. The randomizations in 1-4 mentioned above are motivated by scientific questions concerning the abovementioned engagement strategies in SARA. Specifically, we randomize the push notifications because they can interrupt the participant and are intrusive, thus potentially reducing engagement. The other engagement strategies are not push notifications, so there is low risk of interruption; for example, the participant decides whether to open the app, complete self-reports, and receive the rewards. We randomize the memes and life insights because of the scientific question of how funny and informative contents respectively can influence self-report completion over time [44].

Exit Procedure

Participants are contacted by telephone to complete a 1-month follow-up interview, which includes the identical measures as the screen or baseline or daily survey or Sunday's survey, as well as the following 2 new threads of questions: (1) a 30-day Timeline Followback calendar, which captures past-month alcohol and marijuana consumption [100] and (2) Likert-type and open-ended questions to capture user experience of SARA [101]. For completing the follow-up interview, participants receive a US $30 electronic gift card of their choice (eg, Amazon, Starbucks, etc).

Analysis Plan

In this section, we describe the analysis plan to evaluate SARA. A more detailed version of the analysis plan can be found in a Center for Open Science document [102] (submitted on October 23, 2017).

Outcome Measures

The proximal outcomes of the randomizations are whether participants completed the survey and active tasks. For the inspirational message notifications at 4 pm and the reminder notifications at 6 pm, the proximal outcome is whether the survey or the active tasks are completed in the evening on the same day. For the 2 reinforcement interventions, the proximal outcome is whether the survey or the active tasks, respectively, are completed on the following day. Note that our engagement strategies are primarily designed to encourage self-reporting on the same or next day. Although the strategies may have longer term effects as well [81,82], we are interested in their effects on the same day or the next day; if these effects are consistently higher, then longer term adherence will be higher too. In addition, poststudy open-ended feedback from the participants will be analyzed to refine future versions of SARA.

Primary and Secondary Analyses

Our primary analyses will concern the following 2 hypotheses:

1. Providing the 4 pm reciprocity notification will yield a higher rate of full completion of the survey or active task on the same day than providing no intervention ($P<.025$).
2. Among individuals who complete the survey, providing a postsurvey-completion meme or gif will yield a higher rate of completion of the survey or active task on the next day than not providing meme or gif reinforcement after survey completion ($P<.025$).

We selected these hypotheses as primary since our team found these hypotheses to be the most interesting scientifically. The 4 pm randomization is designed to address the question of whether a notification intended to facilitate reciprocation (by providing an inspirational message before self-report time) is useful, and the randomization of participants upon survey completion is designed to investigate whether providing a postdata collection reward increases data collection. Furthermore, we conjecture that these hypotheses have the greatest potential to be supported. In particular, as the 4 pm notification is proximal in time to the data collection (2 hours prior), there are fewer extraneous distracting circumstances that can occur during this short time window that would reduce the intervention effect. Both hypotheses also consider a contrast between an active agent (ie, the 4 pm notification or the meme) versus nothing. Since these 2 hypotheses are primary, we divide the standard *P* value of .05 by 2.

Our secondary analyses will concern the following 2 hypotheses:

1. The 6 pm reminder notification with an extra persuasive message will yield a higher rate of full completion of the survey or the active task on the same day than not providing the extra persuasive message.
2. Among individuals who complete the active tasks, offering a postactive task-completion life insight will yield a higher rate of the full completion of the survey or active task on the next day than not offering a life insight after active task completion.

We consider the first hypothesis as secondary because the randomization for the 6 pm reminder is between 2 active agents (reminder with vs without a persuasive message). This implies that the additional effect of the persuasive messages may be small; thus, this hypothesis has a more exploratory nature. We consider the second hypothesis as secondary because data must accumulate before the visualizations in life insights become interesting. In addition, this hypothesis has a more exploratory nature because our life insights are only visualizations of past data; in the future, we aim for potentially more potent life insights using prediction tools [103,104].

For both primary and secondary analyses, we will control for the following 3 covariates to reduce variance in the outcome of self-input completion: (1) whether the survey or the active tasks were fully completed on the previous day, (2) whether SMS text messages or phone calls were made in the last 24 hours, and (3) whether the app was opened in the prior 72 hours outside when a survey or active task was completed. We will not include other baseline variables such as gender and age as covariates in the primary analyses because inclusion of a covariate uses up degrees of freedom. We also anticipate that the *within-person* covariate "whether the survey or active tasks were fully completed on the previous day" will capture some of the variance due to baseline gender or age. We will also not include time as a covariate in the primary analyses for the abovementioned reasons as well. Note that the statistical methods [83] will adjust the standard errors of the estimated effects to account for within-person correlation across time in the outcome. For more details about our primary analysis plan, please refer to the Center for Open Science document [102].

