Original Paper

A Personalized Exercise Assistant Using Reinforcement Learning: Results From a Large-Scale Mobile Health Randomized-Controlled Trial

Amy Armento Lee[1,2], MSc; Narayan Hegde[1]; Nina Deliu[1,2,3], PhD; Emily Rosenzweig[1,4], PhD; Arun Suggala[1]; Sriram Lakshminarasimhan[1], PhD; Qian He[1], PhD; John Hernandez[1], PhD; Martin Seneviratne[1], MBBS, MS; Rahul Singh[1]; Pradnesh Kalkar[1]; Karthikeyan Shanmugam[1]; Aravindan Raghuveer[1]; Abhimanyu Singh[1]; Hariharan Manoharan[1]; My Nguyen[1]; James Taylor[1], MD; Jatin Alla[1]; Sofia S Villar[1,2], PhD; Hulya Emir-Farinas[1], PhD

[1]Google (United States), Mountain View, CA, United States

[2]MRC Biostatistics Unit, Cambridge Institute of Public Health, School of Medicine, University of Cambridge, Cambridge, United Kingdom

[3]MEMOTEF Department, Sapienza University of Rome, Rome, Italy

[4]Verily Health Inc., Dallas, TX, United States

Corresponding Author:

Amy Armento Lee, MSc
Google (United States)
1600 Amphitheatre Parkway
Mountain View, CA 94043
United States
Phone: 1-650-390-3895
Email: a.armento.lee@gmail.com

Abstract

Background: Consistent physical inactivity among adults and adolescents poses a major global health challenge. Mobile health (mHealth) interventions, particularly just-in-time adaptive interventions (JITAIs), offer a promising avenue for scalable and personalized physical activity promotion. However, developing and evaluating such adaptive interventions at scale, while integrating robust behavioral science, presents methodological hurdles.

Objective: The Personalized Exercise Assistant Using Reinforcement Learning (PEARL) study aimed to assess the feasibility and effectiveness of a reinforcement learning (RL) algorithm, informed by health behavior change theory (capability, opportunity, and motivation–behavior [COM-B]), to personalize the content and timing of physical activity nudges via the Fitbit app (Google LLC) compared to fixed and random nudging strategies, and to a control group with no nudges.

Methods: We conducted a large-scale, 4-arm randomized controlled trial (RCT) enrolling 13,463 Fitbit users. Participants were randomized to: (1) control (no nudges), (2) random (random content/timing), (3) fixed (logic based on baseline COM-B survey), and (4) RL (adaptive algorithm). The primary outcome was the change in average daily step count from baseline to 2 months. Secondary outcomes included user engagement and survey responses regarding capability, opportunity, and motivation.

Results: A total of 7711 participants were included in the primary analysis (mean age 42.1, SD 9.0 years; 6657/7711, 86.3% female). At 1 month, the RL group showed a significant increase in daily steps compared to control (+296 steps; $P<.001$), random (+218 steps; $P=.005$), and fixed (+238 steps; $P=.002$) groups. For the prespecified primary analyses at 2 months, the RL group sustained a significant increase against the control (+210 steps; $P=.01$), but pairwise differences comparing the RL group against the random and fixed groups were not statistically significant. Generalized estimating equation (GEE) models confirmed a sustained significant increase in the RL group (+208 steps; $P=.002$). In exit surveys, the RL group reported higher favorable responses regarding nudge customization (224 of 605, 37%) compared to other groups.

Conclusions: This study demonstrates the feasibility and early efficacy of using RL to personalize digital health nudges at scale. To our knowledge, this represents one of the largest, nonincentivized randomized trials of this duration to evaluate RL for physical activity. While the magnitude of the observed step increase is modest, a sustained daily improvement of 200-300 steps can contribute to broader public health goals by reducing sedentary behavior in large populations. The adaptive approach outperformed static behavioral rules, showcasing the promise of dynamic personalization in a real-world mHealth setting. Notably, participants receiving nudges based on fixed, theory-informed rules did not show a significant advantage

over the random group, underscoring the challenge of using static baseline assessments to address nonstationary, day-to-day behavioral contexts. However, several limitations warrant consideration: the findings are specific to a 2-month follow-up, the study population was predominantly female (6657/7711, 86.3%) self-selected Fitbit users, and the observed step gains were modest in magnitude.

Trial Registration: ClinicalTrials.gov NCT07783724 https://clinicaltrials.gov/study/NCT07783724 (Retrospectively Registered)

J Med Internet Res 2026;28:e91156; doi: 10.2196/91156

Keywords: physical activity; mHealth; reinforcement learning; JITAI; randomized controlled trial; fitbit; behavioral science; COM-B; digital health; Capability, Opportunity, and Motivation–Behavior; just-in-time adaptive intervention

Introduction

Background

Currently, nearly one-third of adults and 80% of adolescents worldwide fail to meet the World Health Organization (WHO) physical activity recommendations [1,2]. Inactivity rates vary significantly by region and demographics, with higher prevalence among women and in lower-middle-income countries. Alarmingly, global inactivity is projected to reach 35% by 2030 [3]. The global rise in sedentary behavior has placed physical activity at the forefront of public health agendas, calling for scalable and pragmatic strategies. In particular, as emphasized by the Director of Health Promotion at WHO: "We need to find innovative ways to motivate people to be more active, considering factors like age, environment, and cultural background."

There is growing experimental evidence that consumer wearable–based interventions using devices such as Fitbits (Google LLC) are effective in increasing physical activity [4]. By leveraging mobile devices like apps and wearables, mobile health (mHealth) can deliver timely, personalized interventions. Just-in-time adaptive interventions (JITAIs) aim to provide scalable, accessible, context-specific support, adapting to individual needs, preferences, and fitness levels through continuous monitoring [5]. Through sophisticated algorithms, JITAIs can be dynamically adjusted based on user feedback, progress, and evolving behavior. Moreover, they can help identify and target specific population subgroups most vulnerable to inactivity and most likely to benefit from intervention programs.

Prior Work

To our best knowledge, this study represents the first large-scale, real-world, 4-arm randomized controlled trial (RCT) designed to assess the comparative effectiveness of reinforcement learning (RL)-driven personalization. This research builds upon a strong foundation established by previous studies that successfully demonstrated the feasibility and early potential of RL-driven personalization in digital health [6-15]. These pioneering works, such as the HeartSteps and Diabetes and Mental Health Adaptive Notification Tracking and Engagement/Evaluation (DIAMANTE) trials, provided critical proof-of-concept for utilizing RL to manage physical activity and sedentary behavior [6,7,11,12].

While these foundational studies established the methodological groundwork, they typically relied on focused sample sizes ranging from tens to hundreds of participants [6-15]. Many of these important early trials also used monetary compensation to participants, which may introduce different engagement patterns than in a nonincentivized setting [7,9-12]. Additionally, prior work often focused on specific clinical or academic subgroups, such as university staff, students, or specific patient cohorts [8-10,12].

Our study extends this robust lineage of research into a broader, naturalistic environment. By evaluating outcomes from a large sample of nonincentivized participants, this study transitions behaviorally-informed RL from controlled feasibility environments to an unprecedented real-world scale.

Study Objectives

To deliver on the full potential of JITAIs, it is paramount to balance their effectiveness with user burden. This entails effectively integrating into their design: (1) evidence-based behavioral theories, (2) experimental data from mHealth settings, and (3) adaptable, personalized intervention policies.

This work reports and discusses the results of the first large-scale mHealth study that exemplifies this unique 3-way implementation approach: the Personalized Exercise Assistant Using Reinforcement Learning (PEARL) study. First, behavioral change theories are leveraged using the capability, opportunity, and motivation–behavior (COM-B) behavior change model [16] to categorize barriers to physical activity, and the theoretical domains framework [17] to develop nudges that map onto these barriers. These models have been successfully applied across diverse health contexts, from medication adherence to physical activity interventions [18,19]. The selection of COM-B as our theoretical framework was based on systematic evaluation of available behavior change models, including social cognitive theory, the theory of planned behavior, and the transtheoretical model. COM-B was selected due to its robust evidence base in physical activity interventions, clear integration with intervention design tools, and suitability for adaptive digital health applications. A detailed justification for this selection is provided in Multimedia Appendix 1. Second, the study design embeds a microrandomized trial (MRT) [20], which enables researchers to examine the effectiveness of intervention components in real-time and detect any time-varying effects. Notably, this innovative approach may also help uncover patterns of engagement and potential adverse

effects such as habituation and retention. Finally, adaptable personalized interventions are integrated by using machine learning algorithms, specifically, RL [21]. In previous work, these algorithms have shown promising results across different health domains, including physical activity.

The PEARL study was a randomized trial designed to assess the impact of different nudging strategies (including personalized), delivered as pop-up notifications via the Fitbit app, on daily step count. The primary outcome measured was the change in average daily step count, comparing the final month of the study (month 2) with the prestudy baseline month.

Methods

Study Design

Our study had a 60-day randomized trial, with participants randomized into 1 of 4 arms: a control (no nudge), and 3 intervention arms. The intervention arms all drew from the same predefined library of nudges, whose content was informed by behavioral science, and which were mapped to a set of theoretically derived barriers to physical activity. What differed between the arms was the method for selecting the message content and delivery timing for each participant on each day. One condition selected nudges and times of day at random. The second used a fixed logic derived from participants' answers to survey questions designed to measure their barriers to physical activity. The third used an adaptive RL strategy.

The protocol was designed, written, and executed by the investigators (HEF, NH, ER, AAL, and JT), and was registered retrospectively with ClinicalTrials.gov (NCT07783724) [22]. This RCT was conducted and reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online) guidelines (Checklist 1).

