




# BMJ Open mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study

Adrian Aguilera,<sup>1,2</sup> Caroline A Figueroa ,<sup>1</sup> Rosa Hernandez-Ramos,<sup>1</sup> Urmimala Sarkar ,<sup>2</sup> Anupama Cembali,<sup>2</sup> Laura Gomez-Pathak,<sup>1</sup> Jose Miramontes,<sup>2</sup> Elad Yom-Tov,<sup>3</sup> Bibhas Chakraborty,<sup>4,5,6</sup> Xiaoxi Yan ,<sup>4</sup> Jing Xu,<sup>4</sup> Arghavan Modiri,<sup>7</sup> Jai Aggarwal,<sup>7</sup> Joseph Jay Williams,<sup>7</sup> Courtney R Lyles<sup>2</sup>

**To cite:** Aguilera A, Figueroa CA, Hernandez-Ramos R, *et al.* mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. *BMJ Open* 2020;**10**:e034723. doi:10.1136/bmjopen-2019-034723

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-034723>).

AA and CAF are joint first authors.

Received 08 October 2019  
Revised 07 March 2020  
Accepted 02 July 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Caroline A Figueroa;  
c.a.figueroa@berkeley.edu

## ABSTRACT

**Introduction** Depression and diabetes are highly disabling diseases with a high prevalence and high rate of comorbidity, particularly in low-income ethnic minority patients. Though comorbidity increases the risk of adverse outcomes and mortality, most clinical interventions target these diseases separately. Increasing physical activity might be effective to simultaneously lower depressive symptoms and improve glycaemic control. Self-management apps are a cost-effective, scalable and easy access treatment to increase physical activity. However, cutting-edge technological applications often do not reach vulnerable populations and are not tailored to an individual's behaviour and characteristics. Tailoring of interventions using machine learning methods likely increases the effectiveness of the intervention.

**Methods and analysis** In a three-arm randomised controlled trial, we will examine the effect of a text-messaging smartphone application to encourage physical activity in low-income ethnic minority patients with comorbid diabetes and depression. The adaptive intervention group receives messages chosen from different messaging banks by a reinforcement learning algorithm. The uniform random intervention group receives the same messages, but chosen from the messaging banks with equal probabilities. The control group receives a weekly mood message. We aim to recruit 276 adults from primary care clinics aged 18–75 years who have been diagnosed with current diabetes and show elevated depressive symptoms (Patient Health Questionnaire depression scale-8 (PHQ-8) >5). We will compare passively collected daily step counts, self-report PHQ-8 and most recent haemoglobin A1c from medical records at baseline and at intervention completion at 6-month follow-up.

**Ethics and dissemination** The Institutional Review Board at the University of California San Francisco approved this study (IRB: 17-22608). We plan to submit manuscripts describing our user-designed methods and testing of the adaptive learning algorithm and will submit the results of the trial for publication in peer-reviewed journals and presentations at (inter)-national scientific meetings.

**Trial registration number** NCT03490253; pre-results.

## Strengths and limitations of this study

- Novel approach of targeting diabetes and depressive symptoms using a smartphone application.
- Ability to compare adaptive messaging for increasing physical activity to messages delivered with equal probabilities.
- Testing of a smartphone application integrated within primary care settings in a low-income vulnerable patient population.
- Longitudinal design with 6-month follow-up enables assessing intervention effects over time.
- Challenges of this trial include supporting users in key behaviour change in an automated manner with minimal in-person support.

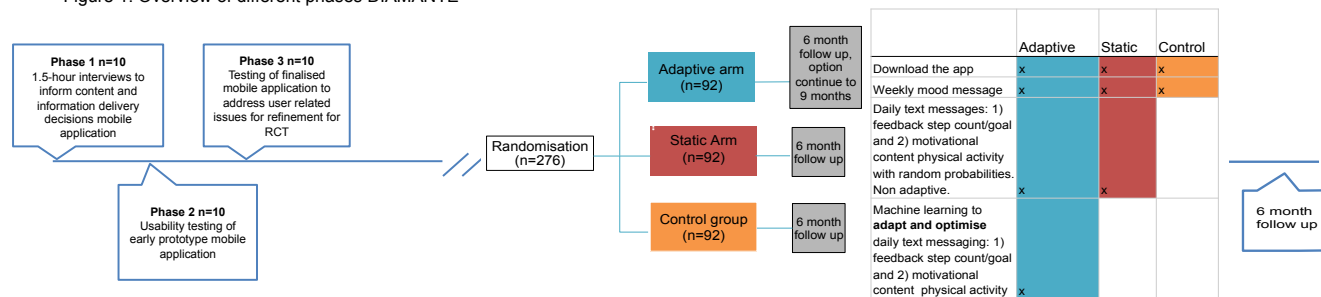
## INTRODUCTION

### Background

Both depression and diabetes are highly prevalent, often co-occurring diseases that are among the major causes of global disability.<sup>1–3</sup> Comorbid diabetes and depression are associated with a worse prognosis of both diseases, including a higher rate of complications of diabetes, greater disability and an increased risk of mortality.<sup>4</sup> Vulnerable populations, including low-income, low health literacy and ethnic minority individuals experience higher prevalences and worse outcomes for both diabetes and depression.<sup>5 6</sup> There is a great need for the development of treatments that can target overlapping risk factors for diabetes and depression. A growing body of evidence suggests that physical activity (PA) is such a risk factor: it is linked to both mental health and diabetic outcomes.<sup>7–11</sup>

