

Effectiveness of a Digital Health Intervention Leveraging Reinforcement Learning to Increase Physical Activity: Results from the DIAMANTE Randomized Clinical trial

Adrian Aguilera, Marvyn Arevalo Avalos, Jing Xu, Bibhas Chakraborty, Caroline Figueroa, Faviola Garcia, Karina Rosales, Rosa Hernandez-Ramos, Chris Karr, Joseph Williams, Lisa Ochoa-Frongia, Urmimala Sarkar, Elad Yom-Tov, Courtney Lyles

Submitted to: Journal of Medical Internet Research
on: May 23, 2024

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 25

 Figures 26

 Figure 1..... 27

 Figure 2..... 28

Effectiveness of a Digital Health Intervention Leveraging Reinforcement Learning to Increase Physical Activity: Results from the DIAMANTE Randomized Clinical trial

Adrian Aguilera^{1,2} PhD; Marvyn Arevalo Avalos¹ PhD; Jing Xu³ PhD; Bibhas Chakraborty³ PhD; Caroline Figueroa⁴ PhD; Faviola Garcia⁵ BA; Karina Rosales¹ BA; Rosa Hernandez-Ramos⁶ BA; Chris Karr⁷ BA; Joseph Williams⁸ PhD; Lisa Ochoa-Frongia⁹ MD; Urmimala Sarkar⁹ MD; Elad Yom-Tov¹⁰ PhD; Courtney Lyles¹¹ PhD

¹School of Social Welfare University of California Berkeley Berkeley US

²Department of Psychiatry and Behavioral Sciences University of California-San Francisco San Francisco US

³Centre for Quantitative Medicine Duke-NUS Medical School National University of Singapore Singapore SG

⁴Faculty of Technology, Policy, and Management Delft Technical University Delft NL

⁵Center for Vulnerable Populations University of California-San Francisco San Francisco US

⁶Department of Psychological Sciences University of California Irvine Irvine US

⁷Audacious Software Chicago US

⁸Department of Computer Science University of Toronto Toronto CA

⁹Department of Medicine University of California-San Francisco San Francisco US

¹⁰Department of Computer Science Bar-Ilan University Ramat Gan IL

¹¹Center for Healthcare Policy and Research Sacramento US

Corresponding Author:

Adrian Aguilera PhD
School of Social Welfare
University of California Berkeley
205 Haviland Hall
Berkeley
US

Abstract

Background: Digital and mobile health interventions using personalization via reinforcement learning algorithms have the potential to reach large numbers of people to support physical activity and help manage diabetes and depression in daily life.

Objective: The DIAMANTE study tested whether a digital physical activity intervention using personalized text messaging via reinforcement learning algorithms could increase step counts in a diverse, multilingual sample of people with diabetes and depression symptoms.

Methods: Design, Setting, and Participants: From January 2020 to June 2022, participants were recruited from four San Francisco-based public primary care clinics and through online platforms to participate in the 6-month randomized clinical trial. Eligibility criteria included English- or Spanish-language preference and a documented diagnosis of diabetes and elevated depression symptoms. Intervention: The trial had three arms: a control group receiving a weekly mood monitoring message, a random messaging group receiving randomly selected feedback and motivational text messages daily, and an adaptive messaging group receiving text messages selected by a reinforcement learning algorithm daily. Randomization was performed with a 1:1:1 allocation. Main Outcomes and Measures: The primary outcome, changes in daily step counts, was passively collected via a mobile app. The primary analysis assessed changes in daily step count using a linear mixed effects model. An a priori sub-analysis compared the primary step count outcome within recruitment samples.

Results: In total, 168 participants were analyzed, including 24% Spanish language preference and 38% from clinic-based recruitment. Participants in the adaptive messaging arm showed a significant step count increase of 19% ($P < 0.001$), in contrast to 1.6% and 3.9% step count increase in the random and control arms, respectively. Intervention effectiveness differences were observed between participants recruited from the San Francisco clinics and those via online platforms, with the significant step count trend persisting across both samples for participants in the adaptive group.

Conclusions: Our study supports the use of reinforcement learning algorithms for personalizing text messaging interventions to

increase physical activity in a diverse sample of people with diabetes and depression. It is the first to test this approach in a large, diverse, and multilingual sample. Clinical Trial: ClinicalTrials.gov Identifier: NCT03490253

(JMIR Preprints 23/05/2024:60834)

DOI: <https://doi.org/10.2196/preprints.60834>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [a JMIR Publications](#)

Original Manuscript

Effectiveness of a Digital Health Intervention Leveraging Reinforcement Learning to Increase Physical Activity: Results from the DIAMANTE Randomized Clinical trial

Adrian Aguilera, PhD^{1, 2}, Marvyn R. Arévalo Avalos, PhD¹, Jing Xu, PhD³, Bibhas Chakraborty, PhD³, Caroline Figueroa, PhD⁴, Faviola Garcia, BA⁵, Karina Rosales, BA¹, Rosa Hernandez-Ramos, BA⁶, Chris Karr, BA⁷, Joseph Williams, PhD⁸, Lisa Ochoa-Frongia⁹, Urmimala Sarkar, MD⁹, Elad Yom-Tov, PhD¹⁰, Courtney R. Lyles, PhD¹¹

¹School of Social Welfare, University of California Berkeley, Berkeley, California, United States

²Department of Psychiatry and Behavioral Sciences, University of California–San Francisco, San Francisco, California, United States

