NEJM Evidence

Published January 22, 2024
NEJM Evid 2024; 3 (2)
DOI: 10.1056/EVIDoa2300164

ORIGINAL ARTICLE | INTERFACE SERIES

Smart Start — Designing Powerful Clinical Trials Using Pilot Study Data

Johannes O. Ferstad, M.Sc.,[1] Priya Prahalad, M.D., Ph.D.,[2] David M. Maahs, M.D., Ph.D.,[2] Dessi P. Zaharieva, Ph.D.,[2] Emily Fox, Ph.D.,[3,4,5] Manisha Desai, Ph.D.,[6] Ramesh Johari, Ph.D.,[1] and David Scheinker, Ph.D.[1,2,7]

Abstract

BACKGROUND Digital health interventions may be optimized before evaluation in a randomized clinical trial. Although many digital health interventions are deployed in pilot studies, the data collected are rarely used to refine the intervention and the subsequent clinical trials.

METHODS We leverage natural variation in patients eligible for a digital health intervention in a remote patient-monitoring pilot study to design and compare interventions for a subsequent randomized clinical trial.

RESULTS Our approach leverages patient heterogeneity to identify an intervention with twice the estimated effect size of an unoptimized intervention.

CONCLUSIONS Optimizing an intervention and clinical trial based on pilot data may improve efficacy and increase the probability of success. (Funded by the National Institutes of Health and others; ClinicalTrials.gov number, NCT04336969.)

Introduction

The use of digital health interventions (e.g., remote patient monitoring) is growing in clinical practice and as an area of research.[1-3] Most digital health interventions are deployed without evaluation in a randomized controlled trial (RCT).[4-6] Of the digital health interventions evaluated in an RCT, many are used in a pilot study before the RCT, but, in most cases, the data from the pilot study are not used to optimize the intervention nor the RCT. When interventions are deployed without being optimized, they may fail to improve patient outcomes sufficiently to warrant widespread use or fail to produce significant results in an RCT. A relatively short RCT of an efficacious multiyear digital health intervention for the management of a chronic condition may fail if the intervention is not well targeted or optimized, leading to low effect sizes and insufficient statistical power. Underpowered randomized trials are wasteful, can be misleading, and may therefore be considered unethical.[6-10] As the use and impact of wearable sensors and

The author affiliations are listed at the end of the article.

Dr. Scheinker can be contacted at dscheink@stanford.edu or at 475 Via Ortega, Stanford, CA 94305.

digital health interventions increase, so does the need for improved methods and practices for their evaluation and optimization.[11,12]

Just-in-time adaptive interventions are designed to provide care at the right time based on estimates of each patient's changing internal state and external context. New methods are emerging for the design and evaluation of just-in-time adaptive interventions. These include the multiphase optimization strategy, sequential multiple assignment randomized trials, and microrandomized trials.[12-16] The multiphase optimization strategy seeks to optimize an intervention in a refining phase based on data from a screening phase before evaluation in a confirming phase. A sequential multiple assignment randomized trial, based on an optimization trial, may be conducted to estimate the short-term effects of components of a digital health intervention and optimize its design.[17] In a microrandomized trial, each participant is randomly selected to receive or not receive a treatment recommendation numerous times over the course of the trial. A microrandomized clinical trial may be used to design an adaptive intervention or as a substitute for a traditional RCT when the primary outcome of interest is closely associated with the short-term effects of individual components of the digital health intervention.

We introduce Smart Start, a three-step method for using retrospective data from a pilot study to design an adaptive health intervention and a subsequent clinical trial. The first step is to define candidate variants of the digital health intervention based on different — potentially adaptive — policies for targeting patients to receive care. The second step is to simulate microrandomized trials of each intervention and estimate the average treatment effect in each simulated trial. In this step, trials of varying size (number of patients) and duration are simulated, and the predicted probability of success (PPOS) is estimated. The final step is to design a standard RCT or a microrandomized trial by choosing the intervention, study size, and study duration.

Smart Start may be used for remote digital health interventions in which: patients wear sensors regularly transmitting data; at regular intervals (e.g., weekly), clinicians or a computer algorithm (e.g., mobile application) review the data and potentially send recommendations to the patient; and short-term observable changes in the patient's data can be used to estimate the impact of the recommendations. In summary, Smart Start requires data from a pilot of the remote intervention with a sufficient number of patients, interactions, and observations. It also requires structured variation in the set of patients who are eligible to receive a recommendation at each decision interval (e.g., weekly). This variation must be sufficient to support simulation and evaluation of the effect of different targeting policies. Such variation is common when interventions are based on patient behavior and characteristics.[18,19] One form of structured variation arises from constraints on how many patients the care team has the capacity to contact in each interval of the study (e.g., weekly).[20] Another form of structured variation arises from constraints on how often to contact or prompt a patient.[21,22] As the use of sensor-based digital health interventions continues to increase, growing constraints on provider availability and patient willingness to engage will expand the interventions to which Smart Start is applicable.