Mobile applications have been found effective in helping patients engage in healthy behaviours including PA. For instance, a recent meta-analysis of nine RCTs concluded

Figure 1. Overview of different phases DIAMANTE



**Figure 1** Overview user-centred design process (UCD) and randomised trial. We first conducted three iterative phases of only UCD with 10 patients each (total n=30). The Diabetes and Mental Health Adaptive Notification Tracking and Evaluation trial will have different intervention groups: adaptive (n=92), static (n=92) and control (n=92). RCT, randomised controlled trial.

that smartphone apps that focus on PA have a moderate positive effect on increasing PA levels, and another meta-analysis including 18 studies moderate-to-large effect in daily step changes.<sup>12 13</sup> These effect sizes are similar to ‘face-to-face’ interventions.<sup>14</sup> However, because around 70% of lower income Americans (including Latinos) currently own smartphones,<sup>15</sup> and the ownership of smartphones is expected to increase in low-income populations globally,<sup>16</sup> mobile apps have great potential to reach individuals that normally do not have access to care. Further, mobile technology can help overcome existing barriers in access to care for vulnerable populations, including lower availability of psychological treatment in primary care settings, language and literacy barriers, stigma, cost and inflexible employment schedules.<sup>17</sup> Latinos in the USA in particular show higher underutilisation rates of mental health treatment than non-Latino whites.<sup>18</sup> Deploying effective mobile applications can therefore decrease existing decrease disparities in health.

However, in the USA, low-income minority patients frequently receive their care in safety net settings (services in the public sector for those unable to attain private health insurance), where novel mobile technologies are not often designed, developed or implemented.<sup>19</sup> This translational gap increases the probability that these interventions will ultimately fail when implemented in actual clinical settings that serve vulnerable populations.

Further, most mobile applications that target behavioural changes are not personalised,<sup>20</sup> which could contribute to lower effect sizes of these interventions for trials with longer study durations (eg, over 3 months).<sup>12</sup> Smartphone interventions allow for data collection by passive sensing technologies, which offer an opportunity for tailoring and personalising interventions to users’ behaviours, preferences and needs. Personalisation can be achieved by computer tailoring: tailoring interventions to observed behaviour and characteristics of the participant. This can include feedback, goal setting or user targeting (ie, conveying that communication is designed specifically for the user).<sup>21</sup> However, more complex forms of personalisation might be needed to

increase and maintain engagement with PA interventions,<sup>22</sup> a requirement for a digital intervention to be effective.

One promising approach is to use adaptive learning, which allows the prediction of which content might be effective for users, learning from its previous actions and participant data collected by mobile phones.<sup>23</sup> For instance, an adaptive learning algorithm might be more effective as it can match motivation needs for each participant. Some intervention studies have attempted cultural tailoring to large groups of people. However, most tailoring approaches occur at the group level whereas we are bringing the tailoring down to the individual level.<sup>24</sup> Research on the efficacy of these adaptive treatments is still in its early stages.

### Aim and hypotheses

The main aim of the ‘Diabetes and Mental Health Adaptive Notification Tracking and Evaluation’ trial (DIAMANTE) is to test a smartphone intervention that generates adaptive messaging, learning from daily patient data to personalise the timing and type of text messages. We will compare the adaptive content to (1) a uniform random messaging intervention, in which the messaging content and timing will be delivered with equal probabilities (ie, not adapted by a learning algorithm), (2) a control condition that only delivers a weekly mood message. The primary outcomes for this aim will be improvements in PA at 6-month follow-up defined by daily step counts.

We will investigate the following specific hypotheses:

### Primary hypothesis

We expect that the group receiving adaptive messaging will have a statistically significant higher increase in step counts over the 6-month intervention period, compared with both the group receiving the uniform random messaging and the control group receiving only a weekly mood message. We expect that both the adaptive and the uniform random messaging groups will have a higher increase in step counts than the control group.

**Table 1** Questionnaire measures obtained during the study

Measures	Baseline	6-month follow-up	Weekly
PHQ-8	x	x	
RM 1-FM	x	x	
Mood-rating (scale 1–9)			x
SEMCD	x	x	
GAD-7	x	x	
BIS/BAS	x		
CAHPS	x	x	
BADS-SF	x	x	
LEIDS-R	x	x	
Neighbourhood Scale*	x		
UCLA Loneliness 3-item	x		
Social Support Scale*	x	x	
Mobile Phone Affinity Scale*	x		
IPAQ*	x	x	
System Usability Scale		x	

The measures above were taken from validated questionnaires.

\*The International Physical Activity Questionnaire (IPAQ), Physical Activity Stages of Change—Questionnaire (RM 1-FM), Self-Efficacy for Managing Chronic Disease (SEMCD), System Usability Scale, Leiden Index Depression Sensitivity-Revised (LEIDS-R), Behavioral Activation for Depression Scale- Short Form (BADS-SF) and Social Support Scale were modified to match the sociodemographic characteristics of our patient population (ie, rephrasing to decrease literacy levels). The Patient Health Questionnaire depression scale (PHQ-8), Generalized Anxiety Disorder 7-item (GAD-7) scale, Behavioural Avoidance/Inhibition Scale (BAS/BIS), Mobile Phone Affinity Scale, Neighbourhood Scale, University of California Los Angeles (UCLA) Loneliness 3-item scale, Consumer Assessment of Healthcare Providers and Systems (CAHPS) scale and System Usability were not modified. We translated the RM 1-FM, the Neighbourhood Scale, UCLA Loneliness 3-item subscale and Mobile Phone Affinity Scale questionnaires into Spanish. For the PHQ-8, GAD-7, BADS-SF, BAS/BIS, SEMCD, IPAQ, LEIDS-R, CAHPS scale and System Usability Scale, we used validated translated versions used in previous studies. The Spanish translations of the BADS-SF, LEIDS-R, SEMCD and IPAQ were modified to match the sociodemographic characteristics of our patient population (ie, rephrasing to decrease literacy levels).