Participants

From February to June 2024, adult Fitbit users (ages 22-60) in the United States, who self-reported an average daily step count of less than 8000 steps in the month prior to recruitment, were enrolled in the 60-day study. Among these, only participants that were verified to have met the inclusion criteria were entered into the final analysis (age 22-60 years, live in the United States, average daily step count <8000 steps in the month prior to recruitment, own a Fitbit device and have a smartphone with the Fitbit app, and wear a Fitbit device regularly). Participants consented to sharing their step count data for the 30-day period prior to study enrollment, which we refer to as the "Prestudy Baseline Month."

The Fitbit app (Google LLC) is publicly available on both the Apple App Store and Google Play Store. A random sample of Fitbit app users were recruited through an in-app card displayed on the Fitbit app's "Steps Dashboard" (a page where users can view prior step-count data as collected by a connected device).

Participants who clicked on the recruitment card were screened for study eligibility, and, if eligible, provided with study welcome materials and informed consent. Upon study onboarding, participants completed a demographics survey, a COM-B survey, and a nudge time preference survey.

On enrollment in the study, participants were still able to modify app settings to turn the notification policy off.

Data Collection Procedures

All study data were collected through the Fitbit app, including: app, wearable device, survey, and app interactions. The study database was locked in August 2024, after final data collection was completed in 2024.

Interventions

The Fitbit app comprises a collection of self-monitoring and self-service offerings for participants. This includes a standard set of notifications related to connected wearable devices (charging, syncing, and reminders to wear). This also includes a standard set of notifications related to physical activity, such as automated reminders to be active and celebrations of physical activity.

The PEARL study introduced a nudge intervention, in addition to the standard *Fitbit* app offerings, which is intended to be a novel and complementary intervention for encouraging increased physical activity.

Nudge Intervention

The theoretical foundation for all intervention content was the COM-B model, selected for its comprehensive approach to behavior change and strong evidence base in physical activity contexts (see Multimedia Appendix 1 for full justification). All nudge themes were mapped to COM-B subconstructs, and assessment instruments were designed to measure capability, opportunity, and motivation as operationalized by COM-B and its associated theoretical domains framework.

The nudge intervention within the PEARL study comprised a combination of a nudge theme and a delivery time, resulting in a total of 12 distinct action classes in each intervention group. The 6 nudge themes are based on subconstructs of the COM-B constructs of capability, opportunity, and motivation [18]. Capability was divided into ability and planning, opportunity was divided into physical and social, and motivation was divided into perceived benefit and prioritization. Delivery times were set for morning (6 AM) or afternoon (3 PM).

Upon selection of a nudge theme, a specific message was randomly sampled from the corresponding theme's repository for display to the user, to ensure diversity and minimize habituation patterns. For each theme, approximately 20 messages were initially developed by behavioral science experts, supplemented by an additional 5-10 messages generated using large language models and subsequently validated by behavioral experts. The full message repository included approximately 155 total nudges across all themes (see Multimedia Appendix 2 for the nudge repository).

Participants in the intervention arms (excluding the control group) received one text-only in-app push notification daily, drawn from this message repository. Nudges were designed to be dismissible and were presented once per day. However, all delivered nudges remained accessible on the study home page within the Fitbit app throughout the day. Furthermore, participants were provided with the option to offer binary feedback (thumbs up/down) and submit free-text comments on each nudge.

Random Arm: Nudge Allocation Policy

The random arm message selection logic was simple. Each day, and for each user, a COM-B theme and delivery time were randomly selected. This provided an approximate 1/6 distribution of nudge themes per user, and a 1/2 distribution of delivery times per user. The combined action class distribution (theme×delivery time) was approximately 1/12 per user.

Fixed Arm: Nudge Allocation Policy

The fixed arm message selection logic was designed to personalize nudge content and timing based on initial onboarding participant surveys. The COM-B survey was designed to measure participants' baseline levels of capability, opportunity, and motivation, and their respective subconstructs [18]. All of the nudges were classified as relating to/acting on one of these 6 themes. In the fixed condition, participants who scored lower on a given theme received more nudges associated with addressing that type of barrier to activity.

For nudge content selection, mean scores for the 6 COM-B subthemes were computed from the user's Likert scale ratings (1-5), with higher scores indicating that these themes represented *less* of a barrier to physical activity. A "barrier score" for each theme was then calculated as 5 minus the Likert score. For each day of the 60-day study, a nudge theme was sampled from a multinomial distribution weighted by these barrier scores, meaning themes associated with higher perceived barriers were more likely to be selected. For nudge time selection, participants' stated preferences (morning, afternoon, or no preference) determined the probability of a morning or afternoon nudge (eg, 70% for preferred time, 30% for the other, or 50% for no preference). A time was then sampled daily from a binomial distribution based on these probabilities. This static policy, derived solely from initial survey responses, remained fixed for each participant throughout the entire study duration, with no further adaptation based on real-time user behavior or feedback.

RL: Nudge Allocation Policy

The adaptive RL arm was designed according to a contextual multiarmed bandit (C-MAB) framework [23]. This specific subclass of RL well suits mHealth settings [24] and has been used in previous studies [7,12]. C-MAB strategies are ideal candidates for capturing rapid changes in an individual user's context and needs, with the scope of efficiently delivering JITAIs in dynamic environments. Furthermore, as opposed to other RL classes, C-MABs enjoy a reduced computational burden from carrying through only the most recent information. We refer to Deliu et al (2024) [23] for the technical details.

In a pilot phase, various C-MAB algorithms have been explored, suggesting e-greedy as the optimal compromise between ease of interpretation, computational feasibility, and a simple yet efficient balance between exploitation and exploration [23]. E-greedy trades-off exploitation and exploration is as follows: with probability $1-\varepsilon$ the policy's recommendation (ie, the action that maximizes the expected reward) is followed, and with probability ε an action is chosen uniformly at random from the set of 12 possible actions. Clearly, a ε lower leads to higher exploration. In this work, we set ε as one of 0.7 or 0.8 throughout the study; the higher exploration value is chosen when the model deployed has high variability in decision across ensemble models.

There are 3 main components that characterize the RL framework: the state, action, and reward variables. In the context of PEARL, the state encapsulates user-specific information, while the action corresponds to the specific combination of the intervention component and delivery time of a nudge. Finally, the reward variable is the proximal outcome defined upon the step counts and serves as a proxy for the user's physical activity level. Below, we provide a formalization of these 3 main ingredients, contextualizing them within the PEARL study, followed by a description of the RL deployment pipeline.

State Variables

These characterize the user's state on each day t and define the feature set used for *personalization* purposes. It comprises static information (demographics, surveys, and baseline steps) as well as the individual time-varying history (past walking trends, app activity, and past nudge feedback, etc). We use $X_{i,t}$ to denote the features of user i on their day t of study. The full set of information is given in Table 1.

Table 1. Feature set (state) for PEARL[a] reinforcement learning algorithm.

User feature class and state class	Cadence
User demographics	
Age, gender, and device type	Static: start of study
Education, area type, and weather	Start of study
User COM-B[b] survey	
20 survey questions with 1-5 Likert scale	Start of study

User feature class and state class	Cadence
User prestudy steps	
Prestudy steps mean and SD	Start of study
Prestudy walk pattern: low vs high walker, regular vs irregular walker across days of week	Start of study
User dynamic features	
Recent steps mean and SD	Last 7 days
Slope of regression line fitting recent steps	Last 7 days
Daily nudge feedback stats	Last 7 days
Day of the week	Each day
Recent study walk pattern: low vs high walker, regular vs irregular walker across days of week	Last 7 days
Average steps in morning (<12 PM), evening (>12 PM)	Last 7 days
Percentage of missing data in steps	Last 7 days
Past week nudge history	Each day

[a]PEARL: Personalized Exercise Assistant using Reinforcement Learning.
[b]COM-B: capability, opportunity, and motivation-behavior

Action or intervention space

This represents the intervention set and is defined upon 6 nudge themes (Ability, Perceived Benefit, Physical Opportunity, Planning, Prioritization, and Social Opportunity) and 2 delivery times (morning and afternoon), for overall 12 actions. Let $a_{i,t}$ denote the specific action (among the 12 unique combinations) chosen by the RL policy user i on their day t of study. For more details, we refer to the subsection Interventions. As opposed to the other study groups, the RL arm is designed to estimate and follow an action-selection policy that optimizes the cumulative reward over time across all users. Such a policy is described in the Policy Learning section.

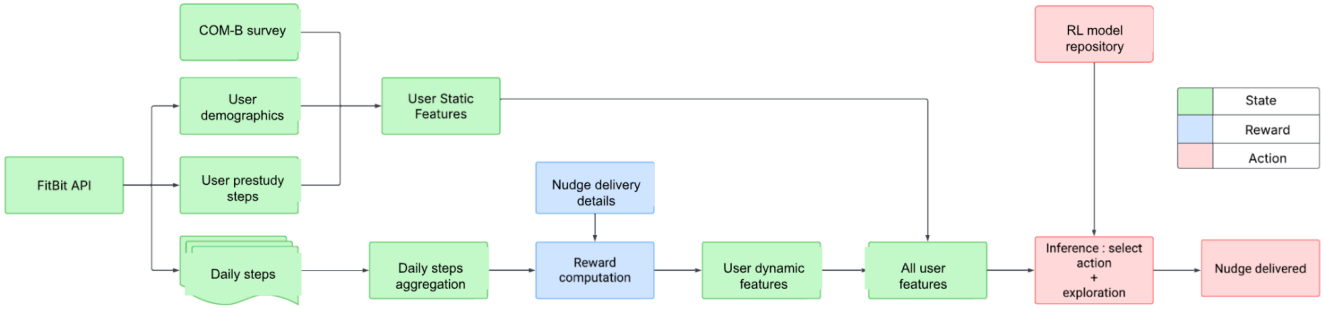
Reward variable

This represents the *proximal outcome* used to guide the *optimization* of the intervention delivery. It is defined as the *relative change* in the number of steps walked between nudge delivery and the following 24 hours, relative to an individual prestudy baseline walking pattern. Formally, denoted by $\mathbb{I}(\text{morning}_{i,t})$ and $\mathbb{I}(\text{weekday}_t)$ the binary variables indicating whether the nudge shown to user i on their day t of study is a morning or an evening nudge, and whether the t day is a weekday or a weekend, respectively, the baseline walking pattern, say $y_{i,0}(\mathbb{I}(\text{morning}_{i,t}), \mathbb{I}(\text{weekday}_t))$. This baseline represents the average daily step count within a 24-hour period starting from morning for user i and is calculated over a 30-day window prior to the start of the study for each combination $\mathbb{I}(\text{morning}_{i,t})$ of and $\mathbb{I}(\text{weekday}_t)$. Now, denoted by $y_{i,t}$ the steps taken by user i on day t of the study within a 24-hour window starting from the time of the nudge on day t, the reward is defined as follows:

$$r_{i,t}(a_{i,t}) = \frac{y_{i,t} - y_{i,0}(\mathbb{I}(\text{morning}_{i,t}), \mathbb{I}(\text{weekday}_t))}{y_{i,0}(\mathbb{I}(\text{morning}_{i,t}), \mathbb{I}(\text{weekday}_t))}$$