We illustrate Smart Start with data collected from individuals with type 1 diabetes using a continuous glucose monitoring system.[23,24] In the study, different sets of patients were eligible each week to receive a recommendation from the clinicians reviewing remote monitoring data.[25] The primary outcome, as is common in the study of interventions for diabetes, was time in range, defined as the proportion of glucose readings between 70 and 180 mg/dl.[22,26,27] The variation in patient eligibility and the short-term effects of interventions on time in range made the study well suited for Smart Start. Incorporating this kind of variation in pilot studies of digital health interventions may facilitate downstream evaluation of various targeting policies with Smart Start.

Methods

STUDY PARTICIPANTS

The data used in this work are from the 4T (Teamwork, Targets, Technology, and Tight Control in Newly Diagnosed Pediatric T1D) study (ClinicalTrials.gov number, NCT03968055), which recruited youth with type 1 diabetes to start continuous glucose monitoring within 1 month of diabetes diagnosis and participate in remote monitoring enabled by algorithmic analysis of patient continuous glucose monitoring data.[26] The Stanford Institutional Review Board approved this study, and written informed consent was obtained. Every week in the first year after diagnosis, continuous glucose monitoring data were reviewed by certified diabetes care and education specialists who decided

whether insulin dose adjustments or other changes to glucose management were necessary and contacted patients using secure messaging within the electronic health record. After the first year, patients who continued remote monitoring were reviewed by a clinician 1 out of every 4 weeks.

PILOT DATA

Our data set combines continuous glucose monitor readings providing up to 288 glucose values daily, messages sent to patients through the electronic health record, and patient information from forms completed by study participants. Included are all study participants who received remote monitoring in the second year after their type 1 diabetes diagnosis, had at least 26 weeks of continuous glucose monitoring data, and received a message based on remote monitoring in their first year. During their first year in the program, each patient's data were included for review by care providers every week. During their second year in the program, their data were included for review 1 week out of every 4 weeks. Each week, data for all patients to be reviewed were shown to a certified diabetes care and education specialist. Patients were prioritized for review according to various glucose metrics: time in range, other aggregated continuous glucose monitoring data, and criteria derived from consensus targets such as percent time below and above range, as well as a personalized metric defined in terms of the patient's change in time in range.[26] After reviewing each patient's data, the certified diabetes care and education specialist decided, based on clinical criteria and judgment, whether to provide care guidance through an asynchronous message to the patient or the patient's parent/caregiver. Thus, to receive a message regarding their glucose data, a patient must first have been selected for review and then the clinician reviewing the data must have decided the data warranted a message. Our goal was to understand which policies for targeting, or selecting, patients for review are most likely to result in a favorable effect being observed in a subsequent clinical trial. The data were split for each patient into training and testing. The training data are each patient's first year in the program (year 1) and the testing data are each patient's second year in the program (year 2).

SMART START

We designed a dynamic intervention based on a targeting policy and a microrandomized trial of the intervention in which the unit of observation is a patient-week. Each week, a targeting policy determines whether a patient is eligible to have their data shown to a certified diabetes care and education specialist for remote review and potential contact. The patient-weeks targeted are randomly assigned in the microrandomized trials into a treatment and control group. For each patient-week in the treatment group, a provider will review their data and potentially contact the patient based on the same criteria as in the pilot study. Patient-weeks in the control group and those that were not targeted according to the policy are not eligible for data review. The effect of each intervention is estimated by comparing the outcomes of patient-weeks in the treatment and control groups. Smart Start consists of three steps.