Exploratory Analyses

We plan to run exploratory analyses to examine how the effectiveness of engagement strategies changes over time (we conjecture that the effectiveness will decrease). We will also run additional exploratory analyses to assess effect moderation. We will examine how the effect of engagement strategies is moderated by gender, weekdays versus weekends, and whether the day is Sunday versus other days of the week (we expect that the completion rate for the longer Sunday's surveys may be lower than that for other days). We initially planned to assess age and phone type (Android or iPhone) as moderators; however, we did not include these variables because recruitment thus far indicates that very few participants in the study have Android phones or are below the age of 18 years.

Sample Size

We started recruiting for the study on August 21, 2017. Recruitment concluded on February 28, 2018. We recruited 73 participants for the study.

Statistical Analyses

In mHealth, it is common to collect time-varying measures of the participants' context (such as stress, mood, and loneliness from the self-report assessments). The provision of engagement strategies is time-varying as well; that is, at each decision point, participants can receive different options of the engagement strategy. A key statistical issue is that covariates (measures of the participant's context) at a time point can be affected by past engagement strategies. For such a setting, Boruvka et al [83] proposed a method to estimate the causal effects of interventions on continuous outcomes. However, in SARA, we are dealing with a binary outcome (whether participants self-reported or not). Multimedia Appendix 8 contains details of a method that we have developed, which extends the work of Boruvka et al to binary outcomes. We will use RStudio 1.1.453 to run the statistical analysis.

The open-ended qualitative data from exit interviews will be coded using thematic analysis [105]. The qualitative analysis will be performed using NVivo 11.

Missing Data

Since our outcome is adherence to self-report completion, not completing self-reports is not missing data for our study. However, missingness can happen in the study if participants uninstall the app in the middle of the study. To account for such missingness, we will conduct the following 3 versions of the primary analysis: (1) only include participants who had the app installed for 30 days, (2) include all participants and include only days when the app was installed, and (3) include all participants and data for both installed and uninstalled days; for the uninstalled days, the 4 pm and 6 pm notifications will be imputed, and the outcome will be considered as "self-report noncompletion." More details on how missingness will be accommodated can be found in the Center of Open Science document [102].

Results

We started recruiting for the study on August 21, 2017. Recruitment concluded on February 28, 2018. We recruited 73 participants for the study. Data analysis is currently underway.

Discussion

Future Work

To the best of our knowledge, this study is the first MRT to systematically explore the efficacy of different engagement strategies on increasing self-reporting of substance use data among adolescents and emerging adults. The results of this trial will answer how different engagement strategies affect self-reporting and to what degree (ie, effect size). We will also learn whether the effectiveness of the strategies varies over time. The qualitative data from the exit interviews will help us to further triangulate and understand the findings of the quantitative analysis. Moreover, the exit interviews' open-ended feedback about the app will help us further fine-tune the app.

For future studies, the analysis of the collected data can be used to initialize machine learning algorithms with the goal of providing engagement strategies in contexts (eg, stress level, loneliness, location, and weekend or weekdays), and at times, they are most effective. In particular, we will use the resulting data to train an initial policy for reinforcement learning algorithms such as "contextual bandits" [106]. As data accumulate on a participant, these algorithms increase the chance of providing the engagement strategy option that is most effective in a particular context and decrease the chance of providing an engagement strategy option that is less effective [107].

Another important direction of future work is the development of therapeutic interventions to prevent substance misuse. The daily surveys and active tasks will provide both subjective and objective data on substance use and related factors (eg, mood) over time. The Sunday's survey will uncover the days when substance use events happened; we will use these data as labels and the daily survey (ie, stress, emotion, or loneliness) or active tasks (ie, spatial memory, reaction time) as features to create machine learning models of impending substance use events. Therapeutic interventions will be provided, with high probability, at the time of impending substance use (or when participants are likely to engage in intervention content). Future versions of SARA will integrate these JITAIs to reduce substance misuse. Note that data-driven JITAIs to reduce substance use require maintenance of sufficient engagement of self-input completion.

Finally, one more future direction is to include sensor data collection from phones and wearables. Sensor data can be useful in multiple ways for SARA, such as (1) reducing self-input burden with predictions, for example, loneliness can be measured by inferring social interactions from the phone [33]; reducing self-input burden may increase engagement [77] and (2) risky times of substance misuse can be preanticipated from sensors; we expect that substance use events co-occur with certain behavioral markers that can be captured using sensors, for example, when participants are close to liquor stores or texted friends with whom they previously engaged in substance misuse [34].

Limitations

Few limitations of this study are as follows: first, this study is not designed to confirm that the SARA is more effective than other apps in collecting substance use-associated data. This trial is only designed to optimize the further development of SARA. We believe that this is a necessary first step since the science of engagement in mHealth is currently in its infancy. This MRT will inform the selection and adaptation of engagement strategies, as well as the development of future versions of SARA, which can be used as an experimental arm in randomized trials.