According to these 3 key components, the RL decision-making process is summarized in Figure 1. The diagram depicts the RL deployment pipeline using the data from the RL arm to estimate the best action (nudge theme and time combination) for user i on day t of the study. This pipeline is run daily at midnight, with the model decision for each user defined by 5 AM While Figure 1 illustrates the daily automated inference pipeline, the underlying model parameters W^*) are recomputed weekly. This periodic retraining involves pooling all cumulative nudge intervention data (including observed states, delivered nudges, and corresponding rewards) to update the policy, which is then validated against held-out test data before deployment.

Figure 1. Personalized Exercise Assistant Using Reinforcement Learning (PEARL) reinforcement learning deployment pipeline. COM-B: capability, opportunity, and motivation–behavior; RL: reinforcement learning.



Cron Name	Data Extraction	Data Aggregation	Reward Compute	Dynamic features	User Data merge	Model inference	Nudge delivery
Start Time	12 AM	2 am	3 am	4 am	5 am	5 am	6 am / 3 pm
Run Duration	90 mins	30 mins	30 mins	30 mins	5 mins	10 mins	6 am / 3 pm

RL decision making pipeline

Policy Learning

A decision-making or action-selection policy maps individual user states or contexts to an action. In RL arm, baseline and time-varying users' data are used to learn nudge policies and guide the action selection so as to optimize the expected reward. That is, given $X_{i,t}$, the state of a user i on day t, we learn a mapping ϕ from the state space to the action space. The mapping is often characterized by a set of unknown parameters W, ie, $\phi = \phi(\cdot; W)$. Denoting by $\phi(a; \mathbf{x}_{i,t}, W)$ the probability of playing action a for a user i with context $X_{i,t}$, learning an optimal policy in this case is equivalent to estimating the unknown parameters such that:

$$W^* = \arg\max_W \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_{a \sim \phi(\cdot; \mathbf{x}_{i,t}, W)}[r_{i,t}(a)]$$

In order to solve the optimization program, the expectation of $r_{i,t}(a_{i,t})$ was estimated using the history of previous data (states observed, nudges delivered, and corresponding rewards) and importance sampling, a classical unbiased estimator of the value of a policy. By weighting the observed rewards by the ratio of the probability of the nudge being sent under the new policy relative to its probability in the historical data, importance sampling provides a robust estimate of the expected rewards for future iterations. This prevents the optimization from becoming biased toward previously frequent actions and allows the agent to accurately evaluate the potential effectiveness of new nudge-allocation strategies.

The optimization problem was tackled with an empirical risk minimization oracle which is readily available for various model classes such as decision trees and neural networks. After estimating the reward model for each of the 12 actions, an optimal action for user i on day t is chosen as $\hat{a}_{i,t} = \arg\max_a \mathbb{E}[\hat{r}_{i,t}(a)]$. Extreme gradient boosting (XGBoost) decision tree is used to implement reward prediction models.

Outcomes

The primary endpoint was to assess the difference in average daily step count change from the prestudy baseline month and the final study month (month 2). A total of 4 pairwise comparisons were prespecified for the primary analysis: random vs control, RL vs random, fixed vs random, and RL vs fixed.

Secondary outcomes of interest include repeating the analysis for study month 1 (compared to prestudy baseline month), and other pairwise comparisons.

Sample Size

The trial was designed to be powered to detect a treatment effect of a 200-step increase in daily step count, which based on prior studies was estimated to require 1392 participants per treatment group. Additional details are available in the study protocol [22].

Randomization

For study arm allocation, simple randomization (unstratified and no blocking) was used. The randomization sequence was generated using a computer-based algorithm and programmed into the study platform to ensure automated, unbiased allocation to the 4 study arms.

Participants were blinded to their specific treatment allocation, although the absence of messages in the control group created an inherent limitation to complete blinding. Data analysts (HEF, NH, ER, AAL, and JT) remained blinded to group assignments during initial data processing to minimize potential observer bias.

Statistical Analysis

Participants were only included in the analysis if they met the following minimum requirements for data collection: for baseline and each study month, the participant must have at least 7 days with at least 500 steps per day [25,26]. We use this threshold as a proxy for wear-time of the Fitbit wearable device, and expect that if a participant wore the

device for a significant portion of the day there would be at least 500 steps. No imputation was applied for missing data for participants that were included in the analysis; however, Multimedia Appendix 3 includes a sensitivity analysis with multiple imputation for the primary outcome. Because of this, the primary analysis operates as a modified intention-to-treat (mITT) or per-protocol framework, rather than a true intention-to-treat (ITT) framework.

The overall demographics of the full sample and by each treatment group were reported. Overall step counts within the sample were also presented descriptively. Balance of baseline characteristics was examined across the 4 study groups. Difference-in-differences (DiD) regression analysis was conducted to determine the impact of the treatment arms on the change in step count for each participant [27]. Benjamini-Hochberg adjustment was applied to control for type I error across these prespecified primary comparisons. Nominal *P* values are reported for the secondary outcomes.

A generalized estimating equation (GEE) model with an exchangeable correlation structure was used to estimate the average effect of each treatment group on the daily step count over time, accounting for the longitudinal nature of the data and within-subject correlation [28,29]. To ensure the validity of our inference, we used robust sandwich variance estimators to calculate the SEs, CIs, and *P* values. These estimators provide consistent SE estimates even under potential misspecification of the correlation structure.

To determine the relative contribution of each contextual feature across the study duration, we computed a time-weighted cumulative feature importance across all retrained policy iterations. For each trained multiclass XGBoost model iteration, the raw feature importance was extracted using the mean split gain metric.

Data analyses were performed using R (version 4.0.1; R Foundation for Statistical Computing) [30] and Python (version 3.14; Python Software Foundation) [31].

Ethical Considerations

Ethical approval for this study (Pro00070641) was granted by Advarra Institutional Review Board (IRB) on April 7, 2023. This study was conducted in accordance with the protocol and IRB requirements. The protocol, informed consent form, and other relevant study documents were submitted and approved by the IRB before the study was initiated. Any amendments to the protocol required IRB approval before implementation of changes made to the study design, except for changes necessary to eliminate an immediate hazard to study participants. Informed consent was obtained from all participants prior to their enrollment in the study. To ensure privacy and confidentiality, all collected data were strictly deidentified before any analysis was conducted. Additionally, participants did not receive financial compensation for their involvement in this research.

Results

Participants

Between February and June 2024, 15,000 Fitbit users were recruited through the Fitbit app's Steps dashboard. Following eligibility screening, 13,463 users were enrolled and randomly assigned to 1 of 4 groups for a 60-day follow-up: a control group (3304/13,463, 24.5%), a random nudge group (3337/13,463, 24.8%), a fixed nudge group receiving nudges selected based on an entrance survey (3419/13,463, 25.4%), and a personalized reinforcement learning group (3403/13,463, 25.3%; Figure 2).

Among 13,463 participants enrolled, 1522 (11.3%) formally withdrew from the study (Figure 1), with the highest withdrawal rate observed in the control group (n=447, 13.5%). Furthermore, a total of 1511 participants were excluded from the primary analysis for failing to meet the minimum data compliance threshold of at least 7 days with 500 steps per month, reflecting patterns of irregular device wear or app engagement common in large-scale digital health trials. Because missing data was determined not to be missing completely at random (MCAR; $\chi^2_{11}=56.9$; $P<.001$), a sensitivity analysis is included in Multimedia Appendix 3.

Of the 13,463 enrolled participants, 7711 (68.0%) participants met the criteria required for analysis of step-count, including meeting study eligibility criteria of less than 8000 steps per day at baseline and availability of step-count data through the 60-day study period (Figure 2). Full eligibility criteria are available in the study protocol [22].

Baseline demographic characteristics are reported in Table 2, with no notable differences across the arms. Eligible participants had a mean age of 42.1 (SD 9.0) years, with 6657 of 7711 women (86.3%). Mean body weight was 90.8 (SD 25.5) kg, and baseline average daily steps was 5618.2 (SD 1502.5).

Figure 2. CONSORT diagram. CONSORT: Consolidated Standards of Reporting Trials; RL: reinforcement learning.