³Centre for Quantitative Medicine, Duke-NUS Medical School, National University of Singapore, Singapore

⁴Faculty of Technology, Policy, and Management, Delft Technical University, Delft, The Netherlands

⁵Center for Vulnerable Populations, University of California–San Francisco, San Francisco, California, United States

⁶Department of Psychological Sciences, University of California Irvine, Irvine, California, United States

⁷Audacious Software, Chicago, Illinois, United States

⁸Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

⁹Department of Medicine, University of California–San Francisco, San Francisco, California, United States

¹⁰Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel

¹¹Center for Healthcare Policy and Research, Sacramento, California, United States

Corresponding Author

Adrian Aguilera

(510) 642-8564
Haviland Hall 205
Berkeley, CA 94709

Abstract

Background: Digital and mobile health interventions using personalization via reinforcement learning algorithms have the potential to reach large numbers of people to support physical activity

and help manage diabetes and depression in daily life.

Objective: The DIAMANTE study tested whether a digital physical activity intervention using personalized text messaging via reinforcement learning algorithms could increase step counts in a diverse, multilingual sample of people with diabetes and depression symptoms.

Methods: *Design, Setting, and Participants:* From January 2020 to June 2022, participants were recruited from four San Francisco-based public primary care clinics and through online platforms to participate in the 6-month randomized clinical trial. Eligibility criteria included English- or Spanish-language preference and a documented diagnosis of diabetes and elevated depression symptoms.

Intervention: The trial had three arms: a control group receiving a weekly mood monitoring message, a random messaging group receiving randomly selected feedback and motivational text messages daily, and an adaptive messaging group receiving text messages selected by a reinforcement learning algorithm daily. Randomization was performed with a 1:1:1 allocation. *Main Outcomes and Measures:* The primary outcome, changes in daily step counts, was passively collected via a mobile app. the primary analysis assessed changes in daily step count using a linear mixed effects model. An *a priori* sub-analysis compared the primary step count outcome within recruitment samples.

Results: In total, 168 participants were analyzed, including 24% Spanish language preference and 38% from clinic-based recruitment. Participants in the adaptive messaging arm showed a significant step count increase of 19% ($P < 0.001$), in contrast to 1.6% and 3.9% step count increase in the random and control arms, respectively. Intervention effectiveness differences were observed between participants recruited from the San Francisco clinics and those via online platforms, with the significant step count trend persisting across both samples for participants in the adaptive group.

Conclusions: Our study supports the use of reinforcement learning algorithms for personalizing text messaging interventions to increase physical activity in a diverse sample of people with diabetes and depression. It is the first to test this approach in a large, diverse, and multilingual sample.

Trial Registration: ClinicalTrials.gov Identifier: NCT03490253



Introduction

Diabetes and depression are significant public health problems individually and interact to worsen outcomes for each other.⁽¹⁾ Depression is associated with unhealthy behaviors such as physical inactivity and poor dietary habits, which are also risk factors for the development of diabetes⁽²⁾. The psychosocial burden of living with a chronic illness like diabetes can contribute to the development and worsening of depression – leading to a cycle of multiplicative negative health effects. Diabetes and depression exhibit racial and socio-economic disparities stemming from known barriers from the individual to the structural levels, such as the stressors due to discrimination and poverty, difficulty in accessing care related to language barriers, competing health demands, and beyond.^{(3),(4)}

Physical activity is a core intervention for both diabetes and depression, as both are associated with low physical activity levels. Increased physical activity has positive effects on both conditions, including improved glycemic control in diabetes and reduced symptoms of depression.⁽⁵⁾ Behavioral activation interventions that incorporate physical activity as a core element are among the most effective for treating depression, relative to other behavioral and psychological approaches.^{(6),(7)} Targeting physical activity is thus an efficient method of improving outcomes in both diabetes and depression.

Digital and mobile health interventions have the potential to reach large numbers of people to support physical activity and help manage diabetes and depression in daily life, with a large research base using digital technology to improve and personalize health behavior programs.^{(8),(9)} Machine learning algorithms can help personalize and optimize interventions by tailoring them to the individual's specific needs, preferences, and capabilities. Reinforcement learning algorithms, in particular, have the potential to create adaptive and dynamic interventions focused on a goal/reward that can adjust based on an individual's responses and behaviors.⁽¹⁰⁾ In previous work these algorithms increased physical activity among patients with diabetes.⁽¹¹⁾ However, novel data science

methods such as personalization via reinforcement learning algorithms typically do not include data from marginalized populations.⁽¹²⁾ It is crucial that digital platforms and newer algorithms are built with diverse populations to increase their generalizability and effectiveness, particularly to make an impact on entrenched health inequities.

The Depression and Mental Health Adaptive Notification and Tracking Evaluation (DIAMANTE) study aimed to address these gaps by developing and evaluating a reinforcement learning-based digital health app co-designed with English- and Spanish speakers with comorbid diabetes and depression symptoms receiving care in a public safety-net healthcare setting. The app utilized reinforcement learning algorithms to tailor daily text messages based on feedback type, motivation type and message timing. Here we present the efficacy of the DIAMANTE intervention, in a three-arm randomized trial, on physical activity outcomes.

Methods

Study Design

We conducted a randomized controlled trial with three arms from January 2020-June 2022. We used the Standard Protocol Items: Recommendations for Interventional Trials checklist for reporting our findings.⁽¹³⁾ The study was approved by the UCSF institutional review board.