### Secondary hypotheses

- We expect that the group receiving adaptive, personalised messages will have significant reduction in haemoglobin A1c (HbA1c) levels, depression scores, weekly mood ratings, higher behavioural activation, lower rumination and anxiety, more positive attitudes about PA and higher self-efficacy to manage chronic diseases at 6 months, compared with the group that

receives the uniform random messaging intervention and the control group.

- Exploratory: because of the unique nature of the patient population enrolled in our study, we expect to find differences by language (English vs Spanish). Language has been shown to be especially influential in mobile technology uptake and effectiveness in our previous work.<sup>25</sup> Spanish speakers may engage more in the mobile intervention and might have different motivations for becoming physically active, which might result in differences in effectiveness of the various messages.

## METHODS AND ANALYSIS

### Design

This study is a randomised, controlled, single-centre superiority trial with three groups and a primary endpoint of increase in daily number of steps during a 6-month intervention delivered by a smartphone app. Randomisation will be performed as block randomisation with a 1:1:1 allocation. Patients will be automatically randomised into groups through our secure server during onboarding of the app, thereby ensuring allocation concealment. Patients will be informed of the nature and frequency of the messages they will be receiving and will be allowed to discuss this with investigators during the course of the study.

Further, if messages are not being sent out appropriately, research assistants will contact the app developer to address errors within 24 hours. If PA data are not coming in, they may need to contact participants to ensure that the app is actively running on their phone, and assist participants with redownloading the app if necessary. The necessity of these steps makes it unfeasible to blind the researchers. We used the Standard Protocol Items: Recommendations for Interventional Trials checklist when writing our report.<sup>26</sup>

### Population

Patients will be recruited within various clinics in the San Francisco Health Network, including the Richard Fine People's Clinic, the Family Health Center and from several endocrinologist providers at the Zuckerberg San Francisco General hospital (ZSFG); as well as the Department of Public Health in San Francisco, California in the USA that will serve as the study sites for this trial. These clinics serve a diverse, chronically ill population and, as in many public hospital settings, they employ multiple part-time providers, creating the risk for lack of continuity of care that mobile health programmes may be able to supplement.

### Inclusion criteria

Patients aged 18–75 years who have been diagnosed with diabetes and documented depressive symptoms (score >5 on the PHQ-8), as determined by their medical health records, will be eligible for this study. If the patients' health records show a diagnosis of diabetes but do not



include information on PHQ-8 scores or depression diagnosis, we will assess patients' PHQ-8 scores by telephone to determine their eligibility. After the assessment, this score will be uploaded into the patient health records, regardless of their eligibility for the study. Further, the researchers will reach out to the patient and his/her primary care physician if participants screen >20 on the PHQ-8 (severe depression). Eligible patients will need to use text messaging and have an iPhone or Android smartphone in order to download the pedometer app onto their phones, but they do not need to be proficient in using mobile phone applications. Concomitant care and interventions are permitted during the trial.

#### Exclusion criteria

We will exclude patients with: an inability to exercise due to physical disability; active psychosis or mania; active suicidal ideation; severe cognitive impairment; inability to read and write in English or Spanish; current pregnancy; plans to leave the country for extended periods of time during the 6-month trial.

#### Recruitment

Patients will be recruited using flyers placed in the waiting rooms, direct provider referrals and in-person recruitment strategies timed to visits. To identify potentially eligible patients, we will ask for permission from known providers to pull patient lists, and review and identify patients with dual diagnosis of diabetes and depression that would fit our inclusion criteria (eg, is able to walk and is not pregnant).

#### Baseline visit

A researcher or healthcare provider will ask patients if they are interested in joining a study to increase PA and help manage their mood using their mobile phone. We will bring interested individuals into our offices at ZSFG for a baseline study visit and to obtain informed consent in Spanish or English. Thereafter, we will collect all baseline survey measures of interests (outlined below) as well as patient demographics and information about current mobile technology familiarity and utilisation. All patients will receive assistance in downloading a pedometer application onto their phone and will send test text messages back to our system. The researcher will construct a plan for PA goals with the patient and instruct patients to have the app open at all times. Thereafter, users will be automatically randomised by the secure server, using a block randomisation with block size 3 to allocate individuals into either the uniform random, adaptive or control group.

#### Partial patient and public involvement

Patients worked with us to design the text messages, mobile app user interface and were asked to assess the burden and time commitment of the study, as part of our user-centred design approach. We did not involve patients in other areas of the study design.

#### Measures

Our primary outcome, change in daily step counts, will be passively collected by a mobile phone application during the time that patients remain in the intervention. For secondary outcomes, we will derive HbA1c, the average plasma glucose over the previous eight to 12 weeks, recommended as a means to diagnose diabetes,<sup>27</sup> from patients' electronic health records (EHR). We will use the most recent, available measurement from a maximum of 12 months before participating in the study. After 6 months, we will again assess the most recent HbA1c (pulling from patients' EHR), ensuring that at least 3 months elapsed between baseline and follow-up HbA1c levels. For additional secondary outcomes, we will administer a survey at baseline and 6-month follow-up.