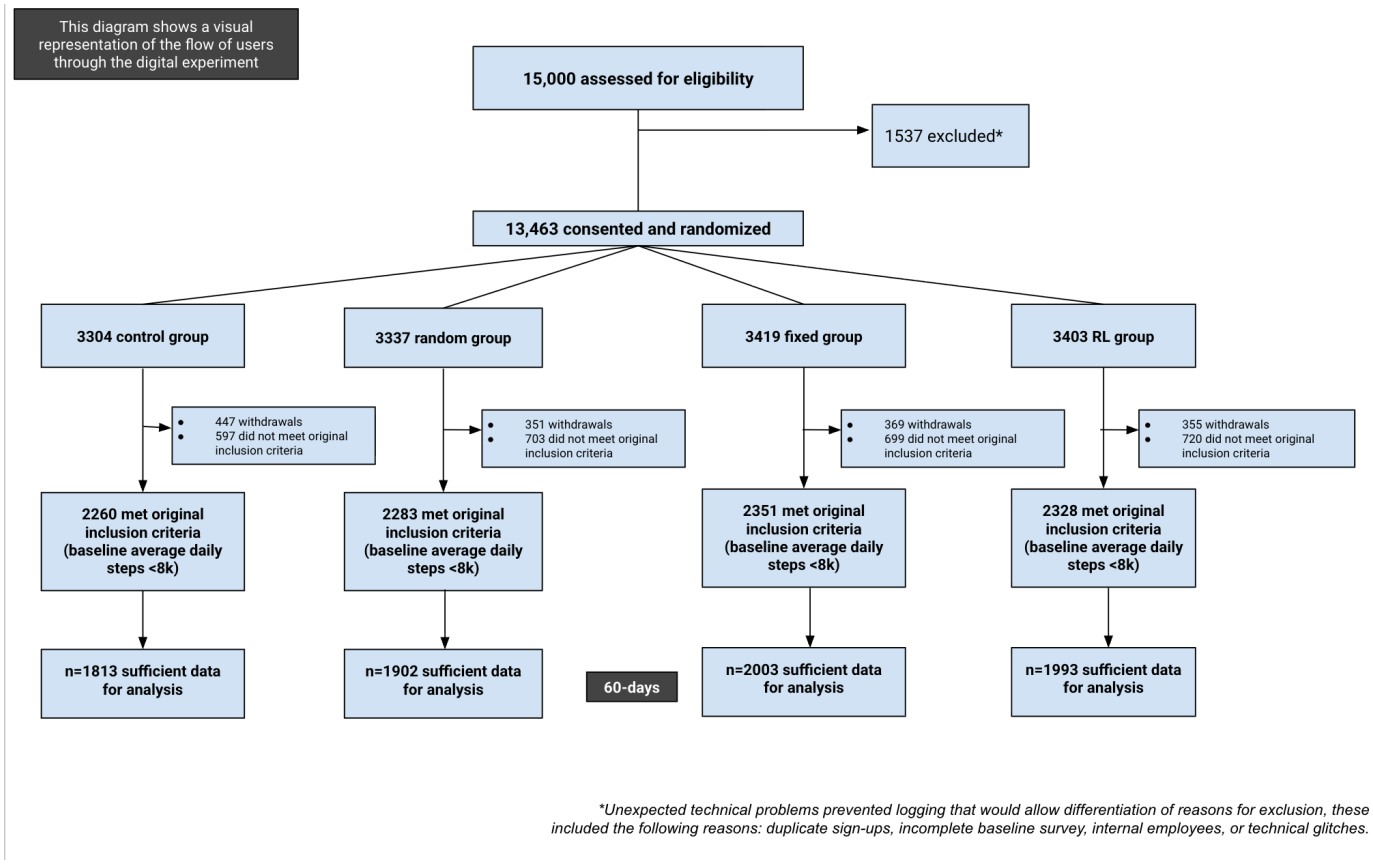


Table 2. Baseline characteristics of study population (n=7711).

Characteristics	Control (n=1813)	Random (n=1902)	Fixed (n=2003)	RL (n=1993)
Age (years), mean (SD)	42.0 (9.1)	42.1 (9.0)	42.0 (8.8)	42.2 (9.0)
Female, n (%)	1578 (87.0)	1641 (86.3)	1705 (85.1)	1733 (87.0)
Weight (kg), mean (SD)	91.1 (29.5)	90.8 (23.6)	90.9 (24.9)	90.4 (23.8)
Baseline average daily steps, mean (SD)	5659.8 (1491.6)	5594.2 (1503.8)	5615.2 (1515.1)	5606.3 (1498.7)
Phase/month of enrollment, n (%)				
Q1[a] 2024 (February/March)	691 (38.1)	766 (40.3)	833 (41.6)	840 (42.2)
Q2 2024 (April, May, and June)	1122 (61.9)	1136 (59.7)	1170 (58.4)	1153 (57.9)
Location (self-report in onboarding survey), n (%)				
Urban	261 (14.4)	295 (15.5)	311 (15.5)	301 (15.1)
Suburban	1153 (63.6)	1216 (63.9)	1267 (63.3)	1240 (62.2)
Rural	376 (20.7)	375 (19.7)	408 (20.4)	435 (21.8)

[a]Quarter.

Outcomes

We analyzed the average daily step count in month 1 and month 2 compared to the month before enrollment (baseline month). Overall, 7711 participants were included in the final analysis (Figure 2).

At the prestudy baseline month, the average daily steps amounted to 5618.2 with no important differences across study arms (Table 3). At month 1, average daily step count increased to 6050 (SD 1922.0) in the control group, 6061.2 (SD 1897.5) in the random group, 6054.8 (SD 1875.9) in the fixed group, and 6282.2 (SD 1956.0) in the RL group. At month 2, average daily step count increased to 5948.6 (SD 2076.9) in the control group, 5947.6 (SD 2068.8) in the random group, 5969.4 (SD 2016.1) in the fixed group, and 6096.3 (SD 2117.4) in the RL group.

Table 3. Step count across study arms at each phase (n=7711).

Study arm	Prestudy baseline month, mean (SD)	Month 1, mean (SD)	Month 2, mean (SD)
Control group	5659.8 (1491.6)	6050.0 (1922.0)	5948.6 (2076.9)

Study arm	Prestudy baseline month, mean (SD)	Month 1, mean (SD)	Month 2, mean (SD)
Random group	5594.2 (1503.8)	6061.2 (1897.5)	5947.6 (2068.8)
Fixed group	5615.2 (1515.1)	6054.8 (1875.9)	5969.4 (2016.1)
RL[a] group	5606.3 (1498.7)	6282.2 (1956.0)	6096.3 (2117.4)
All groups	5618.2 (1502.5)	6114.0 (1915.2)	5991.9 (2070.4)

[a]RL: reinforcement learning.

Difference in differences regression analysis was conducted to compare change in step count across treatment groups; results are presented in (Figure 3, Table 4) [27]. Notably, the RL group demonstrated a statistically significant higher increase in steps at month 1 compared to all other groups. At month 2, the RL group had a higher increase in step count compared to all groups; this difference was only statistically significant for the comparison with the control group. The primary endpoint was to assess the difference in average daily step count change at month 2 for four specific pairwise comparisons, and these are indicated by bolding of the results in Table 4.

Figure 3. Trajectory of average daily step count per group (n=7711). RL: reinforcement learning.

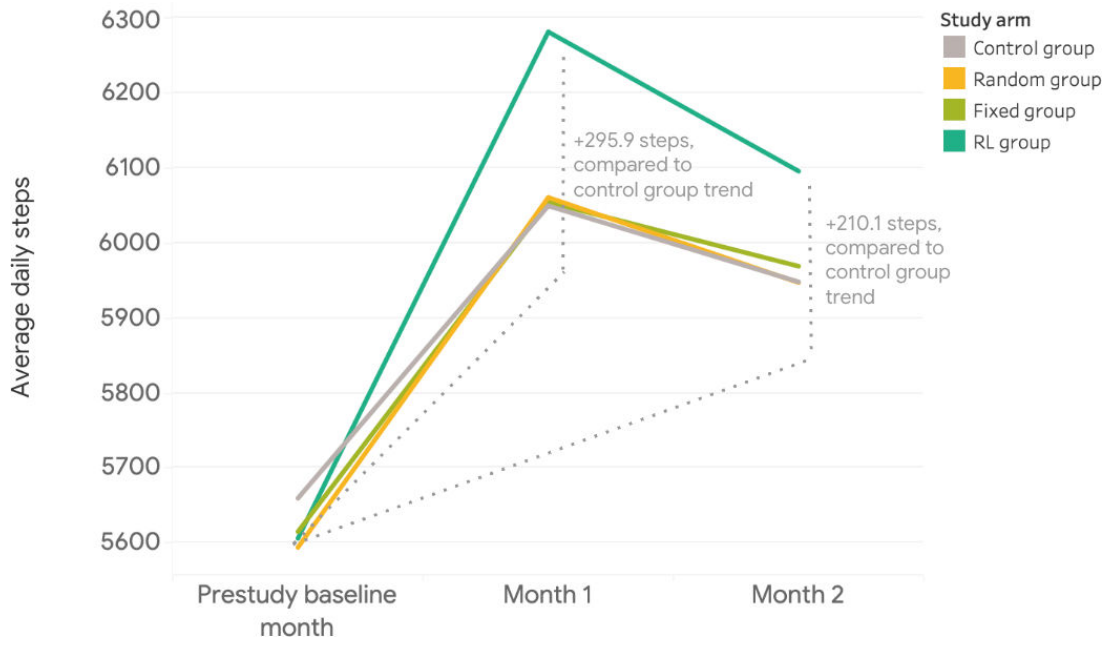


Table 4. Difference in differences regression model results for average daily step count at month 1 and month 2 (n=7711).

	Month 1 vs prestudy baseline month			Month 2 vs prestudy baseline month		
Groups compared	β	SE	*P* value	β	SE	*P* value
Comparisons to control group						
Random vs control	77.6	79.7	.33	63.1	84	.45[a,b]
Fixed vs control	57.5	78.8	.47	75.1	82.7	.36
RL[c] vs control	295.9	79.6	.01	210.1	83.8	.01
Comparisons to random group						
Fixed vs random	20.1	77.7	.80	12	81.7	.88[f]
RL vs random	218.3	78.5	.005	147	82.8	.10[a,e]
Comparisons to fixed group						
RL vs fixed	238.4	77.6	.002	135	81.5	.10[a,f]

[a]Benjamini-Hochberg adjustment used to control for Type 1 error among primary outcomes (primary analyses reported above).
[b]Adjusted *P*=.60.
[c]RL: reinforcement learning.
[d]Adjusted *P*=.88.
[e]Adjusted *P*=.19.
[f]Adjusted *P*=.20.