Step 1: Define Candidate Interventions Based on a Variety of Targeting Policies

Each intervention is defined by a targeting policy for how to select from all patients in the trial, K patients to potentially receive treatment recommendations each week, the value of K, and the number of weeks (W) to run the intervention. The following targeting polices were considered:

- Random: K patients are randomly selected.
- Lowest time in range: the K patients with the lowest time in range in the previous week.
- Most likely to be contacted: the K patients with the highest probability of contact based on their data from that week, estimated using a random forest model trained on year 1 data and predicting contact by the clinic in a given week using continuous glucose monitoring metrics from the previous week. These metrics are the proportion of time in range; the proportion of glucose readings in hypoglycemia (54–70 mg/dl), clinically significant hypoglycemia (<54 mg/dl), hyperglycemia (180–250 mg/dl), and extreme hyperglycemia (>250 mg/dl).[28]
- Largest expected effect: the K patients with the largest expected treatment effect (increase in time in range in the subsequent week). This is defined as the product of the probability of the patient being contacted and the expected change in time in range conditional on the patient being contacted, as estimated by, respectively, the random forest model described earlier and a patient-level mean effect in year 1 data with empirical Bayes adjustments to account for different sample sizes and variances across patients.

Step 2: Simulate Microrandomized Trials of the Interventions

We use year 2 data, when patients were eligible for review only 1 out of every 4 weeks, to simulate microrandomized trials of each intervention. For this step, a variety of methods may be used from the active fields of statistical causal inference and policy evaluation.[29,30] Our process is analogous to emulating target trials of the interventions using their associated candidate targeting policies.[31] We modeled the distributions of outcomes for patients reviewed or not reviewed with the potential outcomes framework (details are provided in Section S1 of the Supplementary Appendix).

The simulated microrandomized trials consist of weekly applications of the dynamic intervention to select K patients, as specified by the targeting policy, from the 200 patients in the trial. Half of these patients are randomly selected for inclusion in the treatment group and the other half for inclusion in the control group (Fig. 1). For each intervention, a trial is simulated, and its average treatment effect is estimated as follows:

Step 1. Sample, with replacement, 100 patient-weeks that were reviewed and 100 patient-weeks that were not reviewed.

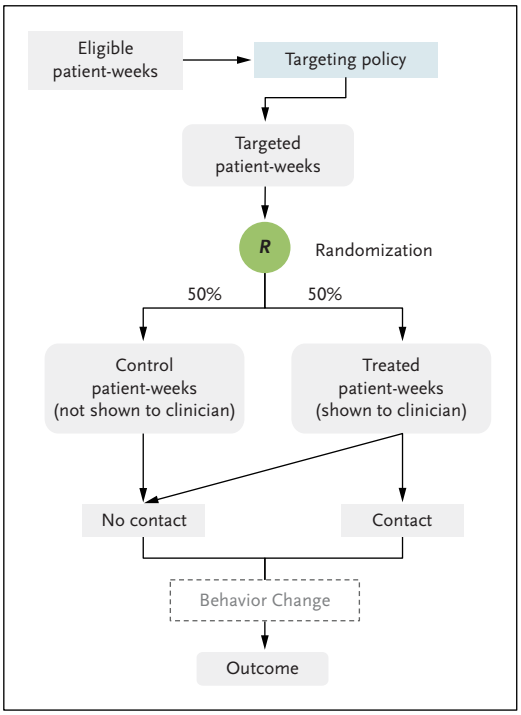


Figure 1. Simulated Microrandomized Trial Design.

Step 2. From the sample of 200 patient-weeks, subsample the top *K/2* reviewed patient-weeks and *K/2* unreviewed patient-weeks according to the targeting policy.

Step 3. Repeat steps #1 and #2 *W* times to simulate the full length of the trial. This results in *W*K* patient-week observations, one half of which were randomly selected for review.

Step 4. Estimate the impact of the intervention on week-to-week changes in time in range with a linear regression of the change in time in range as a function of an indicator variable *reviewed* (whether a patient was reviewed). The average treatment effect is estimated as the coefficient of *reviewed*, and the P value of the coefficient is recorded.

For each intervention, the simulation is repeated 500 times, and the estimated average treatment effect and the P value of each trial are kept.

To test the sensitivity of the results to our assumption that the observations for each patient-week are independent, we performed a simulation respecting the chronological order of observations for each patient. This simulation was done on a transformed version of the original data with synthetic week IDs generated chronologically to achieve a minimum sample size of reviewed and nonreviewed patients per week (results are presented in Section S2 of the Supplementary Appendix). To test whether the sampling process introduced bias by targeting only a subset of patients, we repeated the simulations with the review indicators randomly permuted (results are presented in Section S3 of the Supplementary Appendix).

Step 3: Choose the Intervention to Evaluate in a Randomized Trial and the Size and Duration of the Trial

To summarize the results from the simulated trials of each intervention, we estimate the average treatment effect as the mean estimated average treatment effect across 500 simulations with the same policy and values of K and W. The PPOS is defined as the proportion of the simulated trials with a P value less than 0.05.