The project coordinator and research assistants will be responsible for managing patient data collection. Data will be stored on UCSF Qualtrics and the HealthySMS platform. Once data from all patients are collected, it will be stored on UCSF's Secure Box, a secure cloud-hosted platform. See [table 1](#) for all included questionnaires.

#### Engagement measures

In addition to PA and the measures mentioned above, we will also examine engagement measures, such as (1) times that the app was opened (2) time spent reading the messages (3) usability data, assessed by the Systems Usability Scale<sup>28</sup> and open-ended qualitative questions about their opinions of the app.

#### Procedure

##### Determining content of motivational message categories

We designed motivational text messages according to a behavioural theory framework, using the Capability, Opportunity, Motivation, Behavior (COM-B) model.<sup>29</sup> COM-B is a behavioural change model that proposes that engaging in a particular behaviour depends on the dimensions of capability (physical and psychological), opportunity (social and physical) and motivation (need to engage in the behaviour more than in other behaviours). Messages were designed to fit into these three dimensions. Before the start of the RCT, English and Spanish speaking patients tested the messages in the Amazon Mechanical Turk (MTurk) crowdsourcing platform. Patients on MTurk were given random messages and asked to categorise them into our overarching theoretical categories. MTurk workers and each coauthor also rated messages for their overall quality from 1 to 5.

##### User-centred design testing

We used user-centred design (UCD) methods<sup>30</sup> to iteratively develop the content and text messaging system. We conducted three iterative phases of UCD with 10 patients each (total n=30). The first phase consisted of 1.5 hour individual semistructured interviews that took place at the offices at ZSFG. Findings from phase 1 were used to inform content and information delivery decisions of the final intervention including selecting the thematic

message categories and the design decisions. The second phase ran as a usability test in which patients tested out an early prototype of the mobile application. The third phase tested out the final DIAMANTE intervention including thematic message content and finalised application in order to address any user-related issues prior to launching the randomised control trial (see [figure 1](#) for an overview of the different UCD phases and the RCT).

### Interventions

We employ a mobile phone app, ‘DIAMANTE’ developed by Audacious Software (<https://diamante.healthysms.org/>). This application tracks step counts by pooling from Google Fit, Apple HealthKit or the built-in pedometer on patients’ phones. We use a text-messaging platform HealthySMS, developed by Dr. Aguilera, to send text messages and manage patient responses back to our system. The app only needs to be installed once, but has to remain open consistently. The app is designed in English and Spanish versions and is freely available as a download from the Apple App Store and Android Google Play application.

[Figure 1](#) shows the different intervention groups during the trial period. Briefly, both the adaptive and uniform random group will receive the same types of messages: feedback (four active categories plus no message) and motivation (three active categories plus no message). However, the message categories, timing and frequency will be optimised by a reinforcement learning (RL) algorithm in the adaptive group, and will be delivered with equal probabilities in the uniform random group (following a uniform random distribution). The control group will not receive these messages (only a weekly mood check-in message). All groups will have the app downloaded on their phone and their steps will be passively tracked within the app. See [figure 1](#) for an overview.

### Control condition

Control patients will only instal the app on their phone and will not receive any feedback messages. They will receive one message a week, on a fixed day, asking them to assess their mood in the previous week on a scale of 1–9. The message will be sent weekly at 10:00. Non-responders will receive reminders to submit their mood self-assessments in 2 hours intervals.

### Uniform random message arm

We will send patients up to two messages per day within four randomly selected time intervals. These messages are based on the COM-B framework (examples shown in [table 2A,B](#)). In addition, they will receive one message, on the seventh day, which will ask patients to rate their mood on a scale from 1 to 9. PA (step-count/day) will be passively monitored via the app on their smartphone.

### Adaptive message arm

Patients in the adaptive messaging arm will receive the daily COM-B messages (equal to the uniform random arm, examples shown in [table 2A,B](#)), but the message

**Table 2** Daily motivational and feedback messages DIAMANTE study

**A: Different categories with feedback messages that the algorithm chooses from**

Feedback messages	Examples
0.No feedback message	
1.Reaching goal	“Yesterday, you did not reach your goal”
2.Steps walked yesterday	“Yesterday, you walked 3824 steps”
3.Walked more/less today than yesterday	“Yesterday, you walked more than your goal”
4.Steps walked yesterday, plus a positive/negative motivational message	“You walked 8000 steps yesterday. Great job!”

**B: Different categories with Mmotivational messages that the algorithm chooses from**

Motivational messages	Examples
0.No message	
1.Capability, describes the physical and psychological benefits of walking and exercise	“Doing more physical activity can help reduce feelings of fatigue”
2.Motivation, meant to increase self-confidence and the belief that one is capable of walking even in the face of challenges	“You have made changes to improve your health before, you can do it again”
3.Opportunity, physical and social environment cues that make it more likely to engage in walking	“It there a local park you have been waiting to visit? Use it as an opportunity to get out of the house and do more steps!”

The algorithm can also choose not to send a message (category no feedback).  
The algorithm can also choose not to send a message (category no motivation).

categories, timing and frequency, will not be chosen randomly, but by using an RL algorithm. This allows us to adequately assess whether differences in effects are driven by the use of the RL algorithm. In addition, they will receive one message, on the seventh day, which will ask patients to rate their mood on a scale from 1 to 9. PA (step-count/day) will be actively monitored via the app on their smartphone.