Exploratory Modelling of Effects Over Time

A GEE (exchangeable correlation structure) was used to estimate the average effect of treatment group on the daily step count over time [28,29]; the results are available in Figure 4 and Table 5.

Figure 4. Modeling longitudinal effect by study day (generalized estimating equation; n=7711). RL: reinforcement learning.

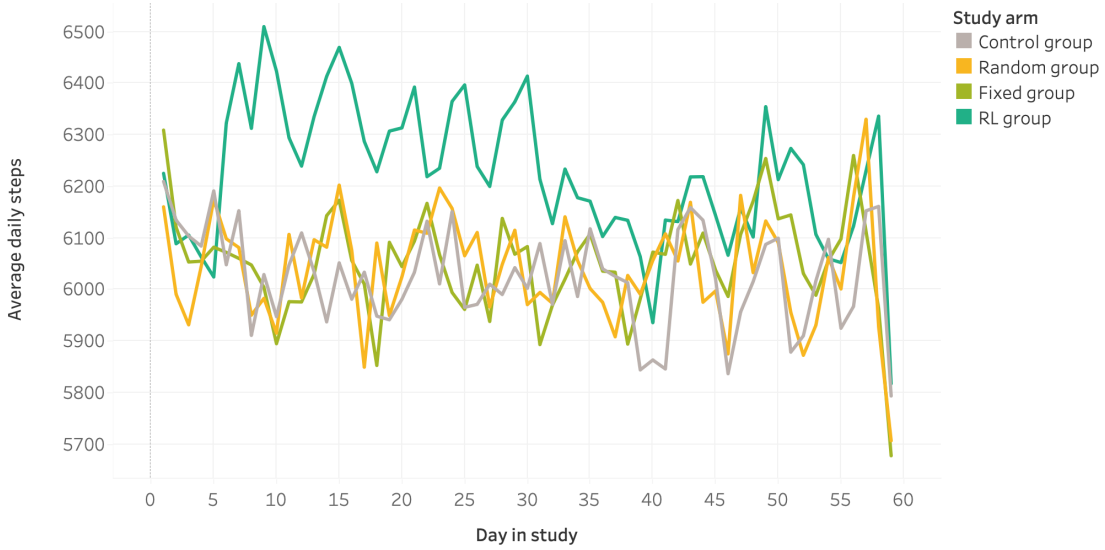


Table 5. Results of the GEE[a] model (n=7711).

Covariate	Estimate (95% CI)	*P* values
(Intercept)	6125.2 (6028.1 to 6222.4)	<.001
Treatment arm		
Random	−14.1 (−148.1 to 119.8)	.84
Fixed	−18.7 (−151.6 to 114.3)	.78
RL[b,c]	208.0 (73.6 to 342.4)	.002
Day in study	−4.5 (−6.6 to −2.5)	<.001
Treatment arm		
Random×study day	0.5 (−2.3 to 3.4)	.71
Fixed×study day	1.3 (−1.5 to 4.1)	.37
RL×study day[b]	−0.6 (−3.4 to 2.2)	.70

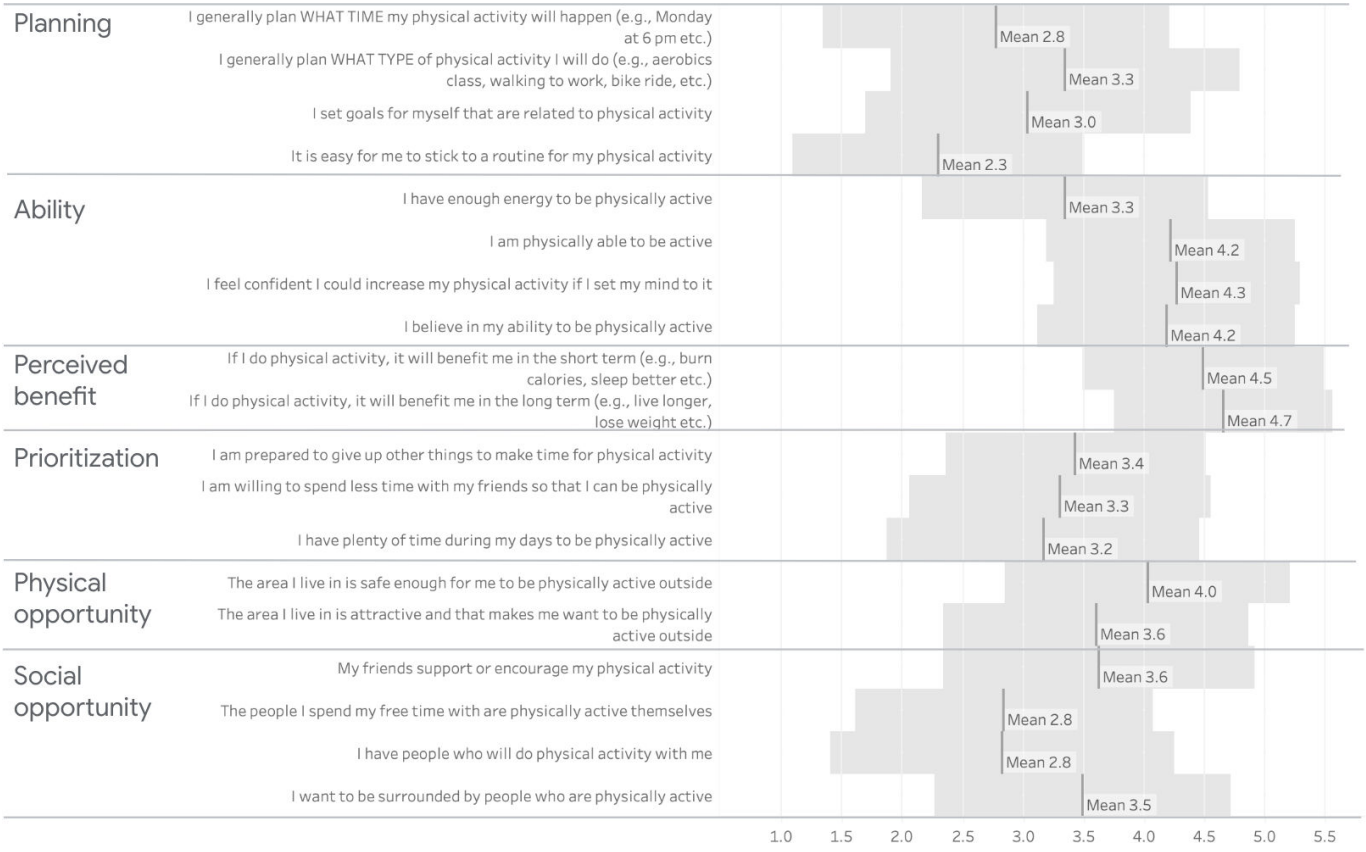
[a]GEE: generalized estimating equation.
[b]Confidence intervals and *P* values for the RL arm are considered exploratory, as statistical certainty may be overstated given the inter-subject dependencies in the pooled RL architecture.
[c]RL: reinforcement learning.

Compared to the control group, the fixed group demonstrated decreased average daily step count by −18.7 (SE 67.9; *P*=.78); this effect increased linearly over time, by a step change of +1.29 (SE 1.4; *P*=.37) compared to the control group. The RL group showed an increase in average daily step count of +208.0 (SE 68.6; *P*=.002), compared to the control group; this effect diminished linearly over time, by a step change of −0.6 (SE 1.4; *P*=.70) compared to the control group. The time effect was negative, with a decrease of −4.5 steps per day; this is consistent with other literature [12].

COM-B Survey

Almost all participants completed the onboarding COM-B survey (7695/N, 99.8%), but only a fraction completed the exit COM-B survey in the final study week (2059/N, 26.7%) (survey is available in Multimedia Appendix 4). At baseline, on average, the study cohort demonstrated the highest scores (reflecting the greatest levels of) perceived benefit and ability, and the lowest scores for the themes planning and social opportunity Figure 5.

Figure 5. COM-B survey questionnaire at study onboarding (mean, SD; n=7695). COM-B: capability, opportunity, and motivation–behavior.



Real-Time User Feedback for Individual Nudges

User feedback was collected at the point of nudge delivery through optional thumbs-up/thumbs-down ratings and open-text comments (92,945 nudges rated, by 5514 participants across all study groups). Thumbs-up and thumbs-down feedback was aggregated by the nudge COM-B theme. Nudges categorized under the ability and perceived benefit themes received higher favorability rates (83,650, 90% positively rated nudges and 81,792, 88% positively rated nudges, respectively), while those in the social opportunity and physical opportunity theme showed markedly lower favorability (77,144, 83% positively rated nudges and 73,426, 79% positively rated nudges, respectively). Nudge feedback was included in the RL model (see Methods section 'Reinforcement Learning Policy'), but was not found to be among the most important features.