Participants

RCT recruitment occurred on two tracks. Clinic-based patient recruitment at the San Francisco Health Network (SFHN, the public healthcare delivery system for the city and county of San Francisco) was our first method of enrolling study participants. These clinic-based strategies were significantly impacted by the COVID-19 pandemic. Prior to March 2020 and after California's reopening in June 2021, patients were recruited by direct provider referrals, with in-person clinic recruitment timed to eligible patient visits. To identify potentially in-person eligible patients, we asked for permission from known providers to access patient lists, and then reviewed and identified patients with documented EHR diagnosis of diabetes and elevated depression symptoms (depression

diagnosis or Patient Health Questionnaire (PHQ-9) score of 5 or higher in the past 5-year period) and were adults 18 years or older. Patients were excluded if they were unable to walk or were currently pregnant. Researchers then contacted patients directly (via in-person visits and via phone calls) to determine interest in joining the 'Diabetes and Mental Health Adaptive Notification Tracking and Evaluation' (DIAMANTE) Study. Interested individuals were invited for a study visit which included obtaining informed consent and completing a baseline survey. All patients were offered assistance in downloading the DIAMANTE application onto their phone and sent test text messages back to our system. The researcher established a baseline plan for physical activity goals with the patient (i.e., average 4000 steps daily) and instructed patients to have the app open at all times. All participants were invited back at 6 months to complete a follow-up survey when the active intervention period ended.

We added a second online recruitment strategy during the COVID-19 pandemic, given that in-clinic recruitment was not possible for many months during the active enrollment period for our RCT. Specifically, between March 2020 and October 2021, we recruited through social media advertisements via ads posted on Facebook, Google Ads, and Craigslist. Potential participants completed a screening survey to assess eligibility (i.e., self-reported diabetes diagnosis and PHQ-8). In addition to the exclusion criteria listed above, we also excluded those that did not have an eligible phone number, were outside of the United States, or those who failed the CAPTCHA requirements to verify their identity online. Online participants were enrolled remotely, offering Zoom meeting and phone assistance to complete all steps outlined above (e.g., onboarding, downloading the app, setting a step count goal) as needed. All informed consent and survey data collection matched our in-clinic recruitment procedures.

The project coordinator and research assistants were responsible for managing patient data collection. Self-reported participant data was stored on REDCap, surveys were stored on UCSF Qualtrics, and daily step count and text messaging data were stored on the HealthySMS platform. All

participants gave written or electronic informed consent to participate in the study.

Randomization

Randomization was performed with block randomization and a 1:1:1 allocation into arms, stratified by patient language preference. Patients were automatically randomized into groups through our secure server during onboarding of the app, thereby ensuring allocation concealment. Patients were informed of the nature of the app and frequency of the messages they would be receiving. The necessity of these steps made it unfeasible to fully blind participants or research staff; however, the final data analyses were completely blinded.

Procedures

In brief, the DIAMANTE study compared the effectiveness of different text messaging strategies on physical activity (measured by step counts on participants' smartphones) over an active 6-month intervention period. We created a custom mobile phone app entitled DIAMANTE and developed by Audacious Software (<https://diamante.healthysms.org/>) for this study. This application passively tracks step counts by pooling from Google Fit, Apple HealthKit, or the built-in pedometer on patients' phones. We used a text-messaging platform HealthySMS, previously developed by Dr. Aguilera and Audacious Software, to integrate with the DIAMANTE app and send text messages by intervention arm. The DIAMANTE app only needed to be installed once and then remain open consistently. The app was designed in English and Spanish versions and was 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. The three comparator arms were: 1) Control arm, who only received one weekly message inquiring about their mood over the past week, 2) Random messaging arm, who received daily randomly selected messages (one feedback message on their step count (e.g., "You walked 4000 steps yesterday") and one motivational message about their health (e.g., "Going for a walk can improve your mood and clear your mind) from a pre-existing bank of messages, and 3) Adaptive messaging arm, who received

daily messages from the same feedback and motivational messaging banks, with message categories and times selected by a reinforcement learning algorithm. The random and adaptive arm also received the weekly mood message. In summary, the feedback messages were selected from 5 message types (step number, step number plus encouragement/reinforcement, relative [walked more/less relative to goal], relative [achieved or did not achieve goal], or no message), and motivation messages could be from 4 categories of messages based on the COM-B (capability, opportunity, motivation, behavior) model (each with 18 message options in these categories, or a “no message” option). Messages were generated from behavioral science theoretical domains, and examples are shown in Table 1. Moreover, patients and end users provided feedback on the text messages, mobile app interface and study protocols during user-centered design phases of the study.⁽¹⁴⁾

Table 1 – Feedback and motivational messages and algorithm message choices

A: Feedback messages	Examples	Random (n=51)	Adaptive (n=53)	P-value
0.No feedback message		19.5%	6.0%	<.001
1.Reaching goal	“Yesterday, you did not reach your goal”	19.9%	20.8%	.35
2.Steps walked yesterday	“Yesterday, you walked 3824 steps”	20.5%	28.1%	<.001
3.Walked more/less today than yesterday	“Yesterday, you walked more than your goal”	19.7%	25.6%	<.001
4.Steps walked yesterday, plus a positive/negative motivational message	“You walked 8000 steps yesterday. Great job!”	20.4%	19.5%	.54
B: Motivational messages	Examples			
0.No message		24.3%	9.3%	<.001
1.Capability, describes the	“Doing more physical activity	24.8%	37.2%	<.001

physical and psychological benefits of walking and exercise	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”	25.6%	19.3%	<.001
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!”	25.2%	34.2%	<.001