Second, the current version of SARA is limited because the design primarily focused on increasing willingness, that is, to the extent a participant is motivated to engage in self-reporting. Prior literature suggests 2 other methods to influence engagement [44,77,108-111]: (1) need, namely an individual's recognition that there is a discrepancy between his or her present state and a preferred future state, and (2) ability, namely the extent to which the individual has the knowledge, experience, skills, and capacity to engage in data collection. Although SARA is not currently designed to experiment with ability or need, SARA may influence them indirectly; for example, the reminder prompts before data collection can address forgetfulness and enhances the ability of participants to collect data. Furthermore, SARA sets clear goals, monitors engagement with aquarium progression, and offers timely feedback [112]. Thus, if participants become engaged with the aquarium or other incentives, they may recognize the need to engage.

A third limitation is the limited funding of the SARA mobile app, which is novel and exploratory in nature. Limited funding constrained our sample size. We also could not implement several features that were requested during user-centered design process; for example, better aesthetics of the fish, interactivity such as touching and interacting with fish, or personalization of such customizable background of the aquarium. In the future, we will use the results of this pilot study and apply for grants that can support larger studies and more resources for app development.

Finally, the study lacks therapeutic interventions to improve substance use outcomes. However, the engagement-only approach provides naturalistic data on substance use and related factors at the daily level, allowing us to study in-the-moment precedents and sequelae of substance use among adolescents and young adults. A better understanding of the in-the-moment precedents and sequelae of substance use is necessary to shape future therapeutic interventions that can be integrated into SARA.

Acknowledgments

This work has been supported by the Michigan Institute for Data Science (PI: SM), the University of Michigan Injury Center (PI: MW), NIDA P50 DA039838 (PI: Linda Collins), NIAAA R01 AA023187 (PI: SM), CDC R49 CE002099 (PI: MW), NIDA R01 DA039901 (PI: IN-S), NIDA K23 #036008 (PI: EEB), NHLBI/NIA R01 HL125440 (PI: PK), NIBIB U54EB020404 (PI: SK). We thank the members of Statistical Reinforcement Learning group at Harvard University and Dynamic Decision-Making Lab at University of Michigan for their input throughout the SARA development process.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questions and active tasks

[PDF File (Adobe PDF File), 265KB - resprot_v7i7e166_app1.pdf]

Multimedia Appendix 2

Fishes and points

[PDF File (Adobe PDF File), 548KB - resprot_v7i7e166_app2.pdf]

Multimedia Appendix 3

Gems and pearls

[PDF File (Adobe PDF File), 208KB - resprot_v7i7e166_app3.pdf]

Multimedia Appendix 4

Screenshots of life-insights

[PDF File (Adobe PDF File), 156KB - resprot_v7i7e166_app4.pdf]

Multimedia Appendix 5

Inspirational quotes.

[PDF File (Adobe PDF File), 32KB - resprot_v7i7e166_app5.pdf]

Multimedia Appendix 6

Protocol for texting and phone calls.

[PDF File (Adobe PDF File), 59KB - resprot_v7i7e166_app6.pdf]

Multimedia Appendix 7

Fun questions in the weekly survey

[PDF File (Adobe PDF File), 16KB - resprot_v7i7e166_app7.pdf]

Multimedia Appendix 8

Analysis steps to assess causal effects

[PDF File (Adobe PDF File), 74KB - resprot_v7i7e166_app8.pdf]

References

1. Erskine HE, Moffitt TE, Copeland WE, Costello EJ, Ferrari AJ, Patton G, et al. A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. Psychol Med 2015 May;45(7):1551-1563. [doi: 10.1017/S0033291714002888] [Medline: 25534496]
2. Volkow ND, Baler RD, Compton WM, Weiss SRB. Adverse health effects of marijuana use. N Engl J Med 2014 Jun 05;370(23):2219-2227 [FREE Full text] [doi: 10.1056/NEJMra1402309] [Medline: 24897085]
3. Benotsch EG, Koester S, Luckman D, Martin AM, Cejka A. Non-medical use of prescription drugs and sexual risk behavior in young adults. Addict Behav 2011;36(1-2):152-155. [doi: 10.1016/j.addbeh.2010.08.027] [Medline: 20863626]
4. Cunningham R, Walton MA, Maio RF, Blow FC, Weber JE, Mirel L. Violence and substance use among an injured emergency department population. Acad Emerg Med 2003 Jul;10(7):764-775 [FREE Full text] [Medline: 12837651]
5. Center FBHSAQ. Behavioral health trends in the United States:. Results from the 2014 National Survey on Drug Use and Health 2015.
6. Center FDCAP. Youth Risk Behavior Surveillance? United States, 2011. Morbidity & Mortality Weekly Report 2012;61(4).