User Feedback in Exit Survey

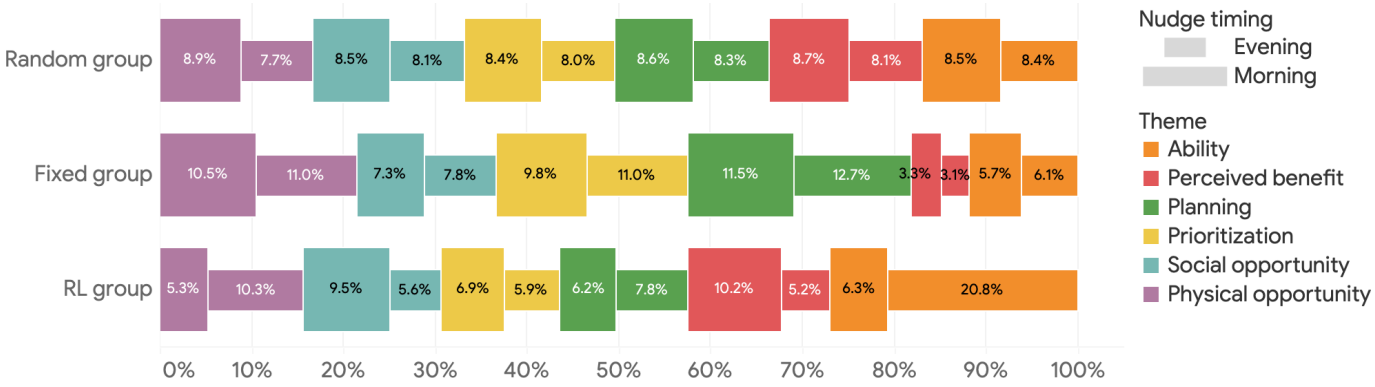
In the final week of the study, participants were invited to complete a study exit survey (see Multimedia Appendix 5 for the full survey). In total, 2985 of 7711 (38.7%) participants completed the survey, with a lower response rate in the no nudge group (325/1813, 17.9%). When asked "Would you recommend nudges to your friends/family starting on a fitness journey?," we observed a higher proportion of positive responses in the RL group (479/617, 77.6%), compared to all other groups: 48.9% (159/325) for the control group, 74.0% (429/580) for the random group, and 70.6% (444/629) for the fixed group.

The exit survey also asked participants: "How would you rate the level of customization and relevance of nudges?." Again, the RL group demonstrated higher rates of favorable responses (224/605, 37%) as compared to all other groups: 35 of 325 (10.8%) for the control group, 155 of 580 (26.7%) for the random group, and 154 of 629 (24.5%) for the fixed group. For both exit survey questions, the control group responded with nonzero rates of positive sentiment, despite receiving no nudges during the study period.

Exploratory Analyses: RL Group and Personalization

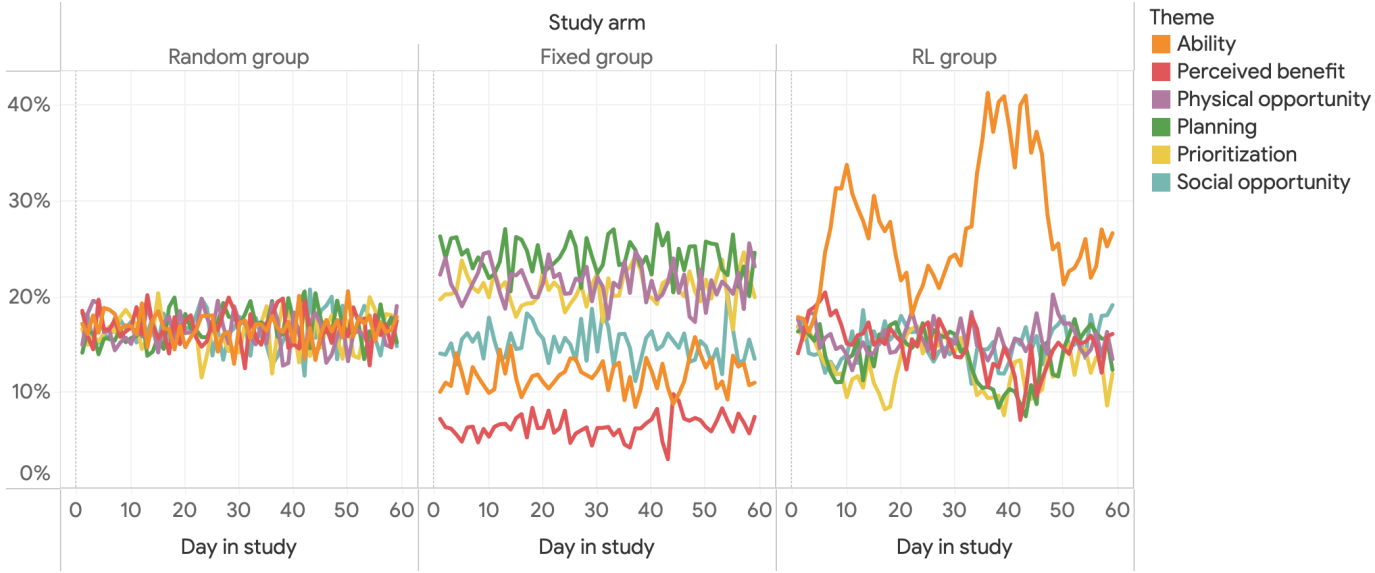
The RL policy for nudge timing and content selection demonstrated distinct patterns compared to the random and fixed arms at the population level. Compared to the random group, the RL group delivered 10% more ability-themed nudges and 7% more evening nudges (Figure 6); note that ability-themed nudges also received the highest favorability rates in the nudge-level feedback. The fixed group delivered the highest rates of planning and physical opportunity-themed nudges (Figure 6), though the physical opportunity theme received the highest rate of negative nudge-level feedback.

Figure 6. Distribution of nudge content and timing selection, across treatment groups. RL: reinforcement learning.



Analysis of nudge distribution over time indicates that in the first ~7 days, the RL policy initially prioritized perceived benefit and planning themes, shifting to the ability theme in later stages, demonstrating temporal adaptation of the policy (Figure 7). As expected, the random group allocated approximately 1/6 of nudges to each nudge theme across participant days.

Figure 7. Distribution of nudge content and timing selection across treatment groups, and over time. RL: reinforcement learning.



Description of the RL Algorithm Learning Process

The most influential features from the RL algorithm across the entire study duration were all fixed characteristics: prestudy baseline step patterns (importance weight 1.23), weight (importance weight 1.19), and age (importance weight 1.17). The behavioral pattern with the largest weight was the previous day step count (importance weight 0.88). Fixed characteristics that carried the most weight within the model were prestudy baseline step patterns, weight, and age. The behavioral pattern with the largest weight was the previous day step count. We refer to the Methods section (RL Policy) for complete details on the features included in the model.

Discussion

Principal Findings

In this large-scale RCT, we investigated the effectiveness of a novel RL algorithm guided by the COM-B behavioral change model, in promoting physical activity through personalized nudges delivered via the Fitbit app. We observed an increase in average daily step count at month 1 and month 2, comparing the RL group to control, random, and fixed groups. This, along with the longitudinal modeling, reveals promising insights into the potential of adaptive, behaviorally-informed interventions. RL holds significant promise as a methodological tool for designing behavioral interventions; this study demonstrated its capacity to leverage user information to optimize both the content and timing of nudges, an adaptive capability that fixed or random designs lack. This integration highlights the primary novelty and innovation of the PEARL trial. While prior mHealth adaptive interventions have successfully demonstrated the

theoretical feasibility of RL in controlled cohorts, PEARL represents a major paradigm shift. It is the first study to scale a behaviorally-grounded (COM-B) contextual multiarmed bandit algorithm within an entirely nonincentivized, organic consumer wearable application framework. By removing artificial financial incentives, which often mask real-world habituation and attrition, this study provides an innovative, ecologically valid evaluation of how dynamic personalization influences natural human behavior at scale.

The observed dropout rate and the proportion of participants with insufficient data for analysis highlight the inherent challenges of maintaining consistent engagement in nonincentivized mHealth interventions (see Statistical Analysis section for data sufficiency requirements). These metrics provide an essential baseline for assessing the feasibility of deploying reinforcement learning algorithms in real-world settings, where irregular app usage and device wear-time are significant variables.

Contextualizing the Observed Step Changes

The significant increase in average daily steps observed in the RL group at one month and the sustained, significant effect over time in the GEE model are noteworthy. While the clinical meaningfulness of step count increases can vary, a consistent increase of 200-300 steps per day, if sustained over time and accompanied by vigorous activity, can contribute to improved health outcomes, particularly in sedentary populations [32]. For instance, studies have shown that even modest increases in daily steps are associated with reduced risks of cardiovascular disease and all-cause mortality [33]. This suggests that while the magnitude of the effect might appear modest, its cumulative impact on a large population could be substantial, aligning with public health goals of reducing global inactivity. Further analysis is needed to understand how the effect is modified by baseline and contextual covariates, particularly among subgroups with very low baseline step-count.

When compared to the existing literature [7-15] (Multimedia Appendix 6), the magnitude of our observed step changes offers an insightful contrast. Foundational trials evaluating RL-driven interventions have typically reported modest initial impacts on physical activity that often attenuate over time. These outcomes were achieved within small, financially compensated cohorts, while PEARL sustained these behavioral benefits across thousands of uncompensated users. This demonstrates that the efficacy of RL-driven JITAIs does not dilute when scaled from tightly controlled academic settings to massive, real-world applications.

While the PEARL study specifically enrolled adults aged 22-60, the underlying global challenge of physical inactivity is particularly acute among adolescents, with 80% currently failing to meet World Health Organization recommendations [2]. Because adolescents are among the most prolific users of mobile technology and wearables, adapting such JITAIs for younger cohorts represents a highly relevant application for future research.

Effectiveness of the Fixed Arm

A distinctly novel aspect of the PEARL study design is the direct comparison between a continuously adaptive algorithm and a static, behavioral science-informed heuristic (the fixed arm). Our findings reveal that participants in the fixed arm, who received nudges selected based on a fixed logic relating to their baseline levels of capability, opportunity, and motivation, did not demonstrate a significant advantage over the group whose nudges were selected at random. This outcome highlights a critical learning point for designing JITAIs: a static, rule-based approach, even when informed by robust behavioral theory, may be insufficient in dynamic, real-world mHealth environments.

The PEARL study exemplified this challenge. A notable observation was the discordance between the high rate of Physical Opportunity nudges delivered by the fixed group (Figure 6) and the predominantly negative user feedback received for this theme. Open-text feedback suggests that Physical Opportunity nudges often relied on inaccurate assumptions about participants' real-world environments (eg, availability of sidewalks or stairs). This underscores a key challenge in applying the baseline COM-B survey to dynamic, in-the-moment interventions. The COM-B framework, as typically applied, may not fully capture the day-to-day or context-specific fluctuations in behavioral barriers (eg, work vs home and weekday vs weekend).