We developed the reinforcement learning algorithm (a type of machine learning) based on previous work.⁽¹⁵⁾ The reinforcement learning algorithm had a ‘cold-start’ beginning with a randomly selected message for each patient and used ongoing participant data for personalizing 1) the feedback message category selection and 2) the motivational message category selection, and 3) the timing of the message delivery daily (into 4 time periods from 8am-8pm) -- upweighting messages and times that had a higher probability of increasing steps. As described previously, we used reinforcement learning algorithms for contextual multi-armed bandit problems,⁽¹⁶⁾ as these can maximize cumulative rewards in sequential decision tasks (as here, which sequences of messages optimally promote the highest step count in the upcoming day). We also used Thompson Sampling, a Bayesian method that can handle small amounts of data,⁽¹⁷⁾ which allowed the algorithm to continuously learn which feedback and motivational messages were effective for a user, based on contextual features like their previous step counts and which messages were sent previously, as well as participant data like demographic and clinical characteristics (such as age, gender, and depressive symptoms scores collected at baseline). In summary, each morning within the adaptive intervention arm, the algorithms evaluated which categories of feedback and motivational messages would likely increase the step count for each participant in the upcoming day, and at which time period the messages should be delivered to increase step count. The rates of algorithm message choices are presented in

Table 1

Patients in all groups received reminders to open the app if no data were being transmitted. Additionally, patients in all groups could reply 'STOP' or 'PARAR' if they wished to stop receiving messages. Finally, the researchers monitored the incoming step data. The researchers contacted the patients by phone for troubleshooting if patients' phones were not transmitting data for more than 3 days (after the automated reminders).

Outcomes

Our primary outcome for the RCT, total change in daily step counts, was passively collected by a mobile phone application during the time that patients remain in the intervention (using Apple Healthkit, GoogleFit, and pedometer data from participants' phones). Prior to analysis, participants were required to have at least 28 days of step count data during the intervention in order to be included in the primary analysis, given that the algorithm needed sufficient step count data to be able to personalize content for each participant. Data quality was evaluated for all participants upon study exit (blinded by study arm), with removal of 5 participants whose data signaled a major discrepancy (>25% of available daily step counts greater than 20,000 steps), as this may have indicated a flaw in the DIAMANTE app step count data extraction.

Statistical Analysis

First, we reported on the overall demographics of the full sample and by intervention arm and overall step counts within the sample. Our intent-to-treat primary analysis then used longitudinal regression to determine the impact of the treatment arms on the change in step count for each participant. We analyzed changes in daily step count using a linear mixed-effects model (LMM) with an intervention variable (Adaptive arm as reference group, Control arm, and Random arm) and a time variable (day, Control*day, Random*day) as fixed effects and a random intercept for each participant. We plotted the raw step count data and the residuals to examine their distributions via histograms. As specified in our protocol, we also completed one *a priori* sub-analysis: the impact of

the intervention by recruitment sample (SFHN clinic-based recruitment vs. online recruitment). We analyzed descriptive data in SPSS v 28 and the LMM analysis using the 'lme4' package in R.