The summarized results are then used to identify the intervention with the best performance given the practical constraints of the expected subsequent trial; for example, the duration of the trial and the number of patients to receive treatment recommendations each period. A microrandomized trial or standard RCT may be used to evaluate the treatment effect of the chosen intervention.

Results

Ninety-five patients met the inclusion criteria. The year 1 training data contained 9934 patient-weeks, 3222 of which had a remote monitoring message sent to the patient; the year 2 testing data contained 4648 patient-weeks, 1106 of which were reviewed and 648 of which had a message sent to the patient as part of the remote monitoring. Of the total 14,582 participant-weeks, 3870 resulted in a message being sent to a patient. There were 304,000 simulated microrandomized trials, 500 each of each permutation of policies, numbers of patients randomly assigned, and numbers of weeks considered.

Across the interventions evaluated, the number of patients selected for the intervention required to exceed a 90% PPOS at a significance level of 0.05 in a 52-week trial was 15 for the Largest Expected Effect, 90 for Lowest Time in Range, 100 for Most Likely to Be Contacted, and 105 for Random. In a 26-week trial, the number of patients selected for the intervention was 90 for the Largest Expected Effect, 135 for Most Likely to Be Contacted, 115 for Lowest Time in Range, and 190 for Random (Fig. 2).

For all interventions, the maximum PPOS is attained in trials with all 200 patients targeted with treatment recommendations each week. The difference between the PPOS for a 52-week trial targeting 200 patients weekly and one targeting 15 patients weekly is 9% for the Largest Expected Effect, 100% for Lowest Time in Range, 25% for Most Likely to Be Contacted, and 74% for Random. The average treatment effect does not change with the number of patients targeted with Random selection, peaks at 100 patients for Lowest Time in Range, and decreases with the number of patients targeted for Most Likely to Be Contacted and Largest Expected Effect (Fig. 3). The main results do not differ significantly from the results of more complex simulations that do not assume independence between patient-weeks (results are presented in the Supplementary Appendix).