Human behavior is inherently contextual and non-stationary. A fixed intervention policy, determined at baseline, struggles to adapt to evolving user needs, preferences, and external circumstances (eg, changes in daily routines, weather, and motivation levels). This finding points to a gap in the current literature regarding the optimal translation of comprehensive behavioral frameworks like COM-B into continuously adaptive digital nudges. While COM-B effectively diagnoses behavioral determinants, the challenge lies in dynamically selecting and delivering the right intervention component (eg, targeting ability vs opportunity) at the opportune moment for a given individual in real-time. Our results suggest that a dynamic algorithm capable of continuous learning and adaptation is crucial for leveraging behavioral science effectively in ongoing digital interventions. Future work should focus on adapting the COM-B framework (or other theoretical models) to account for such temporal and contextual variability, potentially moving beyond general delivery policies to more deeply personalize nudge content by leveraging richer participant information.

Strengths of the PEARL Reinforcement Learning Approach

In contrast to the fixed arm, the RL arm demonstrated a more sustained positive impact on physical activity. This strength stems from RL's inherent ability to learn and adapt over time, balancing exploration (trying different nudges) and exploitation (delivering nudges that have historically worked well for similar contexts/users). The detailed exploratory analyses within the RL arm further illustrate this adaptive capacity.

Furthermore, as highlighted in the Prior Work section, most existing studies on personalized digital health interventions for physical activity are small-scale, with sample sizes typically ranging from 13 to 220 individuals and a median below 100. These studies frequently call for larger, longer-term investigations to fully evaluate the potential of digital health to significantly improve health-related behaviors and outcomes. PEARL directly addresses this gap by providing insights from a population of more than 7000 participants followed over two months, making it one of the largest RL-based studies in physical activity and beyond.

Beyond its unprecedented scale, the PEARL study offers several key advantages. It features a unique four-arm randomized controlled trial design, which includes two embedded microrandomized trials (MRTs), allowing for both overall treatment efficacy assessment and real-time evaluation of intervention components. Unlike some previous studies that relied solely on mobile apps, PEARL leveraged the continuous data collection from Fitbit wrist-worn trackers, providing a more robust measure of physical activity. This robust study design, combining traditional RCT elements with traditional behavior science interventions (random and fixed arms) and an augmented adaptive intervention, enables a deeper understanding of the effects. The ability of the RL algorithm to learn population-level preferences (eg, evening nudges being more effective than morning nudges across themes) and then personalize further based on individual trajectories is a key differentiator. This adaptive nature, coupled with the large-scale deployment, positions the PEARL study as one of the largest and most comprehensive RL-based intervention studies in digital health to date, addressing critical challenges in real-world human subject research and large-scale data-driven interventions.

Comparison With Prior Work

The study's large sample size and real-world recruitment through the Fitbit app are significant strengths. Compared to previous studies that primarily involved university students and staff [9-12], who may not be representative of the broader population, PEARL's sample may have higher generalizability. Furthermore, unlike most existing studies, PEARL did not offer monetary compensation to participants, minimizing potential selection bias.

Limitations

Regarding the attenuation of the effect at 2 months, this likely indicates that the current RL agent requires a broader information set, or a modified reward function, to optimize for sustained engagement rather than solely immediate response. It is important to note that an RL agent strictly optimizes what it is incentivized to treat as "good" via the reward signal. If this signal does not explicitly account for future retention, the algorithm cannot be penalized for failing to prioritize long-term outcomes. The decline in signal suggests that while the agent successfully learned to trigger immediate activity, future iterations must explicitly weight the reward function to balance immediate efficacy with long-term habit maintenance. In addition, the two-month study duration did not allow for an assessment of the longer-term sustainability of physical activity behaviors. Future research with extended follow-up periods is required to determine whether the observed improvements are maintained beyond this initial intervention phase.

The nominal *P* values for secondary outcomes should be interpreted with caution. Future research should consider pre-registering a comprehensive plan for handling multiple comparisons for secondary outcomes. While we report *P* values for the time-varying effects, these secondary findings should be interpreted as exploratory; the standard GEE framework assumes strict independence between participants, an assumption violated by the cross-participant data pooling in the RL arm, which may result in underestimated standard errors and overly confident p-values [34].

Several factors warrant consideration regarding generalizability. The participant population was predominantly female and self-selected as Fitbit users, potentially limiting the direct applicability of findings to broader, more diverse populations, including men or individuals without wearable devices. Future research should aim to recruit more diverse cohorts. In particular, the volunteer Fitbit users may introduce selection bias related to technological access and baseline motivation. Participants in this study already possessed the digital literacy and financial means to utilize wearable devices, which may not reflect the socioeconomic reality of broader, underserved populations. Additionally, as self-selected volunteers for a health intervention, the cohort likely exhibited a higher degree of initial motivation and health consciousness than the general public, potentially influencing the observed effectiveness of the nudges.

The rolling enrollment between February and June 2024 introduces potential seasonality effects, as physical activity patterns can vary with weather and daylight hours. While baseline characteristics were balanced across groups, the impact of enrollment month as a predictor in subgroup analysis was not explicitly significant for the primary outcome, suggesting that randomization effectively distributed this potential confounder. Nevertheless, future studies could consider stratifying recruitment by season or incorporating seasonal variables into the RL algorithm.

The potential for unblinding in the control group, due to the complete absence of messages, is a limitation. While participants were blinded to their specific treatment allocation, the lack of any interaction in the control arm might have led to higher withdrawal rates or altered engagement compared to the intervention arms. This highlights the challenge of maintaining blinding in digital health interventions that involve active engagement. This is mitigated by the inclusion of the random nudge group, which acts as a useful unblinded comparator.

The reliance on Fitbit data for step count, while pragmatic for a large-scale study, means our findings are specific to this measurement method. The reliance on Fitbit data for step count, while pragmatic for a large-scale study, means our findings are specific to this measurement method. While consumer-grade wearables such as Fitbits have been

validated as reliable instruments for measuring step count in healthy populations, technical variations may still arise from differences in device models or usage behavior. Although participants utilized various consumer-grade Fitbit models, device model data was not recorded, preventing statistical adjustment for potential inter-device algorithmic variations. While this represents a minor limitation in measurement precision, it aligns with our objective of evaluating the intervention within a heterogeneous, real-world consumer ecosystem. Factors such as sensor placement, individual gait characteristics, and algorithm versions can influence data precision and consistency, representing a known variable in large-scale mHealth research.

Additionally, the 7-day/500-step compliance criterion served as a proxy for wear-time, but direct wear-time metrics could provide more granular insights in future studies. The primary analysis operates as a modified ITT or Per-Protocol framework, rather than a traditional strict ITT analysis. A sensitivity analysis is available in Multimedia Appendix 3, which evaluates the impact of the insufficient data compliance.

Further research is needed to explore the specific COM-B components that are most amenable to dynamic nudging and how their effectiveness changes over time or across different user contexts. While our RL algorithm adapted some aspects of content and timing, future iterations could expand on these, or explore additional expansions, such as: other intervention parameters (eg, message frequency and intensity, modality), dynamic components within the nudge text itself (eg, reference prior step counts, trends, or patterns), or more diverse data streams (eg, weather, social context, and mood) into the RL state space to further enhance personalization. Future iterations should also explore the use of AI to develop nudges that are uniquely tailored to the individual.

Conclusions

The PEARL study provides compelling evidence for the feasibility and short-term efficacy of a large-scale, behavioral science-informed reinforcement learning approach in digital physical activity interventions. While the 2-month primary outcome was not statistically significant, the significant short-term gains and sustained effects observed in the RL arm highlight the power of adaptive algorithms to design effective digital interventions by personalizing and optimizing health behavior change. Conversely, the lack of significant advantage in the fixed, rule-based arm emphasizes that relying solely on baseline behavioral assessments is insufficient, as these static policies struggle to adapt to the highly contextual and non-stationary nature of human behavior.

To our knowledge, this represents one of the largest, nonincentivized randomized trials of this duration to evaluate reinforcement learning for physical activity, demonstrating that while observed step gains are modest, such consistent improvements are associated with reduced risks of cardiovascular disease and all-cause mortality when applied at a population level. This work contributes significantly to the growing body of literature on JITAIs and underscores the importance of dynamic, data-driven personalization for scalable and effective digital health solutions.

Acknowledgments

The authors thank the study participants and the clinical operations teams for their efforts and support.

Generative AI Use

The authors declare the use of generative AI (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: proofreading and editing. The GAI tool used was: Gemini 2.5 Pro, Gemini 3.0 Pro. The authors used Gemini to proofread and critique the manuscript. We manually reviewed any suggestions that it provided, and any changes that were subsequently made to the manuscript were done by the authors without the assistance of GAI. Please see Multimedia Appendix 7 for the prompts/responses/transcripts. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by AAL.

Funding

This study was funded by Google LLC. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to proprietary restrictions by the collaborating private company. Access to the data is governed by agreements with the company and cannot be shared without their explicit permission. Code and data were reviewed by ND and SSV (both employed by Google at the time of review), who also verified the reproducibility of the main results.

Authors' Contributions

NH, HEF, ER contributed to study idea and design; NH, SL, PK were involved in data collection; data analysis and results interpretation were performed by AAL; RL algorithm design and analysis were performed by NH, AS, KS; AAL led manuscript writing, with key contributions from NH, ND, ER; critical review and editing were carried out by all authors. All authors approved the final manuscript.

Conflicts of Interest

The authors declare the following potential conflicts of interest: AAL, NH, AS, SL, QH, JH, PK, KS, AR, AS, MN, JT, HEF are employees at Google LLC; ER, JA, SL, RS, MS, HM were employees at Google LLC at the time the study was completed; SSV and ND were visiting researchers at Google LLC at the time of study completion and analysis.
SSV also declares a potential conflict of interest with her engagement with Advisory Board PhaseV.