Patients in all groups will receive reminders to open the app if no data are being transmitted. Additionally, patients in all groups can reply ‘STOP’ or ‘PARAR’ if they wish to stop receiving messages. Finally, the researchers will monitor the incoming step data. The researchers will contact the patients by phone for troubleshooting if patients’ phones are not transmitting data for more than 3 days.

### Adaptive learning algorithm

We developed an RL algorithm (an area of machine learning) based on previous work.<sup>31</sup> On a daily basis, this algorithm assesses (1) which feedback message



(table 2A), (2) which motivational message (table 2B) and (3) which time period of the day (in intervals of 2.5 hours from 09:00 to 19:00) is predicted to maximise the number of steps walked the next day. We use algorithms for contextual multiarmed bandit (MAB) problems, as these have been employed in different domains (including mobile health) and show promise to maximise cumulative rewards in sequential decision tasks (here, which sequences of messages optimally promote PA).<sup>32</sup> A contextual MAB problem is an RL setting in which the algorithm chooses between different treatment options which all have different reward functions. The reward functions depend on contextual variables.

We use Thompson Sampling, a Bayesian method, which will allow us to continuously learn which feedback and motivational messages are effective for a user, based on contextual features like their previous PA, demographic and clinical characteristics (such as age, gender and PHQ-8 scores). Thompson sampling can effectively deal with small amounts of data and addresses the exploration/exploitation tradeoff.<sup>33</sup> As such, it frequently picks out from the most rewarding messages and occasionally explores the messages with uncertainty in their reward.

More specifically, each morning, the algorithm evaluates which messages will likely increase the PA for every participant in the upcoming day, and at which time period the messages should be delivered. The algorithm training data consist of the historical data of all participants (contextual variables), which include which messages were sent previously and within which time periods, and select clinical/demographic data (such as age, language and depression scores) to improve prediction abilities.

During the trial, patients will receive messages for 6 months. After 6 months, we will provide patients the possibility to remain receiving messages for up to 9 months.

## Statistical analysis plan

### Baseline characteristics

We will provide descriptive summaries and examine any potential imbalances between the intervention and control groups using  $\chi^2$ -tests for categorical variables and independent samples t-tests for continuous variables.

### Primary and secondary hypothesis

Analyses will be conducted on an intention-to-treat basis. We will use full-information maximum likelihood to handle missing data, which has been shown to be preferred over multiple imputations.<sup>34</sup> We will include patients in the analysis for primary and secondary outcomes that have at least 1 month of data available. The primary outcome measure, increase in daily step counts, and the secondary outcomes measures (HbA1c, PHQ-8 and other questionnaire data, see the Measures section) will be analysed using a conditional growth model, in line with previous work,<sup>35</sup> with intervention (adaptive, uniform random, control) being the fixed effect and time being nested within individuals. Let  $j=1, \dots, n$ ,  $i=1, \dots, t$ ,

where  $t$  is the number of recorded days for the  $j$ th patient and  $Y_{ij}$  is the daily steps count on the  $i$ th day of  $j$ th patient.

Let  $A_j=2, 1$  or  $0$  for patient  $j$  in adaptive, uniform random or control arm, and let  $T_{ij}$  be the  $i$ th day for patient  $j$ .

The daily steps for patients given the treatment, in equation form:

At level 1 (within patient):

$$Y_{ij} = \beta_{0j} + \beta_{1j} \times T_{ij} + \tau_{ij}, \tau_{ij} \sim N(0, \sigma_w^2)$$

At level 2 (between patient):

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}A_j + \alpha_j, \alpha_j \sim N(0, \sigma_b^2) \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}A_j \end{aligned}$$

As this is a randomised trial, we will not adjust for any patient characteristics or patient engagement measures in the primary intention-to-treat analysis unless there is a significant imbalance between the arms of the trial.

### Preplanned subgroup analyses

We will conduct an exploratory subgroup analyses for our primary and secondary hypotheses for English versus Spanish speaking patients. We expect at least 50% of the sample to be Spanish speakers given our previous testing recruitment.

### Power calculations and sample size

Power for growth models were calculated based on Monte Carlo simulations using the 'nlme' package in R.<sup>36</sup> We conducted simulations, using data from our own study team testing the app ( $n=47$ ) and data from pilot patients ( $n=8$ ) to estimate the model parameters for calculating the necessary sample size. We used an estimated mean increase of 1250 steps based on previous work.<sup>37 38</sup> With 5000 Monte Carlo simulation runs, number of days for each patient  $t \sim Uniform(28, 180)$ , we estimate that we need  $n=160$  for 2-arms (the adaptive vs uniform random) for 80% power. Inflating to include the control arm and controlling for 15% drop-out, we aim to recruit a total of 276 patients. We believe that low drop-out is feasible as (1) the primary outcome of interest will be passively assessed and thus requires limited proactive engagement by the user, and (2) we have a comprehensive protocol to automatically text users who have closed the app or are not transmitting data.

### Compensation

Patients will receive a compensation of US\$40 for participation in the baseline part of the study, and an additional US\$70 for completion of the 6-month follow-up.