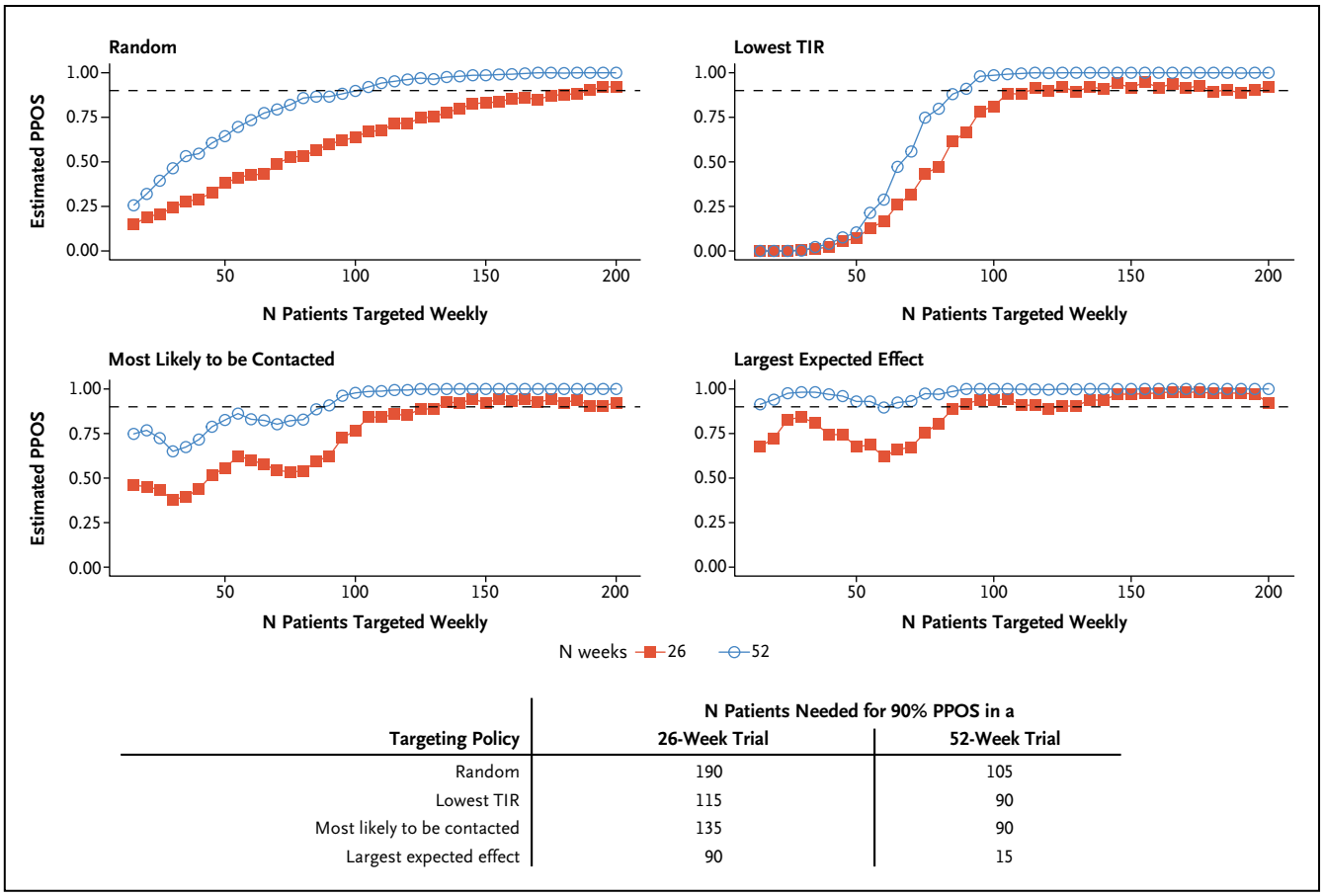
Discussion

We introduced Smart Start, a method to design an adaptive digital health intervention by simulating variants of the intervention before choosing one to evaluate in a randomized clinical trial. We applied it to 14,582 patient-weeks of data from a pilot study of remote patient monitoring for patients with type 1 diabetes using a continuous glucose monitoring system. The adaptive intervention (targeting patients based on the largest expected effect) had higher average treatment effects than the historical intervention (targeting based on lowest time in range), requiring 50% fewer patients selected for the intervention in a 26-week trial and 85% fewer patients selected for the intervention in a 52-week trial to achieve a PPOS of 90% at a significance level of 0.05. Smart Start uses historical patient data to estimate average treatment effects for the design of adaptive interventions that yield greater average treatment effects and greater PPOS than traditional approaches. The improvements in PPOS result from targeting patients in weeks when they are more likely to respond to treatment. In appropriate digital health interventions with pilot data available, using Smart Start may allow for the development of an optimized dynamic intervention and make a randomized clinical trial more likely to succeed.

Existing adaptive trial designs (e.g., adaptive randomization) offer alternative approaches to comparing the effects of various targeting policies. If designed carefully, an adaptive trial can identify a good policy and correctly estimate its long-term effect on patient outcomes. However, to test multiple policies, an adaptive trial will require enough participants receiving care under each policy over a sufficiently long period of time.[32,33] Smart Start avoids having to run an adaptive trial comparing many policies by leveraging variation in pilot data to identify a single good intervention policy for a subsequent confirmatory standard clinical trial.

It is common in digital health interventions for only a subset of patients to receive treatment recommendations at each potential opportunity.[20,22,25] Limits on care provider time may impose limits on how many patients receive an intervention in any given period. A desire to avoid overburdening patients may also lead to limitations on how often any one patient receives an intervention. The Largest Expected Effect policy developed is particularly well suited for interventions in which such constraints are more stringent; its average treatment effect and PPOS improve over those of a random selection policy when the number of patients targeted each week is low.

The candidate policies examined in our example do not consider all indicators that are relevant to the intervention. We evaluated simple targeting policies for illustrative purposes, and some patient populations could be left out of the study unintentionally. Additional work is necessary to identify alternative interventions effective for patients not targeted



Targeting Policy	N Patients Needed for 90% PPOS in a 26-Week Trial	52-Week Trial
Random	190	105
Lowest TIR	115	90
Most likely to be contacted	135	90
Largest expected effect	90	15

Figure 2. Estimated PPOS by Targeting Policy (Panel Titles), Number of Patients Targeted Each Week (*x* Axis), Number of Weeks to Run the Trial (Line Colors), and Number of Patients Needed to Achieve 90% PPOS by Policy (Table).