7. Azofeifa A, Mattson ME, Schauer G, McAfee T, Grant A, Lyerla R. National Estimates of Marijuana Use and Related Indicators - National Survey on Drug Use and Health, United States, 2002-2014. MMWR Surveill Summ 2016 Sep 02;65(11):1-28. [doi: 10.15585/mmwr.ss6511a1] [Medline: 27584586]
8. Johnston L, O'Malley P, Bachman J, Schulenberg J. Monitoring the Future National Survey Results on Drug Use-2005. Volume II: College Students and Adults, Ages. Bethesda, MD: National Institute on Drug Abuse; 1975:06-5884.
9. ElSohly MA, Mehmedic Z, Foster S, Gon C, Chandra S, Church JC. Changes in Cannabis Potency Over the Last 2 Decades (1995-2014): Analysis of Current Data in the United States. Biol Psychiatry 2016 Apr 01;79(7):613-619 [FREE Full text] [doi: 10.1016/j.biopsych.2016.01.004] [Medline: 26903403]
10. Gruber SA, Dahlgren MK, Sagar KA, Gönenc A, Killgore WDS. Age of onset of marijuana use impacts inhibitory processing. Neurosci Lett 2012 Mar 09;511(2):89-94 [FREE Full text] [doi: 10.1016/j.neulet.2012.01.039] [Medline: 22306089]
11. Lopez-Larson MP, Bogorodzki P, Rogowska J, McGlade E, King JB, Terry J, et al. Altered prefrontal and insular cortical thickness in adolescent marijuana users. Behav Brain Res 2011 Jun 20;220(1):164-172 [FREE Full text] [doi: 10.1016/j.bbr.2011.02.001] [Medline: 21310189]
12. Schepis TS, Desai RA, Cavallo DA, Smith AE, McFetridge A, Liss TB, et al. Gender differences in adolescent marijuana use and associated psychosocial characteristics. J Addict Med 2011 Mar;5(1):65-73 [FREE Full text] [doi: 10.1097/ADM.0b013e3181d8dc62] [Medline: 21769049]
13. Harford TC, Grant BF, Yi H, Chen CM. Patterns of DSM-IV Alcohol Abuse and Dependence Criteria Among Adolescents and Adults: Results From the 2001 National Household Survey on Drug Abuse. Alcoholism: Clinical & Experimental Research 2005 May;29(5):810-828. [doi: 10.1097/01.alc.0000164381.67723.76] [Medline: 15897727]
14. Schillie SF, Shehab N, Thomas KE, Budnitz DS. Medication overdoses leading to emergency department visits among children. Am J Prev Med 2009 Sep;37(3):181-187. [doi: 10.1016/j.amepre.2009.05.018] [Medline: 19666156]
15. Abadi MH, Shamblen SR, Thompson K, Collins DA, Johnson K. Influence of risk and protective factors on substance use outcomes across developmental periods: a comparison of youth and young adults. Subst Use Misuse 2011;46(13):1604-1612. [doi: 10.3109/10826084.2011.598598] [Medline: 21899434]
16. Chassin L, Fora DB, King KM. Trajectories of alcohol and drug use and dependence from adolescence to adulthood: the effects of familial alcoholism and personality. J Abnorm Psychol 2004 Nov;113(4):483-498. [doi: 10.1037/0021-843X.113.4.483] [Medline: 15535782]
17. Jackson KM, Sher KJ, Schulenberg JE. Conjoint developmental trajectories of young adult substance use. Alcohol Clin Exp Res 2008 May;32(5):723-737 [FREE Full text] [doi: 10.1111/j.1530-0277.2008.00643.x] [Medline: 18331376]
18. Maggs J, Schulenberg J. Trajectories of alcohol use during the transition to adulthood. Alcohol Res Health 2004;28(4):195-201.
19. Shrier LA, Walls C, Rhoads A, Blood EA. Individual and contextual predictors of severity of marijuana use events among young frequent users. Addict Behav 2013 Jan;38(1):1448-1456 [FREE Full text] [doi: 10.1016/j.addbeh.2012.05.026] [Medline: 23010685]
20. Shrier LA, Rhoads A, Burke P, Walls C, Blood EA. Real-time, contextual intervention using mobile technology to reduce marijuana use among youth: a pilot study. Addict Behav 2014 Jan;39(1):173-180. [doi: 10.1016/j.addbeh.2013.09.028] [Medline: 24139665]
21. Bonar, Cunningham RM, Collins RL, Cranford JA, Chermack ST, Zimmerman MA, et al. Feasibility and Acceptability of Text Messaging to Assess Daily Substance Use and Sexual Behaviors among Urban Emerging Adults. Addict Res Theory 2018;26(2):103-113. [doi: 10.1080/16066359.2017.1310205] [Medline: 29632458]
22. Lenhart A, Duggan M, Perrin A, Stepler R, Rainie L, Parker K. Smartphones facilitate shifts in communication landscape for teens. Teens, Social Media & Technology Overview 2015: Pew Research Center; 2015.
23. Badawy SM, Kuhns LM. Texting and Mobile Phone App Interventions for Improving Adherence to Preventive Behavior in Adolescents: A Systematic Review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
24. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and Adolescents: A Systematic Review of Adolescents' Use of Mobile Phone and Tablet Apps That Support Personal Management of Their Chronic or Long-Term Physical Conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
25. Flurrymobile. App Engagement: The Matrix Reloaded URL: http://flurrymobile.tumblr.com/post/113379517625/app-engagement-the-matrix-reloaded [accessed 2018-01-02] [WebCite Cache ID 6wAn9i4To]
26. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - Or is there? A content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]
27. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A Systematic Review of the mHealth Interventions to Prevent Alcohol and Substance Abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
28. Chan A, Kow R, Cheng J. Adolescents' Perceptions on Smartphone Applications (Apps) for Health Management. JournalMTM 2017 Aug;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]

29. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
30. Gonzales R, Ang A, Murphy DA, Glik DC, Anglin MD. Substance use recovery outcomes among a cohort of youth participating in a mobile-based texting aftercare pilot program. J Subst Abuse Treat 2014 Jul;47(1):20-26 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.010] [Medline: 24629885]
31. McTavish FM, Chih M, Shah D, Gustafson DH. How Patients Recovering From Alcoholism Use a Smartphone Intervention. J Dual Diagn 2012;8(4):294-304 [FREE Full text] [doi: 10.1080/15504263.2012.723312] [Medline: 23316127]
32. Lane N, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell A. A survey of mobile phone sensing. IEEE Commun. Mag 2010 Sep;48(9):140-150. [doi: 10.1109/mcom.2010.5560598]
33. Rabbi M, Ali S, Choudhury T, Berke E. Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors. In: Proc ACM Int Conf Ubiquitous Computing. 2011 Presented at: Proceedings of the 13th International Conference on Ubiquitous Computing - UbiComp; 2011; Beijing, China p. 385-394 URL: http://europepmc.org/abstract/MED/25285324 [doi: 10.1145/2030112.2030164]
34. Bae S, Chung T, Ferreira D, Dey AK, Suffoletto B. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. Addict Behav 2017 Nov 27. [doi: 10.1016/j.addbeh.2017.11.039] [Medline: 29217132]
35. Marques PR, McKnight AS. Field and laboratory alcohol detection with 2 types of transdermal devices. Alcohol Clin Exp Res 2009 Apr;33(4):703-711. [doi: 10.1111/j.1530-0277.2008.00887.x] [Medline: 19170663]
36. SCRAM Systems. SCRAM CAM Provides Accountability and Encourages Compliance URL: https://www.scramsystems.com/products/scram-continuous-alcohol-monitoring/ [accessed 2018-01-02] [WebCite Cache ID 6wAoOlcQ8]
37. BABtrack. Track Your Alcohol Level Directly From Your Wrist URL: https://www.bactrack.com/pages/bactrack-skyn-wearable-alcohol-monitor [accessed 2018-01-02] [WebCite Cache ID 6wAoYGbpx]
38. Choe E, Abdullah S, Rabbi M, Thomaz E, Epstein D, Cordeiro F, et al. Semi-Automated Tracking: A Balanced Approach for Self-Monitoring Applications. IEEE Pervasive Comput 2017 Jan;16(1):74-84. [doi: 10.1109/mprv.2017.18]
39. Van Berkel N, Goncalves J, Hosio S, Kostakos V. Gamification of Mobile Experience Sampling Improves Data Quality and Quantity. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2017 Sep 11 Presented at: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1, no. 3 21; September 11, 2017; Maui, Hawaii, US p. 1-21. [doi: 10.1145/3130972]
40. Lynn P. The Impact of Incentives on Response Rates to Personal Interview Surveys: Role and Perceptions of Interviewers. International Journal of Public Opinion Research 13, no. 3 (September 1) 2001:326-336. [doi: 10.1093/ijpor/13.3.326]
41. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
42. 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 Dec;34 Suppl:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
43. Liao P, Klasnja P, Tewari A, Murphy SA. Sample size calculations for micro-randomized trials in mHealth. Stat Med 2016 Dec 30;35(12):1944-1971 [FREE Full text] [doi: 10.1002/sim.6847] [Medline: 26707831]
44. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
45. Skinner BF. About behaviorism. New York: Vintage Books; 2011.
46. Dragoi V, Staddon J. The dynamics of operant conditioning. Psychological Review 1999;106(1):20-61. [doi: 10.1037/0033-295x.106.1.20]
47. Reynolds, George S. A primer of operant conditioning. Rev 1975.
48. Ferster C, Skinner BF. Schedules of reinforcement. Chicago: Copley Publishing Group; 1997.
49. Gouldner. The Norm of Reciprocity: A Preliminary Statement. American Sociological Review 1960 Apr;25(2):161. [doi: 10.2307/2092623]
50. Blau P. Social exchange. International encyclopedia of the social sciences, 7 1968:452.
51. Cialdini R, Vincent J, Lewis S, Catalan J, Wheeler D, Darby B. Reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. Journal of Personality and Social Psychology 1975;31(2):206-215. [doi: 10.1037/h0076284]
52. Research Kit. Active tasks in Apple Research Kit URL: http://researchkit.org/docs/docs/ActiveTasks/ActiveTasks.html [accessed 2018-01-21] [WebCite Cache ID 6we0NWak8]
53. Nestor L, Roberts G, Garavan H, Hester R. Deficits in learning and memory: parahippocampal hyperactivity and frontocortical hypoactivity in cannabis users. Neuroimage 2008 Apr 15;40(3):1328-1339. [doi: 10.1016/j.neuroimage.2007.12.059] [Medline: 18296071]
54. Celio MA, Usala JM, Lisman SA, Johansen GE, Vetter-O'Hagen CS, Spear LP. Are we drunk yet? Motor versus cognitive cues of subjective intoxication. Alcohol Clin Exp Res 2014 Feb;38(2):538-544 [FREE Full text] [doi: 10.1111/acer.12276] [Medline: 24117655]