### Data statement

We will submit study results for publication in peer-reviewed journals and presentation at (inter)national meetings, taking into account relevant reporting guidelines (eg, Consolidated Standards of Reporting Trials<sup>39</sup>). We will attempt to publish all findings in open-access journals when possible, or in other journals with a concurrent uploading of the manuscript content into PubMed Central for public access. Curated technical appendices,



statistical code and anonymised data will become freely available from the corresponding authors on request.

### Safety protocol

We will review the clinician dashboard for patients in all arms of the trial to identify safety concerns that may emerge over the course of the trial. In addition, we will be automatically notified when key events, such as self-reported suicidality (messages with the words “Kill, Die, Suicide, Don’t want to live, Do not want, Bridge, What is the point, Harm, Hurt, Gun” for English speaking patients; and “Tiro, Matar, Morir, Suicidio, suicidar, no quiero, puente, para que, daño, dañar, dano, danar” for Spanish speaking patients) occur. We will provide immediate phone outreach and will refer patients to urgent or emergency care as necessary. Dr. Aguilera will be available to consult and intervene as necessary for mental health emergencies.

### ETHICS AND DISSEMINATION

The UCSF Institutional Review Board (IRB) has approved this protocol. The informed consent form for this study can be found in the online supplementary file. All protocol amendments will be communicated for approval to the UCSF IRB. We will ensure that our text messaging content is publicly available through a creative commons licensing agreement. The HealthySMS system is available for use on request.

### DISCUSSION

In this randomised controlled trial, we aim to examine the effect of a smartphone app that uses RL to predict the most effective messages for increasing PA in 276 low-income, ethnic and racial minority patients with diabetes and depression in urban public sector primary care clinics. We will compare this intervention to uniform random messages, delivered with equal and unchanging probabilities, and a control group that only receives a weekly mood message.

### Decreasing health disparities

Though the numbers of mHealth pilot studies are increasing in vulnerable populations, many of these fail to follow through with an implementation component to the study design.<sup>40</sup> Here, we are using a blended design: while the intervention is in addition to current care, there are ways we are attempting to make it more a part of patients’ clinical care. For instance, patients are mainly approached through primary care health providers, which recommend eligible patients whom they think are directly interested in a PA intervention. In addition, we will make patients’ data available to providers: a summary of the step increase for that patient at the conclusion of the study and updated PHQ-8 and GAD scores entered into the record. Our study therefore is the first step to addressing this gap because of its integration in primary

care clinics that serve low-income patients. Future work should focus more specifically on implementation of the app as part of routine clinical care.

### PA measure

We chose passively collected daily step count from patients’ preowned digital devices as a measure of PA. Although there are many different ways to measure PA, daily step count seems to be a particularly relevant measure because of: (1) its relative ease to measure, and (2) the clinical importance of individuals’ walking behaviour. Low number of daily step counts have been associated with all-cause mortality in some longitudinal studies<sup>41</sup> and results from pooled population studies show clear dose–response effects of PA to overall mortality.<sup>42</sup> In patients with type 2 diabetes, several studies have now shown that increasing step counts can significantly decrease HbA1c levels. For instance, a 10 000 steps per day walking prescription increased steps and decreased HbA1c in patients with type 2 diabetes.<sup>43</sup> Further, Manjoo *et al* found that each SD increase in daily steps was associated with a 0.21% decrease in HbA1c.<sup>44</sup> However, negative findings have also been reported. For instance, a meta-analysis by Qiu *et al* found that step-counter use was associated with increased steps per day (over 1800 more steps compared with a control group) among people with diabetes, but not with lowering of HbA1c.<sup>45</sup>

In exploratory post hoc analyses, we will also be able to examine the more immediate effect of PA messages, for example, on hourly steps in addition to daily steps, which will help to improve future PA interventions (eg, deliver messages at the right times). For instance, it is possible that one could receive a message in the morning and make plans to walk in the afternoon or evening, or messages could have more of an immediate impact. This information is currently unknown.

### Personalisation of intervention

The results of this RCT will help us understand if adaptive mHealth interventions for depression and diabetes are more beneficial than interventions that do not use learning algorithms. If mHealth interventions are not personalised, their efficacy might be low, due to low engagement and high drop-out rates.<sup>20</sup> The use of machine learning to adapt interventions according to users’ characteristics and behaviours is still in its early stages, but shows promise.<sup>46</sup> For instance, Yov-Tom *et al* using an adaptive learning algorithm found that adaptive feedback messages were more effective in increasing the amount and speed of PA and also reduced HbA1c in sedentary patients with type 2 diabetes.<sup>31</sup> Further, Zhou *et al* showed short-term efficacy of using adaptive weekly step goals determined by RL in healthy patients.<sup>47</sup> The current study, with a relatively large group of patients, will further increase our understanding of the potential of machine-learning-driven text-messaging interventions.



## Limitations and strengths

### Limitations

First, the results of this study might be specific to this population of low-income ethnic minority patients and might therefore not be generalisable to other populations. However, the inclusion of vulnerable patients in a primary care setting increases the likelihood that this intervention will be effective for other underserved populations. Further, our study procedures do not allow a double-blind design, as researchers and patients need to be made aware of the nature and frequency of messages they are receiving. Additionally, as with all digital interventions, technical issues might arise leading to unreliable step-count data and reduced ability for the algorithm to predict the most effective messages. The researchers and technical personnel will frequently check our server for data collection troubleshooting. Patients will also be made aware when their data are not collected correctly.