The horizontal dashed lines indicate 90% PPOS. PPOS denotes predicted probability of success; and TIR, time in range.

with the primary interventions. If the patients who are not expected to benefit significantly from a particular intervention benefit from a complementary intervention, the effect of the new treatment regimen may be greater. It may be unethical not to provide alternative treatment to patients who are less responsive to remote interventions. Future work includes plans to incorporate equity constraints into the adaptive policy we will apply in practice to make sure we do not perpetuate existing inequities in how clinical resources are allocated.[34] One potential solution is to add parity constraints across patient characteristics.[35] Future work will incorporate additional metrics from wearable technology (e.g., wearable activity trackers) that will provide more information on patient outcomes.[36]

Our limitations include working with data from a single clinic. Our results are not meant to estimate any causal effects of an intervention. They estimate the effectiveness of smarter targeting of a dynamic intervention for a subset of a patient population with long, rich data trajectories. This use case will continue to grow in popularity as more clinics adopt the use of algorithm-enabled care and glucose monitoring for people with type 1 and type 2 diabetes, as well as with other remote monitors for other chronic conditions. The main results are from simulations that assume independence between patient-weeks, but these do not differ significantly from the results of a more complex simulation that does not assume independence. Fully accounting for temporal variation in patient responses related to past engagement with recommendations requires running a clinical trial to measure long-term outcomes (last step of Smart Start).

Because we examine the scenario in which the clinical trial follows a pilot study of the same population, we do

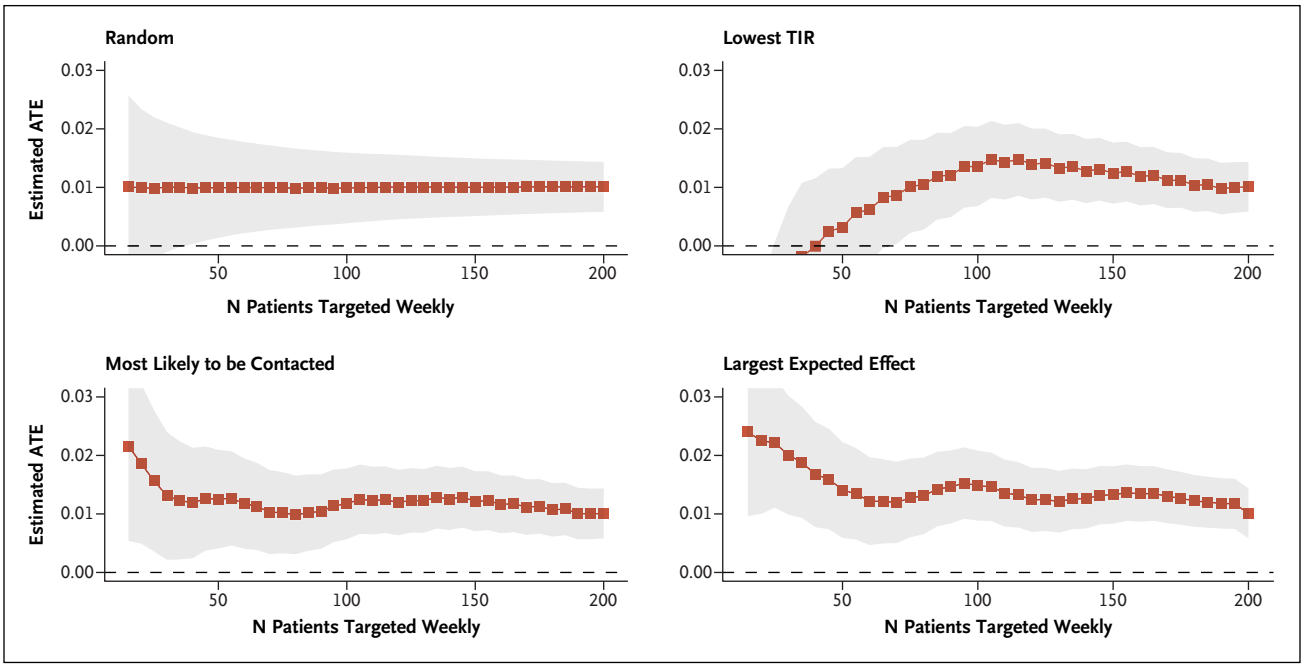


Figure 3. Estimated ATEs by Targeting Policy (Panel Titles) and the Number of Patients Targeted Each Week (*x* Axis).

The gray areas in the first two rows indicate the 95% confidence intervals around the estimated ATEs of the 52-week trials. ATE denotes average treatment effect; and TIR, time in range.

not explicitly quantify the potential of Smart Start to reduce the total number of patients required for a clinical trial or design a policy for an entirely new population. Our results suggest that for trials with fewer patients or new patients, policies designed with Smart Start are likely to achieve a higher PPOS than random policies. This should be explored further in subsequent research.