55. Cranford JA, Shrout PE, Iida M, Rafaeli E, Yip T, Bolger N. A procedure for evaluating sensitivity to within-person change: can mood measures in diary studies detect change reliably? Pers Soc Psychol Bull 2006 Jul;32(7):917-929 [FREE Full text] [doi: 10.1177/0146167206287721] [Medline: 16738025]
56. Ramirez-Valles J, Zimmerman MA, Newcomb MD. Sexual risk behavior among youth: modeling the influence of prosocial activities and socioeconomic factors. J Health Soc Behav 1998 Sep;39(3):237-253. [Medline: 9785696]
57. Hoyle R, Stephenson M, Palmgreen P, Lorch E, Donohew R. Reliability and validity of a brief measure of sensation seeking. Personality and Individual Differences 2002 Feb;32(3):401-414. [doi: 10.1016/s0191-8869(01)00032-0]
58. Lippman L, Moore K, Guzman L, Ryberg R, McIntosh H, Ramos M, et al. Flourishing children: Defining and testing indicators of positive development. Flourishing children: Defining and testing indicators of positive development. Springer; 2014.
59. Ellickson P, Bell R. Prospects for preventing drug use among young adolescents. Conrad N. Hilton Foundation, Rand Corporation 1990.
60. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995 Nov;51(6):768-774. [Medline: 8778124]
61. Stephens RS, Babor TF, Kadden R, Miller M, Marijuana Treatment Project Research Group. The Marijuana Treatment Project: rationale, design and participant characteristics. Addiction 2002 Dec;97 Suppl 1:109-124. [Medline: 12460133]
62. Simons J, Correia C, Carey K, Borsari B. Validating a five-factor marijuana motives measure: Relations with use, problems, and alcohol motives. Journal of Counseling Psychology 1998;45(3):265-273. [doi: 10.1037/0022-0167.45.3.265]
63. Grant VV, Stewart SH, O'Connor RM, Blackwell E, Conrod PJ. Psychometric evaluation of the five-factor Modified Drinking Motives Questionnaire--Revised in undergraduates. Addict Behav 2007 Nov;32(11):2611-2632. [doi: 10.1016/j.addbeh.2007.07.004] [Medline: 17716823]
64. Wray TB, Merrill JE, Monti PM. Using Ecological Momentary Assessment (EMA) to Assess Situation-Level Predictors of Alcohol Use and Alcohol-Related Consequences. Alcohol Res 2014;36(1):19-27 [FREE Full text] [Medline: 26258997]
65. Hursh S. Economic concepts for the analysis of behavior. J Exp Anal Behav 1980;34(2):219-238. [doi: 10.1901/JEAB.1980.34-219]
66. Brown E, Paul C. A Grounded Investigation of Game Immersion. 2004 Presented at: Extended Abstracts of the Conference on Human Factors and Computing Systems - CHI; April 24-29, 2004; Vienna, Austria p. 10. [doi: 10.1145/985921.986048]
67. Lin J, Lena M, Silvia L, Gregory D, Henry B. Fish nSteps: Encouraging Physical Activity with an Interactive Computer Game. UbiComp 2006. [doi: 10.1007/11853565_16]
68. Lane N, Mashfiqui M, Mu L, Xiaochao Y, Hong L, Shahid A, et al. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. 2011 Presented at: Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare; Sep; Dublin, Ireland. [doi: 10.4108/icst.pervasivehealth.2011.246161]
69. Abyssrium. URL: https://www.abyssrium.com/ [accessed 2018-06-11] [WebCite Cache ID 706heYOgW]
70. Miller W, Rose G. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
71. Vorauer JD. An information search model of evaluative concerns in intergroup interaction. Psychol Rev 2006 Oct;113(4):862-886. [doi: 10.1037/0033-295X.113.4.862] [Medline: 17014306]
72. van den Bos K. Making Sense of Life: The Existential Self Trying to Deal with Personal Uncertainty. Psychological Inquiry 2009 Dec 11;20(4):197-217. [doi: 10.1080/10478400903333411]
73. Nahum-Shani I, Henderson MM, Lim S, Vinokur AD. Supervisor support: does supervisor support buffer or exacerbate the adverse effects of supervisor undermining? J Appl Psychol 2014 May;99(3):484-503. [doi: 10.1037/a0035313] [Medline: 24490969]
74. Leonardwell G, Jessica L. The New Adventures of Regulatory Focus: Self-Uncertainty and the Quest for a Diagnostic Self-Evaluation. Handbook of the Uncertain Self 2010. [doi: 10.4324/9780203848753.ch14]
75. Weary G, Jacobson J. Causal uncertainty beliefs and diagnostic information seeking. Journal of Personality and Social Psychology 1997;73(4):839-848. [doi: 10.1037/0022-3514.73.4.839]
76. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]
77. Fogg, BJ. A Behavior Model for Persuasive Design. 2009 Presented at: Proceedings of the 4th International Conference on Persuasive Technology - Persuasive; April; Claremont, California, USA p. 40. [doi: 10.1145/1541948.1541999]
78. Buu A, Massey LS, Walton MA, Cranford JA, Zimmerman MA, Cunningham RM. Assessment methods and schedules for collecting daily process data on substance use related health behaviors: A randomized control study. Drug Alcohol Depend 2017 Sep 01;178:159-164. [doi: 10.1016/j.drugalcdep.2017.05.003] [Medline: 28651151]
79. Rabbi M, Philyaw-Kotov M, Lee J, Mansour A, Dent L, Wang X, et al. SARA: A Mobile App to Engage Users in Health Data Collection. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:781-789 [FREE Full text] [doi: 10.1145/3123024.3125611] [Medline: 29503985]
80. Huang H, Liu K. Efficient key management for preserving HIPAA regulations. Journal of Systems and Software 2011 Jan;84(1):113-119. [doi: 10.1016/j.jss.2010.08.056]