### Strengths

Diabetes and depression are among the top 10 causes of disability in the USA.<sup>48</sup> Developing cost-effective and scalable models of care for patients with common chronic conditions has been postulated as of key importance in improving the performance of healthcare systems.<sup>49</sup> If mHealth apps that target diabetes and depression through their common risk-factor physical inactivity are effective, they can have a major public health impact. Further, because the learning algorithm that we apply in this study is automated and delivers adaptive messages based on patients' behaviours, it can potentially be applied in other patient populations with a wide range of conditions.

To conclude, the outcome of this trial will provide information on the effectiveness of a text-message-based smartphone app that uses machine learning to increase PA in low-income ethnic minority patients in primary care settings. The results will provide key information on the effectiveness of adaptive mobile applications, compared with more traditional static digital interventions. If effective, this application has the ability to decrease healthcare disparities by providing a type of personalised care to a diverse and traditionally hard to reach group of underserved patients.

### Author affiliations

<sup>1</sup>School of Social Welfare, University of California Berkeley, Berkeley, California, USA

<sup>2</sup>UCSF Center for Vulnerable Populations in the Division of General Internal Medicine San Francisco, Zuckerberg San Francisco General Hospital, San Francisco, California, USA

<sup>3</sup>Microsoft Research, Herzliya, Israel

<sup>4</sup>Centre for Quantitative Medicine, Duke-National University of Singapore Medical School, Singapore

<sup>5</sup>Department of Statistics and Applied Probability, National University of Singapore, Singapore

<sup>6</sup>Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA

<sup>7</sup>Computer Science, University of Toronto, Toronto, Ontario, Canada

**Acknowledgements** The authors would like to acknowledge Chris Karr who helped develop and manage the automated text messaging service, passive data

collection and reinforcement algorithm and Patricia Avila Garcia who was involved in the design and implementation of the user-centred design phase of the study.

**Contributors** AA and CRL designed the overall study. RH-R, LG-P, JM, AC, US and CF were involved in the design and implementation of the user-centred design phase and RCT, JJW, JM, JA, CF and EY-T were involved in the design of the reinforcement learning algorithm. CF wrote the first draft of the paper. BC, XY and JX revised the statistical analyses and power calculation sections of the paper. All authors revised the manuscript for relevant scientific content and approved the final version of the manuscript.

**Funding** This trial is funded by an R01 to Dr. Adrian Aguilera and Dr. Lyles, 1R01 HS25429-01 from the Agency for Healthcare Research and Quality.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### ORCID iDs

Caroline A Figueroa <http://orcid.org/0000-0003-0692-2244>

Urmimala Sarkar <http://orcid.org/0000-0003-4213-4405>

Xiaoxi Yan <http://orcid.org/0000-0003-3291-1637>

## REFERENCES

- Vos T, Abajobir AA, Abate KH, *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the global burden of disease study 2016. *Lancet* 2017;390:1211–59.
- Fisher EB, Chan JCN, Nan H, *et al*. Co-occurrence of diabetes and depression: conceptual considerations for an emerging global health challenge. *J Affect Disord* 2012;142:S56–66.
- Chireh B, Li M, D'Arcy C. Diabetes increases the risk of depression: a systematic review, meta-analysis and estimates of population attributable fractions based on prospective studies. *Prev Med Rep* 2019;14:100822.
- Sartorius N. Depression and diabetes. *Dialogues Clin Neurosci* 2018;20:47–52.
- Thomas J, Jones G, Scarinci I, *et al*. A descriptive and comparative study of the prevalence of depressive and anxiety disorders in low-income adults with type 2 diabetes and other chronic illnesses. *Diabetes Care* 2003;26:2311–7.
- Colon E, Giachello A, McIver L, *et al*. Diabetes and depression in the Hispanic/Latino community. *Clin Diabetes* 2013;31:43–5.
- Chekroud SR, Gueorguieva R, Zheutlin AB, *et al*. Association between physical exercise and mental health in 1.2 million individuals in the USA between 2011 and 2015: a cross-sectional study. *Lancet Psychiatry* 2018;5:739–46.
- Diabetes Canada Clinical Practice Guidelines Expert Committee, Sigal RJ, Armstrong MJ, *et al*. Physical activity and diabetes. *Can J Diabetes* 2018;42:S54–63.
- Cooney G, Dwan K, Greig C, *et al*. Exercise for depression Cochrane database. *Syst Rev* 2013;9:CD004366.
- Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. *Am J Prev Med* 2013;45:649–57.
- Avery L, Flynn D, van Wersch A, *et al*. Changing physical activity behavior in type 2 diabetes: a systematic review and meta-analysis of behavioral interventions. *Diabetes Care* 2012;35:2681–9.
- Romeo A, Edney S, Plotnikoff R, *et al*. Can smartphone apps increase physical activity? systematic review and meta-analysis. *J Med Internet Res* 2019;21:e12053.
- Gal R, May AM, van Overmeeren EJ, *et al*. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. *Sports Med Open* 2018;4:42.