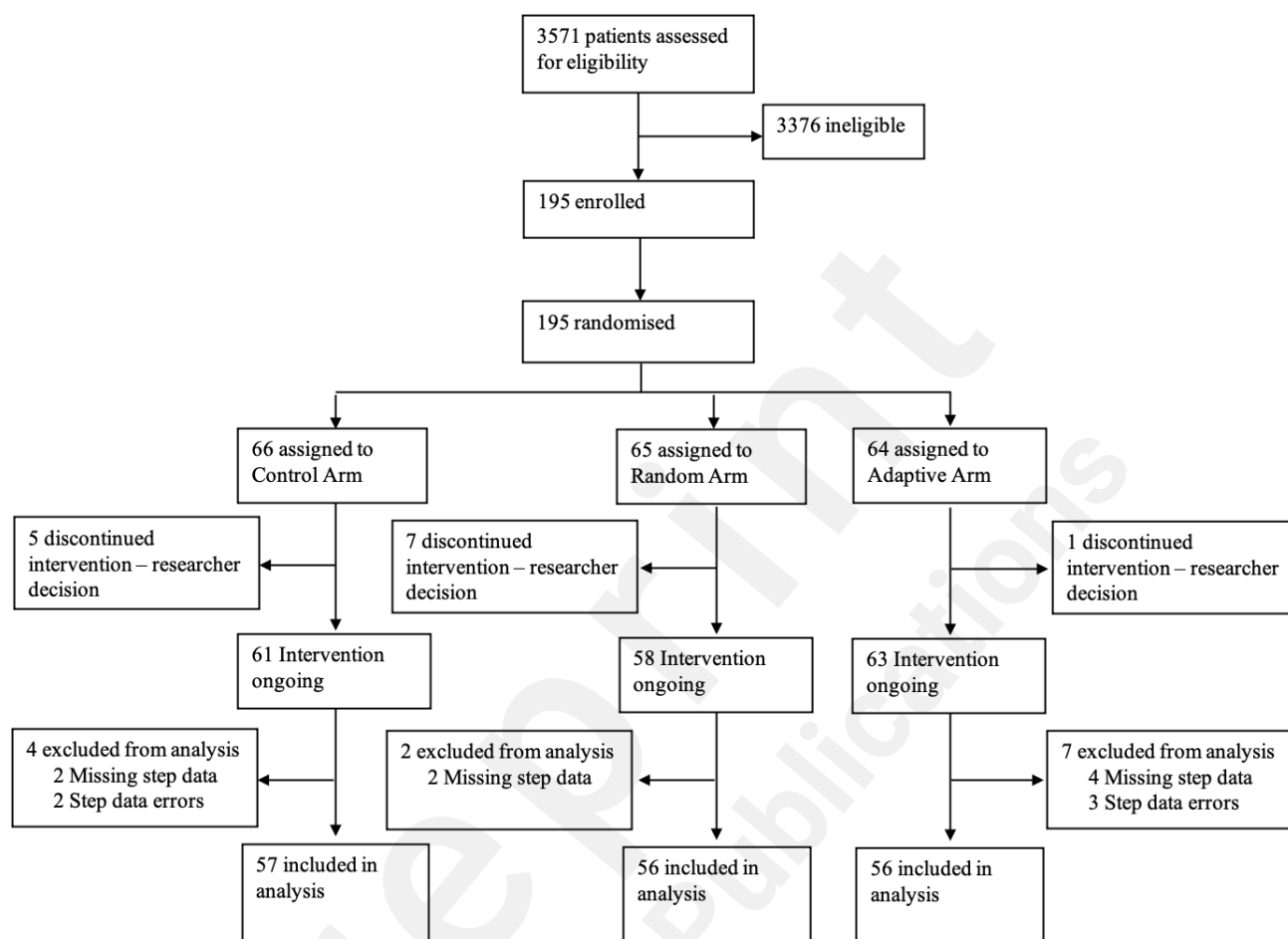


Figure 1. DIAMANTE CONSORT Diagram

Results

The research team recruited a total of 195 participants into the RCT between January 2020-June 2022. The CONSORT diagram is shown in Figure 1. A total of 13 participants were dropped by the researchers due not installing or immediately deleting the DIAMANTE App, and another 14 were excluded because of insufficient (n=9) or incorrect (n=5) step count data explained above. The final analytical sample included 168 participants, each of them are observed 168 days. Over one-third of participants were recruited from the San Francisco Health Network (n = 63, 37.5%) and the remaining recruited online (n = 105, 62.5%). Baseline demographic characteristics stratified by the

intervention arm are reported in Table 2. Overall, the sample participants were middle-aged (age mean = 49 years), predominantly female (62%), and majority English speakers (76%). The sample was diverse in terms of race/ethnicity, education, and employment status. Finally, the average step count within the sample was 3221 steps at baseline and 3783 at follow-up.

Table 2. Baseline Demographics

	Control n = 57	Random n = 56	Adaptive n = 55	Total N = 168
Age in years, Mean (SD)	51 (13.2)	49 (11.0)	48 (12.1)	49 (12.1)
Gender, n (%)				
Female	35 (61%)	38 (68%)	31 (56%)	104 (62%)
Male	20 (35%)	18 (32%)	23 (42%)	61 (36%)
Language, n (%)				
English	42 (74%)	44 (79%)	42 (76%)	128 (76%)
Spanish	15 (26%)	12 (21%)	13 (24%)	40 (24%)
Race/Ethnicity, n (%)				
Asian or Pacific Islander	4 (7%)	4 (7%)	4 (7%)	12 (7%)
Black or African American	11 (19%)	9 (16%)	7 (13%)	27 (16%)
White or Caucasian	14 (25%)	17 (30%)	19 (34%)	50 (30%)
Latinx or Hispanic	22 (39%)	20 (36%)	22 (40%)	64 (38%)
Multi-Racial/Ethnic	6 (11%)	6 (11%)	3 (6%)	15 (9%)
Education, n (%)				
Some high school or less	11 (19%)	10 (18%)	8 (14%)	29 (17%)
High school graduate	7 (12%)	8 (14%)	9 (16%)	24 (14%)
Some college or technical school	9 (16%)	17 (30%)	18 (33%)	44 (26%)
College graduate	24 (42%)	16 (29%)	15 (27%)	55 (33%)
Graduate degree	6 (11%)	5 (9%)	5 (9%)	16 (10%)
Employment, n (%)				
Full time (> 35 hours/week)	19 (33%)	19 (34%)	22 (40%)	60 (36%)
Part-time (< 35 hours/week)	9 (16%)	15 (27%)	5 (9%)	29 (17%)
Homemaker	2 (4%)	3 (5%)	6 (11%)	11 (6%)
Unemployed	8 (14%)	9 (16%)	10 (18%)	27 (16%)
Disabled /on disability	11 (19%)	6 (11%)	10 (18%)	27 (16%)
Retired	7 (12%)	4 (7%)	1 (2%)	12 (7%)

The results of the primary linear effect mixed model are presented in Table 3. The results indicate that on average at the start of the intervention participants in the Adaptive group had 3197 steps, the Control group had 3480, and Random had 3698. There were no statistically significant differences in baseline step count between intervention groups. For the adaptive group there was a significant daily step increase of 3.6 steps ($P<.001$); estimating at 168 days of the intervention, participants in the adaptive group would have gained an average of 606 steps indicating a 19%

increase in daily step count. The Control ($P=.001$) and Random group ($P<.001$) had statistically significantly lower rates of change (albeit positive) in daily step count relative to Adaptive group. The Control group increased linearly by .81 steps per day for a total of 136 steps or an increase of 3.9% daily steps from day 1 to day 168 of the intervention. Finally, the Random group increased linearly by .35 steps per day for a total of 59 steps per day indicating an increase of 1.6% daily steps from day 1 to day 168 of the intervention (Figure 2).