Editorial Notice

This randomized study was retrospectively registered, explained by authors as follows: the registration was conducted retrospectively because the trial evaluated a non-clinical, behavioral mHealth intervention, which the research team originally categorized outside traditional clinical trial mandates. Prospective institutional oversight was maintained via IRB approval and public archiving on the Open Science Framework (OSF). The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1

Detailed justification for the selection of COM-B as the theoretical framework.
[DOCX File (Microsoft Word File), 2382 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Bank of nudge messages, categorized by their respective behavior change theme.
[DOCX File (Microsoft Word File), 2391 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Sensitivity analysis for participants with missing data.
[DOCX File (Microsoft Word File), 20 KB-Multimedia Appendix 3]

Multimedia Appendix 4

COM-B survey, deployed to participants at both onboarding and study exit.
[DOCX File (Microsoft Word File), 2382 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Study exit survey.
[DOCX File (Microsoft Word File), 2893 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Literature review on ongoing and completed studies utilizing mHealth technologies and personalized intervention delivery (through MRT/RL) to promote physical activity.
[DOCX File (Microsoft Word File), 20 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Generative AI prompts/responses/transcripts used in the preparation of the manuscript.
[DOCX File (Microsoft Word File), 21 KB-Multimedia Appendix 7]

Checklist 1

CONSORT e-HEALTH checklist (V 1.6.1)
[DOCX File (Microsoft Word File), 2644 KB-Checklist 1]

References

1. Bull F, Willumsen J, Stevens G, Strain T. Global levels of physical inactivity in adults: off track for 2030. World Health Organization; 2024.
2. Strain T, Flaxman S, Guthold R, et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5·7 million participants. Lancet Glob Health. Aug 2024;12(8):e1232-e1243. [doi: 10.1016/S2214-109X(24)00150-5] [Medline: 38942042]
3. Global action plan on physical activity 2018-2030: more active people for a healthier world. World Health Organization; 2019.

4. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res. Oct 12, 2020;22(10):e23954. [doi: 10.2196/23954] [Medline: 33044175]
5. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med. May 18, 2018;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
6. Spruijt-Metz D, Marlin BM, Pavel M, et al. Advancing behavioral intervention and theory development for mobile health: the HeartSteps II Protocol. Int J Environ Res Public Health. Feb 17, 2022;19(4):2267. [doi: 10.3390/ijerph19042267] [Medline: 35206455]
7. Klasnja P, Smith S, Seewald NJ, et al. Efficacy of contextually tailored suggestions for physical activity: a micro-randomized optimization trial of HeartSteps. Ann Behav Med. May 3, 2019;53(6):573-582. [doi: 10.1093/abm/kay067] [Medline: 30192907]
8. 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. Oct 10, 2017;19(10):e338. [doi: 10.2196/jmir.7994] [Medline: 29017988]
9. 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. Jan 25, 2018;6(1):e28. [doi: 10.2196/mhealth.9117] [Medline: 29371177]
10. Zhou M, Mintz Y, Fukuoka Y, et al. Personalizing mobile fitness apps using reinforcement learning. Presented at: HUMANIZE '18; Mar 11, 2018; Tokyo, Japan. URL: http://ceur-ws.org/Vol-2068/humanize7.pdf [Accessed 2026-10-13]
11. Aguilera A, Arévalo Avalos M, Xu J, et al. Effectiveness of a digital health intervention leveraging reinforcement learning: results from the Diabetes and Mental Health Adaptive Notification Tracking and Evaluation (DIAMANTE) randomized clinical trial. J Med Internet Res. Oct 8, 2024;26:e60834. [doi: 10.2196/60834] [Medline: 39378080]
12. Figueroa CA, Deliu N, Chakraborty B, et al. Daily motivational text messages to promote physical activity in university students: results from a microrandomized trial. Ann Behav Med. Feb 11, 2022;56(2):212-218. [doi: 10.1093/abm/kaab028] [Medline: 33871015]
13. Doherty C, Lambe R, O'Grady B, et al. An evaluation of the effect of app-based exercise prescription using reinforcement learning on satisfaction and exercise intensity: randomized crossover trial. JMIR Mhealth Uhealth. Nov 26, 2024;12:e49443. [doi: 10.2196/49443] [Medline: 39622712]
14. Lauffenburger JC, Yom-Tov E, Keller PA, et al. The impact of using reinforcement learning to personalize communication on medication adherence: findings from the REINFORCE trial. NPJ Digit Med. Feb 19, 2024;7(1):39. [doi: 10.1038/s41746-024-01028-5] [Medline: 38374424]
15. Golbus JR, Gupta K, Stevens R, et al. A randomized trial of a mobile health intervention to augment cardiac rehabilitation. NPJ Digit Med. Sep 14, 2023;6(1):173. [doi: 10.1038/s41746-023-00921-9] [Medline: 37709933]
16. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci. Apr 23, 2011;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
17. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci. Apr 24, 2012;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
18. Park LG, Ng F, Handley MA. The use of the Capability-Opportunity- Motivation Behavior (COM-B) model to identify barriers to medication adherence and the application of mobile health technology in adults with coronary heart disease: a qualitative study. PEC Innov. Dec 15, 2023;3:100209. [doi: 10.1016/j.pecinn.2023.100209] [Medline: 37753273]
19. Webb J, Foster J, Poulter E. Increasing the frequency of physical activity very brief advice for cancer patients. Development of an intervention using the behaviour change wheel. Public Health. Apr 2016;133:45-56. [doi: 10.1016/j.puhe.2015.12.009] [Medline: 26822162]
20. Liu X, Deliu N, Chakraborty B. Microrandomized trials: developing just-in-time adaptive interventions for better public health. Am J Public Health. Jan 2023;113(1):60-69. [doi: 10.2105/AJPH.2022.307150] [Medline: 36413704]
21. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press; 2018.
22. Lee AA, Hegde N, Emir-Farinas H, et al. Protocol: evaluating PEARL - a personalized exercise assistant using reinforcement learning (walkmate study). OSF Registries. URL: https://osf.io/tw7up/ [Accessed 2026-10-31]
23. Deliu N, Williams JJ, Chakraborty B. Reinforcement learning in modern biostatistics: constructing optimal adaptive interventions. Int Statistical Rev. Dec 2025;93(3):385-424. URL: https://onlinelibrary.wiley.com/toc/17515823/93/3 [Accessed 2026-10-13] [doi: 10.1111/insr.12583]
24. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol. Dec 2015;34S:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]

25. Semaan S, Dewland TA, Tison GH, et al. Physical activity and atrial fibrillation: data from wearable fitness trackers. Heart Rhythm. May 2020;17(5 Pt B):842-846. [doi: 10.1016/j.hrthm.2020.02.013] [Medline: 32354448]
26. Block VJ, Lizée A, Crabtree-Hartman E, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol. Feb 2017;264(2):316-326. [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
27. Card D, Krueger AB. Minimum wages and employment: a case study of the Fast-Food Industry in New Jersey and Pennsylvania. Am Econ Rev. 1994;84:772-793. [doi: 10.3386/w4509]
28. Ballinger GA. Using generalized estimating equations for longitudinal data analysis. Organ Res Methods. Apr 2004;7(2):127-150. [doi: 10.1177/1094428104263672]
29. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
30. R: a language and environment for statistical computing. The R Foundation. 2021. URL: https://www.R-project.org [Accessed 2026]1310
31. Python. 2026. URL: python.org [Accessed 2026-10-13]
32. Stamatakis E, Ahmadi MN, Gill JMR, et al. Association of wearable device-measured vigorous intermittent lifestyle physical activity with mortality. Nat Med. Dec 2022;28(12):2521-2529. [doi: 10.1038/s41591-022-02100-x] [Medline: 36482104]
33. Master H, Annis J, Huang S, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. Nat Med. Nov 2022;28(11):2301-2308. [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
34. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc. 2018;113(523):1112-1121. [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]

Abbreviations

C-MAB: contextual multi-armed bandit
COM-B: capability, opportunity, motivation–behavior
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online
DiD: difference-in-differences
GEE: generalized estimating equation
IRB: institutional review board
ITT: intention-to-treat
JITAI: just-in-time adaptive intervention
MCAR: missing completely at random
mHealth: mobile health
mITT: modified intention-to-treat
MRT: microrandomized trial
PEARL: Personalized Exercise Assistant Using Reinforcement Learning
RCT: randomized controlled trial
RL: reinforcement learning
WHO: World Health Organization

Edited by Alicia Stone; peer-reviewed by Jieru Shi, Miloud Chakit; submitted 27.Jan.2026; final revised version received 25.Aug.2026; accepted 25.Aug.2026; published 17.Nov.2026

Please cite as:
Armento Lee A, Hegde N, Deliu N, Rosenzweig E, Suggala A, Lakshminarasimhan S, He Q, Hernandez J, Seneviratne M, Singh R, Kalkar P, Shanmugam K, Raghuveer A, Singh A, Manoharan H, Nguyen M, Taylor J, Alla J, Villar SS, Emir-Farinas H
A Personalized Exercise Assistant Using Reinforcement Learning: Results From a Large-Scale Mobile Health Randomized-Controlled Trial
J Med Internet Res 2026;28:e91156
URL: https://www.jmir.org/2026/1/e91156
doi: 10.2196/91156

© Amy Armento Lee, Narayan Hegde, Nina Deliu, Emily Rosenzweig, Arun Suggala, Sriram Lakshminarasimhan, Qian He, John Hernandez, Martin Seneviratne, Rahul Singh, Pradnesh Kalkar, Karthikeyan Shanmugam, Aravindan Raghuveer, Abhimanyu Singh, Hariharan Manoharan, My Nguyen, James Taylor, Jatin Alla, Sofia S Villar, Hulya Emir-Farinas.

Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.