- 14 Han Y, Yan J. The effect of face-to-face interventions in promoting physical activity. *Am J Nurs* 2014;114:23.
- 15 Schrire V, Beck W. Human heart transplantation--the pre-operative assessment. *S Afr Med J* 1967;41:1263-5.
- 16 Center PR. *Smartphone ownership on the rise in emerging economies*, 2018.
- 17 Vázquez MYG, Sexto CF, Rocha Álvaro, *et al*. Mobile phones and psychosocial therapies with vulnerable people: a first state of the art. *J Med Syst* 2016;40:157.
- 18 AfHRA Q. *National healthcare quality disparities report*, 2016.
- 19 Lyles CR, Handley MA, Ackerman SL, *et al*. Innovative implementation studies conducted in US safety net health care settings: a systematic review. *Am J Med Qual* 2019;34:293-306.
- 20 Polgreen LA, Anthony C, Carr L, *et al*. The effect of automated text messaging and goal setting on pedometer adherence and physical activity in patients with diabetes: a randomized controlled trial. *PLoS One* 2018;13:e0195797.
- 21 Monteiro-Guerra F, Rivera-Romero O, Fernandez-Luque L, Luque LF, *et al*. Personalization in real-time physical activity coaching using mobile applications: a scoping review. *IEEE J Biomed Health Inform* 2020;24:1738-1751.
- 22 Cheung KL, Durusu D, Sui X, *et al*. How recommender systems could support and enhance computer-tailored digital health programs: a scoping review. *Digit Health* 2019;5:205520761882472.
- 23 Nahum-Shani I, Smith SN, Spring BJ, *et al*. Just-in-time adaptive interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018;52:446-62.
- 24 Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. *Curr Diab Rep* 2014;14:534.
- 25 Aguilera A, Berridge C. Qualitative feedback from a text messaging intervention for depression: benefits, drawbacks, and cultural differences. *JMIR Mhealth Uhealth* 2014;2:e46.
- 26 Chan A-W, Tetzlaff JM, Altman DG, *et al*. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med* 2013;158:200-7.
- 27 Organization WH. *Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation*. Geneva: World Health Organization, 2011.
- 28 Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Interact* 2008;24:574-94.
- 29 Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011;6:42.
- 30 McCurdie T, Taneva S, Casselman M, *et al*. *mHealth consumer apps* : the case for user-centered design. *Biomed Instrum Technol* 2012;46:49-56.
- 31 Yom-Tov E, Feraru G, Kozdoba M, *et al*. Encouraging physical activity in patients with diabetes: intervention using a reinforcement learning system. *J Med Internet Res* 2017;19:e338.
- 32 Tewari A, Murphy SA. *From ads to interventions: contextual bandits in mobile health*. Mobile Health: Springer, 2017: 495-517.
- 33 Agrawal S, Goyal N. Analysis of thompson sampling for the multi-armed bandit problem. *Conference on Learning Theory*, 2012.
- 34 Cham H, Reshetnyak E, Rosenfeld B, *et al*. Full information maximum likelihood estimation for latent variable interactions with incomplete indicators. *Multivariate Behav Res* 2017;52:12-30.
- 35 Korinek EV, Phatak SS, Martin CA, *et al*. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *J Behav Med* 2018;41:74-86.
- 36 Pinheiro J, Bates D, DebRoy S, *et al*. *Package 'nlme'. Linear and Nonlinear Mixed Effects Models*, 2017: 3-1.
- 37 Fukuoka Y, Haskell W, Lin F, *et al*. Short- and long-term effects of a mobile phone App in conjunction with brief In-Person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. *JAMA Netw Open* 2019;2:e194281-e.
- 38 Chase J-AD. Interventions to increase physical activity among older adults: a meta-analysis. *Gerontologist* 2015;55:706-18.
- 39 Grant SP, Mayo-Wilson E, Melendez-Torres GJ, *et al*. Reporting quality of social and psychological intervention trials: a systematic review of reporting guidelines and trial publications. *PLoS One* 2013;8:e65442.
- 40 Anderson-Lewis C, Darville G, Mercado RE, *et al*. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. *JMIR Mhealth Uhealth* 2018;6:e128.
- 41 Yamamoto N, Miyazaki H, Shimada M, *et al*. Daily step count and all-cause mortality in a sample of Japanese elderly people: a cohort study. *BMC Public Health* 2018;18:540.
- 42 Cavero-Redondo I, Peleteiro B, Álvarez-Bueno C, *et al*. The effect of physical activity interventions on glycosylated haemoglobin (HbA<sub>1c</sub>) in non-diabetic populations: a systematic review and meta-analysis. *Sports Med* 2018;48:1151-64.
- 43 Fayehun AF, Olowookere OO, Ogunbode AM, *et al*. Walking prescription of 10 000 steps per day in patients with type 2 diabetes mellitus: a randomised trial in Nigerian general practice. *Br J Gen Pract* 2018;68:e139-45.
- 44 Manjoo P, Joseph L, Dasgupta K. Abdominal adiposity and daily step counts as determinants of glycemic control in a cohort of patients with type 2 diabetes mellitus. *Nutr Diabetes* 2012;2:e25.
- 45 Qiu S, Cai X, Chen X, *et al*. Step counter use in type 2 diabetes: a meta-analysis of randomized controlled trials. *BMC Med* 2014;12:36.
- 46 Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. *J Med Internet Res* 2019;21:e12286.
- 47 Zhou M, Fukuoka Y, Mintz Y, *et al*. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. *JMIR Mhealth Uhealth* 2018;6:e28.
- 48 US Burden of Disease Collaborators, Mokdad AH, Ballesteros K, *et al*. The state of US health, 1990-2016: burden of diseases, injuries, and risk factors among US states. *JAMA* 2018;319:1444-72.
- 49 Institute of Medicine Committee on Quality of Health Care in A. *Crossing the quality chasm: a new health system for the 21st century*. Washington DC: National Academies Press (US), 2001.