Smart Start may be used to design and evaluate adaptive policies for digital health interventions based on historical pilot data with some randomness (e.g., from capacity constraints). Smart Start uses a patient-specific and context-dependent likelihood to benefit from an intervention to identify interventions with improved average treatement effects that can achieve a desired predicted probability of success while targeting fewer patients. Improving the design and evaluation of adaptive digital health interventions may improve the impact and rigorous evaluation of digital health interventions.

Disclosures

This work was supported by the National Institutes of Health (NIH; R18DK122422); the Helmsley Charitable Trust (G-2002-04251-2); the ISPAD-JDRF Research Fellowship, Stanford Diabetes Research Center (1P30DK11607401); Lucile Packard Children's Hospital Auxiliaries; the National Science Foundation (2205084); the Stanford Maternal and Child Health Research Institute; the Stanford REDCap Platform (UL1 T001085); and Stanford Data Science. The work was also supported in part by the Stanford Institute for Human-Centered Artificial Intelligence, Air Force Office of Scientific Research Grant FA9550-21-1-0397, and Office of Naval Research Grant N00014-22-1-2110. Dr. Fox is a Chan Zuckerberg Biohub–San Francisco Investigator. The research was partially supported by the following NIH funding source of Stanford's Center for Clinical and Translational Education and Research award, under the Biostatistics, Epidemiology, and Research Design Program: UL1TR003142.

Author disclosures and other supplementary materials are available at evidence.nejm.org.

Restrictions apply to the availability of some, or all data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided. Simulation code is available online at: https://rpubs.com/jferstad/RPM_MRT_Simulations.

We thank the members of the 4T Study team, including Franziska Bishop, Brianna Leverenz, Jeannine Leverenz, Erica Pang, Piper Sagan, and Anjoli Martinez-Singh. We also thank numerous individuals who contributed to the development and deployment of the technology that facilitated remote data review: Angela Gu, Jacqueline Jil Valon,

Daniel Jun, Michael Zikai Gao, Annie Chang, Isha Thapa, Ransalu Senanayake, Jamie Kurtzig, Franziska Bishop, Glenn Loving, Conner Brown, Rick Majzun, and Brendan Watkins.

Author Affiliations

[1] Department of Management Science and Engineering, Stanford University School of Engineering, Stanford, CA

[2] Division of Pediatric Endocrinology, Stanford University School of Medicine, Stanford, CA

[3] Department of Statistics, Stanford University, Stanford, CA

[4] Department of Computer Science, Stanford University, Stanford, CA

[5] Chan Zuckerberg Biohub, San Francisco

[6] Quantitative Sciences Unit, Department of Medicine, Stanford University School of Medicine, Stanford, CA

[7] Clinical Excellence Research Center, Stanford University School of Medicine, Stanford, CA