Table 3 – Linear Mixed Model Daily Step Count

Parameter	Estimate	Lower 95% CI	Upper 95% CI	P-value
Intercept (Adaptive)	3196.79	2568.03	3825.54	<.001
Control	283.32	-598.67	1165.32	0.531
Random	500.92	-384.54	1386.36	0.271
Day-1	3.61	2.45	4.78	<.001
Control \times (Day-1)	-2.80	-4.47	-1.13	.001
Random \times (Day-1)	-3.26	-4.92	-1.60	<.001

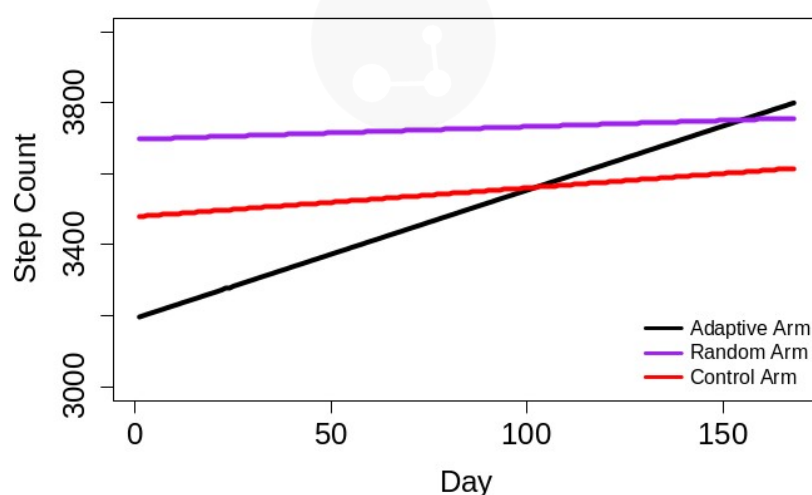


Figure 2. Linear Trajectory of Step Count over 168 days of study

Based on our a priori sub-analysis, the LMM results indicated statistically significant positive rates of change in daily step count for Adaptive group participants in both the SFHN sub-sample (average baseline steps 3278, daily rate of change 6.0 steps ($P<.001$), 30.6% increase in daily step count post intervention) and the online sub-sample (average baseline steps 3156, daily rate of change 2.3 steps ($P=.002$), 12.2% increase in daily step count post intervention). These results were in line with the overall analysis. However, in the SFHN sample the daily steps rate of change for the Random and Control groups were statistically significantly different and negative relative to the Adaptive group. Whereas for the online sample the daily rate of change for both the Random and Control groups were positive but not statistically significantly different than that of the adaptive group.

Discussion

Our study found that applying a reinforcement learning algorithm for personalizing text messaging interventions is an effective approach for increasing physical activity in a diverse sample of people with diabetes and depression. We found significant step count increases over 6 months (168 days) in the adaptive, reinforcement learning intervention arm, as compared to participants that received randomly selected text message content (random messaging arm) or only a weekly mood self-monitoring message (control arm). We found that reinforcement learning algorithms can increase the effectiveness of a physical activity based digital health intervention in a diverse, real-world sample with comorbid diabetes and depression symptoms. Our findings are particularly significant because most digital health studies in vulnerable populations to date have been pilot studies, and the use of machine learning methods for personalization is rarely applied to low-income and Spanish speaking populations. These findings also support previous studies using a similar reinforcement learning algorithms.⁽¹⁰⁾

Personalization (particularly those based on content that might be more meaningful for participants) has been one of the core challenges for the digital health field. Until now, many digital health studies show high drop-out and low-engagement of digital health interventions, also linked to a lack of personalization⁽¹⁸⁾ Here we show that reinforcement learning holds promise for the future of digital health, especially given the rapid recent advancements in artificial intelligence. Our study importantly personalized the content using existing health behavior theory (COM-B), such as motivational content as opportunity cues for behavior change, which are likely important components for future machine learning interventions⁽¹⁹⁾ In this study, it appears that participants in the adaptive messaging arm, received messages that were more relevant in content as evidenced by the varying rates of message selection in the Adaptive arm versus the Random arm of the intervention. Furthermore, the real-world implementation of this study demonstrates the importance of establishing both efficacy and effectiveness of machine learning interventions.⁽²⁰⁾

We did find some differences in our primary step count outcome by recruitment source, with the clinic-based sample receiving more benefit from the intervention compared to the online recruitment sample (30% overall step count gain vs. 12% overall step count gain among Adaptive group participants). The significant clinic-based sample findings are particularly of note, given that this population is primarily insured via Medicaid and receiving primary care within a public healthcare delivery system, which is reflected in the much higher proportion of Spanish speakers and individuals with lower educational attainment who typically do not have access to these types of interventions. There are various possible explanations for these differences that should be studied in future work including reaching a population that has less access to these types of interventions, better integration into care, or benefiting from co-design with similar populations.