81. Thompson R. Habituation. International Encyclopedia of the Social & Behavioral Sciences, Elsevier 2015:480-483. [doi: 10.1016/B978-0-08-097086-8.55025-9]
82. Rankin C, Abrams T, Barry RJ, Bhatnagar S, Clayton DF, Colombo J, et al. Habituation revisited: an updated and revised description of the behavioral characteristics of habituation. Neurobiol Learn Mem 2009 Sep;92(2):135-138 [FREE Full text] [doi: 10.1016/j.nlm.2008.09.012] [Medline: 18854219]
83. Boruvka A, Almirall D, Witkiewitz K, Murphy S. Assessing Time-Varying Causal Effect Moderation in Mobile Health. Journal of the American Statistical Association 2017 Mar 31. [doi: 10.1080/01621459.2017.1305274]
84. Bearman P, Jones J, Udry J. The National Longitudinal Study of Adolescent to Adult Health. 1997. Add Health URL: http://www.cpc.unc.edu/projects/addhealth [accessed 2018-01-10] [WebCite Cache ID 6wN8XoOBP]
85. Centers for Disease Control and Prevention. 2005. Youth Risk Behavior Survey URL: https://www.cdc.gov/healthyyouth/data/yrbs/index.htm [WebCite Cache ID 706icc4LP]
86. Patel SR, Ayas NT, Malhotra MR, White DP, Schernhammer ES, Speizer FE, et al. A prospective study of sleep duration and mortality risk in women. Sleep 2004 May 01;27(3):440-444. [Medline: 15164896]
87. Ali R, Meena S, Eastwood B, Richards I, Marsden J. Ultra-rapid screening for substance-use disorders: the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST-Lite). Drug Alcohol Depend 2013 Sep 01;132(1-2):352-361. [doi: 10.1016/j.drugalcdep.2013.03.001] [Medline: 23561823]
88. Simons JS, Dvorak RD, Merrill JE, Read JP. Dimensions and severity of marijuana consequences: development and validation of the Marijuana Consequences Questionnaire (MACQ). Addict Behav 2012 May;37(5):613-621 [FREE Full text] [doi: 10.1016/j.addbeh.2012.01.008] [Medline: 22305645]
89. Straus M, Gelles R. New Scoring Methods for Violence and New Norms for the Conflict Tactics Scale. Indited with the assistance of Christine Smith, ed. Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families. New Brunswick, NJ: Transaction Publishers; 1999:529-535.
90. Wolfe DA, Scott K, Reitzel-Jaffe D, Wekerle C, Grasley C, Straatman AL. Development and validation of the Conflict in Adolescent Dating Relationships Inventory. Psychol Assess 2001 Jun;13(2):277-293. [Medline: 11433803]
91. Jelalian E, Alday S, Spirito A, Rasile D, Nobile C. Adolescent motor vehicle crashes: the relationship between behavioral factors and self-reported injury. J Adolesc Health 2000 Aug;27(2):84-93. [Medline: 10899468]
92. Spirito A, Jelalian E, Rasile D, Rohrbeck C, Vinnick L. Adolescent risk taking and self-reported injuries associated with substance use. Am J Drug Alcohol Abuse 2000 Feb;26(1):113-123. [Medline: 10718167]
93. Bohlmeijer E, ten KPM, Fledderus M, Veehof M, Baer R. Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. Assessment 2011 Sep;18(3):308-320. [doi: 10.1177/1073191111408231] [Medline: 21586480]
94. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
95. Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ. Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors. The Communities That Care Youth Survey. Eval Rev 2002 Dec;26(6):575-601. [doi: 10.1177/0193841X0202600601] [Medline: 12465571]
96. Hawkins JD, Catalano RF, Miller JY. Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: implications for substance abuse prevention. Psychol Bull 1992 Jul;112(1):64-105. [Medline: 1529040]
97. Cook SH, Bauermeister JA, Gordon-Messer D, Zimmerman MA. Online network influences on emerging adults' alcohol and drug use. J Youth Adolesc 2013 Nov;42(11):1674-1686 [FREE Full text] [doi: 10.1007/s10964-012-9869-1] [Medline: 23212348]
98. Breslin FC, Sobell LC, Sobell MB, Agrawal S. A comparison of a brief and long version of the Situational Confidence Questionnaire. Behav Res Ther 2000 Dec;38(12):1211-1220. [Medline: 11104185]
99. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
100. Hoeppner BB, Stout RL, Jackson KM, Barnett NP. How good is fine-grained Timeline Follow-back data? Comparing 30-day TLFB and repeated 7-day TLFB alcohol consumption reports on the person and daily level. Addict Behav 2010 Dec;35(12):1138-1143 [FREE Full text] [doi: 10.1016/j.addbeh.2010.08.013] [Medline: 20822852]
101. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
102. Rabbi M, Meredith P, Predrag K, Erin B, Inbal N, Maureen W, et al. SARA - Substance Abuse Research Assistant. Open Science Framework 2017 Oct 22 [FREE Full text] [doi: 10.17605/OSF.IO/VWZMD]
103. Rabbi M, Aung MH, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. 2015 Presented at: International Joint Conference on Pervasive and Ubiquitous Computing, ACM; Sep 7-11; Osaka, Japan p. 707-718. [doi: 10.1145/2750858.2805840]
104. Mads F, Doryab A, Maria F, Lars VK, Bardram J. Supporting disease insight through data analysis: refinements of the monarca self-assessment system. 2013 Presented at: International joint conference on Pervasive and ubiquitous computing; Sep 8-12; Zurich, CH p. 133-142. [doi: 10.1145/2493432.2493507]

105. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
106. Li, Lihong. Generalized Thompson sampling for contextual bandits. arXiv preprint arXiv.7163 2013.
107. Sutton R, Batro A. Reinforcement learning: An introduction. In: Vol. 1, no. 1. Cambridge: MIT press; 1998.
108. De Young R. New Ways to Promote Proenvironmental Behavior: Expanding and Evaluating Motives for Environmentally Responsible Behavior. J Social Isssues 2000 Jan;56(3):509-526. [doi: 10.1111/0022-4537.00181]
109. King G, Currie M, Petersen P. Child and parent engagement in the mental health intervention process: a motivational framework. Child Adolesc Ment Health 2012 Dec 14;19(1):2-8. [doi: 10.1111/camh.12015]
110. Lizardi D, Stanley B. Treatment engagement: a neglected aspect in the psychiatric care of suicidal patients. Psychiatr Serv 2010 Dec;61(12):1183-1191. [doi: 10.1176/ps.2010.61.12.1183] [Medline: 21123401]
111. Cavanagh K. Turn on, Tune in and (don't) Drop Out: Engagement, Adherence, Attrition, and Alliance with Internet-Based Interventions. Oxford Clinical Psychology 2015. [doi: 10.1093/med:psych/9780199590117.003.0021]
112. 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 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

Abbreviations

EMA: ecological momentary assessment
HIPAA: Health Insurance Portability and Accountability Act
JITAIs: just-in-time adaptive interventions
MACQ: Marijuana Consequences Questionnaire
MRT: microrandomized trial
RAs: research assistants
SARA: Substance Abuse Research Assistant

Edited by G Eysenbach; submitted 23.01.18; peer-reviewed by T Chung, J Lee, S Badawy, N Unciano; comments to author 08.03.18; revised version received 18.04.18; accepted 08.05.18; published 18.07.18

Please cite as:
Rabbi M, Philyaw Kotov M, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, Walton M, Murphy S
Toward Increasing Engagement in Substance Use Data Collection: Development of the Substance Abuse Research Assistant App and Protocol for a Microrandomized Trial Using Adolescents and Emerging Adults
JMIR Res Protoc 2018;7(7):e166
URL: http://www.researchprotocols.org/2018/7/e166/
doi:10.2196/resprot.9850
PMID:30021714