References

1. Hayes CJ, Dawson L, McCoy H, et al. Utilization of remote patient monitoring within the United States health care system: a scoping review. Telemed J E Health 2023;29:384-394. DOI: 10.1089/tmj.2022.0111.
2. Tang M, Mehrotra A, Stern AD. Rapid growth of remote patient monitoring is driven by a small number of primary care providers. Health Aff (Millwood) 2022;41:1248-1254. DOI: 10.1377/hlthaff.2021.02026.
3. Curtis JR, Willig J. Uptake of remote physiologic monitoring in the US Medicare program: a serial cross-sectional analysis. JMIR Mhealth Uhealth 2023;11:e46046. DOI: 10.2196/46046.
4. The Lancet. Is digital medicine different? Lancet 2018;392:95. DOI: 10.1016/S0140-6736(18)31562-9.
5. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;31:4.
6. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. NPJ Digit Med 2021;41:144.
7. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun 2018;11:156-164. DOI: 10.1016/j.conctc.2018.08.001.
8. Carlisle B, Kimmelman J, Ramsay T, MacKinnon N. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015;12:77-83. DOI: 10.1177/1740774514558307.
9. Zarin DA, Goodman SN, Kimmelman J. Harms from uninformative clinical trials. JAMA 2019;322:813-814. DOI: 10.1001/jama.2019.9892.
10. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002;288:358-362. DOI: 10.1001/jama.288.3.358.
11. Unsworth H, Dillon B, Collinson L, et al. The NICE Evidence Standards Framework for digital health and care technologies — developing and maintaining an innovative evidence framework with global impact. Digit Health 2021;7:20552076211018617. DOI: 10.1177/20552076211018617.
12. Inan OT, Tenaerts P, Prindiville SA, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101. DOI: 10.1038/s41746-020-0302-y.
13. North CM, Barnabas RV. A SMART approach to retention in HIV care. NEJM Evid 2023;2(4). DOI: 10.1056/EVIDe2300040.
14. Kim HM. Get SMART — understanding sequential multiple assignment randomized trials. NEJM Evid 2023;2(4). DOI: 10.1056/EVIDe2300031.
15. Klasnja P, Hekler EB, Shiffman S, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015;34:1220-1228. DOI: 10.1037/hea0000305.
16. 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 2018;52:446-462. DOI: 10.1007/s12160-016-9830-8.
17. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32:(Suppl 5):S112-S118. DOI: 10.1016/j.amepre.2007.01.022.
18. Kario K, Nomura A, Harada N, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021;42:4111-4122. DOI: 10.1093/eurheartj/ehab559.
19. Zhou M, Mintz Y, Fukuoka Y, et al. Personalizing mobile fitness apps using reinforcement learning. CEUR Workshop Proc 2018;2068.
20. Denker K, Dean M, Chapman D, Sump C. Using a centralized nursing team to implement multi-specialty pediatric remote patient monitoring programs. J Pediatr Nurs 2023;69:10-17. DOI: 10.1016/j.pedn.2022.12.028.
21. 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 2019;53:573-582. DOI: 10.1093/abm/kay067.
22. Crossen SS, Romero CC, Lewis C, Glaser NS. Remote glucose monitoring is feasible for patients and providers using a commercially available population health platform. Front Endocrinol (Lausanne) 2023;14:1063290. DOI: 10.3389/fendo.2023.1063290.
23. Scheinker D, Prahalad P, Johari R, Maahs DM, Majzun R. A new technology-enabled care model for pediatric type 1 diabetes. NEJM Catal Innov Care Deliv 2022;3(5). DOI: 10.1056/CAT.21.0438.
24. Scheinker D, Gu A, Grossman J, et al. Algorithm-enabled, personalized glucose management for type 1 diabetes at the population scale: prospective evaluation in clinical practice. JMIR Diabetes 2022;7(2):e27284. DOI: 10.2196/27284.
25. Ferstad JO, Vallon JJ, Jun D, et al. Population-level management of type 1 diabetes via continuous glucose monitoring and algorithm-enabled patient prioritization: precision health meets population health. Pediatr Diabetes 2021;22:982-991. DOI: 10.1111/pedi.13256.

26. Prahalad P, Ding VY, Zaharieva DP, et al. Teamwork, targets, technology, and tight control in newly diagnosed type 1 diabetes: the Pilot 4T Study. J Clin Endocrinol Metab 2022;107:998-1008. DOI: 10.1210/clinem/dgab859.

27. Benhamou P-Y, Adenis A, Tourki Y, et al. Efficacy of a hybrid closed-loop solution in patients with excessive time in hypoglycaemia: a post hoc analysis of trials with DBLG1 System. J Diabetes Sci Technol 2022;29:19322968221128565. DOI: 10.1177/19322968221128565.

28. Battelino T, Alexander CM, Amiel SA, et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. Lancet Diabetes Endocrinol 2023;11:42-57. DOI: 10.1016/S2213-8587(22)00319-9.

29. Yao L, Chu Z, Li S, Li Y, Gao J, Zhang A. A survey on causal inference. ACM Trans Knowl Discov Data 2021;15:74. DOI: 10.1145/3444944

30. Imbens GW, Rubin DB. Causal inference for statistics, social, and biomedical sciences. Cambridge, U.K.: Cambridge University Press, 2015. DOI: 10.1017/CBO9781139025751.

31. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. JAMA 2022;328:2446-2447. DOI: 10.1001/jama.2022.21383.

32. Bhatt DL, Mehta C. Adaptive designs for clinical trials. N Engl J Med 2016;375:65-74. DOI: 10.1056/NEJMra1510061.

33. Pallmann P, Bedding AW, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. BMC Med 2018;16:29. DOI: 10.1186/s12916-018-1017-7.

34. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366:447-453. DOI: 10.1126/science.aax2342.

35. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv 2021;54:1-35.

36. Zaharieva DP, Senanayake R, Brown C, et al. Adding glycemic and physical activity metrics to a multimodal algorithm-enabled decision-support tool for type 1 diabetes care: keys to implementation and opportunities. Front Endocrinol (Lausanne) 2023;13:1096325. DOI: 10.3389/fendo.2022.1096325.