A key strength of this study is the inclusion of a diverse sample in terms of race/ethnicity and language, education, and employment status. As machine learning algorithms continue to be developed and utilized in health interventions, it is crucial that the basis for these algorithmic

approaches include diverse and real-world samples so that algorithmic biases are minimized and their findings and potential benefits can be broadly shared. The results of this RCT provide valuable insights into the potential of adaptive messaging interventions to increase physical activity, especially in groups that are typically excluded from digital health research. If this intervention is widely disseminated and implemented, it can decrease health inequities by providing a type of personalized care for communities historically and presently marginalized.

Our study also highlights the need for further research to understand how digital health interventions can be tailored to different populations to maximize their effectiveness. Our future research will further investigate the parameters that influenced algorithmic decision making, regarding the patterns of sending feedback messages, motivational messages, and timing of message delivery. It is also important to understand if overall categories of messages were more impactful on behavior change and to understand how the algorithm can be refined to be more effective in achieving the goal of increasing physical activity in diverse samples.

Limitations of this study include challenges with data accuracy from phones. We chose to emphasize replicability in the real world and thus did not provide users with wearable devices or standardize the mobile phones that users had to have. This resulted in variability in data collection, and some data collection errors noted above. Furthermore, in order for these types of interventions to scale, we must maximize adoption and relevance (e.g., utilization of devices that users already own) alongside precision (e.g., “better” measurement using new devices).

Another limitation is that we did not specifically assess engagement with the messages. Similar to other SMS interventions, we do not know whether or when people actually read messages, we only know when messages were sent. Finally, due to the unforeseen circumstances from the COVID-19 pandemic, recruitment number for our project was also less than our goal stated in the RCT protocol; however, our effect size was still large enough to be detected in the primary analysis.

The DIAMANTE intervention has shown that reinforcement learning algorithms can be

employed to improve the personalization of physical activity interventions in a diverse, multi-lingual sample with diabetes and depression symptoms. Moving forward, AI and machine learning interventions will rapidly expand, and real-world studies targeting diverse end users are critical, especially when built upon co-design with end users.

Acknowledgments

This study received funding from the Agency for Healthcare Research and Quality (AHRQ, R01HS025429).

Conflicts of Interest. The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Conflict of Interests disclosures: Dr. Aguilera is the owner of HealthySMS and licenses the software to other researchers. Dr. Courtney Lyles has conducted separate research projects with digital health companies and was a visiting researcher at Goglee from 2022-2023 on th health equity team. The rest of the co-authors do not have any conflicts of interest to disclose.

References

1. Bădescu SV, Tătaru C, Kobylinska L, Georgescu EL, Zăhău DM, Zăgrean AM, et al. The association between Diabetes mellitus and Depression. *J Med Life*. 2016;9(2):120–5.
2. Cho YG. The Interrelationship between Diabetes and Depression. *Korean J Fam Med*. 2014;35(3):109.
3. Everson SA, Maty SC, Lynch JW, Kaplan GA. Epidemiologic evidence for the relation between socioeconomic status and depression, obesity, and diabetes. *J Psychosom Res*. 2002 Oct;53(4):891–5.
4. Hansen MC, Cabassa LJ. Pathways to Depression Care: Help-Seeking Experiences of Low-Income Latinos with Diabetes and Depression. *J Immigr Minor Health*. 2012 Dec;14(6):1097–106.
5. Narita Z, Inagawa T, Stickley A, Sugawara N. Physical activity for diabetes-related depression: A systematic review and meta-analysis. *J Psychiatr Res*. 2019 Jun;113:100–7.
6. Lambert JD, Greaves CJ, Farrand P, Price L, Haase AM, Taylor AH. Web-Based Intervention Using Behavioral Activation and Physical Activity for Adults With Depression (The eMotion Study): Pilot Randomized Controlled Trial. *J Med Internet Res*. 2018 Jul 16;20(7):e10112.
7. Dang K, Ritvo P, Katz J, Gratzner D, Knyahnytska Y, Ortiz A, et al. The Role of Daily Steps in the Treatment of Major Depressive Disorder: Secondary Analysis of a Randomized Controlled Trial of a 6-Month Internet-Based, Mindfulness-Based Cognitive Behavioral Therapy Intervention for Youth. *Interact J Med Res*. 2023 Dec 8;12:e46419.
8. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. *BMC Med Inform Decis Mak*. 2018 Dec;18(1):12.
9. De Santis KK, Jahnel T, Matthias K, Mergenthal L, Al Khayyal H, Zeeb H. Evaluation of Digital Interventions for Physical Activity Promotion: Scoping Review. *JMIR Public Health Surveill*. 2022 May 23;8(5):e37820.
10. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System. *J Med Internet Res*. 2017 Oct 10;19(10):e338.
11. Hochberg I, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Yom-Tov E. Encouraging Physical Activity in Patients With Diabetes Through Automatic Personalized Feedback via Reinforcement Learning Improves Glycemic Control. *Diabetes Care*. 2016 Apr 1;39(4):e59–60.
12. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel Determinants of Digital Health Equity: A Literature Synthesis to Advance the Field. *Annu Rev Public Health*. 2023 Apr 3;44(1):383–405.
13. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Ann Intern Med*. 2013 Feb

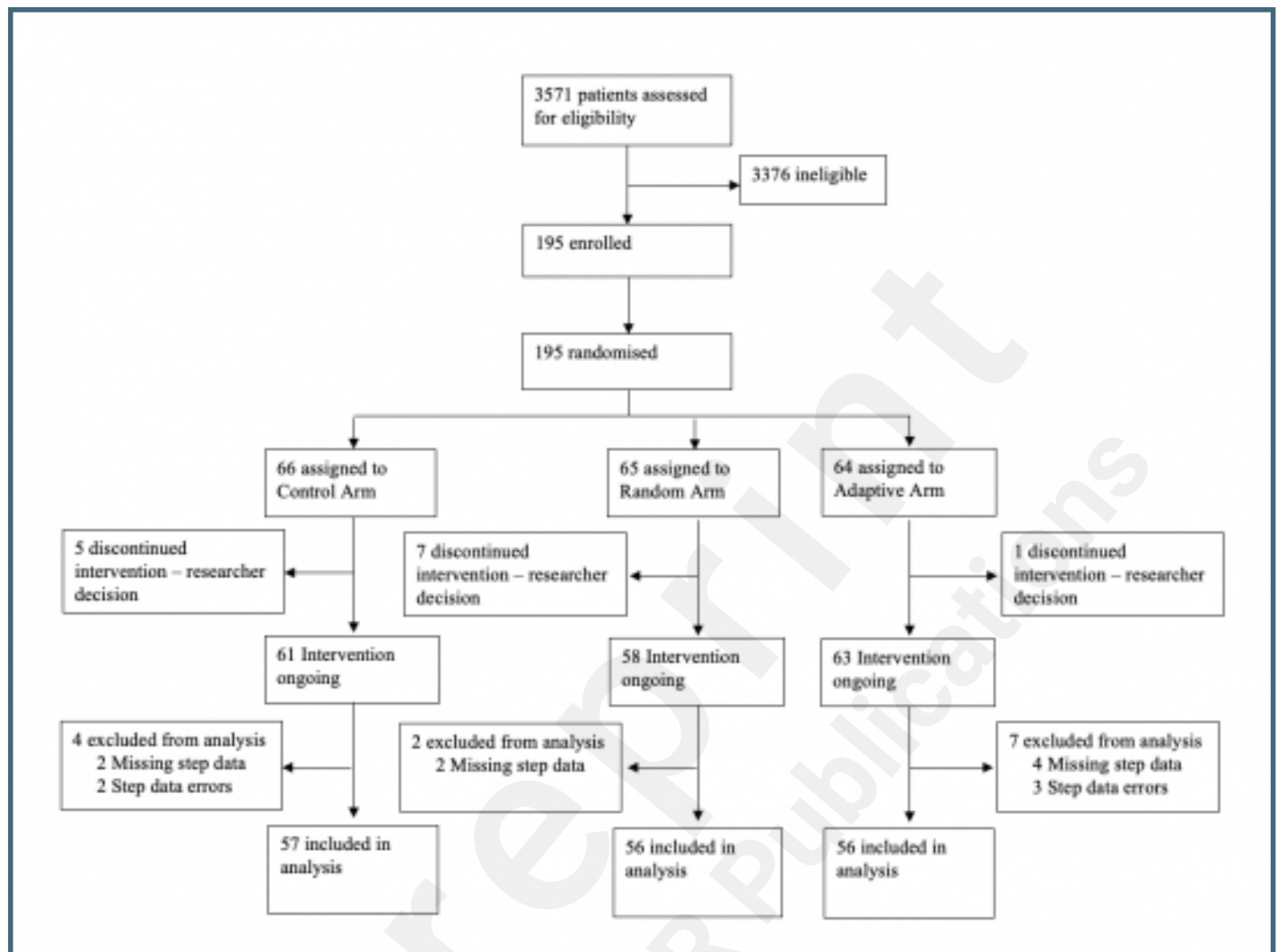
5;158(3):200.

14. Avila-Garcia P, Hernandez-Ramos R, Nouri SS, Cembali A, Sarkar U, Lyles CR, et al. Engaging users in the design of an mHealth, text message-based intervention to increase physical activity at a safety-net health care system. *JAMIA Open*. 2019 Dec 1;2(4):489–97.
15. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cembali A, Gomez-Pathak L, 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 Aug;10(8):e034723.
16. Tewari A, Murphy SA. From Ads to Interventions: Contextual Bandits in Mobile Health. In: Rehag JM, Murphy SA, Kumar S, editors. *Mobile Health [Internet]*. Cham: Springer International Publishing; 2017 [cited 2024 Jan 12]. p. 495–517. Available from: http://link.springer.com/10.1007/978-3-319-51394-2_25
17. Agrawal S, Goyal N. Analysis of Thompson Sampling for the multi-armed bandit problem. 2011 [cited 2024 Jan 12]; Available from: <https://arxiv.org/abs/1111.1797>
18. Gan DZQ, McGillivray L, Larsen ME, Christensen H, Torok M. Technology-supported strategies for promoting user engagement with digital mental health interventions: A systematic review. *Digit Health*. 2022 Jan;8:205520762210982.
19. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. *Health Psychol Rev*. 2019 Jan 2;13(1):1–17.
20. Triantafyllidis AK, Tsanas A. Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. *J Med Internet Res*. 2019 Apr 5;21(4):e12286.

Supplementary Files

Figures

DIAMANTE CONSORT Diagram.



Linear trajectory of step count over 168 days of study.

